H.-D. Ebbinghaus H. Hermes F. Hirzebruch
M. Koecher K. Mainzer A. Prestel R. Remmert
Redaktion: K. Lamotke

Zahlen

Mit 31 Abbildungen

Springer-Verlag
Berlin Heidelberg New York Tokyo
1983

Grundwissen Mathematik 1

Herausgeber

G. Hämmerlin, F. Hirzebruch, M. Koecher,
K. Lamotke (wissenschaftliche Redaktion),
R. Remmert, W. Walter

Heinz-Dieter Ebbinghaus
Mathematisches Institut, Universität Freiburg
Albertstraße 23b, D-7800 Freiburg

Hans Hermes
Mathematisches Institut, Universität Freiburg
Albertstraße 23b, D-7800 Freiburg

Friedrich Hirzebruch
Max-Planck-Institut für Mathematik
Gottfried-Claren-Straße 26, D-5300 Bonn 3

Max Koecher
Mathematisches Institut, Universität Münster
Einsteinstraße 62, D-4400 Münster

Klaus Mainzer
Philosophische Fakultät, Universität Konstanz
Postfach 5560, D-7750 Konstanz

Alexander Prestel
Fakultät für Mathematik, Universität Konstanz
Postfach 5560, D-7750 Konstanz

Reinhold Remmert
Mathematisches Institut, Universität Münster
Einsteinstraße 62, D-4400 Münster

ISBN 3-540-12666-X Springer-Verlag Berlin Heidelberg New York Tokyo
ISBN 0-387-12666-X Springer-Verlag New York Heidelberg Berlin Tokyo

CIP-Kurztitelaufnahme der Deutschen Bibliothek
Zahlen/H.-D. Ebbinghaus ... Red.: K. Lamotke. – Berlin; Heidelberg; New York; Tokyo: Springer, 1983.
(Grundwissen Mathematik; 1)
ISBN 3-540-12666-X (Berlin ...)
ISBN 0-387-12666-X (New York ...)
NE: Ebbinghaus, Heinz-Dieter [Mitverf.]; GT

Satz: Buchdruckerei Dipl.-Ing. Schwarz' Erben KG, Zwettl.
Druck- und Bindearbeiten: Brühlsche Universitätsdruckerei, Gießen
2144/3140-543210

Vorwort

Das *Grundwissen Mathematik*, welches jeder Mathematiker im Laufe seines Studiums erwirbt, wird erst durch die Vielfalt von Bezügen zwischen den einzelnen mathematischen Theorien zu einem einheitlichen Ganzen. Querverbindungen zwischen den Einzeldisziplinen lassen sich oft durch die historische Entwicklung aufzeigen. Es ist ein Leitgedanke dieser Reihe, dem Leser deutlich zu machen, daß Mathematik nicht aus isolierten Theorien besteht, die nebeneinander entwickelt werden, sondern daß vielmehr Mathematik als Ganzes angesehen werden muß.

Das vorliegende Buch über Zahlen weicht von den weiteren Bänden dieser Reihe dadurch ab, daß hier sieben Autoren und ein Redakteur dreizehn Kapitel zusammentrugen. In Gesprächen miteinander stimmten die Verfasser ihre Beiträge aufeinander ab, und der Redakteur bemühte sich, diese Harmonisierung durch kritische Lektüre und Rücksprache mit den Autoren zu fördern. Die anderen Bände dieser Reihe können unabhängig vom vorliegenden Band studiert werden.

Es ist nicht möglich, an dieser Stelle alle Kollegen zu nennen, die uns durch Hinweise unterstützten. Hervorheben möchten wir jedoch Herrn Gericke (Freiburg), der vielfach half, die historische Entwicklung richtig darzustellen.

K. Peters (damals Springer-Verlag) hatte erheblichen Anteil daran, daß die ersten Herausgeber- und Autorentreffen zustande kamen. Diese Zusammenkünfte wurden durch die finanzielle Unterstützung der Stiftung Volkswagenwerk und des Springer-Verlages sowie durch die Gastfreundschaft des Mathematischen Forschungsinstitutes in Oberwolfach ermöglicht.

Ihnen allen gilt unser Dank.

Oberwolfach, im Juli 1983 — Autoren und Herausgeber

Inhaltsverzeichnis

Einleitung

K. Lamotke

Nach überlieferter Auffassung handelt die Mathematik von Zahlen und Figuren. Wenn man nicht wie EUKLID mit den Figuren, das heißt mit der Geometrie beginnt, soll man daher bei den Zahlen anfangen.

Nun hat die mathematische Forschung in den letzten hundert Jahren abstrakte Theorien wie die Mengenlehre, die allgemeine Algebra und Topologie hervorgebracht, die nach neuer Auffassung die Grundlagen bilden und die daher weit in den mathematischen Anfangsunterricht eingedrungen sind. Diese Entwicklung wird von den Autoren dieses Buches keineswegs ignoriert. Sie wollen davon sogar dadurch profitieren, daß beim Leser die Grundbegriffe der (naiven) Mengenlehre und der Algebra unterstellt werden. Andererseits soll ein erster Band über Zahlen betonen, daß die moderne Forschung in der Mathematik und ihre Anwendungen ganz wesentlich an das anknüpfen, was in der Vergangenheit geschaffen wurde, und daß insbesondere das traditionelle Zahlsystem die wichtigste Grundlage aller Mathematik ist.

Das vorliegende Buch hat drei Teile, von denen der erste, der als Herzstück gelten mag, den Aufbau von den natürlichen bis zu den komplexen Zahlen schildert. Der zweite Teil handelt von der Weiterentwicklung zu „hyperkomplexen Zahlen", und im letzten Teil werden zwei verhältnismäßig neue erweiterte reelle Zahlsysteme vorgestellt. Die sechs Kapitel des ersten Teils bringen zum Thema „Zahlen" das, was jeder Mathematiker einmal gehört oder gelesen haben sollte. Die andern beiden Teile sollen eine über das Grundwissen hinausgehende Neugier des Lesers befriedigen. Insgesamt ist „Aufbau von Zahlsystemen" eine genauere Beschreibung für den Inhalt dieses Buches. Der Zahlentheorie im engeren Sinne wird ein späterer Band dieser Reihe gewidmet.

Auf den Inhalt der verschiedenen Beiträge, die Ziele, die die Autoren darin anstreben, und auf die Gründe, die uns bewogen haben, die Beiträge in der vorliegenden Form zusammenzustellen, wird nun genauer eingegangen.

Teil A

Seit dem Ende des vorigen Jahrhunderts beginnt man beim Aufbau des Zahlensystems mit den natürlichen Zahlen und erweitert sie schrittweise zu den ganzen, den rationalen, den reellen und den komplexen Zahlen. Das ist aber nicht die historische Entwicklung des Zahlbegriffs: Zu den natürlichen Zahlen kommen

noch in antiker Zeit die rationalen Zahlen (Brüche, Verhältnisse) und gewisse irrationale Zahlen (die Kreiszahl π und Quadratwurzeln). Das System der (positiven) rationalen und irrationalen Zahlen wird von griechischen Philosophen und Mathematikern auch theoretisch beschrieben, aber es wird als eine eigenständige Lehre von den kommensurablen und inkommensurablen Proportionen dargestellt und nicht als Erweiterung der natürlichen Zahlen aufgefaßt. Erst nachdem man Jahrhunderte lang mit Proportionen wie mit Zahlen gerechnet hatte, setzte sich im 17. Jahrhundert die Erkenntnis durch: Zahl ist etwas, das sich zu Eins so verhält wie eine beliebige Strecke zu einer gegebenen Strecke. Negative Zahlen, deren Gebrauch im 6. Jahrhundert in Indien nachzuweisen ist, und komplexe Zahlen, die CARDANO 1545 als Lösung quadratischer Gleichungen in Erwägung zog, werden noch lange nach ihrem Auftreten angezweifelt. Im Laufe des 19. Jahrhunderts entsteht dann der heute geläufige Aufbau.

Nach einem Grundprinzip dieser Lehrbuchreihe enthält jedes Kapitel einen Beitrag, in dem geschildert wird, wie sich die Grundbegriffe historisch entwickelt haben. Diese Beiträge sollen keine Geschichte des Zahlbegriffs ersetzen, sondern zum besseren Verständnis der modernen Darstellung durch ihre historische Motivation beitragen.

In diesem Sinne beginnt Kapitel 1, § 1, mit den ältesten uns überlieferten Zahldarstellungen und führt bis zum § 2, in dem das Zählen nach DEDEKIND mit mengentheoretischen Begriffen axiomatisch erfaßt wird.

Beim nun folgenden stufenweisen Aufbau des Zahlsystems wiederholen sich einige Überlegungen: (1) Der Schritt von einer Stufe zur nächsten wird jedesmal durch Probleme motiviert, die sich auf der erreichten Stufe formulieren aber nicht lösen lassen. (2) Das Zahlsystem der nächsten Stufe wird mit Hilfe mengentheoretischer Operationen als Erweiterung des vorhandenen Systems so konstruiert, daß die Ausgangsprobleme lösbar werden. (3) Die bereits vorhandenen Rechenoperationen und Relationen werden auf das neu konstruierte System übertragen. (4) Die Gültigkeit aller Rechenregeln muß auf der neuen Stufe nachgewiesen werden. Die Punkte (1)–(3) werden in den folgenden Kapiteln stets ausgeführt. Der Punkt (4) besteht meist aus langwierigen Verifikationen, die bald zur Routine werden. Hier erlauben sich die Autoren, nur weniges exemplarisch auszuführen und die weitere Routinearbeit dem Leser zu überlassen.

So werden bis zum Ende des Kapitels 1 die rationalen Zahlen erreicht. In Kapitel 2, § 2, werden sie mittels Dedekindscher Schnitte zum System der reellen Zahlen erweitert. Der vorangehende § 1 beginnt mit der Entdeckung irrationaler Zahlen durch die Pythagoreer und schildert die philosophischen und mathematischen Ansätze früherer Zeiten, die schließlich zu DEDEKINDS Konstruktion führten. Der § 3 schildert CANTORS Weg zu den reellen Zahlen durch Vervollständigung der rationalen Zahlen mittels Fundamentalfolgen. Hier reichen die historischen Wurzeln nur wenige Jahrzehnte zurück. Aber dieses Verfahren erweist sich später als sehr fruchtbar: Genauso lassen sich bewertete Ringe, metrische Räume, topologische Vektorräume, allgemein uniforme Strukturen vervollständigen. Der dritte Zugang zu den reellen Zahlen in § 4 folgt WEIERSTRASS. Er beruht auf der uralten Idee, schwer faßbare Zahlen in möglichst kleine Intervalle mit rationalen Grenzen einzuschließen, eine Idee, die sich heute noch in Fehlerbetrachtungen numerischer Rechnungen wiederfindet.

Bereits in Kapitel 2, § 2, wird eine Axiomensystem für die reellen Zahlen formuliert. In § 5 wird gezeigt, daß es diese Zahlen bis auf Isomorphie charakterisiert. Ferner wird dort der Aufbau des Zahlsystems aus diesen Axiomen rekonstruiert, und es werden zahlreiche Formulierungen der „Vollständigkeit" der reellen Zahlen miteinander verglichen.

Die Kapitel 3 bis 5 sind den komplexen Zahlen gewidmet. Mit dem Rüstzeug der Linearen Algebra ist es für uns heute leicht, sie als Paare reeller Zahlen zu beschreiben, die vektoriell addiert und gemäß einer explizit angegebenen Regel multipliziert werden. Dieser Definition in Kapitel 3, § 2, geht ein Abriß der geschichtlichen Entwicklung voraus, der zeigt, wie es von der Erfindung der komplexen Zahlen an 300 Jahre dauerte, bis sie durch GAUSS' Eintreten allgemein verstanden und akzeptiert wurden. Ein Grundgedanke durchzieht die Geschichte bis GAUSS: Die komplexen Zahlen machen Unmögliches möglich. Hierzu gehört vor allem die Möglichkeit, alle Gleichungen zweiten und höheren Grades zu lösen. Diesem als Fundamentalsatz der Algebra bekannten Ergebnis ist das Kapitel 4 gewidmet. Es werden zwei Beweise ausgeführt, die auf ARGAND und LAPLACE zurückgehen und die keine komplexe Funktionentheorie benutzen.

Der Leser mag überrascht sein, daß im Zusammenhang mit den komplexen Zahlen der Kreiszahl π das ganze Kapitel 5 gewidmet ist. Nun gehört, wie bereits im Kapitel 3 ausgeführt und im Kapitel 4 benutzt wird, die Darstellung durch Polarkoordinaten zum wesentlichen Bestand des komplexen Zahlsystems. Zum tieferen Verständnis dieser Darstellung wird im Kapitel 5 die komplexe Exponentialfunktion exp betrachtet, die mit π aufs Engste verbunden ist, weil $\exp z = 1$ genau dann eintritt, wenn z ein ganzzahliges Vielfaches von $2\pi i$ ist. Diese Beziehung dient geradezu als Definition von π, und von ihr aus werden alle anderen gängigen Beschreibungen (als Kreiszahl, als Integralwert, als Grenzwert unendlicher Reihen und Produkte) erreicht.

Die komplexen Zahlen bilden den Ausgangspunkt für eine der großen mathematischen Schöpfungen des 19. Jahrhunderts, die Funktionentheorie. Ihren klassischen Bestand werden zwei weitere Bände dieser Reihe behandeln.

Teil B

Mit den komplexen Zahlen ist der Aufbau des Zahlsystems abgeschlossen. Wenn man nach dem Vorbild der komplexen Zahlen, die einen zweidimensionalen reellen Vektorraum bilden, höherdimensionale reelle Vektorräume zu hyperkomplexen Zahlsystemen (heute sagt man kurz „Algebren") machen will, muß man entweder unendliche Dimensionen zulassen, oder man muß vertraute Körperaxiome wie die Kommutativität oder die Assoziativität der Multiplikation oder die Möglichkeit der Division aufgeben. Gibt man zu viel auf, wird man von einer Flut neuer Zahlsysteme überschwemmt. Als Damm wird in Teil B dieses Buches die Endlichkeit der Dimension und die Möglichkeit der Division aufrecht erhalten.

Die vierdimensionale Divisionsalgebra der Quaternionen und die achtdimensionale der Oktaven, die kurz hintereinander im Jahre 1843 erfunden wurden, werden im Kapitel 6 bzw. 8 ausführlich besprochen. So wie die komplexen Zahlen

die euklidische Geometrie der Ebene oft verblüffend einfach beschreiben (Kapitel 3, § 4, enthält einige Kostproben), eignen sich die Quaternionen zur Beschreibung der drei- und vierdimensionalen Geometrie. Auch darauf wird im Kapitel 6 ausführlich eingegangen.

Die übrigen Kapitel des Teils B handeln unter verschiedenen Gesichtspunkten von der Einzigkeit der vier Algebren der reellen und komplexen Zahlen, der Quaternionen und der Oktaven: Wenn man nur die Kommutativität aufgibt, sind die Quaternionen die einzige Möglichkeit (FROBENIUS 1877; Beweis im ersten Teil des Kapitels 7). Wenn man an der Kommutativität festhält, aber bereit ist, die Assoziativität aufzugeben, wird man zwangsläufig auf die reellen und komplexen Zahlen als einzige Möglichkeiten geführt (H. HOPF 1940; Beweis im zweiten Teil des Kapitels 7). Merkwürdig ist nicht nur das Ergebnis sondern auch sein Beweis, der zeigt, daß dies an topologischen Sachverhalten liegt. – Solange man noch ein abgeschwächtes Assoziativgesetz $x(xy) = x^2y$ und $(xy)y = xy^2$ aufrecht erhält (ohne Kommutativität), sind die Oktaven die einzige Möglichkeit (ZORN 1933; Beweis am Ende des Kapitels 8).

Eine weitere Charakterisierung der vier Algebren hat HURWITZ 1898 gefunden: Es sind die einzig möglichen Divisionsalgebren mit Einselement, die gleichzeitig euklidische Vektorräume mit normtreuer Multiplikation ($\|x\| \cdot \|y\| = \|x \cdot y\|$) sind. Dies hängt eng mit der Tatsache zusammen, daß das Produkt zweier natürlicher Zahlen, die beide Summe von 2, 4 oder 8 Quadratzahlen sind, wieder eine Summe von ebenso vielen Quadraten ist und daß dies generell nur bei 2, 4 oder 8 Summanden richtig ist. Das Kapitel 9 behandelt diese Dinge.

Soweit werden alle Ergebnisse mit ausführlichen Beweisen dargestellt, die nur lineare Algebra, Differentialrechnung mehrerer Veränderlicher und Grundkenntnisse aus der Algebra und Topologie voraussetzen.

Im Kapitel 10 geht es um das weitreichendste Ergebnis: Nur in den Dimensionen 1, 2, 4 und 8 sind Divisionsalgebren möglich. Hier wird nichts weiter vorausgesetzt. Dies wurde zur Überraschung der Algebraiker 1958 von BOTT, KERVAIRE und MILNOR bewiesen, und zwar wie schon bei HOPFS Ergebnis mit topologischen Methoden. Diesmal muß aber der ganze umfangreiche Apparat der algebraischen Topologie eingesetzt werden, und im Kapitel 10 kann der Beweis nur skizziert werden.

HAMILTON hielt seine Erfindung der Quaternionen im Jahre 1843 für eines der wichtigsten Ereignisse in der Geschichte der Mathematik. Längst hat man aber eingesehen, daß sie (und erst recht die Oktaven) an Bedeutung weit hinter den komplexen Zahlen zurückstehen. Die Nicht-Kommutativität hat sich als Hindernis nicht überwinden lassen, um eine quaternionale Analysis zu begründen.

Teil C

Der Aufbau des reellen Zahlsystems schien vom Standpunkt der Forschung längst abgeschlossen zu sein, da manifestierten sich in jüngster Zeit neue Ideen.

Im Jahre 1960 entdeckte ROBINSON, wie man nach den Vorbildern aus dem 17. und 18. Jahrhundert Infinitesimalrechnung mit infinitesimalen Größen exakt

betreiben kann. Er erweiterte dazu den reellen Zahlkörper zum angeordneten Körper der Non-Standard-Zahlen, in dem unendlich kleine und große Zahlen Platz haben. Die Konstruktion dieser Erweiterung, die im Kapitel 11 geschildert wird, erfordert nicht mehr Aufwand als beispielsweise CANTORS Konstruktion der reellen Zahlen (vgl. Kapitel 2, § 3); und die Differential- und Integralrechnung mittels infinitesimaler Größen (Kapitel 11, § 4) wird manchem Leser intuitiver und einfacher erscheinen als die herkömmliche Methode. Dazwischen steht aber der Preis, den man zahlen muß: Alle Aussagen, die von den reellen auf die Non-Standard-Zahlen übertragen werden sollen, müssen in einer formalen Sprache ausgedrückt werden; und das erfordert, tiefer in die formale Logik einzudringen, als es die meisten Mathematiker gewöhnt sind.

Noch 10 Jahre jünger ist CONWAYS Einfall, einen großen angeordneten Zahlkörper ohne Zwischenstufen von Anfang an durch iterierte „Dedekindsche" Schnittbildungen aufzubauen, seine Elemente als Spiele zu deuten und die Ordnung mittels Gewinnstrategien zu erklären. All dies wird im Kapitel 12 erklärt.

In den beiden Kapiteln 11 und 12 werden Ideen hauptsächlich präsentiert und nicht in allen Einzelheiten ausgeführt.

Bei Conways Konstruktion reicht eine nur naiv verstandene Mengenlehre nicht mehr ganz aus. Darum werden im letzten Kapitel 13 die Grundzüge der axiomatischen Mengenlehre nach ZERMELO und FRAENKEL dargestellt. Dieses Kapitel ist aber auch für die Leser der ersten beiden Kapitel dieses Buches gedacht, die sich bei der Einführung der natürlichen Zahlen und den darauf aufbauenden Erweiterungskonstruktionen nicht nur auf eine naiv verstandene Mengenlehre stützen wollen. Vom logischen Aufbau her gehört also dieses Kapitel 13 an den Anfang. Aber wir berufen uns auf SCHILLER (Brief an GOETHE vom 5. 2. 1796): „Wo es die Sache leidet, halte ich es immer für besser, nicht mit dem Anfang anzufangen, der immer das Schwerste ist."

Teil A

Von den natürlichen zu den komplexen Zahlen

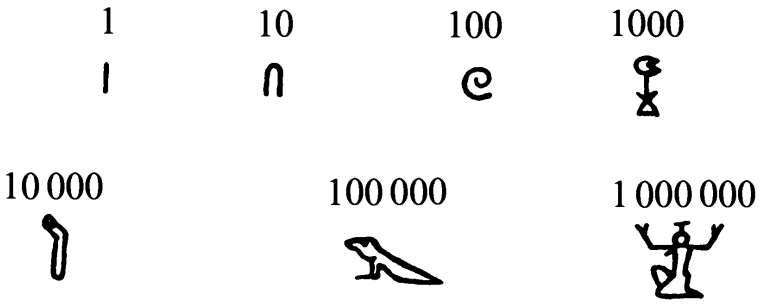
Kapitel 1. Natürliche, ganze und rationale Zahlen

K. Mainzer

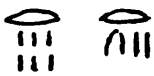
> Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk. (L. KRONECKER, Jahresber. DMV 2, S. 19).
>
> Die Zahlen sind freie Schöpfungen des menschlichen Geistes, sie dienen als ein Mittel, um die Verschiedenheit der Dinge leichter und schärfer aufzufassen. (R. DEDEKIND, Was sind und was sollen die Zahlen? Braunschweig 1887, S. III).

§ 1. Historisches

1. Ägypten und Babylonien. Zahlzeichen gehören zu den ältesten Spuren menschlicher Schrift. Schon in der *Altsteinzeit* finden wir sie als Kerben in Knochen oder Striche auf Felswänden. Es ist die Zeit des Jägerlebens, und wir können heute nur noch spekulieren, ob z. B. mit |||| die Anzahl der erlegten Beutetiere bezeichnet wurde. Zahlensysteme markieren den Beginn der Arithmetik. Ihre ersten Dokumente gehen auf den Beginn der alten *Hochkulturen am Nil, Euphrat und Tigris* zurück. Auf einer Keule des Königs NARMER aus der *1. ägyptischen Dynastie* (ca. 3000 v. Chr.) finden sich Hieroglyphen für die Zahlen 10 000, 100 000 und 1 000 000. Die Zahlen sind in schematischer Wiedergabe:

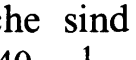


Die verwendeten Bilder könnten Anspielungen auf das praktische Vorkommen der entsprechenden Anzahlen sein: z. B.  als Zeichen für den Meßstrick mit 100 Einheiten. Allerdings ist es auch möglich, daß die Zahlzeichen Gegenstände bezeichnen, die denselben Anfangsbuchstaben haben wie das betreffende Zahlwort. Durch additive Notation entstehen neue Zahlen, z. B.  = 221 000 oder  = 10 010. Addition und Subtraktion bilden dann kein Problem: So ergeben z. B.  = 12 und  = 11 zusammen  = 23. Multiplikation und Division werden auf Verdoppeln und Halbieren zurückgeführt. Die dabei auftretenden Brüche sind aus Stammbrüchen zusammengesetzt, also z. B. $\frac{1}{12}$ =  , wobei das Zeichen  eine Ziffer zum Nenner eines Stammbruchs macht. Für den Bruch $\frac{3}{12}$ wird die Rechnung 3 mal $\frac{1}{12}$ durchgeführt, also

1	$\frac{1}{12}$	(das heißt 1 mal $\frac{1}{12} = \frac{1}{12}$)
2	$\frac{1}{6}$	(verdoppeln)
zusammen	$\frac{1}{6}\frac{1}{12}$,	also  .

Um entsprechende Rechnungen mit Brüchen allgemein durchführen zu können, muß man die Hälfte und das Doppelte eines Stammbruchs als Summe von Stammbrüchen angeben können. Im Pypyrus Rhind (ca. 1650 v. Chr.) gibt es eine Tabelle für $2:n$ bzw. 2 mal $\frac{1}{n}$ für ungerade n. (Zum Ägyptischen Rechnen vgl. man den „*Papyrus Moskau*" [28], den „*Papyrus Rhind*" [23].)

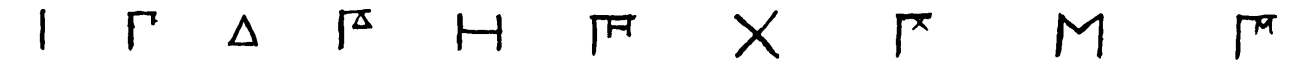
Die *Babylonier* verwendeten Keilsymbole auf Tontäfelchen, die in einem sexagesimalen Positionssystem geordnet wurden: ▼ steht für $1, 60^1, 60^2, \ldots$; $<$ steht für $10, 10 \cdot 60^1, 10 \cdot 60^2, \ldots$. Eine Null wird von den Babyloniern nicht immer gebraucht, eine Markierung wie durch unser Komma nie. Im Positionssystem ist die Rolle der Null die eines „Lückenzeichens". Ein solches, zwei kleine Winkelhaken  , findet sich schon auf einem altbabylonischen Text aus Susa (Text XII, Z. 4), allerdings vereinzelt (J. Tropfke [29], S. 28).

Ohne ein solches Lückenzeichen muß der Stellenwert jeweils aus dem Zusammenhang geschlossen werden. So kann z. B. $<<▼<$ sowohl für $21 \cdot 60 + 10$, als auch $21 \cdot 60^2 + 10 \cdot 60^1$ oder $21 \cdot 60^2 + 10$ etc. stehen. Sexagesimalbrüche sind z. B. $<<<$ für $0;30 = \frac{30}{60} = \frac{1}{2}$ oder  für $0;6,40 = 6\frac{1}{60^1} + 40 \cdot \frac{1}{60^2} = \frac{1}{9}$. (Zum babylonischen Rechnen vgl. O. Neugebauer [20], E. M. Bruins/M. Rutten [7].)

Die Babylonier erweisen sich als hochbegabte Arithmetiker und Algebraiker. Sie entwickeln raffinierte Rechentafeln zur Lösung von Multiplikations- und Divisionsaufgaben, von quadratischen und kubischen Gleichungen, geben eine Rechenvorschrift zur Lösung gemischt-quadratischer Gleichungen durch quadratische Ergänzung und lösen sogar gemischt-kubische Gleichungen mit Tabellen für $x^2(x+1)$. Über ihre Näherungsrechnungen für Wurzeln wird noch in Kap. 2 zu berichten sein. Jedenfalls beeinflussen sie mit ihren geschickten Rechenverfahren maßgebend die weitere Entwicklung der Arithmetik und Algebra.

2. Griechenland. Das Zahlensystem der Griechen war dekadisch, aber kein Positionssystem. Das *frühe System* benutzte Individualzeichen für die dekadischen Stufen, die aus den Anfangsbuchstaben der betreffenden Zahlworte bestanden.

Durch Verbindung des Zeichens für 5 mit den anderen Zeichen wurden die Zwischenstufen 50, 500, ... angegeben:



1	5	10	50	100	500	1000	5000	10 000	50 000

Die *spätere Zahlendarstellung* durch Buchstaben (seit ca. 450 v. Chr.) wird in mathematischen Texten verwendet. Sie besteht aus den 24 griechischen Buchstaben mit drei weiteren Zeichen aus orientalischer Tradition:

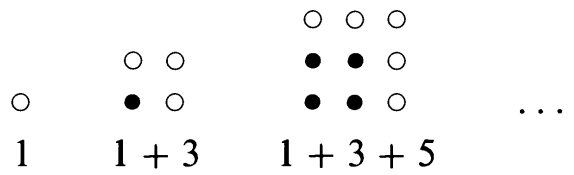
$1-9$	$\alpha, \beta, \gamma, \delta, \varepsilon, \varsigma, \zeta, \eta, \vartheta$	$(\varsigma = 6)$
$10-90$	$\iota, \kappa, \lambda, \mu, \nu, \xi, o, \pi, \varsigma$	$(\varsigma = 90)$
$100-900$	$\rho, \sigma, \tau, \upsilon, \varphi, \chi, \psi, \omega, \text{ϡ}$	$(\text{ϡ} = 900)$
$1000-9000$	$,\alpha, ,\beta, \ldots$	(Strich links unten)
10 000	M	$(M = \mu\upsilon\rho\iota\acute{\alpha}\varsigma)$

Die Zahlen werden durch Aneinanderreihen von Zeichen als Addition gebildet: z. B. $\iota\beta = 10 + 2 = 12$, $\sigma\kappa\beta = 200 + 20 + 2 = 222$, $,\alpha\tau\varepsilon = 1000 + 300 + 5 = 1305$. Die Anzahl der Zehntausender (Myriaden) wird über das Symbol M geschrieben, z. B.

$$\overset{\beta}{M},\varepsilon\mu\gamma = 25\,000 + 40 + 3 = 25\,043.$$

Stammbrüche werden meistens durch den Nenner mit einem Strich rechts oben bezeichnet, allgemeinere Brüche in verschiedener Weise (z. B. durch Untereinanderschreiben von Zähler unter Nenner). Das griechische System war also kein Positionssystem und das Rechnen ziemlich mühsam.

Neben der Rechnung in Ziffernsystemen findet sich auch früh eine Zahldarstellung durch *Rechensteinchen*, mit denen sogar zahlentheoretische Sätze gefunden wurden. So berichtet ARISTOTELES von dem *Pythagoreer* EURYTOS, der bestimmt habe, was die Zahl ($\grave{\alpha}\rho\iota\vartheta\mu\acute{o}\varsigma$) jedes Gegenstandes sei, und der die Formen der Lebewesen durch Spielsteine (gr. $\psi\tilde{\eta}\varphi o\iota$, lt. calculi) nachgeahmt habe „in der Art derjenigen, die Zahlen in die Gestalt eines Dreiecks oder Quadrats bringen." (Aristoteles [1], 1092b, 10–12). So ergeben z. B. die ungeraden Zahlen sukzessive ausgelegt die Quadratzahlen:



Durch Zerlegung der Quadrate parallel zu einer Diagonalen liest man ab

$$2^2 = 1 + 2 + 1, \qquad 3^2 = 1 + 2 + 3 + 2 + 1,$$

allgemein

$$n^2 = 1 + 2 + \cdots + n + \cdots + 2 + 1,$$

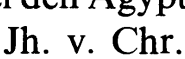
also $1 + 2 + \cdots + (n-1) = \frac{1}{2}(n^2 - n)$ (Aristoteles [2], III 4, 203a, 13–15, O. Becker [3], S. 34ff.).

Während sich Ägypter und Babylonier mit einer – wenn auch hoch entwickelten – Rechentechnik begnügten, gewinnen die Zahlen bei den Pythagoreern philosophische Bedeutung: Das gesamte Universum ist nach pythagoreischer Auffassung durch Zahlen und deren Verhältnisse charakterisiert. Damit entsteht das Problem, allgemein zu definieren, was eine Zahl sei. EUKLID definiert in den „Elementen" VII, 2: „Zahl ist die aus Einheiten zusammengesetzte Menge." Vorher heißt es (El. VII, 1): „Einheit ist das, wonach jedes Ding eines genannt wird." Da die Einheit nicht aus Einheiten zusammengesetzt ist, fassen EUKLID und ARISTOTELES sie nicht als Zahl auf, sondern als „Grundlage des Zählens, der Ursprung der Zahl." Die euklidische Definition klingt wieder an bei G. Cantor, wenn er die Kardinalzahl als eine aus lauter Einsen zusammengesetzte Menge bezeichnet (G. Cantor [8], S. 283).

Neben dieser am Zählen orientierten *Definition der Zahl* findet sich bei ARISTOTELES auch folgende Formulierung: Das, was in diskrete Teile zerlegbar ist, heißt $\pi\lambda\tilde{\eta}\vartheta o\varsigma$ („Vielheit"), und die begrenzte (endliche) Vielheit heißt Zahl (Aristoteles [1], 1020a, 7–14).

Die Griechen fassen also nur die natürlichen Zahlen (ohne die Eins) als Zahlen auf. Brüche behandeln sie als Zahlverhältnisse, irrationale Zahlen als inkommensurable Größenverhältnisse der Geometrie (vgl. Kap. 2).

3. Indisch-arabische Rechenpraxis. In Indien entsteht in der Zeit von 300 v. Chr. bis 600 n. Chr. (vermutlich unter babylonischem Einfluß) die heutige positionelle Dezimalnotation mit 0 und eigenen Ziffern für $1, \ldots, 9$. So entsteht z. B. aus der primitiven Form $-$, $=$ zunächst $\neg$, $\rightleftarrows$, schließlich 1, 2. Die *indische Notation* wurde von den Arabern, nicht zuletzt durch ihre Astronomen, übernommen. Die Inder hatten Bezeichnungen für *positive* und *negative* Zahlen, nämlich „dhana" oder „sva" (Eigentum) und „rina" oder „kśaya" (Verminderung, Schuld). Rechenregeln für positive und negative Zahlen finden sich bei BRAHMAGUPTA (geb. 598) (A. P. Juschkewitsch [15], S. 126). Allerdings finden sich keine Hinweise, daß negative Zahlen generell als Gleichungslösungen anerkannt wurden. So erscheint es sinnlos, die negative Lösung einer Aufgabe anzuerkennen, wenn z. B. nach der Anzahl einer Affenherde gefragt wird. Im Fall einer negativen Lösung für eine Streckenangabe wurde aber einmal richtig die Strecke in entgegengesetzter Richtung gedeutet.

Der indische Mathematiker SRIDHARA (ca. 850–950) formuliert Rechenregeln für die *Null*, deren Symbolisierung bereits bei den Ägyptern (das Zeichen  in einer Inschrift am Tempel von Edfu 2. Jh. v. Chr.), den Griechen (z. B. bei PTOLEMAIOS das Lückenzeichen *o* als möglicherweise Anfangsbuchstabe von *οὐδέν* = nichts) und den Indern (seit dem 5. Jh. n. Chr. das Wort „sunya" für das Leere) auftaucht. Die Araber verwenden die Bezeichnung „al-sifr" für die Null, aus der das Wort „cifra" abgeleitet wurde, was noch bei GAUSS die Bedeutung Null hatte (A. P. Juschkewitz [15], S. 107, R. Lepsius [19] und C. F. Gauß [12], S. 8). Als Zeichen ist in Indien ein Punkt oder ein Kreis seit dem 7. Jahrhundert zu finden.

4. Neuzeit. Die indisch-arabische Rechenpraxis fand Verbreitung durch die berühmten Rechenbücher des 13. bis 16. Jahrhunderts (z. B. LEONARDO VON PISA, A. RIESE, M. STIFEL) und ermöglichte erst die erfolgreichen Gleichungslösungen der italienischen Renaissance-Mathematiker, wie z. B. DEL FERRO, CARDANO, FERRARI u. a. M. STIFEL sagt über negative Zahlen, daß es kein „leeres Geschwätz", sondern „nicht ohne Nutzen" sei, wenn „Zahlen unter Null, das heißt unter dem Nichts, *fingiert*" werden (M. Stifel [27], S. 248 vf.).

In der neuzeitlichen *Algebra* erhalten die Null und die negativen Zahlen eine neue Funktion, da sie die Zusammenfassung mehrerer Gleichungstypen zu einem ermöglichen. Seit R. DESCARTES werden Gleichungen in der Form

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0 = 0$$

(bei DESCARTES allerdings ohne Koeffizientenindizes) geschrieben, wobei die Koeffizienten a_i positiv, negativ oder Null sein können.

Obwohl die Mathematiker seit Beginn ihrer Wissenschaft mit Zahlen operierten und Sätze über Zahlen fanden, werden jedoch erst im *19. Jahrhundert* mathematisch brauchbare Definitionen des Zahlbegriffs angegeben. Dabei standen zunächst die Präzisierungsbemühungen um die Grundlagen der Analysis, also die reellen

Zahlen, im Vordergrund. Erst nachdem R. DEDEKIND, G. CANTOR u. a. die reellen Zahlen mit Mengen von rationalen Zahlen definiert hatten (vgl. Kap. 2), folgten historisch die klassischen Definitionen der natürlichen Zahlen durch Logik und Mengenlehre. Die Einsicht, daß die Erweiterung der natürlichen Zahlen zu den ganzen und rationalen Zahlen nur noch Thema der Algebra sei, war eng mit der Einführung grundlegender algebraischer Begriffe der Ring- und Körpertheorie verbunden.

§ 2. Natürliche Zahlen

Das Zählen mit Zahlzeichen steht historisch am Anfang der Arithmetik. Rechnen setzt Zählen voraus. Bis ins 19. Jahrhundert wurde daher versucht, den Zahlbegriff auf den psychologischen Vorgang des Zählens zurückzuführen. Die dabei verwendete psychologische und philosophische Terminologie stieß jedoch auf Kritik, nachdem mit G. FREGES *Logik* und G. CANTORS *Mengenlehre* logisch-mathematische Grundlagen zur Präzisierung des Zahlbegiffs bereitstanden. R. DEDEKIND, der seit Anfang der 70iger Jahre mit G. CANTOR korrespondierte, stellte in seiner Arbeit „Was sind und was sollen die Zahlen?“ [9] (veröffentlicht 1888, wesentliche Teile bereits 1872–1878 verfaßt) eine mengentheoretische Definition der natürlichen Zahlen vor, der Vorschläge von G. FREGE, G. CANTOR u. a. und schließlich G. PEANOS Axiomatisierung folgten. Daß die so axiomatisierten Zahlen einzig, das heißt, bis auf Isomorphie bestimmt sind, folgt aus DEDEKINDS Rekursionssatz.

Wir setzen von nun an die Grundbegriffe der Mengenlehre als bekannt voraus. Siehe dazu auch das letzte Kapitel dieses Buches.

1. Definition der natürlichen Zahlen. Die *natürlichen Zahlen* bilden eine Menge $\mathbb{N}$, in der ein Element $0 \in \mathbb{N}$ (die *Null*) ausgezeichnet ist und auf der eine Selbstabbildung $S: \mathbb{N} \to \mathbb{N}$ (*Nachfolgerfunktion*; engl. „successor“ Nachfolger) definiert ist, so daß folgende Axiome erfüllt sind:

(S 1) *S ist injektiv.*
(S 2) $0 \notin S(\mathbb{N})$.
(S 3) *Wenn eine Teilmenge* $M \subset \mathbb{N}$ *die Null enthält und durch S in sich abgebildet wird, dann ist* $M = \mathbb{N}$.

Die Abbildung S beschreibt unter Benutzung mengentheoretischer Begriffe den Vorgang des Zählens. Die Vorstellung ist, daß S jeder natürlichen Zahl n die nachfolgende Zahl $S(n)$ zuordnet. $1 := S(0)$, $2 := S(1)$, $3 := S(2)$ usw. Das 1. Axiom präzisiert, daß man beim Zählen nicht mehrmals auf dieselbe Zahl stoßen kann. Im 2. Axiom kommt zum Ausdruck, daß 0 der Ausgangspunkt des Zählens ist, aber auch, daß 0 durch den Zählprozeß nicht erreicht wird. (Manche Mathematiker ziehen es vor, den Zählprozeß mit 1 beginnen zu lassen, z. B. DEDEKIND.) Das 3. Axiom ist die mengentheoretische Formulierung für

Das Prinzip der vollständigen Induktion. Wenn eine Eigenschaft E der Zahl 0 zukommt (Induktionsanfang) und für jede Zahl n, welche die Eigenschaft E hat,

auch der Nachfolger $S(n)$ die Eigenschaft E hat (Induktionsvoraussetzung), dann kommt diese Eigenschaft allen natürlichen Zahlen zu.

Die Äquivalenz dieses Prinzips mit dem 3. Axiom erhält man, indem man die Eigenschaft E durch die Teilmenge M der Zahlen ersetzt, denen sie zukommt. Anstatt „n hat die Eigenschaft E" sagt man auch „Die Aussage E trifft für n zu" oder „$E(n)$ gilt". Das Induktionsprinzip ist keine neue Schlußweise der Mathematiker neben den üblichen Schlußregeln der Logik, sondern nichts anderes als die Anwendung des 3. Axioms, um zu beweisen, daß gewisse Aussagen für alle natürlichen Zahlen gelten.

Beispiel. Wir setzen die Rechenoperationen für natürliche Zahlen (vgl. 3.) als bekannt voraus und beweisen durch vollständige Induktion:

Für alle natürlichen Zahlen n gilt $2(1 + 2 + \cdots + n) = n(n + 1)$. Hier bedeutet $E(0)$ die Aussage $2 \cdot 0 = 0 \cdot 1$, $E(1)$ die Aussage $2 \cdot 1 = 1 \cdot 2$ und $E(n)$ für $n = 2, 3, \ldots$ die Aussage $2(1 + \cdots + n) = n(n + 1)$. Der Induktionsbeginn ist die gültige Aussage $E(0)$. Man beweist die Aussage $E(n + 1)$ durch folgende Rechnung, bei der an der Stelle $*$ die Aussage $E(n)$ als Induktionsvoraussetzung benutzt wird: $2(1 + 2 + \cdots + n + n + 1) = 2(1 + 2 + \cdots + n) + 2(n + 1) \overset{*}{=} n(n + 1) + 2(n + 1) = (n + 1)(n + 2)$.

Für das Induktionsprinzip gibt es andere Formulierungen. Zwei häufig gebrauchte, die die Anordnung der natürlichen Zahlen (vgl. 3) als bekannt voraussetzen, seien hier genannt:

1) Liegt für die Eigenschaft E der Induktionsbeginn nicht bei 0 sondern bei der Zahl n_0, so haben alle Zahlen $n \geqslant n_0$ die Eigenschaft E.

2) Wenn für alle natürlichen Zahlen n die Aussage $E(n)$ aus den Aussagen $E(m)$ für alle $m < n$ folgt, gilt $E(n)$ für alle natürlichen Zahlen n.

Eine Menge M heißt *unendlich*, falls es eine injektive Selbstabbildung $f: M \to M$ mit $f(M) \neq M$ gibt. Inhaltlich bringt diese Definition zum Ausdruck, daß nur unendliche Mengen auf echte Teilmengen injektiv abbildbar sind. Historisch wurde die Definition von R. Dedekind in „Was sind und was sollen die Zahlen?" angegeben. Statt von injektiven Abbildungen spricht Dedekind dort (§ 5, Nr. 64) von „ähnlichen" Abbildungen.

Satz. *Es gibt eine unendliche Menge genau dann, wenn es eine Menge* $\mathbb{N}$ *gibt, welche die Axiome* (S1)–(S3) *erfüllt.*

Beweis. Wenn es eine solche Menge $\mathbb{N}$ gibt, dann existiert nach Axiom (S1) und (S2) auch eine unendliche Menge (man setze $f = S$).

Sei A eine unendliche Menge. Dann existiert eine injektive Selbstabbildung $f: A \to A$ mit $f(A) \neq A$. Also existiert auch ein Element $0 \in A$ mit $0 \notin f(A)$. Sei I die Klasse aller Mengen $M \subset A$ mit $0 \in M$ und $f(M) \subset M$. Nach Voraussetzung ist $I \neq \emptyset$. Also ist der Durchschnitt $\bigcap_{M \in I} M$ definiert. Er erfüllt die Axiome (S1)–(S3), wenn man $f \mid M$ als Nachfolgefunktion S nimmt. □

Bemerkung. Dedekind gibt auch einen Beweis für die Existenz einer unendlichen Menge an, der jedoch den (mengentheoretisch) widerspruchsvollen Begriff der Menge aller Mengen

voraussetzt (§ 5, Nr. 66). Ein analoger Versuch findet sich bei B. BOLZANO [4] in den „Paradoxien des Unendlichen" (§ 13). Wir nehmen nach dem *Unendlichkeitsaxiom* (vgl. Kap. 13) an, daß es unendliche Mengen gibt. $\mathbb{N}$ ist in unserem Beweis eine „kleinste" unendliche Menge, die in einer unendlichen Menge enthalten ist. DEDEKIND spricht deshalb auch von „einfach unendlichen Systemen" (§ 6, Nr. 71). Die im Beweis angegebene Konstruktion von $\mathbb{N}$ hängt von der Wahl von A, von f und von 0 ab. Daß $\mathbb{N}$, die Nachfolgerfunktion S und 0 bis auf Isomorphie eindeutig bestimmt sind, wird in 2 (Einzigkeitssatz) gezeigt. Nach J. v. NEUMANN kommt man zu einem kanonisch definierten mengentheoretischen Modell für $\mathbb{N}$, indem man die Zermelo-Fraenkelsche Mengenlehre zugrunde legt (J. v. NEUMANN [21], vgl. auch Kap. 13).

G. FREGE und G. CANTOR definieren die natürlichen Zahlen als „endliche Mächtigkeiten" bzw. „endliche Kardinalzahlen" (G. FREGE [11], S. 73f., G. CANTOR [8], S. 119, vgl. auch Kap. 13). Dieser Ansatz findet sich auch bei B. RUSSELL [25], S. 116 und N. BOURBAKI [6], I, Chap. III, § 4, Déf. 1.

2. Rekursionssatz und Einzigkeit von $\mathbb{N}$. Neue Begriffe werden im Bereich der natürlichen Zahlen meist rekursiv eingeführt. Man spricht auch von induktiver Definition. Beispielsweise erklärt man die Addition dadurch, daß man nacheinander festlegt: $m + 0 := m$, $m + 1 := S(m)$, $m + 2 := S(m + 1)$, allgemein $m + S(n) := S(m + n)$. Den Nachweis, daß dieses Definitionsverfahren sinnvoll ist, liefert der

Rekursionssatz (DEDEKIND 1888). *Sei eine beliebige Menge A mit einem Element $a \in A$ und eine Selbstabbildung $g: A \to A$ gegeben. Dann gibt es genau eine Abbildung $\varphi: \mathbb{N} \to A$ mit $\varphi(0) = a$ und $\varphi \circ S = g \circ \varphi$.*

Beweis. Um die *Einzigkeit* der Abbildung φ zu zeigen, betrachten wir zwei Abbildungen $\varphi_1, \varphi_2: \mathbb{N} \to A$ mit den behaupteten Eigenschaften. Wir zeigen durch Induktion über n, daß $\varphi_1(n) = \varphi_2(n)$ für alle n.

Induktionsanfang: $\varphi_1(0) = a = \varphi_2(0)$.

Da nach Induktionsvoraussetzung $\varphi_1(n) = \varphi_2(n)$ ist, folgt

$$\varphi_1(S(n)) = g(\varphi_1(n)) = g(\varphi_2(n)) = \varphi_2(S(n)).$$

Um die *Existenz* von φ zu beweisen, betrachtet man alle Teilmengen $H \subset \mathbb{N} \times A$ mit den Eigenschaften (1) $(0, a) \in H$ und (2) Für alle n, b gelte: Wenn $(n, b) \in H$, dann $(S(n), g(b)) \in H$. Da ganz $\mathbb{N} \times A$ solch ein H ist und alle H das Element $(0, a)$ enthalten, ist der Durchschnitt D aller H die kleinste Teilmenge von $\mathbb{N} \times A$, die (1) und (2) erfüllt. Wir behaupten, daß D der Graph einer Abbildung $\varphi: \mathbb{N} \to A$ ist, und zeigen dazu durch vollständige Induktion:

(*) Zu jedem $n \in \mathbb{N}$ gibt es genau ein b, so daß $(n, b) \in D$ ist.

Induktionsanfang: Nach (1) ist $(0, a) \in D$. Wäre auch $(0, c) \in D$, aber $c \neq a$, so könnte man $(0, c)$ aus D entfernen, und die Restmenge $D \setminus \{(0, c)\}$ hätte immer noch die Eigenschaften (1) und (2) im Widerspruch dazu, daß D die kleinste derartige Menge ist.

Induktionsschluß: Nach der Induktionsvoraussetzung gibt es genau ein b, so daß $(n, b) \in D$ ist. Nach (2) ist dann $(S(n), g(b)) \in D$. Wäre auch $(S(n), c) \in D$, aber $c \neq g(b)$, könnte man $(S(n), c)$ aus D entfernen und erhielte wie beim

Induktionsbeginn einen Widerspruch. Nachdem (*) bewiesen ist, kann D als Graph einer Abbildung $\varphi: \mathbb{N} \to A$ geschrieben werden: $D = \{(n, \varphi(n)) \mid n \in \mathbb{N}\}$. Die Eigenschaft (1) von D bedeutet, daß $\varphi(0) = a$, und die Eigenschaft (2), daß $(S(n), g(\varphi(n))) \in D$, also $\varphi \circ S(n) = g \circ \varphi(n)$ für alle n gilt. □

Von der Abbildung φ im Rekursionssatz sagt man, daß $\varphi(n)$ ausgehend von $\varphi(0) = a$ durch die Rekursionsformel $\varphi(n + 1) = g(\varphi(n))$ definiert sei. (Wegen der Bezeichnung „$n + 1$" statt „$S(n)$" vergleiche man 3.)

Beispiel. Die n-te Potenz c^n einer reellen Zahl c wird ausgehend von $c^0 = 1$ durch die Rekursionsformel $c^{n+1} = c^n \cdot c$ definiert. Man wendet hier den Rekursionssatz mit $A = \mathbb{R}$ (Menge der reellen Zahlen), $a = 1$ und $g(b) = b \cdot c$ an.

Als erste Anwendung des Rekursionssatzes folgt der Nachweis der Einzigkeit von $\mathbb{N}$.

Einzigkeitssatz. *Sei $\mathbb{N}'$ eine Menge mit einer Nachfolgerfunktion S', einem ausgezeichneten Element $0'$ und den Axiomen* (S1)–(S3). *Dann sind $\mathbb{N}'$ und $\mathbb{N}$ kanonisch isomorph, das heißt, es gibt genau eine bijektive Abbildung $\varphi: \mathbb{N} \to \mathbb{N}'$ mit $\varphi(0) = 0'$ und $S' \circ \varphi = \varphi \circ S$.*

Beweis. Nach dem Rekursionssatz, angewandt auf $A = \mathbb{N}'$, $a = 0'$ und $g = S'$, gibt es genau eine Abbildung $\varphi: \mathbb{N} \to \mathbb{N}'$ mit $\varphi(0) = 0'$ und $\varphi \circ S = S' \circ \varphi$. Indem man die Rollen von $\mathbb{N}$ und $\mathbb{N}'$ vertauscht, erhält man entsprechend eine Abbildung $\psi: \mathbb{N}' \to \mathbb{N}$ mit $\psi(0') = 0$ und $\psi \circ S' = S \circ \psi$. Um nachzuweisen, daß $\psi \circ \varphi = \mathrm{id}$ (identische Abbildung) ist, benutzt man die Eindeutigkeitsaussage des Rekursionssatzes für $A = \mathbb{N}$, $a = 0$ und $g = S$: Sowohl $\psi \circ \varphi$ als auch id sind Abbildungen $\Phi: \mathbb{N} \to \mathbb{N}$, für die $\Phi(0) = 0$ und $\Phi \circ S = S \circ \Phi$ gilt. Also muß $\psi \circ \varphi = \mathrm{id}$ sein. Entsprechend folgt $\varphi \circ \psi = \mathrm{id}$. □

3. Addition, Multiplikation und Anordnung der natürlichen Zahlen. Für jede feste natürliche Zahl m wird die *Addition* $m + n$, ausgehend von $m + 0 = m$, durch die Rekursionsformel $m + S(n) = S(m + n)$ definiert. Hier wird also der Rekursionssatz für $A = \mathbb{N}$, $a = m$, $g = S$ und $\varphi(n) = m + n$ angewandt. Insbesondere gilt für $1 := S(0)$, daß $m + 1 = S(m)$ der Nachfolger ist.

Alle vertrauten *Rechenregeln der Addition* bedürfen nun eines Beweises. Wir beschränken uns auf den Nachweis des *Assoziativgesetzes* und verweisen für alle weiteren Regeln auf das klassische Werk von E. Landau [23], Kap. 1, § 2:

Satz. *Für alle $k, m, n \in \mathbb{N}$ gilt*: $(k + m) + n = k + (m + n)$.

Beweis. Induktionsbeginn bei $n = 0$: $(k + m) + 0 = k + m = k + (m + 0)$.

Induktionsschluß von n auf $n + 1$: $(k + m) + (n + 1) \overset{*}{=} ((k + m) + n) + 1 \overset{**}{=} (k + (m + n)) + 1 \overset{*}{=} k + ((m + n) + 1) \overset{*}{=} k + (m + (n + 1))$.

Bei $*$ wurde jeweils die Rekursionsformel der Addition, bei $**$ die Induktionsvoraussetzung benutzt. □

Man überzeugt sich in dieser Weise:

> $\mathbb{N}$ *ist bezüglich der Addition eine kommutative Halbgruppe mit Kürzungsregel.*

Die *Kürzungsregel* besagt, daß aus $n + k = m + k$ stets $n = m$ für alle $k, m, n \in \mathbb{N}$ folgt.

Analog wie die Addition wird die *Multiplikation* $m \cdot n$ für festes m, ausgehend von $m \cdot 0 = 0$, rekursiv durch $m \cdot (n + 1) = m \cdot n + m$ definiert. Alle vertrauten Rechenregeln der Multiplikation bedürfen wieder der Beweise, für die auf E. LANDAU [18], Kap. 1, § 4 verwiesen sei.

Eine Ordnungsrelation $\leqslant$ wird wie folgt auf $\mathbb{N}$ definiert: Es gilt $n \leqslant m$ genau dann, wenn es ein $t \in \mathbb{N}$ gibt mit $n + t = m$. Die üblichen Eigenschaften einer Ordnung sind erfüllt, das heißt, für alle $m, n, l \in \mathbb{N}$ gilt:

1) Reflexivität: $n \leqslant n$.
2) Antisymmetrie: Wenn $n \leqslant m$ und $m \leqslant n$ ist, dann gilt $m = n$.
3) Transitivität: Wenn $n \leqslant m$ und $m \leqslant l$ ist, dann gilt $n \leqslant l$.

Man schreibt $m < n$ genau dann, wenn $m \leqslant n$ und $m \neq n$ ist. Die Ordnung ist *linear* (oder *total*): Für alle $m, n \in \mathbb{N}$ ist $n \leqslant m$ oder $m < n$. Die Ordnung ist *monoton* bezüglich der Addition: Für alle $l, m, n \in \mathbb{N}$ folgt aus $m \leqslant n$, daß $m + l \leqslant n + l$ ist (entsprechend mit $<$ statt $\leqslant$). Analog gilt für die Multiplikation: Aus $m \leqslant n$ folgt $m \cdot l \leqslant n \cdot l$ (entsprechend mit $<$ statt $\leqslant$, falls $l \neq 0$ ist).

4. PEANOs Axiome. Nach G. PEANO (ital. Mathematiker, 1858 – 1932) beschreibt man die natürlichen Zahlen auch durch folgende Axiome für die Grundbegriffe $\mathbb{N}$, 0 und S:

(P1) $0 \in \mathbb{N}$.
(P2) Wenn $n \in \mathbb{N}$, dann $S(n) \in \mathbb{N}$.
(P3) Wenn $n \in \mathbb{N}$, dann $S(n) \neq 0$.
(P4) Wenn $0 \in E$ und wenn aus $n \in E$ stets $S(n) \in E$ folgt, ist $\mathbb{N} \subset E$.
(P5) Wenn $m, n \in \mathbb{N}$, folgt aus $S(m) = S(n)$, daß $m = n$ ist.

Wenn man (P1)–(P5) mengentheoretisch interpretiert, sind sie mit der Definition aus § 1.1 äquivalent. Im Unterschied zu DEDEKIND ging es PEANO jedoch nicht primär um eine mengentheoretische Konstruktion der natürlichen Zahlen, sondern um deren Axiomatisierung in einer formalen Sprache. In diesem Sinne sollte man beispielsweise (P4) folgendermaßen lesen: Wenn Null die Eigenschaft E hat und wenn aus der Tatsache, daß n die Eigenschaft E hat, stets folgt, daß der Nachfolger $S(n)$ die Eigenschaft E hat, dann folgt die Eigenschaft E aus der Eigenschaft $\mathbb{N}$, eine natürliche Zahl zu sein. Dieser Gesichtspunkt der formalen Sprache soll hier nicht weiter verfolgt werden. Er wird jedoch an späterer Stelle, beim Übergang von den Standard- zu den Non-Standard-Zahlen im 11. Kapitel, wichtig werden.

Historisch gab PEANO 1889 in den „Arithmetices principia nova methodo exposita" [24] neun Axiome (mit 1 als ausgezeichnetem Element) an. Über die Beziehungen seines Systems zu DEDEKINDS Definition schreibt er ([24], S. 22): „Utilius quoque mihi fuit recens scriptum: R. DEDEKIND, Was sind und was sollen die Zahlen, Braunschweig 1888, in quo quaestiones, quae ad numerorum fundamenta pertinent, acute examinatur."

§ 3. Ganze Zahlen

Die Subtraktion ist im Bereich der natürlichen Zahlen nicht unbeschränkt ausführbar. Nachdem man die negativen bzw. „falschen" (R. DESCARTES) ganzen Zahlen zunächst vorsichtig wie Wurzeln und imaginäre Zahlen als „fiktive" Rechenausdrücke behandelt hatte, bezeichnet L. KRONECKER im 19. Jahrhundert die ganzen Zahlen als den „naturgemäßen Ausgangspunkt für die Entwicklung des Zahlbegriffs" (vgl. J. TROPFKE [29], S. 126; L. KRONECKER [16]). Berühmt wurde KRONECKERS Ausspruch, wonach die ganzen Zahlen der liebe Gott gemacht habe, alles andere in der Mathematik aber Menschenwerk sei. Nach R. DEDEKIND waren aber bereits die positiven ganzen Zahlen nicht bloß „naturgegeben", sondern „freie Schöpfungen des menschlichen Geistes", nämlich mengentheoretische Begriffsbildungen. Algebraisch handelt es sich bei den natürlichen Zahlen um eine additive Halbgruppe, die bei den ganzen Zahlen zu einer Gruppe erweitert wird. Zentral wurde dabei der algebraische Begriff des Integritätsringes, der 1882 von L. KRONECKER [17] in seiner Arbeit „Grundzüge einer arithmetischen Theorie der algebraischen Größen" (§ 5) als sogenannter „Integritätsbereich" eingeführt wurde.

1. Die additive Gruppe $\mathbb{Z}$. Die systematische Einführung der ganzen Zahlen wird durch folgende Betrachtung motiviert: Jede ganze Zahl läßt sich als Differenz $a - b$ zweier natürlicher Zahlen a und b darstellen. Daher liegt es nahe, die ganze Zahl $a - b$ durch das Paar (a, b) zu beschreiben. Man muß allerdings beachten, daß auch andere Paare (c, d) dieselbe ganze Zahl $a - b = c - d$ beschreiben können, nämlich dann, wenn $a + d = b + c$ ist. Man geht deshalb so vor:

Auf $\mathbb{N} \times \mathbb{N}$ betrachtet man die *Relation*

$$(a, b) \sim (c, d) \qquad \text{genau dann, wenn} \qquad a + d = b + c.$$

Man weist nach, daß es sich um eine Äquivalenzrelation handelt. Zum Beispiel zeigt man die Transitivität folgendermaßen: Wenn $(a, b) \sim (c, d)$ und $(c, d) \sim (e, f)$ gelten, hat man $a + d = b + c$ und $c + f = d + e$. Man addiert $a + d + c + f = b + c + d + e$ und kürzt $c + d$ (Kommutativität und Assoziativität werden auch benutzt). Man erhält $a + f = b + e$, das heißt, $(a, b) \sim (e, f)$.

Die *ganzen Zahlen* werden nun als Äquivalenzklassen der Relation $\sim$ definiert. Die durch (a, b) repräsentierte Klasse $\{(x, y) : (x, y) \sim (a, b)\}$ wird mit $[a, b]$ bezeichnet. Genauer müßte man $[(a, b)]$ schreiben. Die Menge aller ganzen Zahlen (eine Menge von Äquivalenzklassen) wird mit $\mathbb{Z}$ bezeichnet.

Auf $\mathbb{N} \times \mathbb{N}$ kann man komponentenweise addieren: $(a, b) + (c, d) := (a + c, b + d)$. Dabei gelten das Kommutativ- und Assoziativgesetz, das Nullelement ist $(0, 0)$. Diese Addition ist mit der Relation $\sim$ verträglich, das heißt, wenn $(a', b') \sim (a, b)$ und $(c', d') \sim (c, d)$ ist, ist $(a' + c', b' + d') \sim (a + c, b + d)$. Daher ist es sinnvoll, auf $\mathbb{Z}$ die *Addition* $\mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$, $[a, b] + [c, d] := [a + c, b + d]$ einzuführen, die ebenfalls kommutativ und assoziativ ist und $[0, 0]$ als Nullelement hat. Durch den Übergang zu den Äquivalenzklassen (ganzen Zahlen) hat man aber mehr erreicht: Jede ganze Zahl $[a, b]$ besitzt eine inverse ganze Zahl, nämlich $[b, a]$, da $[a, b] + [b, a] = [a + b, a + b] = [0, 0]$ ist. Somit gilt der

Satz. *Die ganzen Zahlen bilden bezüglich der Addition eine kommutative Gruppe.*

Das zu $\alpha \in \mathbb{Z}$ inverse Element ist eindeutig bestimmt. Es wird mit $-\alpha$ bezeichnet. Durch $\alpha - \beta := \alpha + (-\beta)$ wird auf $\mathbb{Z}$ die *Subtraktion* eingeführt.

Die Abbildung $\iota: \mathbb{N} \to \mathbb{Z}$, $a \mapsto [a, 0]$, ist injektiv und mit der Addition verträglich. Man identifiziert üblicherweise $\mathbb{N}$ mit der dazu isomorphen Teilmenge $\iota(\mathbb{N}) \subset \mathbb{Z}$. Die ganze Zahl $[a, b]$ schreibt sich dann als $a - b$. Damit ist die zur Motivation benutzte Darstellung gerechtfertigt. Wenn man $\mathbb{N}^+ = \mathbb{N} \setminus \{0\}$ benutzt, kann man $\mathbb{Z} = -N^+ \cup \{0\} \cup N^+$ als disjunkte Vereinigung darstellen: Je nachdem, ob $a > b$, $a = b$ oder $a < b$ ist, liegt $[a, b] = a - b$ in $\mathbb{N}^+$, in $\{0\}$ oder in $-\mathbb{N}^+$.

Die Konstruktion der ganzen Zahlen ist algebraisch: Anstatt von $\mathbb{N}$ kann man von irgendeiner kommutativen Halbgruppe H ausgehen und dazu wie oben eine kommutative Gruppe G konstruieren. Wenn H die Kürzungsregel nicht erfüllt, muß man etwas modifizieren: Man definiert $(a, b) \sim (c, d)$ genau dann, wenn es ein e mit $a + d + e = b + c + e$ gibt. Allerdings ist dann $\iota: H \to G$ nicht mehr injektiv.

2. Der Integritätsring $\mathbb{Z}$. Die Differenzendarstellung der ganzen Zahlen motiviert die Definition ihrer Multiplikation. Wir wollen ja $(a - b) \cdot (c - d) = (ac + bd) - (ad + bc)$ rechnen und werden daher auf folgende Definition geführt:

$$[a, b] \cdot [c, d] = [ac + bd, ad + bc] \qquad \text{für} \qquad a, b, c, d \in \mathbb{N}.$$

Diese Definition ist unabhängig von der Wahl der repräsentierenden Paare.

Satz. *Die ganzen Zahlen bilden bezüglich ihrer Addition und Multiplikation einen Integritätsring (kommutativen, nullteilerfreien Ring mit Einselement).*

Übrigens ist $\mathbb{Z}$ der kleinste Integritätsring, der $\mathbb{N}$ umfaßt: Zu jedem Integritätsring $R \supset \mathbb{N}$ gibt es genau einen Monomorphismus (injektive, mit $+$ und $\cdot$ verträgliche Abbildung) $\varphi: \mathbb{Z} \to R$ mit $\varphi \mid \mathbb{N} =$ Inklusion von $\mathbb{N}$ in R.

3. Die Anordnung in $\mathbb{Z}$ wird definiert durch

$$a \leqslant b \qquad \text{genau dann, wenn} \qquad b - a \in \mathbb{N}.$$

Satz. *Der Ring $\mathbb{Z}$ der ganzen Zahlen wird durch $\leqslant$ linear (total) geordnet. Für alle $a, b, c \in \mathbb{Z}$ mit $a \leqslant b$ gilt $a + c \leqslant b + c$ und, falls $c > 0$, auch $a \cdot c \leqslant b \cdot c$.*

Die natürlichen Zahlen $\neq 0$ sind also die ganzen Zahlen > 0, die sogenannten *positiven* Zahlen. Man nennt a *negativ*, wenn $-a$ positiv ist.

Bemerkungen. Jeder kommutative Ring R, der sich als disjunkte Vereinigung $R = -P \cup \{0\} \cup P$ darstellen läßt, wobei P additiv und multiplikativ abgeschlossen ist, läßt sich durch „$a \leqslant b$ genau dann, wenn $b - a \in P \cup \{0\}$“ total anordnen.

Historisch führte auch R. Dedekind die ganzen Zahlen durch Zahlenpaare aus $\mathbb{N} \times \mathbb{N}$ ein. In einem Brief des 82jährigen (also 1913) an einen ehemaligen Studenten liefert

DEDEKIND ([10], S. 490) eine „Erweiterung des Reiches N der natürlichen Zahlen zu dem Bereich G der ganzen rationalen Zahlen." E. LANDAU [18] konstruiert aus $\mathbb{N}$ erst die rationalen Zahlen $\geqslant 0$, ergänzt sie durch die negativen rationalen Zahlen zum Körper $\mathbb{Q}$ (vgl. § 4) und erhält $\mathbb{Z}$ als Unterring von $\mathbb{Q}$.

§ 4. Rationale Zahlen

1. Historisches. Die Division als Umkehrung der Multiplikation ist im Bereich der ganzen Zahlen nicht unbeschränkt ausführbar. Brüche, die diese Division immer möglich machen, werden schon in früher Zeit betrachtet. Sie waren nie so von Geheimnissen umwittert wie die negativen Zahlen, die man sich unterhalb von „Nichts" vorstellte, oder wie die irrationalen und imaginären Zahlen, von denen noch zu berichten sein wird. Die erste systematische Darstellung findet man in Buch VII der „Elemente" EUKLIDS, das von den Verhältnissen natürlicher Zahlen handelt. Die uns geläufige Vorstellung, Zahlenverhältnisse als Brüche zu deuten und durch sie den Bereich der ganzen Zahlen zu erweitern, entsteht erst in der Neuzeit. Die ersten theoretischen Abhandlungen stammen aus dem 19. Jahrhundert.

B. BOLZANO [5] entwickelt in einer im Nachlaß entdeckten „Reinen Zahlenlehre" eine Theorie der rationalen Zahlen, und zwar als Theorie derjenigen Zahlenmenge, die gegenüber den *vier elementaren Rechenoperationen* abgeschlossen ist. Auch in einer Abhandlung von M. OHM [22] (einem Bruder des bekannten Physikers) finden wir die Absicht, die rationalen Zahlen durch „die einzigen Grundwahrheiten für die Addition, Subtraktion, Multiplikation und Division" zu bestimmen.

In den Vordergrund tritt also die Untersuchung der Eigenschaften bestimmter Verknüpfungen – und nicht die Frage nach dem „Wesen" der Zahl. Bei H. HANKEL ([13], S. 2) heißt es schließlich 1867 in seiner „Theorie der complexen Zahlensysteme", daß die Gesetze dieser Operationen „das System der Bedingungen" bestimmen, „welche nöthig und ausreichend sind, um die Operation formal zu definieren." Außer bei den rationalen Zahlen wird der Körperbegriff der Sache nach (wenn auch nicht unter diesem Namen) bereits bei N. H. ABEL und E. GALOIS diskutiert, wenn z. B. zum System der rationalen Zahlen eine Wurzel adjungiert wird und die möglichen Ausdrücke aus dieser und den rationalen Zahlen mit Addition, Subtraktion, Multiplikation und Division untersucht werden. L. KRONECKER spricht 1853 in seiner Theorie algebraischer Größen von „Rationalitätsbereichen" (L. KRONECKER [17], § 1), R. DEDEKIND zunächst von „rationalen Gebieten", schließlich 1871 von „Körpern" bei reellen und komplexen Zahlen (R. DEDEKIND [12], S. 224). Zahlenkörper werden auch von H. WEBER [30] und D. HILBERT [14] untersucht. E. STEINITZ [26] gibt 1910 eine abstrakte Definition dieses algebraischen Grundbegriffs an. STEINITZ stellt auch heraus, daß hinter der Erweiterung der ganzen zu den rationalen Zahlen eine allgemeine algebraische Konstruktion steht, nämlich die Einbettung eines Integritätsringes in einen Körper durch Quotientenbildung.

2. Der Körper $\mathbb{Q}$. Wie H. WEBER in seinem „Lehrbuch der Algebra" von 1895 führen wir die Brüche als Äquivalenzklassen ganzer Zahlen ein, und zwar gehen wir, motiviert durch

$$\frac{a}{b} = \frac{c}{d} \qquad \text{genau dann, wenn} \qquad ad = bc,$$

von der Äquivalenzrelation

$$„(a,b) \sim (c,d) \qquad \text{genau dann, wenn} \qquad ad = bc"$$

auf $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ aus. Die Äquivalenzklassen werden *Brüche* oder *rationale Zahlen* genannt. Der durch (a,b) repräsentierte Bruch wird mit $\frac{a}{b}$ bezeichnet. Die *Addition* und *Multiplikation* von Brüchen wird durch die bekannten Beziehungen

$$\frac{a}{b} + \frac{c}{d} := \frac{ad + bc}{bd} \qquad \text{und} \qquad \frac{a}{b} \cdot \frac{c}{d} := \frac{ac}{bd}$$

definiert. Diese Definitionen sind von der Wahl der Repräsentanten unabhängig. Bei E. LANDAU [18], Kap. 2, § 3–4 wird ausführlich bewiesen:

Satz. *Die Menge $\mathbb{Q}$ der rationalen Zahlen zusammen mit der oben definierten Addition und Multiplikation ist ein Körper.*

Durch $\iota: \mathbb{Z} \to \mathbb{Q}$, $a \mapsto \frac{a}{1}$, wird $\mathbb{Z}$ auf den Unterring $\iota(\mathbb{Z}) \subset \mathbb{Q}$ isomorph abgebildet. Üblicherweise wird $\mathbb{Z}$ mit $\iota(\mathbb{Z})$ identifiziert. Der Körper $\mathbb{Q}$ ist der kleinste Körper, der $\mathbb{Z}$ als Unterring enthält.

3. Die Anordnung in $\mathbb{Q}$. Ein Bruch $\frac{a}{b}$ heißt positiv, wenn a und b beide positiv oder beide negativ sind. Die Menge P der positiven Brüche ist bezüglich $+$ und $\cdot$ abgeschlossen. Man hat die Darstellung $\mathbb{Q} = -P \cup \{0\} \cup P$ als disjunkte Vereinigung. Wie in der Bemerkung in 3.3. ist dann durch „$r \leqslant s$ genau dann, wenn $s - r \in P \cup \{0\}$" eine totale Anordnung auf $\mathbb{Q}$ definiert, die auf $\mathbb{Z}$ mit der in 3.3. definierten Anordnung übereinstimmt.

Die Ordnungsrelation in $\mathbb{Q}$ ist *archimedisch*, das heißt, für alle positiven rationalen Zahlen $r, s \in \mathbb{Q}$ existiert eine natürliche Zahl n mit $s < n \cdot r$. Zum Beweis schreiben wir $s = p/h$ und $r = q/h$ als Brüche natürlicher Zahlen mit einem gemeinsamen Nenner h. Die Behauptung folgt dann, sobald $p < n \cdot q$ für natürliche Zahlen > 0 bewiesen ist. Letzteres zeigt man bei festen $q \geqslant 1$ durch Induktion über $p = 1, 2, \ldots$. Eine bemerkenswerte Eigenschaft, welche den geordneten Körper $\mathbb{Q}$ vom geordneten Integritätsring $\mathbb{Z}$ unterscheidet, ist seine *Dichte*: Für alle $r, s \in \mathbb{Q}$ mit $r < s$ läßt sich immer ein $t \in \mathbb{Q}$ angeben mit $r < t < s$. Man wähle z. B. $t := \frac{1}{2}(r + s)$ als arithmetisches Mittel.

Literatur

[1] ARISTOTELES: Metaphysik, Aristotelis Opera, ed. I. Bekker, Berlin 1831, repr. Darmstadt 1960

[2] Aristoteles: Physik, Aristotelis Opera, ed. I. Bekker, Berlin 1831, repr. Darmstadt 1960
[3] Becker, O.: Grundlagen der Mathem. in geschichtlicher Entwicklung, Freiburg/München 1954, ^{2}1964, Frankfurt 1975
[4] Bolzano, B.: Paradoxien des Unendlichen, ed. F. Přihonský, Leipzig 1851, Berlin ^{2}1889, ed. A. Höfler, Leipzig 1920, mit Einl., Anm., Reg. u. Bibliographie ed. B. van Rootselaar, Hamburg 1955, ^{2}1975
[5] Bolzano, B.: Reine Zahlenlehre, in: B. Bolzano – Gesamtausgabe (eds. E. Winter, J. Berg, F. Kambartel, J. Loužil, B. v. Rootselaar), Reihe II Nachlaß, A. Nachgelassene Schriften, Bd. 8 Größenlehre II, Reine Zahlenlehre, Stuttgart/Bad Cannstatt 1976
[6] Bourbaki, N.: Eléments de mathématique, Paris, seit 1939
[7] Bruins, E. M., Rutten, M.: Textes mathématiques de Suse, Mémoires de la Mission Archéologique en Iran, Tome 34, Paris 1961
[8] Cantor, G.: Gesam. Abh. mathem. u. philos. Inhalts, Berlin 1932, repr. Berlin 1980
[9] Dedekind, R.: Was sind und was sollen die Zahlen? Braunschweig 1888, ^{10}1965, repr. 1969
[10] Dedekind, R.: Mathem. Werke Bd. 3, Braunschweig 1932, repr. New York 1969
[11] Frege, G.: Die Grundlagen der Arithmetik. Eine logisch mathematische Untersuchung über den Begriff der Zahl, Breslau 1884, repr. Darmstadt/Hildesheim 1961
[12] Gauss, C. F.: Werke Bd. 3, Göttingen 1876
[13] Hankel, H.: Theorie der complexen Zahlensysteme, Leipzig 1867
[14] Hilbert, D.: Über den Zahlbegriff, in: Jahresber. d. Deutschen Math. Verein. 1900, 180–184
[15] Juschkewitsch, A. P.: Geschichte der Mathematik im Mittelalter, dt. Leipzig 1964
[16] Kronecker, L.: Über den Zahlbegriff, in: Journ. f. d. reine u. angew. Mathem. 101 1887, 339, in: Math. Werke Bd. 3, Leipzig 1899/1931, repr. New York 1968, 249–274
[17] Kronecker, L.: Grundzüge einer arithmetischen Theorie der algebraischen Größen, in: J. f. d. reine u. angew. Mathem. 1882, 1–122, in: Math. Werke Bd. 2, Leipzig 1897, repr. New York 1968, 237–387
[18] Landau, E.: Grundlagen der Analysis, Leipzig 1930, repr. Darmstadt 1963
[19] Lepsius, R.: Über eine Hieroglyphische Inschrift am Tempel in Edfu, in: Abh. d. Kgl. Akad. d. Wiss., Berlin 1855, 69–111
[20] Neugebauer, O.: Mathem. Keilschrifttexte, Quellen u. Studien A3, Berlin I, II 1935, III 1937
[21] Neumann, J. v.: Zur Einführung der transfiniten Zahlen, in: Acta Szeged 1 1923, 199–202, repr. in: *A. H. Taub* (ed.), Collected Works, Oxford/London/Paris Bd. 1 1961, 24–33
[33] Ohm, M.: Die reine Elementarmathematik Bd. I, Berlin ^{2}1834
[23] Papyrus Rhind, (Hrsg. A. Eisenlohr) Leipzig 1877; A. B. Chace, The Rhind Mathem. Papyrus, Oberlin I 1927, II 1929
[24] Peano, G.: Arithmetices principia nova exposita, in: Opere scelte Bd. II, Rom 1958, 20–55
[25] Russell, B.: The principles of mathematics, London 1903, ^{7}1956
[26] Steinitz, E.: Algebraische Theorie der Körper, in: J. f. d. reine u. angew. Math. 137 1910, 167–309
[27] Stifel, M.: Arithmetica integra, Nürnberg 1544
[28] Struwe, W. W.: Papyrus des staatl. Museums der schönen Künste in Moskau, Quellen u. Studien A1 1930
[29] Tropfke, J.: Geschichte der Elementarmathematik, Bd. 1 Arithmetik und Algebra, vollst. neu bearb. von H. Gericke, K. Reich u. K. Vogel, Berlin ^{4}1980
[30] Weber, H.: Lehrbuch der Algebra Bd. 1 1895, repr. der 3. Aufl. New York 1961

Kapitel 2. Reelle Zahlen

K. Mainzer

> *λέγω δ' εἶναι συνεχὲς ὅταν ταὐτὸ γένηται καὶ ἓν τὸ ἑκατέρου πέρας οἷς ἅπτονται, καὶ ὥσπερ σημαίνει τοὔνομα, συνέχηται**)
> (ARISTOTELES, Physik 227a, 11–12).
>
> Continuum est totum, cuius duae quaevis partes cointegrantes (seu quae simul sumtae toti coincidunt) habent aliquid commune, ... saltem habent communem terminum**) (G. W. LEIBNIZ, Mathem. Schr. VII, 284).
>
> Zerfallen alle Punkte der Geraden in zwei Klassen von der Art, daß jeder Punkt der ersten Klasse links von jedem Punkt der zweiten Klasse liegt, so existiert ein und nur ein Punkt, welcher diese Einteilung aller Punkte in zwei Klassen, diese Zerschneidung der Geraden in zwei Stücke, hervorbringt (R. DEDEKIND, Stetigkeit und irrationale Zahlen, Braunschweig 1872, 10).

§ 1. Historisches

1. HIPPASUS und das Pentagon. Wenn wir heute die reellen Zahlen als Elemente eines vollständig geordneten Körpers definieren, so ist uns nicht mehr gegenwärtig, wie sehr die Entdeckung, daß sich nicht alles durch rationale Zahlen erfassen läßt, einst Bildungs- und Weltanschauungskrisen auslöste – ja, wenn man späteren Legenden trauen darf – ihrem Entdecker die Strafe der Götter einbrachte. Gemeint ist die *Entdeckung inkommensurabler Streckenverhältnisse* vermutlich durch den Pythagoreer HIPPASUS VON METAPONT im 5. vorchristlichen Jahrhundert, die in pythagoreischen Kreisen einen Schock ausgelöst haben soll. Schließlich stellte diese Entdeckung die Annahme in Frage, auf der die Philosophie der *Pythagoreer* zunächst beruht hatte, nämlich daß alle Dinge in ganzen Zahlen ausgedrückt werden könnten. (Zur Quellenlage über die Entdeckung der Inkommensurabilität und die Person des HIPPASUS vgl. K. v. Fritz [10]; S. Heller [11]; J. Tropfke [23], S. 132; B. L. van der Waerden [25]). Um die Auswirkungen dieser Krise zu verstehen, muß berücksichtigt werden, daß die Pythagoreer nicht nur als eine einflußreiche mathematische Schule wirkten, die als erste die Forderung nach exakter mathematischer Wissenschaft erhoben und von ihren Mitgliedern eine strenge Ausbildung in Arithmetik, Geometrie, Astronomie und Musik verlangten. Sie verpflichteten sich außerdem zu einem ordensmäßig geregelten Lebenswandel und beherrschten im zweiten Viertel des 5. Jahrhunderts bis zum Aufstand von 445 ganz Unteritalien. In diesen politischen Wirren spielte HIPPASUS vermutlich eine wichtige Rolle (vgl. Iamblichus [14], S. 77, 6f.; ebenso K. v. Fritz [10], S. Heller

---

*) Ich sage aber, es sei etwas zusammenhängend, wenn allemal die Grenze eines jeden von zwei [Teilen], die einander berühren, ein und dieselbe wird, und, wie der Name bedeutet, es zusammenhängt.

**) Zusammenhängend ist alles, bei dem zwei beliebige Teile, die das Ganze ausmachen, (sei es nun, daß sie in gleicher Weise genommen ganz übereinstimmen) etwas gemeinsam haben, ... einen gemeinsamen Sprung als Grenze haben.

[11]). Die Betrachtung der Streckenverhältnisse ging von einer seit altersher geübten Meßpraxis aus. Danach wurde eine Strecke a gemessen, indem eine Maßeinheit e auf der Strecke m-mal hintereinander angelegt wurde:

$$a = \underbrace{e + \cdots + e}_{m\text{-mal}} = m \cdot e.$$

Zwei Strecken a_0 und a_1 heißen *kommensurabel*, wenn sie in diesem Sinne mit derselben Maßeinheit e gemessen werden können: $a_0 = m \cdot e$ und $a_1 = n \cdot e$ für zwei natürliche Zahlen m und n. In diesem Falle ist das Streckenverhältnis $a_0 : a_1$ ein Verhältnis $m : n$ natürlicher Zahlen. Die Methode, ein gemeinsames Maß zweier Strecken a_0 und a_1 zu bestimmen, wurde bereits vor aller griechischen Philosophie und Wissenschaft von den Handwerkern als Verfahren der Wechselwegnahme ausgeübt. EUKLID hat es in den „Elementen" durch den heute nach ihm benannten Algorithmus beschrieben: Man trage die kleinere Strecke a_1 auf der größeren a_0 so oft wie möglich ab. Der Rest sei a_2, das heißt

$$a_0 = n_1 a_1 + a_2 \qquad \text{mit} \qquad a_2 < a_1.$$

Dann fahre man entsprechend fort:

$$\begin{aligned} a_1 &= n_2 a_2 + a_3 \qquad \text{mit} \qquad a_3 < a_2, \\ a_2 &= n_3 a_3 + a_4 \qquad \text{mit} \qquad a_4 < a_3, \\ &\vdots \end{aligned}$$

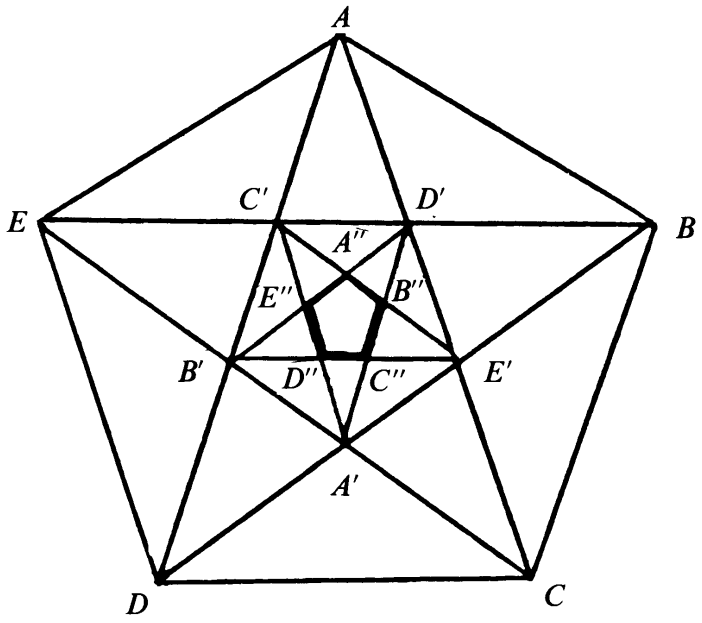
Wenn a_0 und a_1 ein gemeinsames Maß besitzen, bricht dieses Verfahren nach endlich vielen Schritten ab, das heißt, es gibt ein k mit $a_{k-1} = n_k a_k$, und a_k ist ein gemeinsames Maß von a_0 und a_1.

Anschaulich war man vielleicht zunächst davon überzeugt, daß das Verfahren der „Wechselwegnahme" immer abbricht und daher ein gemeinsames Maß immer vorhanden ist. Modern gesprochen zeigt dieses Verfahren allerdings nur, daß sich jedes Streckenverhältnis in einem *Kettenbruch*

$$\begin{aligned} a_0 : a_1 &= n_1 + a_2 : a_1 \\ &= n_1 + \frac{1}{a_1 : a_2} = n_1 + \frac{1}{n_2 + a_3 : a_2} \\ &= n_1 + \frac{1}{n_2 + \dfrac{1}{a_2 : a_3}} = \cdots = n_1 + \frac{1}{n_2 + \dfrac{1}{n_3 + \cdots}} \end{aligned}$$

entwickeln läßt, der endlich ist, wenn a_0 und a_1 kommensurabel sind.

Das Ordenssymbol der Pythagoreer war das *Pentagramm*, das seine magische Wirkung noch in der mittelalterlichen Astrologie behielt und mit dem Faust den Mephisto gebannt haben soll. Vieles spricht dafür, daß Hippasus ausgerechnet an diesem Symbol feststellte, daß zwei Strecken nicht kommensurabel sind (vgl. Iamblichus [15], S. 132, 11–12; zur Quellenlage vgl. K. v. Fritz [10], S. Heller [11], J. Tropfke [23].) Betrachten wir zunächst das reguläre Pentagon $ABCDE$, in das sämtliche Diagonalen eingezeichnet sind:



Die Diagonalen erzeugen in der Mitte ein kleineres reguläres Fünfeck A', B', C', D', E'. Je eine Seite und eine Diagonale am regulären Fünfeck sind aus Symmetriegründen parallel. Die Dreiecke AED und $BE'C$ haben daher parallele Seiten, sind also ähnlich. Somit ist $AD : AE = BC : BE'$. Es gilt $BE' = BD - BC$, denn $BC = AE = DE'$, da EA und DB bzw. DE und AC parallel sind. Am Pentagon gilt also:

$$\text{Diagonale} : \text{Seite} = \text{Seite} : (\text{Diagonale} - \text{Seite}).$$

Bezeichnen wir die Diagonale mit a_0, die Seite mit a_1 und ihre Differenz mit $a_2 = a_0 - a_1$, so ist $a_0 : a_1 = a_1 : a_2$, insbesondere $a_2 < a_1$. Bildet man wieder die Differenz $a_3 = a_1 - a_2$, so erhält man dieselbe Verhältnisgleichung $a_1 : a_2 = a_2 : a_3$, insbesondere $a_3 < a_2$. Man kann das Verfahren unendlich fortsetzen:

$$a_2 = a_0 - a_1, \qquad a_3 = a_1 - a_2, \qquad a_4 = a_2 - a_3 \cdots$$

$$a_0 : a_1 = a_1 : a_2 = a_2 : a_3 = a_3 : a_4 = \cdots .$$

Der Euklidische Algorithmus für a_0 und a_1

$$\begin{aligned} a_0 &= 1 \cdot a_1 + a_2, \\ a_1 &= 1 \cdot a_2 + a_3, \\ a_2 &= 1 \cdot a_3 + a_4 \\ &\cdots\cdots \end{aligned}$$

bricht nicht ab, Seite a_1 und Diagonale a_0 des Pentagons sind also nicht kommensurabel.

Als Kettenbruch erhält man

$$a_0 : a_1 = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}}}$$

Aus $a_0 : a_1 = a_1 : (a_0 - a_1)$ folgt $a_0 : a_1 = \frac{1}{2}(1 + \sqrt{5})$. Man nennt dieses Verhältnis den *goldenen Schnitt*. Anschaulich erkennt man den unendlichen Euklidischen Algorithmus an den unendlich vielen Fünfecken, die, wie in der Figur angedeutet ist, ineinander geschachtelt sind. Die Seiten dieser Fünfecke sind $a_1, a_3, a_5, \ldots$ und ihre Diagonalen $a_0, a_2, a_4, \ldots$.

2. EUDOXOS und die Proportionenlehre. Die Babylonier rechneten zwar mit endlichen Näherungswerten für *irrationale* (inkommensurable) Verhältnisse, z. B. den Sexagesimalbrüchen 1; 25 und 1; 24, 51, 10 für $\sqrt{2}$. Aber die grundsätzliche Erkenntnis, daß $\sqrt{2}$, das Verhältnis von Diagonale und Seite eines Quadrates, inkommensurabel ist, verdankt man erst der griechischen Mathematik. In EUKLIDS „Elementen" X, § 115a, findet man folgenden Beweis: Es sei a die Seite und d die Diagonale eines Quadrates. Wenn sie kommensurabel wären, müßte herauskommen, daß dieselbe Zahl gerade und ungerade wäre: Offenbar ist $d^2 = 2a^2$. Da d und a als kommensurabel angenommen werden, wäre $d : a = m : n$ das Verhältnis zweier (natürlicher) Zahlen; hier seien m und n die kleinstmöglichen Zahlen. Es wäre auch $d^2 : a^2 = m^2 : n^2$. Aber $d^2 = 2a^2$, also wäre auch $m^2 = 2n^2$, also m^2 gerade. Folglich wäre m selbst gerade: $m = 2l$. Da m und n die kleinstmöglichen Zahlen sind, wären sie teilerfremd, also n ungerade. Da $m = 2l$, wäre $m^2 = 4l^2$. Nun war $m^2 = 2n^2$, also wäre $n^2 = 2l^2$ eine gerade Zahl und damit n gerade.

Die Erkenntnis der Irrationalität ist allerdings älter als EUKLID. Nach dem Zeugnis PLATONS (Theaetet 147d) wurde die Irrationalität einzelner Quadratwurzeln wie $\sqrt{3}, \sqrt{5}, \ldots, \sqrt{17}$ bereits von THEODOROS VON KYRENE gezeigt. In den „Nomoi" (819–820) begeistert er sich für das Problem der Inkommensurabilität: „Ihr wackren Hellenen, das ist eins von den Dingen, davon gesagt wird, es sei eine Schande, wenn man's nicht wisse und wenn man das Notwendige weiß, ist's erst noch keine sonderliche Ehre".

Entscheidend für den Fortschritt der griechischen Mathematik war die ausgeprägte Logik. Die Schlußform „reductio ad absurdum" (Beweis durch Widerspruch) erlaubte die ersten Unmöglichkeitsbeweise und die ersten exakten Aussagen über das „Unendliche." Nach HERMANN WEYL wird die Mathematik bei den Griechen erstmals zur „Wissenschaft vom Unendlichen".

Es ist die geniale Leistung des EUDOXOS VON KNIDOS, des Zeitgenossen und Bekannten PLATONS, eine geometrische Proportionenlehre auch für inkommensurable Größenverhältnisse geschaffen zu haben. Uns ist diese Lehre im Buch V der „Elemente" EUKLIDS überliefert. EUDOXOS geht von (positiven) Größen derselben Art aus (z. B. Strecken $a, b, \ldots$ oder Flächen $A, B, \ldots$). Er nimmt an, daß gleichartige Größen addiert werden können, wobei das Assoziativ- und Kommutativgesetz stillschweigend vorausgesetzt werden. Die Größen gleicher Art werden geordnet: $a < b$, wenn es ein c mit $a + c = b$ gibt. Es wird angenommen, daß für $a \neq b$ entweder $a < b$ oder $b < a$ gilt. Ganzzahlige Vielfache werden durch wiederholte Addition definiert: $m \cdot a = a + \cdots + a$ (m Summanden). Das heute meist nach ARCHIMEDES benannte Axiom wird vorausgesetzt: Zu jedem a und b gibt es eine natürliche Zahl n mit $a < n \cdot b$. Damit sind unendlich kleine Größen ausgeschlossen. (Diese doch zuzulassen, blieb einer späteren Zeit vorbehalten, siehe dazu das Kap. 11.)

Es werden *Verhältnisse* gleichartiger Größen (Streckenverhältnisse, Flächenverhältnisse usw.) betrachtet, die nicht kommensurabel sein müssen. Um solche

Verhältnisse miteinander zu vergleichen, wird definiert (EUKLIDS „Elemente" V Definition 5): „Man sagt, daß Größen in demselben Verhältnis stehen, die erste zur zweiten wie die dritte zur vierten, wenn bei beliebiger Vervielfältigung die Gleichvielfachen der ersten und dritten den Gleichvielfachen der zweiten und vierten gegenüber, paarweise entsprechend genommen, entweder zugleich größer oder zugleich gleich oder zugleich kleiner sind." Modern ausgedrückt bedeutet dies: Man definiert $a:b = A:B$, wenn für alle natürlichen Zahlen n und m gilt: $n \cdot a > m \cdot b$ genau dann, wenn $n \cdot A > m \cdot B$, $n \cdot a = m \cdot b$ genau dann, wenn $n \cdot A = m \cdot B$, $n \cdot a < m \cdot b$ genau dann, wenn $n \cdot A < m \cdot B$.

Zahlreiche Sätze der Proportionenlehre können wir heute als Rechengesetze für reelle Zahlen deuten. Man muß allerdings beachten, daß die Griechen nicht einmal rationale, geschweige denn irrationale Verhältnisse als Erweiterungen des Bereichs der natürlichen Zahlen auffaßten, sondern als Begriffe eigener Art ansahen. Das Ziel der Proportionenlehre sind geometrische Ergebnisse, wie beispielsweise die exakte Begründung zahlreicher Formeln zur Flächen- und Inhaltsberechnung. Die geometrischen Beweise dafür, die meist durch Widerspruch geführt werden, mögen uns, die wir elegante Kalküle kennen, umständlich erscheinen. Aber erst im 19. Jahrhundert gelang es, erfolgreiche Kalküle, die vor allem seit Beginn der Neuzeit entwickelt wurden, mit der in der griechischen Mathematik üblichen Strenge zu begründen.

3. Irrationalzahlen in der neuzeitlichen Mathematik. Nach der geometrischen Proportionenlehre der Griechen wird für die neuzeitliche Entwicklung der arithmetische Aspekt wichtig. Er geht auf das praktische Berechnen von Näherungswerten zurück, wie es von den an Technik und Astronomie interessierten Mathematikern bereits früh geübt wurde: Nach den Babyloniern ist besonders an ARCHIMEDES zu erinnern, der π bei der Kreisumfangsbestimmung zwischen $3\frac{1}{7}$ und $3\frac{10}{71}$ einschließt, und an PTOLEMAIOS (ca. 150 n. Chr.), dem großen Astronomen der antiken und mittelalterlichen Welt, der den Sexagesimalbruch 3;8,30 als Mittelwert von $3\frac{1}{7} = 3;8{,}34$ und $3\frac{10}{71} = 3;8{,}27$ wählt. Das Verfahren der *Intervallschachtelung* wird hier angewendet. Während jedoch in der griechischen Mathematik das Interesse am Rechnen mit Zahlen zugunsten von logischer Beweisführung und geometrischer Konstruktion im Hintergrund stand, erhält die Entwicklung des Zahlbegriffs durch den Einfluß der indisch-arabischen Algebra einen entscheidenden Auftrieb. So rechnet der arabische Mathematiker ABŪ KĀMIL (ca. 850–930) mit Quadratwurzelausdrücken u. a. nach der Regel

$$\sqrt{p} + \sqrt{q} = \sqrt{p + q + 2 \cdot \sqrt{pq}}$$

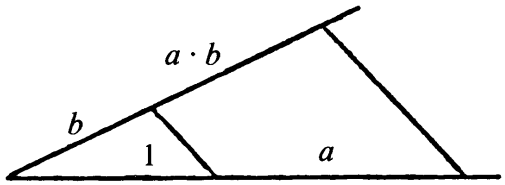
(Tropfke [23], S. 135): Man beginnt mit neuen Ausdrücken zu rechnen, ohne sie schon als neue Zahlen zu begreifen. Durch die Lösungsformeln für Gleichungen 3. und 4. Grades, die im 16. Jahrhundert entdeckt wurden, erhielt dieses Vorgehen großen Auftrieb. Mehr darüber findet der Leser in Kap. 3, § 1.

M. STIFEL [22] schreibt noch in seiner „Arithmetica integra" von 1544: „So wie eine unendliche Zahl keine Zahl ist, so ist eine irrationale Zahl keine wahre Zahl, weil sie sozusagen unter einem Nebel der Unendlichkeit verborgen ist."

Diesen „Nebel der Unendlichkeit" präzisiert S. STEVIN (1548–1620) bereits als unendliche Folge von Dezimalbrüchen, die er durch Intervallschachtelung bei der

Lösungsapproximation von z. B. $x^3 = 300x + 33\,900\,000$ entwickelt: „Et procedant ainsi infiniment, l'on approche infiniment plus pres au requis" (S. STEVIN [21], S. 353). STEVIN verwendet hier den Zwischenwertsatz, den allerdings erst B. BOLZANO ausdrücklich formuliert und bewiesen hat, wenn man von kurzen Erwähnungen bei LEIBNIZ und EULER absieht.

In R. DESCARTES „Géométrie" von 1637 werden die Operationen der Addition, Subtraktion, Multiplikation, Division und des Wurzelziehens mit Strecken so eingeführt, daß das Ergebnis wieder eine Strecke ergibt. Während nämlich bisher das Produkt zweier Strecken als Rechteck interpretiert wurde, erhält DESCARTES das Ergebnis bei festgewählter Einheitsstrecke als 4. Proportionale nach dem Strahlensatz:



Einen neuen Schub erfährt die Entwicklung des Zahlbegriffs durch die *Infinitesimalrechnung* im 17. und 18. Jahrhundert. Hier liefert insbesondere die Theorie der Reihen seit LEIBNIZ und den Gebrüdern BERNOULLI eine neue Möglichkeit der Zahlendarstellung. Bereits in der „Arithmetica infinitorum" (1655) des J. WALLIS (1616–1703), dem großen englischen Mathematiker vor NEWTON, findet sich z. B. eine unendliche Produktenentwicklung $\frac{\pi}{2} = \frac{2}{1} \cdot \frac{2}{3} \cdot \frac{4}{3} \cdot \frac{4}{5} \cdot \frac{6}{5} \cdot \frac{6}{7} \cdot \ldots$.

Zahlendarstellungen durch unendliche Summen bzw. Produkte wurden jedoch nicht – wie seit CAUCHY und WEIERSTRASS üblich – als konvergierende Folgen mit dem Grenzwertbegriff definiert. Man sagte vielmehr, daß sich z. B.

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)}$$

von 1 um eine „infinitesimal kleine" Größe unterscheidet. L. EULER [9] formuliert 1734 ein Konvergenzkriterium für Reihen in der Sprache der infinitesimalen Größen. Neben den „endlichen" und „wirklichen" (reellen) Zahlen, die als Meßwerte Anwendung fanden, schien es also noch „infinitesimale" und „ideale" Zahlen zu geben. Im 19. Jahrhundert wurden sie jedoch als ungenaue und psychologisierende Redeweisen aus der Mathematik verbannt und nach Einführung des Grenzwertbegriffs als überflüssig empfunden. Erst in der Non-Standard-Analysis (vgl. Kap. 11) kamen die infinitesimal kleinen Zahlen wieder zu neuen Ehren.

4. Präzisierungen des 19. Jahrhunderts. CAUCHY formuliert im „Cours d'Analyse" (1821) das nach ihm benannte Konvergenzkriterium und setzt es mit den bekannten Rechengesetzen als evidente Eigenschaft der reellen Zahlen voraus. Die hier zum Ausdruck kommende Vollständigkeit der reellen Zahlen wurde jedoch auch vor CAUCHY vorausgesetzt. So nahm z. B. G. W. LEIBNIZ für eine stetige Linie, die auf einer Fläche zum Teil innerhalb und zum Teil außerhalb eines Teils der Fläche liegt, an, daß sie den Rand dieses Flächenstücks schneidet.

Unter Voraussetzung des Cauchy-Kriteriums beweist B. BOLZANO [4] 1817 den *Zwischenwertsatz*. Dabei verfügt er bereits vor CAUCHY über das genannte Konvergenzkriterium, das BOLZANO in einer vor einigen Jahren in seinem Nachlaß entdeckten „Größenlehre" durch Intervallfolgen näher zu begründen sucht. Mit K. WEIERSTRASS werden die Überlegungen zur Begründung der reellen Zahlen in die mathematischen Grundvorlesungen aufgenommen. Allerdings sind uns davon nur von WEIERSTRASS zum Teil kritisch beurteilte Schülernachschriften überliefert. Die zentrale Vorstellung vom Begriff der reellen Zahl kommt für WEIERSTRASS [24] im Intervallschachtelungsprinzip zum Ausdruck, das er auch voraussetzt, um seinen bekannten Satz vom Häufungspunkt zu beweisen (vgl. auch P. DUGAC [8]). Eine systematische Definition der reellen Zahlen durch Intervallschachtelung wird 1892 von P. BACHMANN [1] angegeben.

Mit G. CANTORS Theorie der *Fundamentalfolgen* entstand eine weitere Definition der reellen Zahl (vgl. 2). Kurz vorher hatte C. MERAY (1835–1911), aber ohne Wissen von CANTOR, diesen Ansatz zur Definition der irrationalen Zahlen als „fiktive" Grenzwerte konvergenter Folgen bzw. – mit Blick auf die antike Entdeckung – als „nombres incommensurables" verwendet.

Die Proportionenlehre des EUDOXOS wird schließlich von R. DEDEKIND (1831–1916) in seiner berühmten Schrift „*Stetigkeit und Irrationalzahlen*" [7] von 1872 erneut und mit vorbildlicher Schärfe aufgegriffen. Die Dedekindsche Definition bringt die seit der Antike tief verwurzelte geometrische Anschauung vom Kontinuum zum Ausdruck, daß die Punkte der Geraden durch „Zerschneidung der Geraden in zwei Teile" (DEDEKIND), „durch die gemeinsame Grenze zweier Teile, die das Ganze ausmachen" (LEIBNIZ) oder „durch die Grenze zweier Stücke, die sich berühren" (ARISTOTELES), bestimmt sind (vgl. § 1). Die Frage, ob EUDOXOS bzw. EUKLID mit ihrer *Proportionenlehre* die *Theorie der Irrationalzahlen* erledigt haben, hat noch im Anschluß an DEDEKINDS Werk von 1872 zu einer *Kontroverse* geführt. So schreibt 1876 der Mathematiker R. LIPSCHITZ an DEDEKIND: „... Ich kann nur sagen, daß (ich) die von Euclid V,5 aufgestellte Definition ... für genauso befriedigend halte, als Ihre Definition. Aus diesem Grunde würde ich wünschen, daß namentlich die Behauptung wegfiele, daß solche Sätze wie $\sqrt{2} \cdot \sqrt{3} = \sqrt{6}$ bisher nicht wirklich bewiesen seien". Charakteristisch ist LIPSCHITZS Bemerkung: „Was Sie an der Vollständigkeit des Gebietes erwähnen, die aus Ihren Principien abgeleitet wird, so fällt dieselbe in der Sache mit der Grundeigenschaft einer Linie zusammen, ohne die kein Mensch sich eine Linie vorstellen kann". Während LIPSCHITZ also einen Standpunkt zum Ausdruck bringt, der an Mathematiker früherer Jahrhunderte erinnert, denen ein intuitives Verständnis der Grundlagen ihrer Wissenschaft häufig ausreichte, steht DEDEKIND am Anfang einer neuen methodischen Einstellung, der es – wie G. CANTOR, G. FREGE, G. PEANO u. a. – um die präzise und explizite Formulierung der mathematischen Grundlagen geht. Und so schreibt DEDEKIND an LIPSCHITZ besonders mit Blick auf den Vollständigkeitsbegriff: „... Aber Euklid schweigt vollständig über diesen, für die Arithmetik wichtigsten Punkt, und deshalb kann ich Ihrer Ansicht nicht zustimmen, daß bei Euklid die vollständigen Grundlagen für die Theorie der irrationalen Zahlen zu finden seien".

Problematisiert wurde der Begriff der reellen Zahl noch einmal in der Grundlagendiskussion der 20er Jahre zwischen HILBERT und BROUWER, nachdem

RUSSELL aus der sogenannten „naiven" Mengenlehre CANTORS und FREGES einen Widerspruch abgeleitet hatte und auch für die axiomatisierten Versionen der Mengenlehre kein Widerspruchsfreiheitsbeweis vorgelegt werden konnte und, wie K. Gödel zeigte, auch mit finiten Mitteln nicht vorgelegt werden kann. Diese Überlegungen führten bis heute vor allem im Rahmen der *mathematischen Logik* zu einer interessanten Diskussion eingeschränkter Begriffsbildungen wie z. B. berechenbare und konstruktive reelle Zahlen (vgl. E. BISHOP [3], H. HERMES [12], P. LORENZEN [18]).

§ 2. Dedekindsche Schnitte

Die Unvollständigkeit des Körpers $\mathbb{Q}$ der rationalen Zahlen läßt sich nach R. DEDEKIND beheben, indem in $\mathbb{Q}$ „Schnitte" eingeführt werden, die in natürlicher Weise vollständig und total (= linear) geordnet werden. Für diese neuen Objekte werden Addition und Multiplikation definiert, so daß sie einen Körper bilden. Insgesamt werden diese Schnitte folgende Eigenschaften (R1)–(R3) haben, die heute üblicherweise als Axiome für die reellen Zahlen genommen werden.

Eine Menge $(K, +, \cdot, \leqslant)$ mit den beiden (inneren) Verknüpfungen $+$ und $\cdot$ und der zweistelligen Relation $\leqslant$ heißt *Menge der reellen Zahlen* genau dann, wenn folgende Axiome erfüllt sind:

(R 1) $(K, +, \cdot)$ *ist ein Körper.*
(R 2) $\leqslant$ *ist eine lineare Anordnung auf* K, *die mit Addition und Multiplikation verträglich ist.*
(R 3) *Vollständigkeitsaxiom: Jede nicht leere, nach unten beschränkte Teilmenge* $M \subset K$ *hat ein Infimum in* K.

Eine untere Schranke s einer geordneten Menge M heißt *Infimum* (Abk. $\inf M$), wenn alle unteren Schranken $\leqslant s$ sind. Man sieht, daß $\inf M$ die größte untere Schranke von M ist.

1. Die Menge $\mathbb{R}$ der Schnitte. Ein Dedekindscher Schnitt ist ein geordnetes Paar (α, β) von Mengen α („Untermenge") und β („Obermenge") mit $\alpha, \beta \subset \mathbb{Q}$, die folgende Forderungen erfüllen:

(D 1) *Jede rationale Zahl liegt in genau einer der Mengen* α, β.
(D 2) α *und* β *sind nicht leer.*
(D 3) *Jedes Element von* α *ist kleiner als jedes Element von* β.
(D 4) β *hat kein kleinstes Element* („*Minimum*").

Jeder Schnitt ist durch seine Unter- und Obermenge je für sich eindeutig bestimmt. Er wird daher im folgenden mit seiner Obermenge β identifiziert, die folgende Eigenschaften besitzt:

(D'1) β und die Komplementärmenge $\bar{\beta} = \mathbb{Q} \setminus \beta$ sind nicht leer.
(D'2) Aus $r \in \beta$, $s \in \mathbb{Q}$ und $r < s$ folgt $s \in \beta$.
(D'3) β hat kein kleinstes Element („Minimum").

Griechische Buchstaben $\alpha, \beta, \ldots$ bezeichnen im folgenden Obermengen. Ein Dedekindscher Schnitt wird reelle Zahl genannt. Die Menge aller Dedekindscher Schnitte bezeichnen wir mit $\mathbb{R}$.

Jede rationale Zahl s bestimmt den Schnitt $\underline{s} := \{r : r \in \mathbb{Q},\ s < r\}$, welcher *rational* genannt wird. Ein Schnitt α ist genau dann rational, wenn $\bar{\alpha}$ ein größtes Element (Maximum) besitzt. Durch $\mathbb{Q} \to \mathbb{R}$, $s \mapsto \underline{s}$, wird $\mathbb{Q}$ in $\mathbb{R}$ eingebettet.

Nicht alle Schnitte sind rational. Beispielsweise ist $\sqrt{2}$, das heißt der Schnitt $\alpha := \{r : r \in \mathbb{Q}, r > 0, r^2 > 2\}$, nicht rational. Die ersten beiden Schnittaxiome sind für α einfach nachzuweisen. Für das dritte Axiom muß man für jedes $r \in \alpha$ ein $s \in \alpha$ mit $s < r$ angeben: Dazu wähle man $s := \dfrac{2r+2}{r+2} \geqslant 0$. Wegen $r - s = \dfrac{r^2-2}{r+2}$ und $r^2 > 2$ mit $r \geqslant 0$ ist $s < r$. Wegen $s^2 - 2 = \dfrac{2(r^2-2)}{(r+2)^2}$ und $r^2 > 2$ ist $s^2 > 2$. Der Schnitt α ist *nicht rational*, da $\bar{\alpha}$ kein Maximum besitzt: Für $r \in \bar{\alpha}$ mit $r \geqslant 0$ (also $r^2 < 2$) wähle man s wieder wie oben. Dann folgt wegen $s^2 < 2$ auch $s \in \bar{\alpha}$ und $r < s$.

2. Die Anordnung in $\mathbb{R}$. Für zwei Schnitte (Obermengen) wird die Ordnungsrelation $\alpha < \beta$ durch die mengentheoretische Inklusion $\beta \subset \alpha$ definiert. Die Reflexivität, Transitivität und Antisymmetrie dieser Relation weist man leicht nach. Die Ordnung ist total (linear): Sei nämlich $\alpha \neq \beta$ und etwa $r \in \alpha$ mit $r \notin \beta$. Dann ist $r \in \bar{\beta}$, und für jedes $s \in \beta$ folgt $r < s$, also $s \in \alpha$, das heißt, $\beta \subset \alpha$. Die Ordnung ist vollständig im Sinne des Axioms R3: Es sei A eine nach unten beschränkte Menge von Schnitten. Dann ist $\beta = \bigcup_{\alpha \in A} \alpha$ ein Schnitt. (Da A nach unten beschränkt ist, gibt es ein $c \in \mathbb{Q}$ mit $c \notin \beta$.) Das 2. und 3. Schnittaxiom für β sind leicht nachzuweisen, ebenso die Tatsache, daß β ein Infimum von A ist.

Führt man die Dedekindsche Schnittbildung erneut über $\mathbb{R}$ durch, erhält man nichts Neues: Zu jedem Schnitt a in $\mathbb{R}$ gibt es ein $\gamma \in \mathbb{R}$, so daß $a = \{\alpha \in \mathbb{R} : \gamma < \alpha\}$ ist. Man nimmt nämlich das Infimum $\gamma = \bigcup_{\alpha \in a} \alpha$ von a.

Diesen Sachverhalt drückt DEDEKIND durch den Satz aus, der als drittes Motto über diesem Kapitel steht. Die anderen beiden Motti (ARISTOTELES und LEIBNIZ) zeigen, daß die zugrunde liegende Vorstellung des zusammenhängenden Kontinuums sehr alt ist.

Die Einbettung von $\mathbb{Q}$ in $\mathbb{R}$, siehe 1., verträgt sich mit der Anordnung. Die rationalen Zahlen liegen dicht in $\mathbb{R}$: Zu je zwei Schnitten (reellen Zahlen) $\alpha < \beta$ gibt es ein $r \in \mathbb{Q}$, so daß $\alpha < \underline{r} < \beta$ ist.

3. Die Addition in $\mathbb{R}$. Für zwei Schnitte α und β aus $\mathbb{R}$ definiert man die *Summe* $\alpha + \beta$ als die Menge $\{r + s : r \in \alpha, s \in \beta\}$. Die drei Schnitteigenschaften von $\alpha + \beta$ folgen sofort aus den entsprechenden Eigenschaften von α und β, das heißt, $\alpha + \beta \in \mathbb{R}$. Auf der Teilmenge $\mathbb{Q}$ von $\mathbb{R}$ stimmt die Summe mit der üblichen Addition rationaler Zahlen überein. Für die Ordnungsrelation sieht man sofort, daß für zwei Schnitte $\alpha < \beta$ auch $\alpha + \gamma < \beta + \gamma$ für jedes γ aus $\mathbb{R}$ folgt.

Satz. *Die Menge $\mathbb{R}$ ist bezüglich der Addition $+$ eine geordnete kommutative Gruppe mit dem Nullschnitt als neutralem Element.*

Beweis. Assoziativität, Kommutativität und $\alpha + \underline{0} = \alpha$ folgen sofort aus der Definition der Addition. Als inversen Schnitt $-\alpha$ zu $\alpha \in \mathbb{R}$ definieren wir $-\alpha := \{-r : r \in \alpha, r \neq \max \alpha\}$. (Man muß $-\max \alpha$ entfernen, damit (D'3) erfüllt wird.) Beweis von $\alpha + (-\alpha) = \underline{0}$: Die Inklusion $\subset$ ist einfach nachzuweisen. Umgekehrt sei $r \in \underline{0}$, also $r > 0$. Zu zeigen ist $r \in \alpha + (-\alpha)$. Da sich $\bar{\alpha}$ und α beliebig nahe kommen, gibt es ein $s \in \bar{\alpha}$ und ein $t \in \alpha$, so daß $0 < t - s < r$ ist. Es sei o. B. d. A. $s \neq \max \bar{\alpha}$. (Man ersetzt sonst s durch $s - \frac{1}{2}(r - (t - s))$.) Dann ist $-s \in -\alpha$, also $t - s \in \alpha + (-\alpha)$ und, weil $r > t - s$, auch $r \in \alpha + (-\alpha)$. □

4. Die Multiplikation in $\mathbb{R}$. Falls die Schnitte α, β beide $\geqslant 0$ sind, wird das Produkt in naheliegender Weise durch $\alpha \cdot \beta = \{r \cdot s : r \in \alpha, s \in \beta\}$ definiert. Man kann dann routinemäßig nachweisen, daß $\alpha \cdot \beta$ die Schnittaxiome (D'1)–(D'3) erfüllt, diese Multiplikation assoziativ und kommutativ ist, $\underline{1}$ ein Einselement ist, das Distributivgesetz gilt und die Multiplikation ordnungstreu ist.

Die Schwierigkeiten beginnen bei der Existenz multiplikativ-inverser Elemente: Wenn $\alpha > 0$ ein Schnitt ist, definiert man

$$\alpha^{-1} := \{r^{-1} : r \in \bar{\alpha}, r > 0, r \neq \max \bar{\alpha}\}.$$

Dem Leser sei überlassen nachzuprüfen, daß α^{-1} tatsächlich ein Schnitt ist und daß $\alpha \cdot \alpha^{-1} \subset \underline{1}$ ist. Zum Beweis von $\alpha \cdot \alpha^{-1} = \underline{1}$ fehlt dann noch $\underline{1} \subset \alpha \cdot \alpha^{-1}$, was man folgendermaßen einsieht: Es sei $r \in \underline{1}$, also $r - 1 > 0$. Es sei $q \in \alpha^{-1}$. Nach dem Archimedischen Prinzip für rationale Zahlen (Kap. I, § 4.2) gibt es eine natürliche Zahl n mit $q < n \cdot (r - 1)$. Nun verfährt man ähnlich wie im Beweis von $\alpha + (-\alpha) = \underline{0}$, siehe 3.: Da sich α und $\bar{\alpha}$ beliebig nahe kommen, findet man ein $s \in \bar{\alpha}$ und ein $t \in \alpha$ mit $0 < t - s < n^{-1}$, wobei man o. B. d. A. annehmen kann, daß $s \neq \max \bar{\alpha}$ und $q^{-1} < s$ ist. Dann ist $s^{-1} \in \alpha^{-1}$, also $t \cdot s^{-1} \in \alpha \cdot \alpha^{-1}$. Nun ist $t \cdot s^{-1} < (s + n^{-1})s^{-1} = 1 + n^{-1}s^{-1} < 1 + n^{-1}q < r$, also $r \in \alpha \cdot \alpha^{-1}$.

Eine weitere Schwierigkeit liegt darin, daß die oben angegebene Definition $\alpha \cdot \beta = \{r \cdot s : r \in \alpha, s \in \beta\}$ nur sinnvoll ist, wenn $\alpha \geqslant 0$ und $\beta \geqslant 0$ ist. Sonst läge gar kein Schnitt vor. Um auch mit negativen Schnitten zu multiplizieren, geht man wie bei der Definition der Multiplikation ganzer Zahlen vor, siehe Kap. I, § 3.2. Man zeigt zunächst, daß sich jeder Schnitt γ als Differenz $\gamma = \alpha - \beta$ nicht-negativer Schnitte $\alpha \geqslant 0$ und $\beta \geqslant 0$ schreiben läßt. Man definiert sodann das Produkt von $\gamma = \alpha - \beta$ und $\gamma' = \alpha' - \beta'$, wobei auch $\alpha', \beta' \geqslant 0$ sind, durch Ausmultiplizieren:

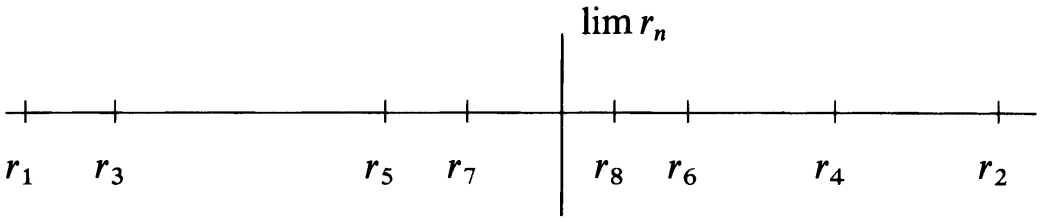
$$\gamma \cdot \gamma' = (\alpha - \beta) \cdot (\alpha' - \beta') := \alpha \cdot \alpha' + \beta \cdot \beta' - \alpha \cdot \beta' - \beta \cdot \alpha'.$$

Man weist nach, daß diese Definition nur von γ und γ' und nicht von der Differenzdarstellung abhängt. Falls γ und γ' beide $\geqslant 0$ sind, stimmt diese Definition mit der alten Definition überein. Das sieht man schnell aufgrund der Differenzdarstellungen $\gamma = \gamma - 0$, $\gamma' = \gamma' - 0$ ein. Aber es bleibt ein langwieriges, wenn auch routinemäßiges Geschäft, nunmehr alle Körperaxiome nachzuweisen. E. LANDAU, der dies alles in [16] ausführt, schreibt darüber in seinem „Vorwort für den Kenner": „Ein anderer hat sich meine zum Teil langweilige Mühe nicht gemacht." Im „Vorwort für den Lernenden" heißt es dagegen: „Bitte vergiß alles, was Du auf der Schule gelernt hast; denn Du hast es nicht gelernt." In der Tat, wenn man das Ziel hat, den von der Schule wohlvertrauten Umgang mit den Zahlen zu begründen,

muß man aufpassen, daß man wirklich nur schon Bewiesenes benutzt und nicht das, was einem sonst alles vertraut ist.

§ 3. Fundamentalfolgen

1. Historisches. Die auf G. CANTOR und C. MERAY [19] zurückgehende Definition der reellen Zahlen macht davon Gebrauch, daß jede reelle Zahl Grenzwert einer Folge von rationalen Zahlen ist, bei der die Differenzen der Folgeglieder mit wachsenden Indizes beliebig klein werden („Fundamentalfolge"):



CANTORS Beitrag zur Theorie der irrationalen Zahlen bildet einen Teil (§ 9) seiner größeren Publikation „*Grundlagen einer allgemeinen Mannigfaltigkeitslehre*" von 1883, in denen er die Grundlagen seiner neuen Mengenlehre entwickelt. Außer seiner Definition nennt CANTOR noch den Ansatz von K. WEIERSTRASS und die Arbeit von R. DEDEKIND. Der logischen Klarheit der Dedekindschen Definition steht nach CANTOR „der große Nachteil" entgegen, „daß die Zahlen in der Analysis sich niemals in der Form von ‚Schnitten' darbieten, in welche sie erst mit großer Kunst und Umständlichkeit gebracht werden müssen." Demgegenüber läßt CANTOR keinen Zweifel, daß er seine Definitionsform für die „einfachste und natürlichste von allen" hält. Historisch nennt er als Beiträge zu diesem Ansatz eine eigene Arbeit von 1871 (Math. Ann. 5, S. 123) und R. LIPSCHITZ [17].

Unabhängig von der Definition der reellen Zahlen hat sich die Cantorsche Konstruktion mit Fundamentalfolgen insofern als die fruchtbarste erwiesen, als sie auch zur Vervollständigung metrischer Räume verwendet werden kann. In diesem Sinne ist CANTOR zuzustimmen, wenn er von seiner Konstruktion feststellt: „Man hat an ihr den Vorteil, daß sie sich dem analytischen Kalkül am unmittelbarsten anpaßt." Im folgenden Abschnitt werden Grundkenntnisse über Folgen vorausgesetzt.

2. Das Cauchysche Konvergenzkriterium. Gemäß CANTORS Grundidee (reelle Zahlen sind Grenzwerte rationaler Folgen) lassen sich reelle Zahlen durch konvergente rationale Folgen beschreiben. Zwei rationale Folgen (r_n) und (s_n) haben genau dann denselben (reellen) Grenzwert, wenn die Differenzenfolge $(r_n - s_n)$ nach Null konvergiert. Es bietet sich also an, die reellen Zahlen als die Äquivalenzklassen konvergenter rationaler Folgen zu definieren, wobei zwei Folgen äquivalent genannt werden, wenn ihre Differenzenfolge nach Null konvergiert.

Damit diese Definition sinnvoll wird, muß man die Konvergenz einer Folge charakterisieren, ohne ihren Grenzwert zu benutzen. Das leistet das *Cauchysche*

Konvergenzkriterium, welches hier zur Definition der betrachteten Folgen genommen wird:

Eine Folge (r_n) rationaler Zahlen heißt *Fundamentalfolge* oder *Cauchysche Folge*, wenn es zu jedem rationalen $\varepsilon > 0$ einen Index k gibt, so daß für alle $m, n \geqslant k$ gilt, daß $|r_m - r_n| < \varepsilon$ ist.

Die rationale Folge (r_n) heißt *rational konvergent*, wenn es eine rationale Zahl r gibt, so daß zu jedem $\varepsilon > 0$ ein Index k existiert mit $|r_n - r| < \varepsilon$ für alle $n \geqslant k$. Dann ist r eindeutig bestimmt, und man schreibt $r = \lim r_n$. Jede rational konvergente Folge ist eine Fundamentalfolge.

Umgekehrt gibt es Fundamentalfolgen, welche nicht rational konvergieren. Jeder nicht periodische Dezimalbruch, wie beispielsweise derjenige für $\sqrt{2}$, ist ein Beispiel:

$$r_0 = 1; \qquad r_1 = 1{,}4; \qquad r_2 = 1{,}41; \qquad r_3 = 1{,}414; \qquad r_4 = 1{,}4142; \ldots .$$

Um auch ein Beispiel zu haben, wo das Bildungsgesetz der Folge angegeben ist, betrachten wir die Kettenbruchentwicklung für das Verhältnis $\frac{1}{2}(1 + \sqrt{5})$ des goldenen Schnittes, vgl. 1.1. Dieser Kettenbruch ist die durch $r_0 = 1$, $r_{n+1} = 1 + \dfrac{1}{1 + r_n}$ rekursiv definierte Folge. Um nachzuweisen, daß es eine Fundamentalfolge ist, geht man so vor: Man zeigt, daß $|r_{n+1} - r_n| < \frac{1}{2}|r_n - r_{n-1}|$ ist. Denn

$$r_{n+1} - r_n = 1 + \frac{1}{1 + r_n} - \left(1 + \frac{1}{1 + r_{n-1}}\right) = \frac{r_{n-1} - r_n}{(1 + r_n)(1 + r_{n-1})} \quad \text{und} \quad r_{n-1}, r_n \geqslant 1.$$

Durch vollständige Induktion folgt sodann

$$|r_{n+1} - r_n| < 2^{-n}|r_1 - r_0| = 2^{-n-1},$$

somit

$$\begin{aligned} |r_{n+k} - r_n| &\leqslant |r_{n+k} - r_{n+k-1}| + |r_{n+k-1} - r_{n+k-2}| + \cdots + |r_{n+1} - r_n| \\ &< 2^{-n-k} + 2^{-n-k-1} + \cdots + 2^{-n-1} < 2^{-n}. \end{aligned}$$

Zu vorgegebenem $\varepsilon > 0$ wählt man also l, so daß $2^{-l} \leqslant \varepsilon$. Für alle $n \geqslant 1$ und alle k ist dann $|r_{n+k} - r_n| < \varepsilon$.

3. Der Ring der Fundamentalfolgen. Die Menge F aller Fundamentalfolgen wird zu einem Ring, indem man die Addition und Multiplikation gliedweise definiert:

$$(r_n) + (s_n) := (r_n + s_n) \qquad \text{und} \qquad (r_n) \cdot (s_n) = (r_n \cdot s_n).$$

Man weist folgendermaßen nach, daß Summe und Produkt wieder Fundamentalfolgen sind: Zu vorgegebenem $\varepsilon > 0$ wählt man k so groß, daß $|r_m - r_n| < \frac{1}{2}\varepsilon$ und $|s_m - s_n| < \frac{1}{2}\varepsilon$ für alle $m, n \geqslant k$ ist. Dann ist $|r_m + s_m - r_n - s_n| \leqslant |r_m - r_n| + |s_m - s_n| < \varepsilon$. Beim Produkt benutzt man zunächst, daß Fundamentalfolgen beschränkt sind: Es gibt ein $c \geqslant 1$, so daß $|r_n|, |s_n| \leqslant c$ ist. Zu vorgegebenem $\varepsilon > 0$ wählt man k so groß, daß $|r_m - r_n|, |s_m - s_n| < \frac{1}{2}\frac{\varepsilon}{c}$ für alle $m, n \geqslant k$ ist. Dann ist $|r_m s_m - r_n s_n| = |r_m(s_m - s_n) + s_n(r_m - r_n)| \leqslant |r_m||s_m - s_n| + |s_n||r_m - r_n| < c\frac{1}{2}\frac{\varepsilon}{c} + c\frac{1}{2}\frac{\varepsilon}{c} = \varepsilon$.

Man bettet $\mathbb{Q}$ als Unterring in F ein, indem man jedem $r \in \mathbb{Q}$ die konstante Folge $(r, r, r, \ldots)$ zuordnet.

4. Der Restklassenkörper F/N der Fundamentalfolgen modulo den Nullfolgen. Eine rationale Folge (r_n) heißt *Nullfolge*, wenn $\lim r_n = 0$ ist. Die Menge N der Nullfolgen ist ein *Ideal* in F, das heißt: (1) Wenn (r_n) und (s_n) Nullfolgen sind, ist $(r_n + s_n)$ eine Nullfolge. (2) Wenn (r_n) eine Nullfolge ist und (s_n) eine beliebige Fundamentalfolge ist, ist $(r_n \cdot s_n)$ eine Nullfolge.

Man nennt zwei Fundamentalfolgen äquivalent, wenn ihre Differenz eine Nullfolge ist. (Der Leser möge nachprüfen, daß tatsächlich eine Äquivalenzrelation vorliegt.) Die von (r_n) repräsentierte Äquivalenzklasse ist $(r_n) + N := \{(r_n + h_n) : (h_n) \in N\}$. Man nennt dies die Restklasse von r_n modulo N. Da N ein Ideal ist, kann man Restklassen addieren und multiplizieren: $((r_n) + N) + ((s_n) + N) = (r_n + s_n) + N$ und $((r_n) + N) \cdot ((s_n) + N) = (r_n \cdot s_n) + N$. Die Menge F/N der Restklassen wird dadurch zu einem kommutativen Ring mit 1. Er umfaßt $\mathbb{Q}$, wenn man jede rationale Zahl r mit der Klasse der zugehörigen konstanten Folge modulo N identifiziert.

Satz. *Die Restklassen der Fundamentalfolgen modulo den Nullfolgen bilden einen Körper F/N.*

Beweis. Zu jedem $(r_n) + N$ mit $(r_n) \notin N$ muß eine multiplikativ inverse Klasse angegeben werden. Hier bietet sich $\left(\frac{1}{r_n}\right) + N$ an. Allerdings muß dazu immer $r_n \neq 0$ sein. Das kann man in der Tat annehmen: Da $(r_n) \notin N$, können höchstens endlich viele Folgeglieder $= 0$ sein. Man ersetzt diese durch 1. Das ändert die Klasse $(r_n) + N$ nicht. Man muß nun noch zeigen, daß $\left(\frac{1}{r_n}\right)$ eine Fundamentalfolge ist: Da $(r_n) \notin N$ und alle $r_n \neq 0$ sind, gibt es ein $\delta > 0$, so daß $|r_n| > \delta$ für alle n gilt. Zu vorgegebenem $\varepsilon > 0$ wählt man den Index k so groß, daß $|r_m - r_n| < \delta^2 \varepsilon$ für alle $m, n \geqq k$ ist. Dann ist

$$\left|\frac{1}{r_m} - \frac{1}{r_n}\right| = \frac{|r_m - r_n|}{|r_m r_n|} < \frac{\delta^2 \varepsilon}{\delta \delta} = \varepsilon.$$

Nach CANTOR definiert man nunmehr den *Körper der reellen Zahlen* als $\mathbb{R} := F/N$. □

5. Der vollständig geordnete Restklassenkörper F/N. Eine rationale Fundamentalfolge (r_n) heißt *positiv*, wenn es ein rationales $\varepsilon > 0$ gibt, so daß $r_n > \varepsilon$ für fast alle (alle bis auf endlich viele) Indizes n gilt. Es sei P die Menge der positiven Fundamentalfolgen. Es ist $P + N \subset P$, $P + P \subset P$ und $P \cdot P \subset P$. Die Menge aller Fundamentalfolgen ist disjunkt zerlegt in $F = -P \cup N \cup P$. Man erhält daher eine wohlbestimmte *totale Ordnung* auf F/N, indem man definiert:

$$(r_n) + N \geqq (s_n) + N \qquad \text{genau dann, wenn} \qquad (r_n - s_n) \in P \cup N.$$

Summe und Produkt positiver Elemente in F/N sind wieder positiv. Auf der Teilmenge $\mathbb{Q} \subset F/N$ stimmt die Ordnung mit der üblichen Ordnung der rationalen Zahlen überein.

Aus der Definition der positiven rationalen Fundamentalfolgen ergibt sich: Zu jedem $\rho \in F/N$ mit $\rho > 0$ gibt es ein $r \in \mathbb{Q}$ mit $0 < r < \rho$. Es ist daher bei der Definition der Konvergenz in F/N egal, ob man alle positiven $\varepsilon \in F/N$ oder nur diejenigen $\in \mathbb{Q}$ zuläßt. Es gilt auch, daß es zu jedem $\sigma \in F/N$ ein $s \in \mathbb{Q}$ mit $s \geqslant \sigma$ gibt. (Das ist für $\sigma < 0$ trivial. Sonst wählt man $r \in \mathbb{Q}$, so daß $0 < r \leqslant \sigma^{-1}$ ist, und nimmt $s = r^{-1}$.)

Die Ordnung von F/N ist *archimedisch*: Wenn $\alpha, \beta \in F/N$ vorgegeben sind und beide > 0 sind, findet man folgendermaßen eine natürliche Zahl n, so daß $n\alpha > \beta$ ist: Man wählt $a, b \in \mathbb{Q}$, so daß $0 < a < \alpha$ und $\beta < b$. Da $\mathbb{Q}$ archimedisch geordnet ist, gibt es ein n mit $na > b$. Dann ist auch $n\alpha \geqslant na > b \geqslant \beta$.

Der Körper F/N wurde so konstruiert, daß (1) jedes $\rho \in F/N$ Grenzwert einer rationalen Folge (r_n) ist und (2) jede rationale Fundamentalfolge in F/N konvergiert. Dabei läßt sich (2) verbessern zum

Satz. *In F/N gilt das Cauchysche Konvergenzkriterium*: Eine Folge (ρ_n) in F/N konvergiert genau dann, wenn gilt: Zu jedem $\varepsilon > 0$ gibt es einen Index k, so daß $|\rho_m - \rho_n| < \varepsilon$ für alle $m, n \geqslant k$ ist.

Beweis. Wegen (1) gibt es zu jedem ρ_n ein $r_n \in \mathbb{Q}$, so daß $|\rho_n - r_n| < \frac{1}{n}$ ist. Dann ist (r_n) eine rationale Fundamentalfolge: Zu vorgegebenem $\varepsilon > 0$ wählt man den Index k so groß, daß $\frac{1}{k} < \frac{1}{3}\varepsilon$ ist und $|\rho_m - \rho_n| < \frac{1}{3}\varepsilon$ für alle $m, n \geqslant k$ gilt. Dann ist $|r_m - r_n| \leqslant |r_m - \rho_m| + |\rho_m - \rho_n| + |\rho_n - r_n| < \frac{1}{m} + \frac{1}{3}\varepsilon + \frac{1}{n} < \varepsilon$. Wegen (2) konvergiert (r_n) nach einem $\rho \in F/N$. Dann konvergiert auch (ρ_n) nach ρ. Denn zu vorgegebenem $\varepsilon > 0$ wählt man den Index l so groß, daß $\frac{1}{l} < \frac{1}{2}\varepsilon$ und $|\rho - r_n| < \frac{1}{2}\varepsilon$ für alle $n \geqslant l$ ist. Dann ist $|\rho - \rho_n| \leqslant |\rho - r_n| + |r_n - \rho_n| < \frac{1}{2}\varepsilon + \frac{1}{n} \leqslant \varepsilon$ für alle $n \geqslant l$. □

In 5.2 werden zahlreiche Formulierungen für die Vollständigkeit von total angeordneten Körpern miteinander verglichen. Dabei wird sich unter anderem das Vollständigkeitsaxiom (R3), welches am Anfang von § 2 angegeben wurde, als äquivalent erweisen zu: Die Ordnung ist archimedisch, und das Cauchysche Konvergenzkriterium gilt. Daher *erfüllt der Cantorsche Körper F/N alle Axiome für die reellen Zahlen*. Irgend zwei Körper, die diese Axiome erfüllen, werden sich in 5.3 als kanonisch isomorph erweisen. Insbesondere ist also F/N zum Körper der Dedekindschen Schnitte isomorph.

§ 4. Intervallschachtelungen

1. Historisches. Die Idee der Intervallschachtelung ist sehr alt und findet sich vor allem in einer technisch orientierten Mathematik, der es um die Berechnung von Größen durch Näherungswerte ging. Bei den *Babyloniern* finden sich bereits die Sexagesimalbrüche $1;25 = 1 + \frac{25}{60}$ und $1;24,51,10 = 1 + \frac{24}{60} + \frac{51}{60^2} + \frac{10}{60^3}$ als Näherungswerte für $\sqrt{2}$ (vgl. O. Neugebauer/A. Sachs [20], S. 42), die sich nach folgendem allgemeinen Verfahren der Intervallschachtelung für $\sqrt{a}$ erschließen lassen, wenn $a > 1$ ist:

$$a > \sqrt{a} > 1,$$

$$x_0 = \frac{1}{2}(a+1) > \sqrt{a} > \frac{a}{x_0},$$

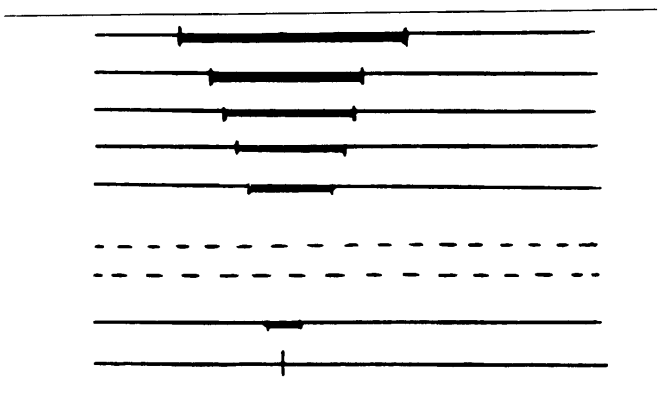
$$x_1 = \frac{1}{2}\left(x_0 + \frac{a}{x_0}\right) > \sqrt{a} > \frac{a}{x_1},$$

$$x_2 = \frac{1}{2}\left(x_1 + \frac{a}{x_1}\right) > \sqrt{a} > \frac{a}{x_2}.$$

In der Tat ergeben sich für $a = 2$ die Werte $x_0 = \frac{3}{2} = 1;30$, $x_1 = \frac{1}{2}(\frac{3}{2} + \frac{4}{3}) = \frac{17}{12} = 1;25$ und $x_2 = \frac{1}{2}(\frac{17}{12} + \frac{24}{17}) = \frac{577}{408} = 1;24,51,10$. Allerdings findet sich das allgemeine Verfahren nicht in den babylonischen Texten, so daß wir auf Vermutungen angewiesen sind. Man kann dieses Verfahren als den Spezialfall $b = 1$ für die Einschachtelung des geometrischen Mittels durch das harmonische und arithmetische Mittel auffassen: $\frac{2ab}{a+b} < \sqrt{a \cdot b} < \frac{a+b}{2}$. Bereits den Pythagoreern waren diese Mittel bekannt, wie ein Fragment des ARCHYTAS VON TARENT zeigt (vgl. O. Becker [2], S. 78f.).

Auch die Bestimmung der Kreisfläche durch ein- und umbeschriebene Polygone ist ein Verfahren der Intervallschachtelung. Es war S. STEVIN, der um 1594 die Rechentechnik der Dezimalbruchentwicklung benutzte und eine reelle Zahl durch Intervallschachtelung bestimmt, ohne dieses Verfahren jedoch allgemein herauszustellen (vgl. 1.3). Im 19. Jahrhundert wurden Intervallschachtelungen beim Beweis einiger zentraler Sätze der Analysis benutzt. Auf B. BOLZANO [4] geht ein Versuch zurück, die reellen Zahlen durch gewisse Intervallfolgen zu definieren, um damit CAUCHYS Konvergenzkriterium zu beweisen. K. WEIERSTRASS [21] benutzt Intervallschachtelungen, um den Satz vom Häufungspunkt zu beweisen. Schließlich führt P. BACHMANN in seinen „Vorlesungen über die Theorie der Irrationalzahlen" (Leipzig 1892) die reellen Zahlen systematisch über Intervallschachtelungen ein.

2. Intervallschachtelungen und Vollständigkeit. Die Einführung der reellen Zahlen durch Intervallschachtelungen wird durch folgende Gedanken motiviert: Man betrachtet auf der Zahlengeraden eine Folge von Intervallen $I_1, I_2, \ldots, I_n, \ldots$, von denen jedes in dem vorhergehenden enthalten ist und von der Art, daß die Länge des n-ten Intervalls I_n mit wachsendem n gegen Null strebt. (Für Dezimalintervalle ist die Länge von I_n gleich 10^{-n}, und die Endpunkte von I_n sind ganzzahlige Vielfache von 10^{-n}.) Gefordert ist, daß zu jeder Intervallschachtelung genau ein Punkt auf der Zahlengeraden existiert, der in allen Intervallen enthalten ist:



Eine rationale Intervallschachtelung (I_n) ist eine Folge abgeschlossener Intervalle $[r_n, s_n]$ mit $r_n, s_n \in \mathbb{Q}$, so daß $I_n \supset I_{n+1}$ für alle n gilt und $\lim(s_n - r_n) = 0$ ist. *Eine Schachtelung (J_n) heißt feiner als (I_n), falls $J_n \subseteq I_n$ für alle n. Man nennt (I_n) und (I'_n) äquivalent, wenn sie eine gemeinsame Verfeinerung (J_n) besitzen.* Das ist genau dann der Fall, wenn $r''_n = \max(r_n, r'_n) \leqslant s''_n = \min(s_n, s'_n)$ ist; (I''_n) mit $I''_n = [r''_n, s''_n]$ ist eine gemeinsame Verfeinerung. Als *reelle Zahlen* definiert man die Äquivalenzklassen der Schachtelungen. Die rationalen Zahlen werden in diese reellen Zahlen eingebettet, indem man jedem $r \in \mathbb{Q}$ die Äquivalenzklasse der konstanten Schachtelung (I_n) mit $I_n := [r, r]$ zuordnet.

Ein Beispiel einer Intervallschachtelung ist $([e_n, e'_n])$ mit $e_n := (1 + \frac{1}{n})^n$ und $e'_n := (1 + \frac{1}{n})^{n+1}$. Diese Schachtelung bestimmt die nach L. EULER benannte reelle Zahl $e = 2{,}71828\ldots$, die in der Analysis beim Logarithmus und der Exponentialfunktion grundlegend wichtig ist, siehe auch Kap. 5.

Es wäre nunmehr nötig, für die Äquivalenzklassen der Schachtelungen eine Addition, eine Multiplikation und eine Anordnung zu definieren und nachzuweisen, daß die Axiome (R1)–(R3), die zu Beginn des § 2 ausgesprochen wurden, erfüllt sind. Das soll hier jedoch nicht ausgeführt werden. Statt dessen werden wir eine direkte Beziehung zu den Dedekindschen Schnitten (§ 2) und zu den Fundamentalfolgen (§ 3) herstellen.

Zu einer Schachtelung $([r_n, s_n])$ bildet man $\alpha := \{x : x \in \mathbb{Q} \text{ und } x \leqslant s_n \text{ für alle } n\}$ und $\beta' := \{y : y \in \mathbb{Q} \text{ und } y > r_n \text{ für alle } n\}$. Falls β' ein Minimum enthält, entfernt man es, $\beta := \beta' - \{\min \beta'\}$. Dann hat (α, β) die Eigenschaften (D1)–(D4) eines Dedekindschen Schnittes, siehe 2.1. Wenn man die Schachtelung verfeinert, ändert sich der Schnitt nicht. Umgekehrt: Zu jedem Dedekindschen Schnitt (α, β) gibt es Schachtelungen $([r_n, s_n])$ mit $r_n \in \alpha$ und $s_n \in \beta$: Man beginnt mit irgendwelchen $r_0 \in \alpha$, $s_0 \in \beta$ und geht rekursiv vor: Wenn r_n, s_n bereits gewonnen sind, bildet man den Mittelwert $d_n = \frac{1}{2}(r_n + s_n)$ und definiert

$$[r_{n+1}, s_{n+1}] = \begin{cases} [d_n, s_n], & \text{falls } d_n \in \alpha, \\ [r_n, d_n], & \text{falls } d_n \in \beta. \end{cases}$$

Alle Schachtelungen $[r_n, s_n]$ mit $r_n \in \alpha$ und $s_n \in \beta$ sind äquivalent. Man ordnet (α, β) diese Äquivalenzklasse zu. Beide Zuordnungen sind zueinander inverse Abbildungen. Wenn man die rationalen Zahlen einmal als Äquivalenzklassen konstanter Schachtelungen auffaßt und das andere Mal als rationale Schnitte, werden sie in obiger Weise einander identisch zugeordnet.

Die direkte Beziehung zwischen Intervallschachtelungen und Fundamentalfolgen beruht auf folgenden Ergebnissen: (1) Jede beschränkte, monotone Folge ist eine Fundamentalfolge. (2) Zu jeder rationalen Fundamentalfolge (a_n) gibt es eine monoton steigende rationale Folge (r_n) und eine monoton fallende rationale Folge (s_n), so daß $(r_n - a_n)$ und $(s_n - a_n)$ Nullfolgen sind. Wenn nun eine Schachtelung $([r_n, s_n])$ vorgegeben ist, sind (r_n) und (s_n) Fundamentalfolgen, und $(s_n - r_n)$ ist eine Nullfolge. Wenn man die Schachtelung zu $([r'_n, s'_n])$ verfeinert, ist $(r'_n - r_n)$ eine Nullfolge. Die Zuordnung $([r_n, s_n]) \mapsto (r_n)$ induziert also eine wohldefinierte Abbildung der Äquivalenzklassen rationaler Intervallschachtelungen in den Cantorschen Körper F/N der Fundamentalfolgen modulo den Nullfolgen. Umgekehrt wählt man zu einer vorgegebenen Fundamentalfolge (a_n) je eine monotone steigende und fallende Folge (r_n) und (s_n) gemäß (2). Dann ist $([r_n, s_n])$ eine Intervallschachtelung. Wenn man

anstatt von (a_n) von einer anderen Fundamentalfolge (a'_n) ausgeht, so daß $(a'_n - a_n)$ eine Nullfolge ist, und dazu (r'_n) und (s'_n) gemäß (2) wählt, ist $([r'_n, s'_n])$ zu $([r_n, s_n])$ äquivalent. Man hat also eine wohlbestimmte Abbildung von den Fundamentalfolgen modulo den Nullfolgen in die Äquivalenzklassen der Schachtelungen. Diese Abbildung ist zu der oben beschriebenen invers.

Praktische Vorteile der Intervallschachtelung gegenüber Schnitten oder Fundamentalfolgen sind folgende: Wenn (I_n) die reelle Zahl x beschreibt, wird durch jedes I_n die Lage von x auf der Zahlengeraden innerhalb bestimmter Schranken fixiert. Hingegen weiß man bei einer Fundamentalfolge (r_n) durch die Angabe nur eines r_n noch nichts über die Lage von x. Die Beschreibung von x als Schnitt $(\bar{\alpha}, \alpha)$ kann durch Eigenschaften der Menge α erfolgen, die über die Lage von x direkt nichts aussagen.

Theoretischer Nachteil der Intervallschachtelungen: Es macht Mühe, für die Äquivalenzklassen von Intervallschachtelungen die $\leqslant$-Beziehung direkt einzuführen (was für Schnitte leicht ist – vgl. § 2.1) und für Addition und Multiplikation die Körpereigenschaften nachzuweisen (was für Fundamentalfolgen über eine kanonische algebraische Konstruktion einfach gelang – vgl. § 3.4).

§ 5. Axiomatische Beschreibung der reellen Zahlen

Während die axiomatische Methode zunächst nur in der Geometrie angewendet wurde (vgl. EUKLIDS „Elemente"), wird sie spätestens seit D. HILBERTS „Grundlagen der Geometrie" [13] auch für die reellen Zahlen benutzt. Im folgenden wird allerdings nicht das Axiomensystem HILBERTS zugrunde gelegt (in [13], § 13 mit Blick auf EUKLIDS „Elemente" die „Lehre von den Proportionen" genannt), sondern die Axiome (R1)–(R3) von § 2.

1. Die natürlichen, ganzen und rationalen Zahlen im reellen Zahlkörper sollen wiedergefunden werden, wenn letzterer durch (R1)–(R3) axiomatisch eingeführt wurde. Man braucht dazu nur (R1) und (R2). Es sei also K ein total angeordneter Körper, das heißt, K erfülle die Axiome (R1) und (R2) von § 2. Man nennt eine Teilmenge $M \subset K$ induktiv, wenn $0 \in M$ ist und wenn mit jedem $x \in M$ auch $x + 1 \in M$ ist. Beispielsweise sind ganz K oder $K^+ = \{x : x \in K, x \geqslant 0\}$ induktiv. Der Durchschnitt N aller induktiven Teilmengen in K ist die kleinste induktive Teilmenge in K. Sie erfüllt zusammen mit der Nachfolgefunktion $S(x) := x + 1$ die Axiome (S1)–(S3) für die natürlichen Zahlen, die in Kap. 1, 2.1 formuliert wurden. Nach dem Einzigkeitssatz (Kap. 1, 2.2) kann man daher $\mathbb{N}$ und $N \subset K$ in eindeutiger Weise miteinander identifizieren.

Es sei $Z \subset K$ der kleinste Unterring, der die 1 enthält. Durch vollständige Induktion folgt, daß $\mathbb{N} \subset Z$ ist. Daher ist Z als der kleinste Ring, welcher $\mathbb{N}$ umfaßt, in eindeutiger Weise zu $\mathbb{Z}$ isomorph, siehe Kap. 1, 3.2.

Es sei $Q \subset K$ der kleinste Unterkörper. Er umfaßt den kleinsten Unterring $\mathbb{Z}$. Daher ist Q zu $\mathbb{Q}$ in eindeutiger Weise isomorph, siehe Kap. 1, § 4.2.

Genau dann, wenn der Körper K archimedisch angeordnet ist (zu je zwei Elementen $a, b > 0$ in K gibt es ein $n \in \mathbb{N}$ mit $na > b$), liegt $\mathbb{Q}$ dicht in K, das heißt, zwischen je zwei Elementen $x < y$ in K gibt es ein $r \in \mathbb{Q}$ mit $x < r < y$.

Eine Richtung (wenn $\mathbb{Q}$ dicht in K liegt) wurde bereits in § 3.5 (mit $K = F/N$) bewiesen. Umgekehrt: Zu $a = 1$ und $b = (y - x)^{-1}$ gibt es ein $n \in \mathbb{N}$ mit $(y - x)^{-1} < n$. Man findet ferner ein $m \in \mathbb{Z}$, so daß $\frac{m}{n} \leqslant x < \frac{m+1}{n}$ ist. Dann ist $x < \frac{m+1}{n} \leqslant x + \frac{1}{n} < y$, letzteres wegen $(y - x)^{-1} < n$.

2. Vollständigkeitssätze. Den drei Konstruktionen für die reellen Zahlen durch Schnitte, Fundamentalfolgen und Intervallschachtelungen liegt als Idee jeweils eine andere Formulierung für die Vollständigkeit zugrunde. Wir zeigen, daß jede zum Vollständigkeitsaxiom (R3) von § 2 äquivalent ist.

Es sei K ein total angeordneter Körper, das heißt, für K seien die Axiome (R1) *und* (R2) *von § 2 erfüllt. Dann sind folgende Aussagen äquivalent:*

(a) *Jede nach unten beschränkte Teilmenge von K besitzt ein Infimum.*

(a′) *Jede nach oben beschränkte Teilmenge von K besitzt ein Supremum.*

(b) *Wenn (α, β) ein Schnitt in K ist (die Axiome* (D1)–(D4) *von § 2 sind erfüllt, wenn man statt rationaler Zahlen die Elemente von K nimmt), besitzt α ein Maximum.*

(c) *Jede nach unten beschränkte, monoton fallende Folge konvergiert in K.*

(c′) *Jede nach oben beschränkte, monoton wachsende Folge konvergiert in K.*

(d) *Der Körper K ist archimedisch geordnet, und jede Fundamentalfolge (Cauchysche Folge) von Elementen aus K konvergiert in K.*

(e) *Der Körper K ist archimedisch geordnet, und zu jeder Intervallschachtelung $I_0 \supset I_1 \supset \cdots \supset I_n \cdots$ in K, bei der die Länge von I_n mit wachsendem n nach Null konvergiert, gibt es genau ein s, welches in allen I_n liegt.*

Offenbar sind (a) und (a′) äquivalent: Genau dann, wenn M nach unten beschränkt ist, ist $-M = \{-x : x \in M\}$ nach oben beschränkt und $-\inf M = \sup(-M)$. Entsprechend sind (c) und (c′) äquivalent. Die Äquivalenz wird durch einen Ringschluß (a) → (b) → (c) → (d) → (e) bewiesen.

(a) → (b): Die Menge β ist nach unten beschränkt, jedes $a \in \alpha$ ist eine untere Schranke. Nach (a) hat β ein Infimum. Da β kein Minimum hat, muß $\inf \beta \in \alpha$ sein. Da $a < b$ für alle $a \in \alpha$ und $b \in \beta$ gilt, gilt $a \leqslant \inf \beta$ für alle $a \in \alpha$, das heißt, $\inf \beta$ ist das Maximum von α.

(b) → (c): Sei (b_n) eine nach unten beschränkte, monoton fallende Folge. Man erhält durch $\alpha = \{x : x \leqslant b_n \text{ für alle } n\}$ und $\beta = \{y : \text{es gibt ein } n \text{ mit } b_n < y\}$ einen Schnitt (α, β). Nach (b) hat α ein Maximum s. Dann konvergiert (b_n) nach s. Beweis dazu: Wenn $\varepsilon > 0$ vorgegeben ist, existiert ein Index k, so daß $b_k < s + \varepsilon$ ist. Denn wenn $s + \varepsilon \leqslant b_k$ für alle k wäre, wäre $s + \varepsilon \in \alpha$ im Widerspruch zu $s = \max \alpha$. Da (b_n) monoton fällt, ist $b_m \leqslant b_k$ für alle $m \geqslant k$. Außerdem ist $s \leqslant b_m$ für alle m, also $s \leqslant b_m \leqslant b_k < s + \varepsilon$ für alle $m \geqslant k$.

(c) → (d): Die archimedische Ordnung wird folgendermaßen bewiesen: Es sei $a, b > 0$. Angenommen, für alle $n \in \mathbb{N}$ wäre $na \leqslant b$. Dann wäre (na) eine durch b nach oben beschränkte, monoton wachsende Folge, die gemäß (c) nach einem s konvergieren würde. Es gäbe also einen Index k, so daß $s - a < na < s$ für alle $n \geqslant k$ gälte. Zwischen $s - a$ und s hat aber höchstens *ein* Folgenglied na Platz.

Um nachzuweisen, daß jede Fundamentalfolge konvergiert, werden zwei Hilfssätze benötigt:

(1) Jede Folge (a_n) besitzt eine monotone Teilfolge.
(2) Jede Fundamentalfolge ist beschränkt.

Wir stellen den Beweis von (1) und (2) einen Augenblick zurück und beweisen zunächst, daß jede Fundamentalfolge (a_n) konvergiert: Es sei (a_{n_j}) eine monotone Teilfolge. Sie ist beschränkt. Somit existiert $s = \lim_{j\to\infty} a_{n_j}$. Wir behaupten, daß dann auch $s = \lim_{n\to\infty} a_n$ ist. Denn wenn ein $\varepsilon > 0$ vorgegeben ist, wählt man den Index k so, daß $|a_m - a_n| < \frac{1}{2}\varepsilon$ für alle $m, n \geqslant k$ ist. Es gibt ein j, so daß $n_j \geqslant k$ ist und $|a_{n_j} - s| < \frac{1}{2}\varepsilon$ ist. Dann gilt für alle $n \geqslant k$, daß $|a_n - s| \leqslant |a_n - a_{n_j}| + |a_{n_j} - s| < \varepsilon$ ist.

Beweis des Hilfsatzes (1): Man sagt, daß die Folge beim Index k eine Spitze a_k hat, wenn $a_k \geqslant a_n$ für alle $n \geqslant k$ gilt. Wenn es unendlich viele Spitzen gibt, bilden sie eine monoton fallende Teilfolge. Wenn es gar keine oder nur endlich viele Spitzen gibt, möge beim Index k die letzte liegen. Man beginnt dann die Teilfolge mit $n_0 = k + 1$. Da a_{n_0} keine Spitze ist, gibt es ein $n_1 > n_0$, so daß $a_{n_1} > a_{n_0}$ ist. Da a_{n_1} keine Spitze ist, gibt es ein $n_2 > n_1$, so daß $a_{n_2} > a_{n_1}$ ist usw.. Rekursiv findet man eine monoton wachsende Teilfolge (a_{n_j}).

Beweis des Hilfsatzes (2): Es sei (a_n) eine Fundamentalfolge. Es gibt einen Index k, so daß $|a_m - a_n| < 1$ für alle $m, n \geqslant k$ ist. Insbesondere liegen also alle Folgenglieder a_n für $n \geqslant k$ im beschränkten Intervall $(a_k - 1, a_k + 1)$. Die endlich vielen Folgenglieder $a_0, \ldots, a_{k-1}$ bilden natürlich auch eine beschränkte Menge. Daher ist die Menge aller Folgenglieder a_n mit $n \in \mathbb{N}$ beschränkt.

(d) → (e): Es sei $([a_n, b_n])$ eine Intervallschachtelung. Dann ist (a_n) eine Fundamentalfolge. Denn für jedes k und alle $m, n \geqslant k$ liegen a_m, a_n in $[a_k, b_k]$, also $|a_m - a_n| < b_k - a_k$. Wegen $\lim(b_n - a_n) = 0$ kann man also $|a_m - a_n| < \varepsilon$ erreichen, indem man k groß genug wählt. Wegen d) existiert $s = \lim a_n$. Da (a_n) monoton wächst, ist $a_n \leqslant s$ für alle n. Da $a_k \leqslant b_n$ für alle k und n gilt, ist auch $s \leqslant b_n$ für alle n, also $s \in [a_n, b_n]$ für jedes n. Da $b_n - a_n$ mit wachsendem n beliebig klein wird, ist s eindeutig bestimmt.

(e) → (a): Es sei M eine nach unten beschränkte, nicht-leere Teilmenge von K. Folgendermaßen konstruiert man eine Intervallschachtelung $([a_n, b_n])$, so daß alle a_n untere Schranken von M sind und alle b_n keine unteren Schranken von M sind. Man beginnt mit irgendeiner unteren Schranke a_0 und einem b_0, das keine untere Schranke ist. Dann geht man rekursiv vor: Wenn $[a_n, b_n]$ bereits definiert ist, bildet man den Mittelwert $d_n = \frac{1}{2}(a_n + b_n)$ und definiert

$$[a_{n+1}, b_{n+1}] = \begin{cases} [d_n, b_n], & \text{falls } d_n \text{ untere Schranke ist,} \\ [a_n, d_n], & \text{falls } d_n \text{ keine untere Schranke ist.} \end{cases}$$

Dann ist $b_{n+1} - a_{n+1} = \frac{1}{2}(b_n - a_n)$, also $b_n - a_n = 2^{-n}(b_0 - a_0)$. Da die Ordnung archimedisch ist, ist $\lim(b_n - a_n) = 0$. Wegen e) gibt es genau ein s, welches in allen $[a_n, b_n]$ liegt. Dann ist c eine untere Schranke von M. Denn sonst gäbe es ein $x \in M$ mit $x < c$. Da jedes $a_n \leqslant x$ ist, wäre $b_n - a_n \geqslant c - a_n \geqslant c - x$ im Widerspruch zu $\lim(b_n - a_n) = 0$. Dieses c ist die größte untere Schranke. Denn wäre $b > c$ eine untere Schranke, müßte jedes $b_n > b$ sein, also $b_n - a_n > b - a_n > b - c$ im Widerspruch zu $\lim(b_n - a_n) = 0$. □

Die Liste der äquivalenten Formulierungen ist durch (a)–(e) keineswegs erschöpft. Man könnte beispielsweise noch die Überdeckungseigenschaft von Heine und Borel anfügen oder die Tatsache, daß jede beschränkte, unendliche Teilmenge eine Häufungspunkt besitzt. Als Konsequenz der Vollständigkeit lernt man diese und andere Ergebnisse in jeder Einführung in die Analysis kennen.

Es gibt total angeordnete Körper, in denen jede Fundamentalfolge konvergiert, ohne daß die Ordnung archimedisch ist. Beispielsweise werden in Kap. 11 die reellen Zahlen zum Körper $^*\mathbb{R}$ der Non-Standard-Zahlen erweitert. Es gibt dann unendlich große und unendlich kleine Zahlen. Deswegen ist $^*\mathbb{R}$ nicht archimedisch, und jede Fundamentalfolge ist konstant, also konvergent. Wie sehr das Archimedische Axiom allein die Möglichkeiten einschränkt, zeigt folgendes Ergebnis von H. CARTAN [6]: Eine angeordnete Gruppe ist genau dann archimedisch, wenn sie zu einer Untergruppe der additiven Gruppe der reellen Zahlen isomorph ist. Man braucht gar nicht vorauszusetzen, daß die Gruppe kommutativ ist; es folgt.

3. Einzigkeit und Existenz der reellen Zahlen. Wir zeigen nun, daß das Axiomensystem (R1)–(R3) die reellen Zahlen eindeutig kennzeichnet. Es sei F/N der Cantorsche Körper der Fundamentalfolgen modulo den Nullfolgen.

Satz. *Jeder angeordnete Körper K, der die Axiome* (R1)–(R3) *erfüllt, ist in eindeutiger Weise zu F/N isomorph.*

Beweis. Folgendermaßen wird eine Abbildung $\varphi: K \to F/N$ definiert: Es sei $x \in K$. Da $\mathbb{Q}$ dicht in K liegt, gibt es eine rationale Fundamentalfolge (x_n) mit $\lim x_n = x$. Man setzt $\varphi(x) = (x_n) \bmod N$. Diese Definition hängt von der Wahl von (x_n) nicht ab, da bei jeder anderen Wahl (x'_n) die Differenzen $x'_n - x_n$ eine Nullfolge bilden. Da der Limes mit der Summe und dem Produkt verträglich ist, ist φ ein Homomorphismus. Offenbar bildet φ die rationalen Zahlen identisch ab. Insbesondere ist φ nicht der Nullhomomorphismus, und das Ideal $\operatorname{Kern}\varphi$ muß $= 0$ sein, das heißt, φ ist injektiv. Soweit wurde nur benutzt, daß K archimedisch ist. Aus der Voraussetzung, daß in K jede (rationale) Fundamentalfolge konvergiert, folgt, daß φ auch surjektiv, also ein Isomorphismus ist.

Die *Einzigkeit* von φ ergibt sich aus folgendem, auch für sich interessanten Ergebnis.

Der Körper der reellen Zahlen gestattet außer der Identität keinen Automorphismus.

Mit „Körper der reellen Zahlen" ist hier jeder Körper K gemeint, der die Axiome (R1)–(R3) erfüllt. Zum Beweis gehen wir davon aus, daß K den Körper $\mathbb{Q}$ der rationalen Zahlen umfaßt. Jeder Automorphismus σ von K bildet $\mathbb{Q}$ identisch ab. Denn es ist $\sigma(0) = 0$, $\sigma(1) = 1$. Daraus folgt durch vollständige Induktion, daß $\sigma \mid \mathbb{N} = \mathrm{id}_{\mathbb{N}}$ ist. Da sich jedes Element aus $\mathbb{Q}$ als $(a - b)/c$ mit $a, b, c \in \mathbb{N}$ darstellen läßt, ist dann auch $\sigma \mid \mathbb{Q} = \mathrm{id}_{\mathbb{Q}}$.

Die Anordnung in K läßt sich allein aufgrund der Körperstruktur beschreiben: Es ist $x \geqslant y$ genau dann, wenn es ein $z \in K$ mit $z^2 = x - y$ gibt. Daher erhält jeder Automorphismus σ die Anordnung. Wenn nun eine Folge (x_ν) nach x in K konvergiert, muß somit die Bildfolge $(\sigma(x_\nu))$ nach $\sigma(x)$ konvergieren. (Anders ausgedrückt: σ ist stetig.) Da $\mathbb{Q}$ dicht in K liegt, gibt es zu jedem $x \in K$ eine Folge in $\mathbb{Q}$, welche nach x konvergiert. Die Folge wird durch σ identisch abgebildet. Als Bildfolge aufgefaßt konvergiert sie nach $\sigma(x)$. Da der Grenzwert eindeutig bestimmt ist, ist $\sigma(x) = x$. □

Bezüglich der *Existenz der reellen Zahlen* ist zu bemerken: In den Kapiteln 1 und 2 wurde $\mathbb{R}$ ausgehend von einer unendlichen Menge auf dem Wege über $\mathbb{N}$, $\mathbb{Z}$ und $\mathbb{Q}$

mit den Methoden der Mengenlehre konstruiert. Die Existenz von $\mathbb{R}$ ist also gesichert, falls man die Existenz dieser Mengenlehre akzeptiert. Anders ausgedrückt: Die Axiome (R1)–(R3) sind widerspruchsfrei, falls die benutzte Mengenlehre widerspruchsfrei ist. Wegen des Problems der Widerspruchsfreiheit der Mengenlehre wird auf das letzte Kapitel verwiesen.

Literatur

[1] Bachmann, P.: Vorlesungen über die Theorie der Irrationalzahlen, Leipzig 1892
[2] Becker, O.: Grundlagen der Mathematik in geschichtlicher Entwicklung, Freiburg/München 1954
[3] Bishop, E.: Foundations of Constructive Analysis, New York 1967
[4] Bolzano, B.: Rein analytischer Beweis des Lehrsatzes, daß zwischen je zwei Werthen, die ein entgegengesetztes Resultat gewähren, wenigstens eine reelle Wurzel der Gleichung liege, Prag 1817, Ostwalds Klassiker Nr. 153, Leipzig 1905
[5] Bolzano, B.: Reine Zahlenlehre, 7. Abschnitt, in: Bernard Bolzano – Gesamtausgabe, eds. E. Winter/J. Berg/F. Kambartel/I. Loužil/B. v. Rootselaar, Reihe II. A. Nachgelassene Schriften Bd. 8, Stuttgart/Bad Cannstatt 1976
[6] Cartan, H.: Un théorème sur les groupes ordonnés, in: Bull. Sci. Math. 63 1939, 201–205
[7] Dedekind, R.: Stetigkeit und Irrationalzahlen, Braunschweig 1872, ^{7}1965
[8] Dugac, P.: Eléments d'analyse de Karl Weierstraß, in: Arch. hist. ex. Sciences 10 1973, 41–176
[9] Euler, L.: De progressionibus harmonicis observationes (1734/35), in: Op. omn. I, 14, 73–86
[10] Fritz, K. v.: The Discovery of Incommensurability by Hippasus of Metapontum, in: Annals of Mathematics 46 1945, 242–264
[11] Heller, S.: Die Entdeckung der stetigen Teilung, Abh. d. Dt. Ak. Wiss. Berlin, Klasse für Mathematik, Physik u. Technik 1958 Nr. 6, Berlin 1958
[12] Hermes, H.: Aufzählbarkeit, Entscheidbarkeit, Berechenbarkeit. Einführung in die Theorie der rekursiven Funktionen, Berlin/Heidelberg/New York 1971
[13] Hilbert, D.: Grundlagen der Geometrie, Leipzig 1899, ed. mit Supplementen P. Bernays, Stuttgart ^{11}1972
[14] Iamblichi de communi mathematica scientia liber (ed. N. Festa), Leipzig 1891
[15] Iamblichi de vita Pythagorica liber (ed. L. Deubner), Leipzig 1937
[16] Landau, E.: Grundlagen der Analysis (Das Rechnen mit ganzen, rationalen, irrationalen, komplexen Zahlen). Ergänzung zu den Lehrbüchern der Differential- und Integralrechnung, Leipzig 1930 (repr. Frankfurt 1970)
[17] Lipschitz, R.: Grundlagen der Analysis, Bonn 1877
[18] Lorenzen, P.: Differential und Integral. Eine konstruktive Einführung in die klassische Analysis, Frankfurt 1965
[19] Meray, C.: Remarques sur la nature des quantités définies par le condition de servir de limites à des variables données, in: Revue des Sociétés savantes. Sciences mathém., phys. et naturelles, 2^{e} séries, t. IV, 1869
[20] Neugebauer, O./Sachs, A.: Mathematical Cuneiform Texts, New Haven 1945
[21] Stevin, S.: La practique d'arithmetique, Leiden 1685
[22] Stifel, M.: Arithmetica integra, Nürnberg 1544, Buch II, Kap. 1
[23] Tropfke, J.: Geschichte der Elementarmathematik Bd. 1 Arithmetik und Algebra, Berlin/New York 1980

[24] WEIERSTRASS, K.: Einleitung in die Theorie der analytischen Funktionen. Vorlesung 1880/81. Nachschrift von A. Kneser

[25] VAN DER WAERDEN, B. L.: Die Pythagoreer: religiöse Bruderschaft und Schule d. Wiss., Zürich 1979

[26] VAN DER WAERDEN, B. L.: Algebra II, Berlin/Heidelberg/New York 1967

Kapitel 3. Komplexe Zahlen

*R. Remmert**)

> Ex irrationalibus oriuntur quantitates impossibiles seu imaginariae, quarum mira est natura, et tamen non contemnenda utilitas (G. W. Leibniz).

Die quadratische Gleichung $x^2 + 1 = 0$ hat im Körper $\mathbb{R}$ der reellen Zahlen keine Lösung, da jede Quadratsumme $r^2 + 1$, $r \in \mathbb{R}$, positiv ist. Eine bahnbrechende Erkenntnis der Mathematik der Neuzeit war, daß sich diese Unvollkommenheit der reellen Zahlen durch eine einfache nochmalige Erweiterung des Zahlenbereiches zum Körper $\mathbb{C}$ der komplexen Zahlen beheben läßt.

Die Entwicklung der Theorie der komplexen Zahlen ist ein eindrucksvolles Beispiel zur Ideengeschichte der Mathematik. Bei ihrem ersten Auftreten in der Renaissance bezeichnete man diese Größen – wie ehemals auch die negativen Zahlen – als *unmögliche Zahlen* (quantitates impossibiles); man begann mit ihnen vorsichtig zu rechnen, ohne sie anzuerkennen. Bis zum Ende des 18. Jahrhunderts gelang keine exakte Begründung der Theorie der imaginären Zahlen. Eine Größe $i = \sqrt{-1}$, deren Quadrat $i^2 = -1$ negativ ist, blieb unvorstellbar. Nichtsdestoweniger arbeitete man seit Bombelli und erst recht seit Euler immer sicherer und erfolgreicher mit imaginären Zahlgrößen. Die alle Erwartungen übertreffende Verwendbarkeit mit ihren unanfechtbaren Resultaten, vor allem die Gültigkeit des Fundamentalsatzes der Algebra (vgl. hierzu Kapitel 4), verhalfen diesen Zahlen schließlich zur vollen Anerkennung, ganz besonders nachdem jeder sie durch die Darstellung als Punkte der Ebene bildlich vor sich sehen konnte.

Im § 1 dieses Kapitels wird die Genesis der komplexen Zahlen beschrieben; in den §§ 2 bis 5 wird *die elementare Theorie* dieser Zahlen, soweit sie *ohne Hilfsmittel der Analysis* auskommt, entwickelt. Im § 6 wird die Polarkoordinatendarstellung komplexer Zahlen

$$z = |z|e^{i\varphi} = |z|(\cos\varphi + i\sin\varphi)$$

behandelt; dabei müssen Eigenschaften der Exponentialfunktion und der trigonometrischen Funktionen herangezogen werden, deren Beweise tiefer liegen. Insbesondere muß die Kreiszahl π zur Verfügung stehen; diese Zahl wird im Kapitel 5 ausführlich diskutiert.

Komplexe Zahlen bilden die Grundlage der Funktionentheorie. Diese Theorie wird in Grundwissen Mathematik, Bd. 5, Funktionentheorie I, behandelt.

---

*) Der Stiftung Volkswagenwerk danke ich für die Bereitstellung eines Akademie-Stipendiums im Wintersemester 1980/81; dadurch wurden die Arbeiten an den Kapiteln 3, 4 und 5 dieses Buches ganz wesentlich gefördert.

§ 1. Genesis der komplexen Zahlen

Es ist heute, wo man schon in der Schule etwas von $i = \sqrt{-1}$ als Lösung der Gleichung $x^2 + 1 = 0$ hört, kaum noch verständlich, welche Schwierigkeiten die komplexen (= imaginären) Zahlen den Mathematikern und Philosophen in der Vergangenheit bereitet haben. Wir stellen im folgenden wichtige historische Daten zusammen; an Sekundärliteratur wurde u. a. benutzt:

ARNOLD, W. und WUSSING, H. (Herausgeber): Biographien bedeutender Mathematiker, Aulis Verlag Deubner u. Co KG, Köln 1978

BOYER, C. B.: A History of Mathematics, John Wiley and Sons, Inc., New York, London, Sidney 1968

CARTAN, E.: Nombres complexes. Exposé, d'après l'article allemand de E. Study (Bonn), Encyclop. Sci. Math. édition française 15, 1908; vgl. E. Cartan Œuvres II, 1, 107–247

HANKEL, H.: Theorie der complexen Zahlensysteme, Leipzig 1867

KLINE, M.: Mathematical Thought from Ancient to Modern Times, Oxford University Press, New York 1972

MARKUSCHEWITSCH, A. I.: Skizzen zur Geschichte der Analytischen Funktionen, VEB Deutscher Verlag der Wissenschaften, Berlin 1955

STUDY, E.: Theorie der gemeinen und höheren complexen Grössen, Encykl. Math. Wiss. I.1, 147–183, Teubner Verlag Leipzig, 1898–1904

TROPFKE, J.: Geschichte der Elementarmathematik, 4. Aufl., Bd. 1: Arithmetik und Algebra, Vollständig neu bearbeitet von Kurt Vogel, Karin Reich und Helmuth Gericke; Walter de Gruyter, Berlin, New York 1980

Der Artikel von CARTAN ergänzt und vertieft den Artikel von STUDY wesentlich.

1. CARDANO (1501–1576). Imaginäre Größen kommen erstmals in der Renaissance vor. Girolamo CARDANO (1534 Mathematiker und berühmter Arzt in Mailand; lernte 1539 von TARTAGLIA ein Verfahren zur Lösung Gleichungen dritten Grades, brach 1545 seinen Schwur, das Geheimnis niemandem zu verraten; 1570 eingekerkert unter der Anklage, Christus das Horoskop gestellt zu haben; 1571 Günstling von Papst Pius V., der ihm eine Leibrente aussetzte. Zitat nach „*Dictionary of Scientific Biography*", Bd. 3.) versucht 1545 in seinem in Nürnberg erschienenen Werk „Artis magnae sive de regulis algebraicis liber unus", bei quadratischen Gleichungen mit imaginären Wurzeln zu operieren: im Kapitel 37 schreibt er z. B. kühn, aber nur „unter Überwindung geistiger Qualen", für die Gleichung $x(10 - x) = 40$ die Lösungen $5 + \sqrt{-15}$ und $5 - \sqrt{-15}$ hin; er sagt: „Manifestum est, quod casus seu quaestio est impossibilis, sic tamen operabimus ...". Da ihm das Hingeschriebene sinnlos erscheint, nennt er $\sqrt{-15}$ eine „quantitas sophistica", was man vielleicht mit „formale Zahl" übersetzen sollte.

Es scheint, daß CARDANO über die Theorie der *kubischen* (*und nicht der quadratischen*!) *Gleichungen* zu komplexen Zahlen gelangt ist. Während nämlich quadratische Gleichungen $x^2 + b = ax$ (mit der Lösungsformel $x = \frac{1}{2}a \pm \sqrt{\frac{1}{4}a^2 - b}$) im Falle $a^2 < 4b$ *keine reellen* Lösungen haben (*unmögliche* Gleichungen), so erhält CARDANO bei kubischen Gleichungen $x^3 = px + q$ *reelle*

*Lösungen als Summen imaginärer Kubikwurzeln**) : Er sagt im Kapitel 12, daß seine Lösungsformel

$$x = \sqrt[3]{q/2 + \sqrt{d}} + \sqrt[3]{q/2 - \sqrt{d}} \qquad \text{mit} \qquad d := (q/2)^2 - (p/3)^3$$

im Fall $(p/3)^3 > (q/2)^2$ *versagt*; er gibt als Beispiele die Gleichungen $x^3 = 20x + 25$ und $x^3 = 30x + 36$ an (die er aus der Identität $x^3 = (x^2 - x)x + x^2$ durch Einsetzen von 5 bzw. 6 gewinnt!): seine Formel führt auf Wurzeln aus negativen Zahlen, die Gleichungen sind aber nicht *unmöglich*, da die Lösungen $x = 5$ bzw. $x = 6$ evident sind. Wahrscheinlich hat CARDANO sich erst nach Erkenntnis dieses Sachverhalts mit komplexen Zahlen befaßt.

2. BOMBELLI (1526–1572). Die Algebra von CARDANO wurde von Rafael BOMBELLI, dessen 1572 in Bologna veröffentlichte „L'algebra" wahrscheinlich zwischen 1557 und 1560 entstand, weiter entwickelt. BOMBELLI hat, ohne sich große Gedanken über das Wesen komplexer Zahlen zu machen, acht fundamentale Rechenregeln zusammengestellt; die letzte ist (in heutiger Notation): $(-i)(-i) = -1$. BOMBELLI führt einige Rechnungen korrekt durch, er weiß z. B.

$$(2 \pm i)^3 = 2 \pm 11i, \qquad \text{also} \qquad \sqrt[3]{2 \pm \sqrt{-121}} = 2 \pm \sqrt{-1}.$$

Diese Identität wendet er auf die Gleichung $x^3 = 15x + 4$, wo die CARDANOsche Lösungsformel

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}$$

liefert, an: die evidente Lösung 4 wird durch $(2 + \sqrt{-1}) + (2 - \sqrt{-1})$ gegeben; er kommt also mit Hilfe komplexer Zahlen zu reellen Lösungen. BOMBELLI hat als erster das formal korrekte Rechnen mit komplexen Zahlen gelehrt.

3. DESCARTES (1596–1650), NEWTON (1643–1727) und LEIBNIZ (1646–1716). René DESCARTES stellt in seiner „La Géométrie" (Leyden 1637) den Gegensatz reell-imaginär heraus; er sagt in etwa: Man kann sich bei jeder Gleichung so viel Wurzeln einbilden (imaginer) wie der Grad angibt, nur entspricht diesen eingebildeten Wurzeln manchmal keine reelle Größe. Im übrigen gesteht DESCARTES unumwunden ein, daß man sich von imaginären Größen noch keine Vorstellung machen könne.

Isaac NEWTON betrachtet komplexe Wurzeln als Indiz für die Unlösbarkeit eines Problems, er hat sich dazu wie folgt geäußert (Universal Arithmetic, 2. Aufl. 1728, S. 193): „But it is just that the Roots of Equations should be impossible, lest they should exhibit the cases of Problems that are impossible as if they were possible". In der NEWTON-Zeit treten komplexe Zahlen in der Physik noch nirgends auf.

---

*) Heute weiß man, daß es grundsätzlich unmöglich ist, eine über $\mathbb{Q}$ irreduzible kubische Gleichung, deren drei Wurzeln alle reell sind, durch reelle Radikale zu lösen (sogenannter Casus irreducibilis). Näheres hierzu vgl. VAN DER WAERDEN, *Algebra*, Erster Teil; Springer-Verlag Berlin-Heidelberg-New York, 7. Auflage 1966, S. 194.

Gottfried Wilhelm LEIBNIZ bereichert in einem 1674 oder 1675 an HUYGENS gerichteten Brief (vgl. LEIBNIZ' *Math. Schriften*, ed. GERHARDT, Bd. 1, II, S. 12) die Lehre vom Imaginären durch die verblüffende Beziehung

$$\sqrt{1+\sqrt{-3}}+\sqrt{1-\sqrt{-3}}=\sqrt{6};$$

1702 nennt er in der von ihm 1682 mitbegründeten ersten deutschen wissenschaftlichen Zeitschrift, den Leipziger *Acta Eruditorum* (= Berichte der Gelehrten)*) (vgl. auch *Math. Schriften*, ed. GERHARDT, Bd. 5, S. 357), imaginäre Wurzeln eine „feine und wunderbare Zuflucht des göttlichen Geistes, beinahe ein Zwitterwesen zwischen Sein und Nichtsein (inter Ens et non Ens Amphibio). LEIBNIZ hat bereits 1712 dafür plädiert, daß $\log(-1)$ eine imaginäre Zahl sei.

4. EULER (1707–1783). Dieser große schweizerische Mathematiker rechnet unbefangen, aber intuitiv richtig und souverän mit komplexen Zahlen; er kennt bereits 1728 die transzendente Beziehung

$$i\log i=-\tfrac{1}{2}\pi \qquad \text{oder, was dasselbe ist,} \qquad i^i=e^{-\frac{1}{2}\pi},$$

doch macht er keinen Versuch einer strengen Herleitung. In seinem berühmten 1748 erschienenen Lehrbuch „Introductio in Analysin infinitorum" (deutsche Übersetzung „Einleitung in die Analysis des Unendlichen" von H. MASER, Springer-Verlag Berlin 1885, Nachdruck 1983) treten imaginäre Zahlgrößen im § 30 ganz plötzlich und völlig unmotiviert auf; eine entscheidende Rolle spielen sie im § 138 bei der Herleitung der „EULERschen Formeln"

$$\cos x=\tfrac{1}{2}(e^{ix}+e^{-ix}) \qquad \text{und} \qquad \sin x=\tfrac{1}{2i}(e^{ix}-e^{-ix}).$$

Erst 1768 erscheint in Petersburg zuerst in russischer und 1770 in deutscher Sprache Leonhard EULERS Lehrbuch**) „Vollständige Anleitung zur Algebra" (Opera Omnia 1, 1–498, ed. H. WEBER, auch Reclams Univ.-Bibl. Nr. 1802–06, Stuttgart 1959). Es bereitet EULER große Schwierigkeiten zu erklären, was imaginäre Zahlen, mit denen er seit Jahrzehnten meisterhaft gerechnet hatte, eigentlich sind. Er betont, daß die Quadratwurzel aus einer negativen Zahl nicht größer als Null, nicht kleiner als Null und auch nicht gleich Null sein kann, und schreibt dann in Capitel 13, Abschnitt 143: „so ist klar, daß die Quadrat-Wurzeln von Negativ-Zahlen nicht einmahl unter die möglichen Zahlen können gerechnet werden: folglich müssen wir sagen, daß dieselben ohnmögliche Zahlen sind. Und dieser Umstand leitet uns auf den Begriff von solchen Zahlen, welche ihrer Natur nach ohnmöglich sind, und gemeiniglich *imaginäre Zahlen*, oder *eingebildete* Zahlen genennt werden, weil sie bloss allein in der Einbildung statt finden." Man würde heute über solche Sätze lächeln, wenn sie nicht vom großen EULER wären. In

*) Der eigentliche Gründer dieser nach dem Vorbild des „Journal des Savants" gegründeten Zeitschrift war O. MENCKE. Die *Acta Eruditorum* stellte 1782 ihr Erscheinen ein.

**) Der schon erblindete EULER hat dieses Buch einem Sekretär, der ehemals Schneider war, diktiert. Angeblich ließ EULER das Manuskript erst dann unverändert, wenn er überzeugt war, daß sein Schreiber den Text voll begriffen hatte (Fernziel aller angewandten Didaktik).

seinem Algebrabuch macht EULER auch Fehler, so rechnet er z. B. $\sqrt{-1}\sqrt{-4} = \sqrt{4} = 2$, weil ja $\sqrt{a}\sqrt{b} = \sqrt{ab}$.

5. WESSEL (1745–1818) und ARGAND (1768–1822). Zum ersten Mal ist die Darstellung komplexer Zahlen als Punkte der Ebene 1797 von dem norwegischen Feldmesser Caspar WESSEL vorgeschlagen worden. WESSEL, der Autodidakt war, schrieb eine Arbeit „Über die analytische Darstellung der Richtung; ein Versuch", die sich 1798 in den Abhandlungen der Dänischen Akademie findet. WESSEL will primär mit gerichteten Strecken rechnen und kommt dabei auf komplexe Zahlen, er stellt gerichtete Strecken durch komplexe Zahlen dar (und nicht umgekehrt!). WESSEL führt senkrecht zur reellen Geraden eine imaginäre Achse ein (er schreibt ε für $\sqrt{-1}$) und deutet Vektoren in der Ebene als komplexe Zahlen. Er definiert die Rechenoperationen für Vektoren und damit komplexe Zahlen *geometrisch* einwandfrei. Trotz ihres großen Verdienstes blieb WESSELS Arbeit unbeachtet, bis 1897 eine französische Übersetzung erschien.

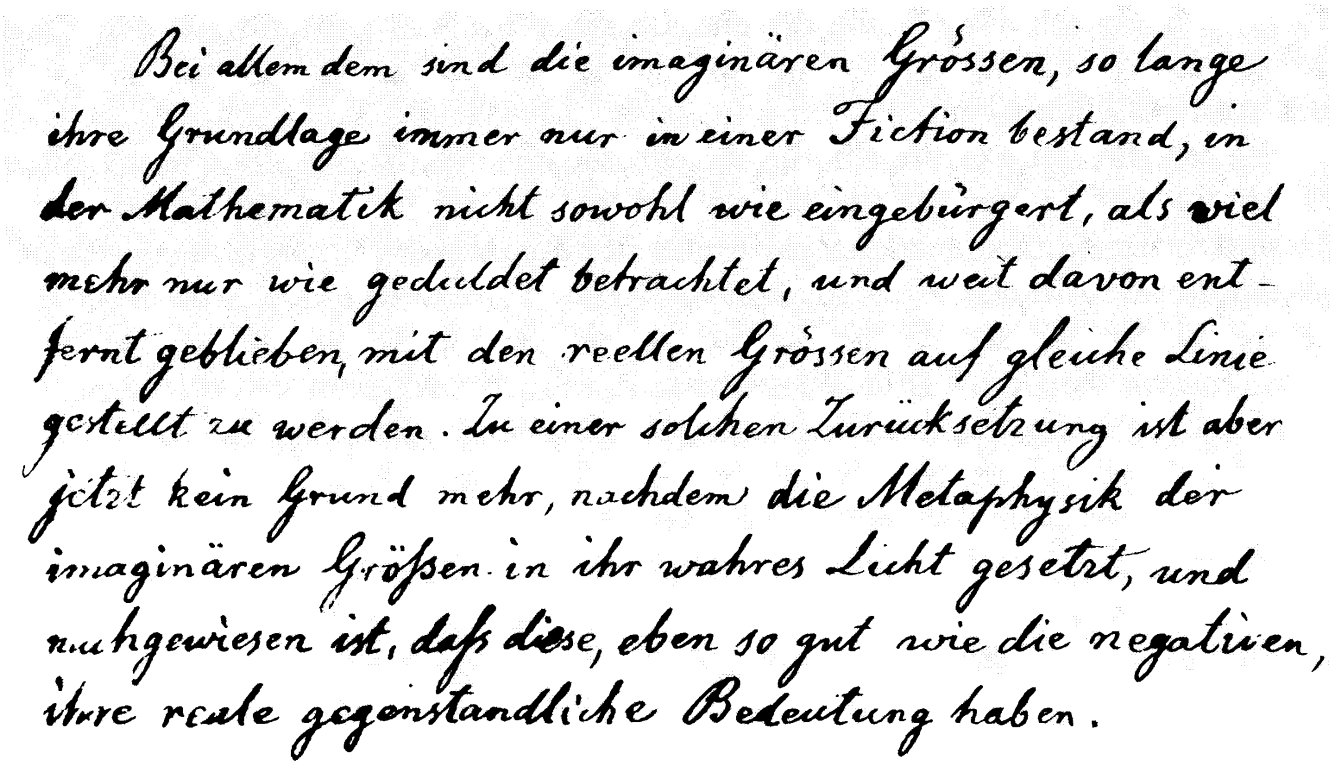
Eine etwas andere geometrische Deutung komplexer Zahlen gibt 1806 der schweizerische Buchhalter Jean Robert ARGAND in seinem „Essay sur une manière de représenter les quantités imaginaires dans les constructions géométriques". ARGAND, der wie WESSEL ebenfalls Amateur war, deutet $\sqrt{-1}$ als *Drehung* um 90° in der Ebene und motiviert dies damit, daß zweimalige Drehung, das heißt das Produkt $\sqrt{-1}\sqrt{-1} = -1$, dann gerade zur Drehung um 180° (Spiegelung) führt, (wir werden diese Deutung in 6.2 präzisieren). Auch ARGANDS Arbeit blieb weitgehend ohne Einfluß, allerdings spricht man in der älteren Literatur häufig von der ARGANDschen Ebene.

Es sprechen gute Gründe dafür, daß EULER sich bereits 1749 komplexe Zahlen als Punkte einer Ebene vorstellte. In seiner Arbeit „De la controverse entre Mrs. LEIBNIZ et BERNOULLI sur les logarithmes des nombres negatives et imaginaires" (Mémoires de l'academie des sciences de Berlin [5], (1749), 1751, 139–179; Opera Omnia, 1. Ser. XVII, 195–232) sagt er (in französisch, S. 230): „.... In jedem anderen Fall wird die Zahl x imaginär sein; man braucht, um sie zu finden, nur einen Bogen g des Einheitskreises zu nehmen und seinen sinus und cosinus zu bestimmen, die gesuchte Zahl ist dann

$$x = \cos g + \sqrt{-1} \cdot \sin g."$$

6. GAUSS (1777–1855). Die Ansichten über komplexe Zahlen ändern sich erst durch das Wirken von Carl Friedrich GAUSS. Er kennt die Interpretation komplexer Zahlen als Punkte der Zahlenebene etwa seit 1796, er benutzt sie 1799 in seiner Dissertation, wo er den Fundamentalsatz der Algebra beweist (vgl. hierzu Kapitel 4), allerdings noch in vorsichtiger Verhüllung. Im Jahre 1811 schreibt GAUSS an BESSEL (Werke 8, S. 90): „So wie man sich das ganze Reich aller reellen Grössen durch eine unendliche gerade Linie denken kann, so kann man das *ganze* Reich aller Grössen, reeller und imaginärer Grössen sich durch eine unendliche Ebene sinnlich machen, worin jeder Punct, durch Abscisse = a Ordinate = b bestimmt, die Grösse $a + bi$ gleichsam repräsentirt." Das ist die Darstellung durch reelle Zahlenpaare in geometrischer Sprache.

Spätestens 1815 war GAUSS im vollen Besitz der geometrischen Theorie. Doch zur echten Verbreitung gelangte die komplexe Zahlenebene erst ab 1831 durch die GAUSSsche Abhandlung „Theoria Residuorum Biquadraticorum, Commentatio Secunda" (Werke 2, 93–148). In der klassischen Selbstanzeige zu dieser zweiten Abhandlung (Werke 2, 169–178) legt er seine alle logischen Bedenken überwindenden Ansichten klar dar. Er prägt den Ausdruck „komplexe Zahl" und beschreibt die Einstellung seiner Zeitgenossen zu diesen Zahlen wie folgt: „allein die den reellen Grössen gegenübergestellten imaginären – ehemals, und hin und wieder noch jetzt, obwohl unschicklich, *unmögliche* genannt – sind noch immer weniger eingebürgert als nur geduldet, und erscheinen also mehr wie ein an sich inhaltsleeres Zeichenspiel, dem man ein denkbares Substrat unbedingt abspricht, ohne doch den reichen Tribut, welchen dieses Zeichenspiel zuletzt in den Schatz der Verhältnisse der reellen Grössen steuert, verschmähen zu wollen". Bezüglich des immer noch vorhandenen Anhauchs von Mystik, der den komplexen Zahlen anhaftet, sagt er (S. 177/178): „Hat man diesen Gegenstand bisher aus einem falschen Gesichtspunkt betrachtet und eine geheimnisvolle Dunkelheit dabei gefunden, so ist dies grossentheils den wenig schicklichen Benennungen zuzuschreiben. Hätte man $+1$, -1, $\sqrt{-1}$ nicht positive, negative, imaginäre (oder gar unmögliche) Einheit, sondern etwa directe, inverse, laterale Einheit genannt, so hätte von einer solchen Dunkelheit kaum die Rede sein können." Und später (nach 1831, Werke 10, 1, S. 404) sagt er rückblickend:


Bei allem dem sind die imaginären Grössen, so lange ihre Grundlage immer nur in einer Fiction bestand, in der Mathematik nicht sowohl wie eingebürgert, als viel mehr nur wie geduldet betrachtet, und weit davon entfernt geblieben, mit den reellen Grössen auf gleiche Linie gestellt zu werden. Zu einer solchen Zurücksetzung ist aber jetzt kein Grund mehr, nachdem die Metaphysik der imaginären Größen in ihr wahres Licht gesetzt, und nachgewiesen ist, daß diese, eben so gut wie die negativen, ihre reale gegenständliche Bedeutung haben.

Abdruck mit freundlicher Genehmigung der Niedersächsischen Staats- und Universitätsbibliothek Göttingen

Erst die Autorität von GAUSS hat den komplexen Zahlen allen Hauch von Mystizismus genommen: seine einfache Deutung der komplexen Zahlen als Punkte der Zahlenebene befreite diese fiktiven Größen von allem Geheimnisvollen und Spekulativen und gab ihnen neben den reellen Zahlen das völlig gleiche Bürgerrecht in der Mathematik. „Sie haben das Unmögliche möglich gemacht" steht 1849 in einer Glückwunschadresse des Collegium Carolinum von Braunschweig (der heutige Technischen Universität) an GAUSS anläßlich seines 50jährigen

Doktor-Jubiläums. Die deutsche Bundespost gab 1977 anläßlich der 200. Wiederkehr des GAUSSschen Geburtstages eine Briefmarke mit der GAUSSschen Zahlenebene heraus.

7. CAUCHY (1789–1857). Der französische Mathematiker Augustin-Louis CAUCHY hat die geometrische Deutung komplexer Zahlen nicht als das letzte Wort angesehen; er schreibt 1821 in seinem richtungweisenden und Maßstäbe für mathematische Strenge setzenden Buch „Cours d'Analyse de l'Ecole Royale Polytechnique" (Œuvres 3, 2. Ser., 17–331, S. 155): „On appelle expression imaginaire toute expression symbolique*) de la forme $a + b\sqrt{-1}$, a, b désignant deux quantités réelles ... toute équation imaginaire n'est que la représentation symbolique de deux équations entre quantités réelles." Diese Auffassung der imaginären Ausdrücke als symbolische Darstellungen von zwei reellen Größen ist im Unterschied zur geometrischen Deutung von GAUSS rein algebraisch.

CAUCHY war noch 1847, also lange nach HAMILTON (vgl. den nächsten Abschnitt), unzufrieden mit der Deutung des Symbols i. In einer Comptes Rendues Note mit dem bezeichnenden Titel „Mémoire sur une nouvelle théorie des imaginaires, et sur les racines symboliques des équations et des équivalences" (Œuvres 10, 1. Ser., 312–323) gibt er eine Definition, die es ermöglicht, „á reduire les expressions imaginaires, et la lettre i elle même, à n'être plus que des quantités réelles". Unter Verwendung des Äquivalenzbegriffes (mit einem ausdrücklichen Hinweis auf die Arbeiten von GAUSS über Klassen von quadratischen Formen) interpretiert er jetzt das Rechnen mit komplexen Zahlen als Rechnen mit reellen Polynomen modulo des Polynoms $X^2 + 1$. In heutiger Sprache heißt dies, daß er den Körper $\mathbb{C}$ der komplexen Zahlen als den Zerfällungskörper von $X^2 + 1$ deutet: $\mathbb{C} = \mathbb{R}[X]/(X^2 + 1)$. CAUCHY beweist hier also im Spezialfall den heute nach KRONECKER benannten Satz, daß für jeden (abstrakten) Körper K und jedes irreduzible Polynom $f \in K[X]$ der Restklassenring $L = K[X]/(f)$ ein endlicher Erweiterungskörper von K ist, in dem f wenigstens eine Nullstelle besitzt.

8. HAMILTON (1805–1865). So hilfreich die geometrische Deutung komplexer Zahlen als Punkte bzw. Vektoren in der Ebene auch ist („Sehen heißt Glauben"), die geometrische Begründung des Rechnens mit solchen Größen ist anschaulich und – zumindest für Algebraiker – nicht recht befriedigend („On ne cherche pas à voir, mais à comprendre"). Der wichtige (wenn auch heute trivial anmutende)

---

*) CAUCHY versucht auch zu erklären, was ein *symbolischer Ausdruck* ist. Er sagt (S. 153): „En analyse, on appelle expression symbolique ou symbole toute combinaison de signes algébriques qui ne signifie rien par elle-même ou à laquelle on attribue une valeur différente de celle qu'elle doit naturellement avoir." Diese wunderliche Definition nennt HANKEL 1867 in seinem Buch „Theorie der complexen Zahlensysteme", wo er selbst mit der Metaphysik mathematischer Grundbegriffe ringt, ein Gaukelspiel und (S. 73) einen *Galimatias* (sinnloses, verworrenes Gerede; Deutung unsicher: wahrscheinlich „Wissen eines Gallus" zu nlat. „galli", einer Bezeichnung für bestimmte Disputanten an der Sorbonne, und griech. „-mátheia" [= das Erlernen]; Quelle: Meyers Enz. Lexik. 1973); er schreibt aggressiv (S. 14): „Ich glaube nicht zu viel zu sagen, wenn ich dies ein unerhörtes Spiel mit Worten nenne, das der Mathematik, die auf die Klarheit und Evidenz ihrer Begriffe stolz ist und stolz sein soll, schlecht ansteht."

Schritt zur formalen Definition als *geordnete Paare reeller Zahlen* blieb noch zu tun. Dies geschah erst 1835 durch Sir William Rowan HAMILTON, wahrscheinlich bei seinen Vorstudien zu Quaternionen. In seiner Arbeit mit dem seltsamen Titel*) „Theory of Conjugate Functions, or Algebraic Couples, with a Preliminary and Elementary Essay on Algebra as the Science of Pure Time" (Math. Papers 3, 3–96) findet sich erstmals die Definition komplexer Zahlen als geordnete reelle Zahlenpaare (S. 81). HAMILTON definiert Addition und Multiplikation so, daß die bekannten Rechenregeln (Distributivgesetze, Assoziativ- und Kommutativgesetz) erhalten bleiben. Wir führen in 2.1 komplexe Zahlen nach HAMILTONschem Vorbild ein. GAUSS sagt 1837 in einem Brief an Wolfgang BOLYAI, daß ihm die Darstellung durch geordnete Paare bereits seit 1831 geläufig war.

9. Ausblick. Die komplexen Zahlen traten im letzten Jahrhundert einen ungestümen Siegeszug durch alle Gebiete der Mathematik an. Für Bernhard RIEMANN (1826–1866) sind sie bereits etwas Selbstverständliches; in seiner berühmten, 1851 verfaßten Göttinger Inauguraldissertation „Grundlagen für eine allgemeine Theorie der Funktionen einer veränderlichen complexen Größe" (Werke 5–43) philosophiert er (S. 37/38): „Die Einführung der complexen Größen in die Mathematik hat ihren Ursprung und nächsten Zweck in der Theorie einfacher durch Größenoperationen ausgedrückter Abhängigkeitsgesetze zwischen veränderlichen Größen. Wendet man diese Abhängigkeitsgesetze in einem erweiterten Umfang an, indem man den veränderlichen Größen, auf welche sie sich beziehen, complexe Werte gibt, so tritt eine sonst versteckt bleibende Harmonie und Regelmäßigkeit hervor." Auf der anderen Seite urteilt noch 1854 der 23jährige Richard DEDEKIND (1831–1916, Freund RIEMANNS, Mathematiker in Braunschweig; „Dedekind occupied a relatively obscure position for fifty years while men who were not fit to lace his shoes filled important and influential university chairs", Zitat nach E. T. BELL „Men of Mathematics", S. 518) in seinem in Göttingen gehaltenen Habilitationsvortrag „Über die Einführung neuer Funktionen in der Mathematik" (Math. Werke 3, S. 434), bei dem GAUSS anwesend war: „Bis jetzt ist bekanntlich eine vorwurfsfreie Theorie der imaginären ... Zahlen entweder nicht vorhanden, oder doch wenigstens noch nicht publiziert."

Komplexe Zahlen drangen auch bald in die Physik ein: Schon 1823 benutzte A. FRESNEL komplexe Zahlen in seiner Theorie der Totalreflexion (veröffentlicht 1831). Heute findet ein Physiker nichts mehr dabei, von komplex-wertigen physikalischen Objekten zu sprechen; Grundgleichungen der Quantenmechanik schreibt man bedenkenlos in der Gestalt

$$pq - qp = \frac{h}{2\pi i}, \qquad \frac{h}{2\pi i}\frac{\partial \Psi}{\partial t} = -H\Psi.$$

*) Der merkwürdige Titel ist durch KANT verursacht. Reelle Zahlen wurden damals üblicherweise als Verhältnisse von Strecken zu einer festgelegten Einheitsstrecke definiert. Nun hatte aber KANT gesagt, daß Geometrie zum Raum und Arithmetik, also Zahlen, zur Zeit gehöre. Deshalb erklärt HAMILTON unter Berufung auf KANT Zahlen als Verhältnisse von Zeitintervallen. Mathematisch ist damit natürlich nichts gewonnen; interessant ist aber, daß er (lange vor WEIERSTRASS und in Unkenntnis von BOLZANO) versucht, reelle Zahlen neu zu erklären.

Auch in der Elektrotechnik verwendet man seit langem komplexe Zahlen; dabei schreibt der Ingenieur j statt i (da i für die Stromstärke vergeben ist). Es ist weitgehend unbekannt, daß einer der ersten Computer (vor K. ZUSES programmgesteuertem Rechenautomaten und vor dem ENIAC-Computer in Princeton) ein „Complex Number Computer" zum Multiplizieren und Dividieren komplexer Zahlen war: er wurde in den Jahren 1938 bis 1940 in den „Bell Telephone Laboratories" von dem Ingenieur G. R. STIBITZ entwickelt und von 1940 bis 1949 für Berechnungen bei der Netzwerkanalyse, vor allem von Telefonschaltungen, erfolgreich eingesetzt.

So haben sich die „numeri impossibiles" im Laufe der Jahrhunderte ihren festen Platz in Naturwissenschaften und Technik erobert; man rechnet widerspruchsfrei und unbekümmert mit ihnen, Mathematiker denken überhaupt nicht mehr über Ens oder non-Ens von $i = \sqrt{-1}$ nach.

§ 2. Der Körper $\mathbb{C}$

Komplexe Zahlen*) werden nach HAMILTON (vgl. 1.8) als geordnete reelle Zahlenpaare eingeführt, sie bilden einen kommutativen 2-dimensionalen Oberkörper $\mathbb{C}$ des Körpers $\mathbb{R}$. Es gibt ein Element $i \in \mathbb{C}$ mit $i^2 + 1 = 0$, jede komplexe Zahl z schreibt sich eindeutig in der Form $z = x + iy$, $x, y \in \mathbb{R}$. Komplexe Zahlen lassen sich auch durch *reelle* 2×2 *Matrizen* elegant beschreiben.

1. Definition durch reelle Zahlenpaare. Die Menge $\mathbb{R} \times \mathbb{R}$ aller geordneten reellen Zahlenpaare $z := (x, y)$ ist bezüglich der natürlichen *Addition*

$$(1) \qquad (x_1, y_1) + (x_2, y_2) := (x_1 + x_2, y_1 + y_2)$$

eine *abelsche Gruppe*. Man führt vermöge

$$(2) \qquad (x_1, y_1) \cdot (x_2, y_2) := (x_1 x_2 - y_1 y_2, x_1 y_2 + y_1 x_2)$$

eine (auf den ersten Blick gekünstelt anmutende) *Multiplikation* in $\mathbb{R} \times \mathbb{R}$ ein. Dann gelten, wie man sofort nachprüft, das Kommutativgesetz, das Assoziativgesetz und das Distributivgesetz; das Element $e := (1, 0)$ ist das Einselement. Eine direkte Rechnung zeigt: *Ist* $z = (x, y) \neq 0$, *so ist* $z^{-1} := \left(\frac{x}{x^2 + y^2}, \frac{-y}{x^2 + y^2} \right)$ *das Inverse von* z, *das heißt*, $z z^{-1} = e$.

Die Menge $\mathbb{R} \times \mathbb{R}$ ist somit bezüglich der Rechenoperationen (1) und (2) ein kommutativer Körper: er heißt *der Körper* $\mathbb{C}$ *der komplexen Zahlen.*

Da $(x_1, 0) + (x_2, 0) = (x_1 + x_2, 0)$ und $(x_1, 0)(x_2, 0) = (x_1 x_2, 0)$, so ist die Abbildung $\mathbb{R} \to \mathbb{C}$, $x \mapsto (x, 0)$ eine *Körpereinbettung*. Man identifiziert die reelle

*) Das Adjektiv „komplex" ist erst seit GAUSS (1831) ein fester terminus technicus; bis dahin hatte auch er das Wort „imaginär" gebraucht. 1773 hat BÉZOUT in seinem in Paris publizierten „Cours de mathématiques à l'usage des gardes du pavillon et de la marine. I. Partie. Éléments d'arithmétique" auf S. 105f. den Ausdruck „komplexe Zahl" in ganz anderer Weise als Bezeichnung für benannte Zahlen benutzt, in denen verschiedene Maßeinheiten vorkommen, z. B. Tage, Stunden, Minuten.

Zahl x mit der komplexen Zahl $(x, 0)$. Dadurch wird $\mathbb{C}$ zu einem *Oberkörper von* $\mathbb{R}$ mit dem Einselement $e = (1, 0) = 1$. Da $\mathbb{C}$ ein 2-dimensionaler, reeller Vektorraum ist, so hat $\mathbb{C}$ in der Sprache der Algebra über $\mathbb{R}$ den Grad 2.

Die Menge $\mathbb{C} \setminus \{0\}$ aller komplexen Zahlen $\neq 0$ wird mit $\mathbb{C}^\times$ bezeichnet.

$\mathbb{C}^\times$ ist bezüglich der Multiplikation in $\mathbb{C}$ eine abelsche Gruppe mit dem Einselement 1 als neutralem Element (multiplikative Gruppe des Körpers $\mathbb{C}$).

Man kann gut motivieren, warum gerade gemäß (2) multipliziert wird: im $\mathbb{R}$-Vektorraum $\mathbb{R}^2$ mit der natürlichen Basis $(1,0)$, $(0,1)$ soll der erste Vektor Einselement werden und der zweite Vektor die Eigenschaft haben, daß sein Quadrat das negative Einselement ist: $(0,1)^2 = -(1,0)$. Dann folgt, wenn die gewohnten Rechenregeln gelten sollen:

$$\begin{aligned}(x_1, y_1)(x_2, y_2) &= [x_1(1,0) + y_1(0,1)][x_2(1,0) + y_2(0,1)] \\ &= x_1x_2(1,0) + (x_1y_2 + y_1x_2)(0,1) + y_1y_2(0,1)^2 \\ &= (x_1x_2 - y_1y_2)(1,0) + (x_1y_2 + y_1x_2)(0,1) \\ &= (x_1x_2 - y_1y_2, x_1y_2 + y_1x_2).\end{aligned}$$

Bemerkung. Die Motivierung für (2) bei HAMILTON ist anders: zunächst hält er es für naheliegend, Produkte mit reellen Zahlen r durch $r(x_1, y_1) := (rx_1, ry_1)$ zu definieren ($\mathbb{R}$-Vektorraumstruktur!). Dann gilt bereits

$$(x_1, y_1) = x_1e + y_1\varepsilon \qquad \text{mit} \qquad e := (1,0), \qquad \varepsilon := (0,1).$$

Soll nun e Einselement werden und sollen die distributiven Gesetze gelten, so ist

$$(*) \qquad (x_1e + y_1\varepsilon)(x_2e + y_2\varepsilon) = x_1x_2e + (x_1y_2 + y_1x_2)\varepsilon + y_1y_2\varepsilon^2$$

zwingend. Die Multiplikation ist daher bestimmt, sobald man ε^2, das von der Form $pe + q\varepsilon$ sein muß, kennt. *Es gibt aber unendlich viele Möglichkeiten, p und q so zu wählen, daß die Multiplikation eindeutig umkehrbar wird!* (Der Leser gebe Beispiele an.) HAMILTON postuliert daher (wie später bei seinen Quaternionen, vgl. 6.E.2) die *Produktregel*: nennt man $|z| := \sqrt{x^2 + y^2}$ die *Länge von $z = (x, y)$*, so soll die *Länge eines Produktes gleich dem Produkt der Längen der Faktoren* sein. Dann braucht man nur auf

$$\varepsilon^2 = pe + q\varepsilon \qquad \text{und} \qquad (e + \varepsilon)(e - \varepsilon) = e - \varepsilon^2 = (1 - p)e - q\varepsilon$$

diese Produktregel anzuwenden: wegen $|\varepsilon^2| = |\varepsilon||\varepsilon| = 1$ und $|e + \varepsilon| = |e - \varepsilon| = \sqrt{2}$ erhält man schnell $p = -1$, $q = 0$, womit $(*)$ zur Gleichung (2) wird.

Zur Produktregel siehe 3.4.

2. Die imaginäre Einheit *i*. Traditionsgemäß benutzt man die seit EULER (1777) übliche und durch GAUSS Gemeingut gewordene Notation

$$i := (0,1) \in \mathbb{C}.$$

Man nennt i gern *die imaginäre Einheit* von $\mathbb{C}$, es gilt: $i^2 = -1$. Im Körper $\mathbb{C}$ hat das reelle Polynom $X^2 + 1$ die beiden Nullstellen i und $-i$. Im komplexen Polynomring zerfällt also $X^2 + 1 = (X - i)(X + i)$ in Linearfaktoren.

Für alle $z = (x, y) \in \mathbb{C}$ besteht die Gleichung $(x, y) = (x, 0) + (0, 1)(y, 0)$; damit erhält man die gebräuchliche Schreibweise für komplexe Zahlen:

$$z = x + iy, \qquad x, y \in \mathbb{R}.$$

Realteil und *Imaginärteil* von $z = x + iy$ werden durch $\operatorname{Re} z := x$, $\operatorname{Im} z := y$ definiert; zwei komplexe Zahlen z_1, z_2 sind genau dann gleich, wenn sie gleichen Realteil und Imaginärteil haben:

$$z_1 = z_2 \Leftrightarrow \operatorname{Re} z_1 = \operatorname{Re} z_2 \quad \text{und} \quad \operatorname{Im} z_1 = \operatorname{Im} z_2.$$

Eine Zahl $z \in \mathbb{C}$ heißt *reell* bzw. *rein imaginär*, wenn $\operatorname{Im} z = 0$ bzw. $\operatorname{Re} z = 0$, letzteres bedeutet $z = iy$. Die Abbildungen $\operatorname{Re}: \mathbb{C} \to \mathbb{R}$, $\operatorname{Im}: \mathbb{C} \to \mathbb{R}$ sind *linear unabhängige Linearformen* des $\mathbb{R}$-Vektorraumes $\mathbb{C}$.

3. Geometrische Darstellung. Man veranschaulicht sich seit WESSEL, ARGAND und GAUSS (vgl. 1.5 und 1.6) die komplexen Zahlen geometrisch in der *Gaußschen Zahlenebene* mit einem rechtwinkligen Koordinatensystem (Fig. a), die Addition komplexer Zahlen ist dann die übliche *Vektoraddition* (*Parallelogrammregel*, Fig. b).

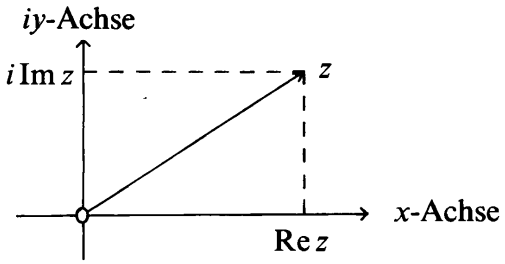


Fig. a

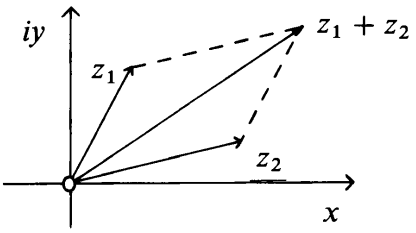


Fig. b

Die Multiplikation komplexer Zahlen wird vollständig beherrscht durch die *eine* Gleichung $i^2 = -1$. Daraus fließt automatisch (vgl. Abschnitt 1)

$$(x_1 + iy_1)(x_2 + iy_2) = (x_1x_2 - y_1y_2) + i(x_1y_2 + y_1x_2).$$

Die geometrische Interpretation der Multiplikation komplexer Zahlen mittels Polarkoordinaten ist nicht mehr ganz elementar und wird erst in 6.2 gegeben. □

Die eindeutige Darstellbarkeit komplexer Zahlen in der Form $x + iy$ zusammen mit der Gleichung $i^2 = -1$ besagt in der Sprache der Algebra:

Der Körper $\mathbb{C}$ ist eine 2-dimensionale (algebraische) Körpererweiterung des Körpers $\mathbb{R}$ und isomorph zum Zerfällungskörper des über $\mathbb{R}$ irreduziblen Polynoms $X^2 + 1 \in \mathbb{R}[X]$.

Wir können hier bereits eine erste Eindeutigkeitsaussage für $\mathbb{C}$ beweisen.

Satz. *Jeder kommutative, nullteilerfreie, 2-dimensionale Oberring K von $\mathbb{R}$ mit Eins ist zum Körper $\mathbb{C}$ isomorph.*

Beweis. Wegen $\dim_{\mathbb{R}} K = 2$ existiert ein $u \in K \setminus \mathbb{R}$, so daß $1 \in \mathbb{R} \subset K$ und u eine Basis des $\mathbb{R}$-Vektorraumes K bilden. Es gilt $u^2 = c + 2du$ mit Zahlen $c, d \in \mathbb{R}$. Für $v := u - d \notin \mathbb{R}$ folgt $v^2 = r$ mit $r := c + d^2 \in \mathbb{R}$. Es ist $r < 0$, denn sonst wäre $\sqrt{r} \in \mathbb{R}$, und man hätte $v = \pm\sqrt{r} \in \mathbb{R}$. Es gibt somit ein $s \in \mathbb{R}$ mit $s^2 = -r^{-1}$. Für $w := sv \in K \setminus \mathbb{R}$ folgt: $w^2 = -1$. Die Abbildung $\mathbb{C} \to K$, $x + iy \mapsto x + wy$ ist nun ein Körperisomorphismus. □

In 4.3.5 wird der vorangehende Satz unter Benutzung des Fundamentalsatzes der Algebra wesentlich verallgemeinert.

4. Nichtanordbarkeit des Körpers $\mathbb{C}$. Der Körper $\mathbb{R}$ der reellen Zahlen ist ein *angeordneter* Körper (vgl. Kap. 2, § 2). *Es ist unmöglich, den Körper der komplexen Zahlen anzuordnen,* das heißt, es gibt keine Relation „> 0" des „positiv sein", so daß folgende zwei Anordnungsregeln erfüllt sind:

1) *Für jedes* $z \in \mathbb{C}$ *gilt genau eine der Beziehungen*: $z > 0$, $z = 0$, $-z > 0$.
2) *Aus* $w > 0$ *und* $z > 0$ *folgt* $w + z > 0$ *und* $wz > 0$.

Beweis. Angenommen, man hätte eine Anordnungsrelation „> 0" auf $\mathbb{C}$. Dann müßte (wie im Reellen) für jedes $z \neq 0$ gelten: $z^2 > 0$. Insbesondere wäre also $1^2 > 0$, $i^2 > 0$ und folglich auch $0 = i^2 + 1 > 0$, was absurd ist. □

Die Nichtanordbarkeit von $\mathbb{C}$ ist ein weiterer Grund für die Schwierigkeiten, die man im 18. und 19. Jahrhundert mit komplexen Zahlen hatte. Hierfür liefern die in 1.4 zitierten Sätze aus EULERS Anleitung zur Algebra ein beredtes Zeugnis.

5. Darstellung durch reelle 2 × 2 Matrizen. Anstelle von reellen Zahlenpaaren kann man zur Einführung komplexer Zahlen $z = x + iy$ auch reelle 2×2 Matrizen verwenden. Jede komplexe Zahl $c = a + ib$ wird als $\mathbb{C}$-lineare Abbildung

$$T_c : \mathbb{C} \to \mathbb{C}, \qquad z \mapsto cz = ax - by + i(bx + ay)$$

von $\mathbb{C}$ in sich interpretiert (sogenannte linksreguläre Darstellung im Sinne der Algebra): dies präzisiert und verallgemeinert die ARGANDsche Deutung komplexer Zahlen, so ist z. B. die zu i gehörende Abbildung $z \mapsto iz$ die Drehung um 90° um 0 im Gegenuhrzeigersinn (1 geht in i, i geht in -1 über usw., vgl. auch 1.5). Identifiziert man $\mathbb{C}$ mit $\mathbb{R}^2$ vermöge $z = x + iy = \binom{x}{y}$, so folgt

$$T_c\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax - by \\ bx + ay \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Die durch $c = a + ib$ bestimmte Abbildung T_c wird daher durch die reelle 2×2 Matrix $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ beschrieben. Man wird so dazu geführt, folgende Abbildung

$$F : \mathbb{C} \to \mathrm{Mat}(2, \mathbb{R}), \qquad c = a + ib \mapsto \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

des Körpers $\mathbb{C}$ in den *nichtkommutativen* Ring $\mathrm{Mat}(2, \mathbb{R})$ der reellen 2×2 Matrizen zu studieren (man vergesse die Motivierung mittels T_c). Diese Abbildung ist $\mathbb{R}$-*linear* und *multiplikationstreu*:

$$F(rc + r'c') = rF(c) + r'F(c'), \quad F(cc') = F(c)F(c'), \quad r, r' \in \mathbb{R}; \quad c, c' \in \mathbb{C};$$

dabei ist $F(c)F(c')$ das *Matrizenprodukt* . Es gilt $F(1) = E := \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Man sieht:

Die Menge $\mathscr{C} := \{\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right) : a, b \in \mathbb{R}\}$ *ist bezüglich der Matrizenaddition und Matrizenmultiplikation ein kommutativer Körper mit der Einheitsmatrix* E

als Einselement; die $\mathbb{R}$*-lineare Abbildung*

$$F: \mathbb{C} \to \mathscr{C}, \quad a + bi \mapsto \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \quad \text{mit} \quad I := F(i) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad I^2 = -E,$$

ist ein Körperisomorphismus; die Matrix I ist die „imaginäre Einheit" in $\mathscr{C}$.

Die Einführung komplexer Zahlen durch reelle 2×2 Matrizen hat gegenüber der Einführung durch geordnete Paare reeller Zahlen den Vorteil, daß nicht ad hoc eine Multiplikation eingeführt werden muß. In den gängigen Lehrbüchern werden komplexe Zahlen nicht als reelle 2×2 Matrizen erklärt; eine Ausnahme macht das 1935 erschienene Buch von E. T. Copson „An introduction to the theory of functions of a complex variable", Oxford, At the Clarendon Press.

Es gibt neben $\mathscr{C}$ *unendlich viele weitere zu* $\mathbb{C}$ *isomorphe Unterkörper in* $\mathrm{Mat}(2, \mathbb{R})$. Einen Überblick verschafft der folgende

Satz. a) *Für jede invertierbare reelle* 2×2 *Matrix W ist die Abbildung*

$$g_W: \mathbb{C} \to \mathrm{Mat}(2, \mathbb{R}), \qquad a + bi \mapsto W \begin{pmatrix} a & -b \\ b & a \end{pmatrix} W^{-1}$$

ein Monomorphismus reeller Algebren (vgl. R.3).

(b) *Jeder Homomorphismus* $g: \mathbb{C} \to \mathrm{Mat}(2, \mathbb{R})$, $g \neq 0$, *hat die Gestalt* g_W.

Beweis. a) Der Fall $W := E =$ Einheitsmatrix wurde oben betrachtet. Da die Abbildung $T_W: \mathrm{Mat}(2, \mathbb{R}) \to \mathrm{Mat}(2, \mathbb{R})$, $A \to WAW^{-1}$ ein $\mathbb{R}$-Algebra-Automorphismus ist, folgt wegen $g_W = T_W \circ g_E$ die Behauptung.

b) Für $A := g(1), B := g(i) \in \mathrm{Mat}(2, \mathbb{R})$ gilt $A^2 = A, BA = AB = B, B^2 = -A$. Da $\mathbb{C}$ ein Körper ist, so ist g injektiv, also $A \neq 0$. Wir wählen einen Spaltenvektor $v \in \mathbb{R}^2$ mit $w := Av \neq 0$. Es gilt:

$$(*) \qquad Aw = A^2v = Av = w, \qquad A(Bw) = BAw = Bw, \qquad B^2w = -Aw = -w.$$

Wegen der letzten Gleichung sind w, Bw linear unabhängig, da sonst eine Gleichung $Bw = \lambda w$, $\lambda \in \mathbb{R}$, bestände, die zum Widerspruch $\lambda^2 = -1$ führt. Die Matrix $W := (w, Bw) \in \mathrm{Mat}(2, \mathbb{R})$ ist somit invertierbar, wegen $(*)$ folgt $AW = W$, also $A = E$, und weiter $BW = (Bw, -w) = (w, Bw)\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) = WI$. Damit ist gezeigt $g(1) = E = g_W(1)$, $g(i) = WIW^{-1} = g_W(i)$. Aus der $\mathbb{R}$-Linearität von g und g_W folgt nun $g = g_W$. □

Für $W := \left(\begin{smallmatrix} 2 & 3 \\ 1 & 2 \end{smallmatrix}\right)$ gilt z. B.

$$g_W(\mathbb{C}) = \left\{ \begin{pmatrix} a + 8b & -13b \\ 5b & a - 8b \end{pmatrix} : a, b \in \mathbb{R} \right\} \simeq \mathbb{C};$$

in diesem Beispiel ist $\left(\begin{smallmatrix} 8 & -13 \\ 5 & -8 \end{smallmatrix}\right)$ die „imaginäre Einheit".

§ 3. Algebraische Eigenschaften des Körpers $\mathbb{C}$

Der Körper $\mathbb{C}$ besitzt den *Konjugierungsautomorphismus* $\mathbb{C} \to \mathbb{C}$, $z \mapsto \bar{z}$, der für viele Überlegungen grundlegend ist. Durch

$$\langle w, z \rangle := \mathrm{Re}(w\bar{z}) = ux + vy, \qquad |z| := \sqrt{z\bar{z}} = \sqrt{x^2 + y^2},$$

wobei $w = u + iv$, $z = x + iy$, wird das natürliche *Skalarprodukt* in $\mathbb{C}$ nebst zugehöriger *Betragsfunktion* eingeführt; mit Hilfe dieser Funktion wird im Abschnitt 5 elementar gezeigt, daß jede quadratische Gleichung $z^2 + az + b = 0$, $a, b \in \mathbb{C}$, in $\mathbb{C}$ lösbar ist. Diese Aussage ist ein erstes Indiz dafür, daß der Körper $\mathbb{C}$ vollkommener ist als der Körper $\mathbb{R}$. Der Satz von der Lösbarkeit aller quadratischen Gleichungen in $\mathbb{C}$ war schon lange vor EULER bekannt; er ist ein Spezialfall des berühmten und tiefliegenden Fundamentalsatzes der Algebra, der besagt, daß im Körper $\mathbb{C}$ jedes nichtkonstante komplexe Polynom Nullstellen hat; diesen Satz werden wir im Kapitel 4 diskutieren.

1. Die Konjugierung $\mathbb{C} \to \mathbb{C}$, $z \mapsto \bar{z}$. Bekanntlich besitzt der Körper $\mathbb{R}$ außer der Identität keinen Automorphismus (vgl. Kapitel 2, 5.3). Im Gegensatz hierzu gibt es unendlich viele Automorphismen des Körpers $\mathbb{C}$. Unter diesen ist einer dadurch ausgezeichnet, daß er $\mathbb{R}$ in sich und i in die (prinzipiell gleichberechtigte) zweite Nullstelle $-i$ von $X^2 + 1$ überführt.

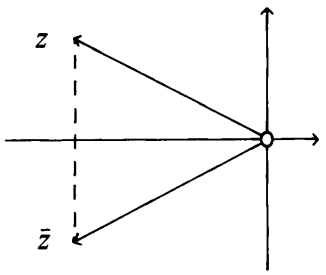
Für jede komplexe Zahl $z = x + iy$, $x, y \in \mathbb{R}$, heißt

$$\bar{z} := x - iy = 2\,\mathrm{Re}\, z - z$$

die *zu z konjugiert komplexe* Zahl*). In der GAUSSschen Zahlenebene erhält man $\bar{z}$ aus z durch Spiegelung an der reellen Achse (Figur). Es gilt

$$\mathrm{Re}\, z = \frac{1}{2}(z + \bar{z}), \quad \mathrm{Im}\, z = \frac{1}{2i}(z - \bar{z}), \quad z\bar{z} = x^2 + y^2 \in \mathbb{R}, \quad z\bar{z} > 0 \quad \text{für} \quad z \neq 0,$$

insbesondere ist z genau dann reell bzw. rein imaginär, wenn $z = \bar{z}$ bzw. $z = -\bar{z}$.



Das Rechnen mit konjugiert komplexen Zahlen regelt folgender

Satz. *Die Konjugierungsabbildung $\mathbb{C} \to \mathbb{C}$, $z \mapsto \bar{z}$, ist ein Automorphismus des Körpers $\mathbb{C}$, das heißt, es gilt $\bar{1} = 1$ und*

$$\overline{w + z} = \bar{w} + \bar{z}, \qquad \overline{wz} = \bar{w}\bar{z} \qquad \textit{für alle} \qquad w, z \in \mathbb{C}.$$

Es gilt stets $\bar{\bar{z}} = z$. Die Fixpunktmenge $\{z \in \mathbb{C} : \bar{z} = z\}$ ist der Körper $\mathbb{R}$.

Alle diese Behauptungen folgen mühelos aus der Definition von $\bar{z}$; wir verifizieren hier lediglich die Multiplikationsregel. Sei $w = u + iv$, $z = x + iy$. Dann gilt $wz = ux - vy + i(vx + uy)$ und also

$$\overline{wz} = ux - vy - i(vx + uy) = (u - iv)(x - iy) = \bar{w}\bar{z}. \qquad \square$$

*) Die Bezeichnung „konjugiert" (conjugué) wurde 1821 von CAUCHY im „Cours d'Analyse" eingeführt.

Aufgabe. Zeigen Sie, daß für alle $a, b, c, d \in \mathbb{C}$ mit $a\bar{a} = b\bar{b} = c\bar{c}$ gilt:

$$(a-b)(c-d)(\bar{a}-\bar{d})(\bar{c}-\bar{b}) - i(c\bar{c}-d\bar{d})\,\mathrm{Im}(c\bar{b}-c\bar{a}-a\bar{b}) \in \mathbb{R}. \qquad \square$$

Man beweist direkt:

Kriterium für lineare Abhängigkeit: *Zwei Zahlen $w, z \in \mathbb{C}$ sind genau dann linear abhängig über $\mathbb{R}$, wenn $w\bar{z} \in \mathbb{R}$.*

Die Konjugierungsabbildung läßt sich vorteilhaft verwenden, um alle $\mathbb{R}$-*linearen Abbildungen* $T: \mathbb{C} \to \mathbb{C}$ zu beschreiben. $\mathbb{R}$-Linearität bedeutet, daß für $z = x + iy$ gilt $T(z) = xT(1) + yT(i)$. Hieraus ergibt sich unmittelbar:

Folgende Aussagen über eine Abbildung $T: \mathbb{C} \to \mathbb{C}$ sind äquivalent:

i) *T ist $\mathbb{R}$-linear.*
ii) *Es gilt $T(z) = az + b\bar{z}$ mit Konstanten $a, b \in \mathbb{C}$.*

Eine $\mathbb{R}$-lineare Abbildung $T: \mathbb{C} \to \mathbb{C}$ ist genau dann $\mathbb{C}$-linear, wenn $T(i) = iT(1)$; dies trifft genau dann zu, wenn gilt: $T(z) = az$.

Für den in 2.5 eingeführten Isomorphismus $F: \mathbb{C} \to \mathscr{C}$ gilt, wenn A^t die zu A *transponierte* Matrix bezeichnet:

$$F(\bar{c}) = F(c)^t \qquad \text{für} \qquad c \in \mathbb{C}.$$

Die Konjugierungsabbildung in $\mathbb{C}$ ist also nichts anderes als die Transpositionsabbildung in $\mathscr{C}$.

2. Körperautomorphismen von $\mathbb{C}$. Die Abbildung $z \mapsto \bar{z}$ ist einfach charakterisierbar.

Satz. *Die Konjugierungsabbildung ist der einzige Körperautomorphismus von $\mathbb{C}$, der $\mathbb{R}$ in sich abbildet und von der Identität verschieden ist.*

Beweis. Ist $f: \mathbb{C} \to \mathbb{C}$ ein Automorphismus mit $f(\mathbb{R}) \subset \mathbb{R}$, so gilt zunächst $f(x) = x$ für alle $x \in \mathbb{R}$. Für alle Zahlen $z = x + iy$, $x, y \in \mathbb{R}$, folgt dann

$$f(z) = f(x + iy) = f(x) + f(i)f(y) = x + f(i)y.$$

Wegen $i^2 = -1$ ist $f(i)^2 = f(i^2) = f(-1) = -1$, das heißt, $f(i) = \pm i$. Der Fall $f(i) = i$ führt zu $f = id$, der Fall $f(i) = -i$ führt zur Konjugierung. $\square$

Kein Geringerer als R. DEDEKIND hat noch zu Beginn dieses Jahrhunderts (1901) geschrieben: „die Zahlen des reellen Körpers scheinen mir durch die Stetigkeit so unlöslich miteinander verbunden zu sein, daß ich vermute, er könne außer der identischen gar keine andere Permutation [= Automorphismus] besitzen, und hieraus würde folgen, daß der Körper aller Zahlen [= Körper $\mathbb{C}$] nur die beiden genannten Permutationen besitzt. Nach einigen vergeblichen Versuchen, hierüber Gewißheit zu erlangen, habe ich diese Untersuchung aufgegeben; um so mehr würde es mich erfreuen, wenn ein anderer Mathematiker mir eine entscheidende Antwort auf diese Frage mitteilen wollte." (Math. Werke 2,

S. 277). Heute weiß man, daß es unendlich viele weitere Automorphismen von $\mathbb{C}$ gibt (die dann notwendig $\mathbb{R}$ *nicht* in sich abbilden). Man konstruiert solche Abbildungen unter Heranziehung des Auswahlaxioms; wirklich gesehen hat einen solchen Automorphismus noch niemand, vgl. Grundwissen Mathematik, Bd. 2, Lineare Algebra und analytische Geometrie, S. 44.

3. Das natürliche Skalarprodukt $\mathrm{Re}(w\bar{z})$ und die euklidische Länge $|z|$. Das *euklidische* Skalarprodukt im reellen Vektorraum $\mathbb{C} = \mathbb{R}^2$ wird gegeben durch

$$\langle w, z\rangle := \mathrm{Re}(w\bar{z}) = ux + vy, \qquad \text{wobei} \qquad w = u + iv, \qquad z = x + iy.$$

Da stets $z\bar{z} = x^2 + y^2 \geqslant 0$, so existiert die nicht negative reelle Quadratwurzel

$$|z| := +\sqrt{\langle z, z\rangle} = +\sqrt{z\bar{z}} = \sqrt{x^2 + y^2};$$

sie mißt den *euklidischen Abstand* des Punktes z vom Nullpunkt der Gaussschen Zahlenebene, das heißt, die Länge von z; man nennt $|z|$ den (*absoluten*) *Betrag**) von z. Für reelle Zahlen $z \in \mathbb{R}$ stimmt $|z|$ mit dem üblichen Betrag für reelle Zahlen überein. Es gilt:

$$|z| = |\bar{z}| \qquad \text{für alle} \qquad z \in \mathbb{C}.$$

Wegen $z\bar{z} = |z|^2$ hat man folgende elegante Darstellung des Inversen:

$$z^{-1} = \frac{\bar{z}}{|z|^2} \qquad \text{für alle} \qquad z \in \mathbb{C}^\times.$$

Die Abbildung $\mathbb{C} \times \mathbb{C} \to \mathbb{R}$, $(w, z) \mapsto \langle w, z\rangle$ ist $\mathbb{R}$-*bilinear*, *symmetrisch und positiv-definit*, das heißt, für alle $w, w', z \in \mathbb{C}$ gilt:

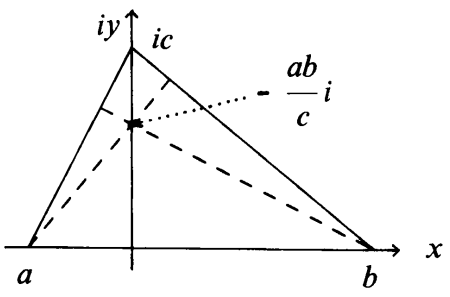
$$\langle w + w', z\rangle = \langle w, z\rangle + \langle w', z\rangle; \qquad \langle aw, z\rangle = a\langle w, z\rangle, \qquad a \in \mathbb{R};$$

$$\langle w, z\rangle = \langle z, w\rangle; \qquad \langle z, z\rangle > 0, \qquad \text{falls} \qquad z \neq 0;$$

diese Rechenregeln folgen unmittelbar aus der Definition von $\langle\ ,\ \rangle$. □

Zwei Vektoren w, z heißen *orthogonal* (*stehen senkrecht aufeinander*), wenn gilt: $\langle w, z\rangle = 0$. Die Vektoren iz und z stehen wegen $\mathrm{Re}(iz\bar{z}) = |z|^2\,\mathrm{Re}(i) = 0$ immer senkrecht aufeinander; wegen $z\bar{z} \in \mathbb{R}$ gilt allgemein:

Genau dann sind $z, cz \in \mathbb{C}^\times$ orthogonal, wenn c rein imaginär ist.



Der Leser benutze dies zu einem einfachen Beweis des *Höhensatzes für Dreiecke* (der gemeinsame Höhenschnittpunkt ist $-\frac{ab}{c}i$, vgl. Figur).

---

*) K. Weierstrass verwendet in seinen Vorlesungen die Bezeichnung „absoluter Betrag"; vorher war die Redeweise „Modulus" gebräuchlich.

Es ist amüsant, das Skalarprodukt $\mathrm{Re}(w\bar{z})$ im Körper $\mathscr{C}$ der reellen Matrizen $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ zu deuten; wir stellen folgende

Aufgabe. Zeigen Sie, daß

$$\langle A, B\rangle := \tfrac{1}{2}\mathrm{Spur}(A \cdot B^t), \qquad A, B \in \mathscr{C},$$

eine *positiv-definite, symmetrische Bilinearform* ist. Zeigen Sie, daß für den Isomorphismus $F: \mathbb{C} \to \mathscr{C}$ gilt: $\langle F(w), F(z)\rangle = \langle w, z\rangle$ (Längentreue). Zeigen Sie weiter:

$$\langle A, A\rangle = \det A.$$

Vergleichen Sie hierzu auch 6.2.2.

4. Produktregel und „Zwei-Quadrate-Satz". Für das Rechnen mit Beträgen gilt

$$|wz| = |w|\,|z| \qquad \textit{für alle} \qquad w, z \in \mathbb{C} \qquad (\textit{Produktregel}).$$

Zum Beweis schreibt man $|wz|^2 = wz(\overline{wz}) = w\bar{w}z\bar{z} = |w|^2|z|^2$. □

Die Produktregel enthält den berühmten, bereits DIOPHANTOS VON ALEXANDRIA (griechischer Mathematiker der 2. Hälfte des 3. Jh. n. Chr.) bekannten

„Zwei-Quadrate-Satz". *Für alle $u, v, x, y \in \mathbb{R}$ gilt*

$$(u^2 + v^2)(x^2 + y^2) = (ux - vy)^2 + (uy + vx)^2.$$

Beweis. Man wende die Produktregel an auf $w := u + iv, z := x + iy$. □

Die komplexen Zahlen dienen nur dazu, den Zwei-Quadrate-Satz zu entdecken. Im Nachhinein läßt er sich durch Ausmultiplizieren *für jeden kommutativen Ring* bestätigen, insbesondere für den Ring $\mathbb{Z}$ der ganzen Zahlen; diese Tatsache ist in der elementaren Zahlentheorie bedeutsam, so ist z. B. eine natürliche Zahl $n > 1$ stets dann Summe zweier Quadrate natürlicher Zahlen, wenn jeder Primfaktor von n es ist. In der elementaren Zahlentheorie zeigt man, daß genau die Primzahlen der Form $4k + 1$ die Form $l^2 + m^2$, $l, m \in \mathbb{N}$, haben.

Verallgemeinerungen des „Zwei-Quadrate-Satzes" werden in den späteren Kapiteln dieses Bandes eine große Rolle spielen, vgl. z. B. 6.2.3, 8.2.4 und Kapitel 9. □

Die Produktregel impliziert die

Divisionsregel: $$\left|\frac{w}{z}\right| = \frac{|w|}{|z|} \qquad \textit{für alle} \qquad w, z \in \mathbb{C}^\times.$$

Die Produktregel hat ferner als unmittelbare Konsequenz:

Die Menge $S^1 := \{z \in \mathbb{C} : |z| = 1\}$ aller komplexen Zahlen der Länge 1 ist bezüglich der Multiplikation in $\mathbb{C}$ eine Untergruppe von $(\mathbb{C}^\times, \cdot)$.

In der GAUSSschen Ebene ist S^1 die „Peripherie der Einheitskreisscheibe" um den Nullpunkt. Wir nennen S^1 kurz *die Kreislinie* oder die *Kreisgruppe*, sie wird in 5.2 zur Beschreibung der orthogonalen Gruppe $O(\mathbb{C})$ herangezogen und spielt bei der Einführung von Polarkoordinaten im § 6 eine entscheidende Rolle.

5. Quadratische Gleichungen. Zu jeder reellen Zahl $r \geqslant 0$ gibt es genau eine reelle Zahl $s \geqslant 0$ mit $s^2 = r$; man nennt s *die* nichtnegative Quadratwurzel aus r und schreibt $s = \sqrt{r}$ (wir haben dies bereits bei der Definition von $|z|$ benutzt!). Aus negativen reellen Zahlen lassen sich keine reellen Quadratwurzeln ziehen. Im Komplexen existieren stets komplexe Quadratwurzeln.

Satz. *Es sei* $c = a + ib$, $a, b \in \mathbb{R}$, *irgendeine komplexe Zahl. Man setze*

$$\zeta := \sqrt{\frac{1}{2}(a + |c|)} + i\frac{b}{|b|}\sqrt{\frac{1}{2}(-a + |c|)}, \qquad \textit{wenn} \qquad b \neq 0;$$

$$\zeta := \sqrt{|c|}, \quad \textit{wenn} \quad b = 0, \quad a \geqslant 0; \qquad \zeta := i\sqrt{|c|}, \quad \textit{wenn} \quad b = 0, \quad a < 0.$$

Dann gilt stets: $\zeta^2 = c$.

Der Beweis erfolgt durch direkte Verifikation. Wie im Reellen heißt ζ eine *Quadratwurzel* aus c, in Zeichen $\sqrt{c}$. Neben ζ ist $-\zeta$ die einzige weitere Quadratwurzel aus c; das Symbol $\sqrt{c}$ ist also zweiwertig. Es lassen sich nun auch sofort alle normierten quadratischen Gleichungen

$$z^2 + az + b = 0, \qquad a, b \in \mathbb{C},$$

lösen. Mittels des uralten Tricks der Babylonier von der quadratischen Ergänzung geht man zur *reinen* Gleichung

$$z^2 + az + b = (z + \tfrac{1}{2}a)^2 + \tfrac{1}{4}(4b - a^2) = 0$$

über und liest die beiden Lösungen z_1, z_2 ab:

$$z_1 := -\tfrac{1}{2}a + \tfrac{1}{2}\sqrt{a^2 - 4b}, \qquad z_2 := -\tfrac{1}{2}a - \tfrac{1}{2}\sqrt{a^2 - 4b},$$

wobei $\sqrt{a^2 - 4b}$ in beiden Fällen *dieselbe* Quadratwurzel bedeutet. Man hat die lineare Faktorisierung

$$z^2 + az + b = (z - z_1)(z - z_2)$$

und speziell die aus dem Schulunterricht bekannte

*Viètasche Wurzelregel**): $\qquad z_1 + z_2 = -a, \qquad z_1 z_2 = b.$

In 6.3 werden die Lösungen quadratischer Gleichungen in Polarkoordinaten angegeben; dort werden allgemeiner n-te Wurzeln mit Hilfe von Polarkoordinaten bestimmt. Es ist nicht möglich, elementar zu zeigen, daß bei vorgegebenem $n \in \mathbb{N}$, $n \geqslant 3$, zu jeder Zahl $w \in \mathbb{C}$ ein $z \in \mathbb{C}$ mit $z^n = w$ existiert.

In 5.2 werden wir benutzen:

Zu jeder Zahl $z = x + iy \in S^1$ *mit* $x \geqslant 0$ *gibt es ein* $w = u + iv \in S^1$ *mit*

$$w^2 = z, \qquad u \geqslant 0, \qquad |v| \leqslant \frac{1}{\sqrt{2}}|y|.$$

---

*) François Viète (1540–1603, Paris; Staatsbeamter). Er führte die Buchstabenrechnung ein: Vokale für unbekannte, Konsonanten für bekannte Größen.

Beweis. Es gibt ein $w = u + iv \in \mathbb{C}$ mit $w^2 = z$. Wegen $|w|^2 = |z| = 1$ gilt $w \in S^1$; wegen $(-w)^2 = z$ dürfen wir $u \geqslant 0$ annehmen. Aus $u^2 + v^2 = 1$, $u^2 - v^2 = x$ und $x \geqslant x^2$ (wegen $0 \leqslant x \leqslant 1$) folgt nun $2v^2 = 1 - x \leqslant 1 - x^2 = y^2$, das heißt, $\sqrt{2}|v| \leqslant |y|$.

§ 4. Geometrische Eigenschaften des Körpers $\mathbb{C}$

In diesem Paragraphen stehen das Skalarprodukt $\langle w, z\rangle$, die Längenfunktion $|z|$ und das *Doppelverhältnis von vier Punkten* in $\mathbb{C}$ im Mittelpunkt der Betrachtungen. Wir beweisen u. a. den fast 2000 Jahre alten Sehnenviereckssatz von PTOLEMÄUS und den Satz von der SIMSONschen Geraden; es sei deutlich gesagt, daß diese geometrischen Anwendungen aus historischen Gründen gewählt sind; es lassen sich leicht viele weitere, ebenso nette und weniger bekannte Anwendungen finden. Wir verweisen diesbezüglich auf I. M. YAGLOM: Complex Numbers in Geometry, Academic Press, New York 1968.

1. Die Identität $\langle w, z\rangle^2 + \langle iw, z\rangle^2 = |w|^2|z|^2$. Da stets $\mathrm{Re}(iz) = -\mathrm{Im}\, z$, so gilt immer $\langle iw, z\rangle = -\mathrm{Im}\, w\bar{z}$. Hieraus ergibt sich mit Hilfe der Produktregel 3.4 die nützliche Identität (vgl. auch Grundwissen Mathematik, Bd. 2, Lineare Algebra und analytische Geometrie, S. 133)

$$\langle w, z\rangle^2 + \langle iw, z\rangle^2 = |w|^2|z|^2, \qquad w, z \in \mathbb{C}. \tag{1}$$

Beweis. $\langle w, z\rangle^2 + \langle iw, z\rangle^2 = (\mathrm{Re}\, w\bar{z})^2 + (-\mathrm{Im}\, w\bar{z})^2 = |w\bar{z}|^2 = |w|^2|z|^2$. □

Als Korollar erhält man die

Ungleichung von CAUCHY-SCHWARZ. $|\langle w, z\rangle| \leqslant |w|\,|z|$ *für alle* $w, z \in \mathbb{C}$; *das Gleichheitszeichen gilt genau dann, wenn* w, z *reell linear abhängen.*

Beweis. Die Ungleichung ist in (1) enthalten, Gleichheit besteht nach (1) genau dann, wenn $\langle iw, z\rangle = -\mathrm{Im}\, w\bar{z} = 0$, das heißt, wenn $w\bar{z} \in \mathbb{R}$. □

Wir geben einen zweiten Beweis, der die Produktregel und die per definitionem klaren Ungleichungen $|\mathrm{Re}\, z| \leqslant |z|$, $|\mathrm{Im}\, z| \leqslant |z|$, $z \in \mathbb{C}$, benutzt: Man hat $|\langle w, z\rangle| = |\mathrm{Re}(w\bar{z})| \leqslant |w\bar{z}| = |w|\,|\bar{z}| = |w|\,|z|$; dabei gilt $|\mathrm{Re}(w\bar{z})| = |w\bar{z}|$ genau dann, wenn $w\bar{z} \in \mathbb{R}$.

2. Cosinussatz und Dreiecksungleichung. Wie für jedes Skalarprodukt gilt

$$|w + z|^2 = |w|^2 + |z|^2 + 2\,\mathrm{Re}(w\bar{z}) \qquad (\textit{Cosinussatz}).$$

Beweis. Wegen der Additivität und der Symmetrie von $\langle w, z\rangle$ hat man

$$\begin{aligned} |w + z|^2 = \langle w + z, w + z\rangle &= \langle w, w\rangle + \langle w, z\rangle + \langle z, w\rangle + \langle z, z\rangle \\ &= |w|^2 + 2\,\mathrm{Re}(w\bar{z}) + |z|^2. \end{aligned}$$

□

Wir kommen auf den Cosinussatz in 6.2 zurück, wo auch die Wortwahl ihre Erklärung finden wird. Mit Hilfe der Ungleichung von CAUCHY-SCHWARZ folgt die

Dreiecksungleichung. *Für alle* $w, z \in \mathbb{C}$ *gilt*: $|w + z| \leqslant |w| + |z|$, *das Gleichheitszeichen gilt genau dann, wenn* $w\bar{z} \geqslant 0$.

Beweis. $|w + z|^2 = |w|^2 + 2\langle w, z\rangle + |z|^2 \leqslant |w|^2 + 2|w|\,|z| + |z|^2 = (|w| + |z|)^2$. Nach CAUCHY-SCHWARZ gilt: $|\langle w, z\rangle| = |w|\,|z| \Leftrightarrow w\bar{z} \in \mathbb{R}$. Daher tritt der Fall $\langle w, z\rangle = |w|\,|z|$ genau dann ein, wenn $w\bar{z} \geqslant 0$.

Eine Abbildung $|\ |: K \to \mathbb{R}$ eines (kommutativen) Körpers K in $\mathbb{R}$ heißt *Bewertung von* K, wenn für alle $w, z \in K$ folgendes gilt:

1) $\quad |z| \geqslant 0, \qquad |z| = 0 \quad \Leftrightarrow \quad z = 0,$

2) $\quad |wz| = |w|\,|z| \qquad$ (*Produktregel*),

3) $\quad |w + z| \leqslant |w| + |z| \qquad$ (*Dreiecksungleichung*).

Ein Körper zusammen mit einer Bewertung heißt ein *bewerteter Körper*. Die Körper $\mathbb{Q}$ und $\mathbb{R}$ sind bewertete Körper. Wir haben gesehen, daß $\mathbb{C}$ durch die Betragsfunktion $|\ |: \mathbb{C} \to \mathbb{R}$, $z \mapsto |z|$ bewertet wird, und daß diese Bewertung eine Fortsetzung der Bewertung von $\mathbb{R}$ durch den Absolutbetrag ist.

Ein schönes Wechselspiel zwischen Betragsfunktion und den Körperoperationen zeigt sich im

Satz vom Dreiparteiensystem. *Es seien* z_1, z_2, z_3 *verschiedene komplexe Zahlen mit* $|z_1| = |z_2| = |z_3|$. *Dann sind folgende Aussagen äquivalent*:

i) z_1, z_2, z_3 *sind die Eckpunkte eines gleichseitigen Dreiecks*: $|z_1 - z_2| = |z_1 - z_3| = |z_2 - z_3|$.

ii) $z_1 + z_2 + z_3 = 0$.

iii) z_1, z_2, z_3 *annullieren ein Polynom* $z^3 - c$ *mit* $c \in \mathbb{C}^\times$.

Stellt man sich unter z_1, z_2, z_3 politische Parteien vor und interpretiert man „längengleich" als „*gleichmächtig*", so motiviert die Implikation i) $\Rightarrow$ ii) die Bezeichnung des Satzes.

Den Beweis überlassen wir dem Leser: man reduziere auf den Fall $z_1 z_2 z_3 = 1$ und betrachte zum Beweis von ii) $\Rightarrow$ iii) den Ausdruck $z_1 z_2 z_3(\bar{z}_1 + \bar{z}_2 + \bar{z}_3)$. □

Definiert man den *Schwerpunkt eines Dreiecks* mit den Eckpunkten z_1, z_2, z_3 durch $\frac{1}{3}(z_1 + z_2 + z_3)$, so besagt die Äquivalenz von i) und ii), daß der Schwerpunkt genau dann der Mittelpunkt des Umkreises ist, wenn das Dreieck gleichseitig ist. □

In Analogie zum Vorangehenden hat man für vier verschiedene komplexe Zahlen $z_1, \ldots, z_4 \in \mathbb{C}$ mit $|z_1| = \cdots = |z_4|$ folgende Äquivalenz:

i) z_1, z_2, z_3, z_4 *sind die Eckpunkte eines Rechtecks.*

ii) $z_1 + z_2 + z_3 + z_4 = 0$.

iii) $z_1, \ldots, z_4$ *annullieren ein Polynom* $(z^2 - a^2)(z^2 - b^2)$ *mit* $|a| = |b| \neq 0$.

3. Zahlen auf Geraden und Kreisen. Doppelverhältnis. Zwei Zahlen $a, b \in \mathbb{C}$ liegen genau dann auf einer Geraden durch 0, wenn $a\bar{b} \in \mathbb{R}$ (vgl. 3.1). Allgemeiner gilt:

Drei Zahlen $a, b, c \in \mathbb{C}$, $a \neq b$, liegen genau dann auf einer Geraden, wenn

$$\text{(1)} \qquad \frac{c-a}{b-a} \in \mathbb{R}, \qquad \textit{das heißt, wenn} \qquad c\bar{b} - c\bar{a} - a\bar{b} \in \mathbb{R}.$$

Der Beweis ist trivial, da die durch a und b laufende Gerade die Parameterdarstellung $a + (b-a)s$, $s \in \mathbb{R}$, hat. □

Sind $a, b, c, d \in \mathbb{C}$ mit $a \neq d$, $b \neq c$, so heißt

$$\text{(2)} \qquad DV(a,b,c,d) := \frac{a-b}{a-d} : \frac{c-b}{c-d} = \frac{(a-b)(c-d)}{(a-d)(c-b)}$$

$$= \frac{(a-b)(c-d)(\bar{a}-\bar{d})(\bar{c}-\bar{b})}{|a-d|^2|c-b|^2} \in \mathbb{C}$$

das *Doppelverhältnis* von a, b, c, d. Diese Zahl hängt von der Reihenfolge der Zahlen $a, \ldots, d$ ab; bei zyklischer Vertauschung erhält man z. B. reziproke Werte:

$$DV(b,c,d,a) = DV(a,b,c,d)^{-1}.$$

Wir zeigen nun:

Satz. *Vier Zahlen $a, b, c, d \in \mathbb{C}$, $a \neq d$, $b \neq c$, die nicht alle auf einer Geraden liegen, liegen genau dann auf einer Kreislinie, wenn ihr Doppelverhältnis reell ist.*

Beweis. Seien etwa a, b, c nicht auf einer Geraden. Da diese Bedingung und das Doppelverhältnis *translationsinvariant* sind, dürfen wir annehmen, daß der Umkreis des Dreiecks mit den Eckpunkten a, b, c den Nullpunkt zum Mittelpunkt hat. Dann gilt $|a| = |b| = |c|$ und also

$$(a-b)(c-d)(\bar{a}-\bar{d})(\bar{c}-\bar{b}) - i(|c|^2 - |d|^2)\,\mathrm{Im}(c\bar{b} - c\bar{a} - a\bar{b}) \in \mathbb{R}$$

nach Aufgabe 3.1. Da a, b, c nicht auf einer Geraden liegen, gilt $\mathrm{Im}(c\bar{b} - c\bar{a} - a\bar{b}) \neq 0$ nach (1). Damit folgt:

$$(a-b)(c-d)(\bar{a}-\bar{d})(\bar{c}-\bar{b}) \in \mathbb{R} \Leftrightarrow |c| = |d|.$$

Dies ist aufgrund von (2) die Behauptung. □

In der Theorie der gebrochen linearen Transformationen $z \mapsto \dfrac{az+b}{cz+d}$ spielt das Doppelverhältnis eine zentrale Rolle; dabei wird auch der Fall, daß ein Argument ∞ ist, zugelassen. Das Doppelverhältnis ist invariant gegenüber gebrochen linearen Transformationen, dies ermöglicht einen neuen Beweis des vorstehenden Satzes, vgl. z. B. Fischer/Lieb: Funktionentheorie, S. 236 (Vieweg-Verlag 1980).

4. Sehnenvierecke und Doppelverhältnis. Je vier verschiedene Punkte $a, b, c, d \in \mathbb{C}$ bestimmen ein Viereck *abcd* in $\mathbb{C}$ mit Ecken a, b, c, d, dessen Seiten die Verbindungsstrecken von a nach b, von b nach c usw. sind. Ein Viereck heißt ein *Sehnenviereck*, wenn seine Ecken alle auf einer Kreislinie liegen und wenn zwei verschiedene Seiten sich höchstens in einem Eckpunkt schneiden (die Figur des nächsten Abschnitts zeigt ein Sehnenviereck *abcd*; das durch Vertauschung der Ecken b und c entstehende Viereck *acbd* ist kein Sehnenviereck).

Satz. *Ein Viereck abcd ist genau dann ein Sehnenviereck, wenn das Doppelverhältnis* $DV(a, b, c, d)$ *negativ ist.*

Beweis (mittels eines Stetigkeitsargumentes). Es sei S^1 die gegebene Kreislinie. Die Quadrate Q bzw. Q' mit den Ecken $1, i, -1, -i$ bzw. $1, -i, -1, i$ sind Sehnenvierecke, das Doppelverhältnis ihrer Eckpunkte ist -1. „Ersichtlich" ist ein Viereck V mit Ecken auf S^1 genau dann ein Sehnenviereck, wenn V aus Q bzw. Q' durch *stetiges* Verschieben der Ecken längs S^1 so entsteht, daß dabei niemals zwei Ecken zusammenfallen.

Da das (nach Satz 3) reelle Doppelverhältnis von vier verschiedenen Punkten auf S^1 eine stetige Funktion ihrer Argumente ist, die nie verschwindet, so ist nach dem Zwischenwertsatz ein Viereck mit Ecken $a, b, c, d \in S^1$ genau dann ein Sehnenviereck, wenn $DV(a, b, c, d) < 0$.

5. Satz von PTOLEMÄUS. Der ägyptische Mathematiker Claudius Ptolemäus (Alexandria, um 150 n. Chr.) bewies in seinem Almagest, Buch 1, Kap. 10, folgenden Satz, der heute noch gelegentlich im Schulunterricht besprochen wird:

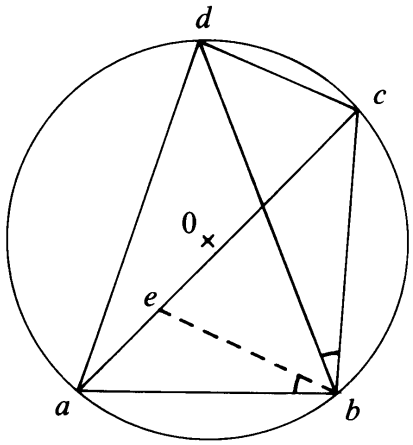
In jedem Sehnenviereck abcd ist die Summe der Produkte gegenüberliegender Seiten gleich dem Produkt der Diagonalen:

$$|a-b| \cdot |c-d| + |a-d| \cdot |c-b| = |a-c| \cdot |b-d|.$$

Ptolemäus stellte diesen Satz in den Dienst der Astronomie und benutzte ihn als Hilfsmittel bei der Berechnung seiner berühmten Sehnentafeln: ist nämlich eine Seite ein Kreisdurchmesser, so entsteht leicht das Additionstheorem

$$\sin(\alpha - \beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta.$$

Ptolemäus bewies seinen Satz durch einen eleganten elementargeometrischen Trick (vgl. Figur): er bestimmt den Punkt e auf der Strecke ac so, daß gilt $\measuredangle\, abe = \measuredangle\, cbd$. Dann sind die Dreiecke *abe* und *bcd* ähnlich (!), jetzt führen einfache Überlegungen zum Ziel. □



Um den Satz von Ptolemäus und mehr mit komplexen Methoden zu beweisen, ordnen wir jedem Viereck *abcd* in $\mathbb{C}$ die „Ptolemäuszahl"

$$P(abcd) := |(a-b)(c-d)| + |(a-d)(c-b)| - |(a-c)(b-d)|$$

zu. Da $(a-b)(c-d)-(a-d)(c-b)=(a-c)(b-d)$ für jeden kommutativen Ring gilt, und da $DV(a,b,c,d)=(a-b)(c-d)(a-d)^{-1}(c-b)^{-1}$, so zeigt eine direkte Verifikation:

$$P(abcd)=|DV(a,b,c,d)|+1-|DV(a,b,c,d)-1|.$$

Da $|w-1|=|w|+1$ nach der Dreiecksungleichung 4.2 genau dann gilt, wenn w reell und $\leqslant 0$ ist, so haben wir aufgrund von Satz 4 gezeigt:

Satz. *Folgende Aussagen über ein Viereck abcd in* $\mathbb{C}$ *sind äquivalent*:

i) *Für abcd gilt die Aussage des Satzes von PTOLEMÄUS*: $P(abcd)=0$.
ii) *Das Viereck abcd ist ein Sehnenviereck.*

Die Umkehrung der Ptolemäischen Aussage, das heißt die Implikation i) $\Rightarrow$ ii), wurde 1832 im CRELLEschen Journal Bd. 8, S. 320, als Aufgabe gestellt. Lösungen findet man in den Bänden 10, S. 41; 11, 264–271 und 13, 233–236; u. a. gab CLAUSEN eine elegante Lösung.

6. SIMSONsche Gerade. Es seien $a,b,u\in\mathbb{C}$, $a\neq b$. Der *Fußpunkt* v des *Lotes* von u auf die durch a und b gehende Gerade $L:=\{z=a+s(b-a):s\in\mathbb{R}\}$ ist, da $i(b-a)$ auf $b-a$ senkrecht steht, der Schnittpunkt von L mit der Geraden $L':=\{z=u+it(a-b)\}$, vgl. Fig. a. Dies gibt für s,t die Bedingung $s-ti=(u-a)(b-a)^{-1}$, also $2s=(u-a)(b-a)^{-1}+(\bar{u}-\bar{a})(\bar{b}-\bar{a})^{-1}$ und daher

$$v=\frac{1}{2}\left[a+u+(\bar{u}-\bar{a})\frac{b-a}{\bar{b}-\bar{a}}\right].$$

Im Falle $|a|=|b|$ gilt $(b-a)(\bar{b}-\bar{a})^{-1}=-b(\bar{a})^{-1}$ und folglich

$$(*)\qquad v=\frac{1}{2}\left(a+b+u-\bar{u}\frac{ab}{|a|^2}\right),\qquad \text{falls}\qquad |a|=|b|.$$

Wir benutzen (*) zum Beweis einer in der Elementargeometrie wenig bekannten Aussage über „drei merkwürdige*) Punkte“ eines Dreiecks.

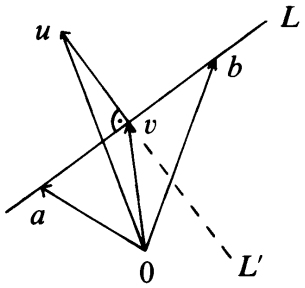


Fig. a

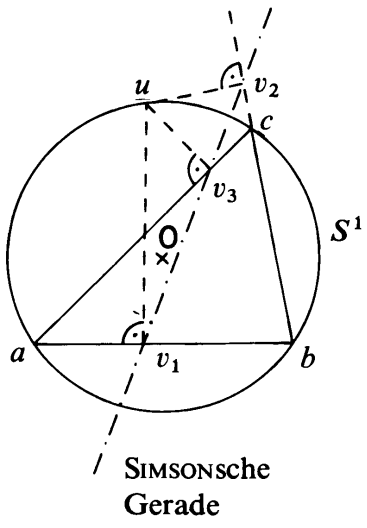


Fig. b

*) In der klassischen Elementargeometrie verwendet man das Wort „merkwürdig“ im Sinne von „des Merkens würdig“.

Satz. *Es seien $a, b, c \in \mathbb{C}$ Eckpunkte eines Dreiecks, und es seien v_1, v_2, v_3 die Lotfußpunkte eines beliebigen Punktes $u \in \mathbb{C}$ auf die durch a, b bzw. b, c bzw. c, a bestimmte Gerade. Dann sind folgende Aussagen äquivalent (vgl. Fig. b):*

i) *Die Punkte v_1, v_2, v_3 liegen auf einer Geraden.*
ii) *Der Punkt u liegt auf dem Umkreis des Dreiecks mit den Ecken a, b, c.*

Beweis. Wir dürfen annehmen, daß S^1 der Umkreis ist. Aufgrund von (*) gilt dann, wenn wir zunächst $v_2 \neq v_3$, $u \neq 0$ unterstellen

$$\frac{v_1 - v_3}{v_2 - v_3} = \frac{b - c - \bar{u}ab + \bar{u}ac}{b - a - \bar{u}bc + \bar{u}ac} = \frac{(c-b)(\bar{u}a - 1)}{(a-b)(\bar{u}c - 1)} = \frac{c-b}{c - \bar{u}^{-1}} : \frac{a-b}{a - \bar{u}^{-1}}$$

$$= DV(c, b, a, \bar{u}^{-1}).$$

Die Äquivalenz i) $\Leftrightarrow$ ii) ergibt sich nun aufgrund der Resultate des Abschnittes 3, da $\bar{u}^{-1} \in S^1$ mit $u \in S^1$ äquivalent ist.

Der Fall $v_2 = v_3$ ist wegen $a \neq b$ nur möglich, wenn $\bar{u}c = 1$, das heißt, wenn $u \in S^1$. Im Fall $u = 0$ gilt $(v_1 - v_3) : (v_2 - v_3) = (c - b) : (a - b)$, daher liegen jetzt v_1, v_2, v_3 auf keiner Geraden, da a, b, c es nicht tun. □

Falls u auf dem Umkreis liegt, so heißt die Gerade durch v_1, v_2, v_3 die SIMSONsche Gerade nach dem englischen Mathematiker Robert SIMSON (1687–1768, versuchte in England erfolgreich die Wiederbelebung der antiken griechischen Geometrie); der Satz war aber bereits früher von SIMSONS Landsmann John WALLIS (1616–1703) bewiesen worden.

§ 5. Die Gruppen $O(\mathbb{C})$ und $SO(2)$

In diesem Paragraphen zeigen wir u. a., daß die Kreisgruppe S^1 zur orthogonalen Gruppe $SO(2)$ der reellen eigentlich orthogonalen 2×2 Matrizen bezüglich der Abbildung $F: \mathbb{C} \to \mathscr{C}$, $a + bi \mapsto \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ isomorph ist. Wir gewinnen weiter eine klassische Parameterdarstellung der Gruppe $SO(2)$.

1. Abstandstreue Abbildungen von $\mathbb{C}$. Eine (nicht notwendig $\mathbb{R}$-lineare) Abbildung $f: \mathbb{C} \to \mathbb{C}$ heißt *abstandstreu*, wenn gilt

$$|f(w) - f(z)| = |w - z| \qquad \text{für} \qquad w, z \in \mathbb{C}.$$

Satz. *Folgende Aussagen über $f: \mathbb{C} \to \mathbb{C}$ sind äquivalent:*

i) *Es gilt $f(z) = f(0) + cz$ oder $f(z) = f(0) + c\bar{z}$ mit $c \in S^1$.*
ii) *f ist abstandstreu.*

Beweis. i) $\Rightarrow$ ii): Trivial, da $f(w) - f(z) = c(w - z)$ bzw. $= c(\bar{w} - \bar{z})$.

ii) $\Rightarrow$ i): Da $c := f(1) - f(0) \in S^1$, so ist $g: \mathbb{C} \to \mathbb{C}$, $z \mapsto c^{-1}(f(z) - f(0))$ abstandstreu! Wegen $g(0) = 0$ und $g(1) = 1$ gilt $|g(z)|^2 = |z|^2$ und $|g(z) - 1|^2 = |z - 1|^2$.

Hieraus folgt $\operatorname{Re} g(z) = \operatorname{Re} z$, speziell $g(i) = \pm i$. Im Fall $g(i) = i$ ist $\hat{g}(z) := -ig(iz)$ abstandstreu mit $\hat{g}(0) = 0$, $\hat{g}(1) = 1$, daher gilt (nach dem schon Bewiesenen) $\operatorname{Re}(-ig(iz)) = \operatorname{Re} z$, das heißt, $\operatorname{Im} g(z) = \operatorname{Im} z$; also $g(z) = z$ und $f(z) = f(0) + cz$. Im Fall $g(i) = -i$ folgt mit $\hat{g}(z) := ig(iz)$ analog wie eben $\operatorname{Re}(ig(iz)) = \operatorname{Re} z$, das heißt, $\operatorname{Im} g(z) = -\operatorname{Im} z$, also $f(z) = f(0) + c\bar{z}$. □

Speziell ist jede abstandstreue Abbildung von $\mathbb{C}$ in sich, die den Nullpunkt festläßt, $\mathbb{R}$-linear.

In der Linearen Algebra heißt jede abstandstreue Abbildung eines euklidischen Vektorraumes V in sich eine *Bewegung*; die eben hergeleitete Aussage ist dann ein Spezialfall des allgemeinen Satzes, daß jede Bewegung $f: V \to V$ die Form $x \mapsto f(0) + h(x)$ hat, wo $h: V \to V$ orthogonal ist (vgl. Grundwissen Mathematik, Bd. 2, Lineare Algebra und analytische Geometrie, S. 173).

2. Die Gruppe $O(\mathbb{C})$. Eine $\mathbb{R}$-*lineare* Abbildung $f: \mathbb{C} \to \mathbb{C}$ heißt *orthogonal*, wenn stets $\langle f(w), f(z)\rangle = \langle w, z\rangle$. Jede orthogonale Abbildung $f: \mathbb{C} \to \mathbb{C}$ ist *längentreu*: $|f(z)| = |z|$, und also wegen der $\mathbb{R}$-Linearität auch abstandstreu.

Satz. *Eine Abbildung $f: \mathbb{C} \to \mathbb{C}$ ist genau dann orthogonal, wenn gilt*

$$f(z) = cz \quad \text{oder} \quad f(z) = c\bar{z} \quad \text{mit} \quad c \in S^1.$$

Beweis. Die angeschriebenen Abbildungen sind orthogonal, z. B. gilt im zweiten Falle:

$$\langle f(w), f(z)\rangle = \operatorname{Re}(c\bar{w}\overline{(c\bar{z})}) = |c|^2 \operatorname{Re}(\bar{w}z) = \langle w, z\rangle$$

wegen $c \in S^1$.

Ist umgekehrt f orthogonal, so ist f abstandstreu, und die Behauptung folgt aus Satz 1 wegen $f(0) = 0$. □

Aufgabe. Beweisen Sie den Satz direkt, indem Sie die Charakterisierung $\mathbb{R}$-linearer Abbildungen aus 3.1 verwenden und durch Nachrechnen zeigen:

$$|az + b\bar{z}| = |z| \text{ für alle } z \in \mathbb{C} \Leftrightarrow a \in S^1 \text{ und } b = 0 \text{ oder } a = 0 \text{ und } b \in S^1.$$

Die orthogonalen Abbildungen von $\mathbb{C}$ bilden bei Komposition eine nicht abelsche *Gruppe*, die sogenannte *orthogonale Gruppe* $O(\mathbb{C})$. Die orthogonalen Abbildungen der Form $T_c(z) = cz$, $c \in S^1$, heißen *Drehungen* oder auch *eigentlich orthogonal*, sie bilden einen *Normalteiler* $SO(\mathbb{C})$ von $O(\mathbb{C})$. Aus dem Vorangehenden folgt:

Die Abbildung $S^1 \to SO(\mathbb{C})$, $c \mapsto T_c$, ist ein Gruppenisomorphismus.

Speziell ist die Gruppe $SO(\mathbb{C})$ *abelsch*. Die Abbildungen $f(z) = c\bar{z}$, $c \in S^1$, heißen *Spiegelungen*, sie bilden neben $SO(\mathbb{C})$ die einzige Nebenklasse in $O(\mathbb{C})$ bezüglich $SO(\mathbb{C})$.

3. Die Gruppe $SO(2)$ und der Isomorphismus $S^1 \to SO(2)$. Die Menge

$$O(2) := \{A \in GL(2, \mathbb{R}) : AA^t = E\} \tag{1}$$

aller reellen *orthogonalen* 2×2 Matrizen ist eine wichtige Untergruppe der Gruppe $GL(2, \mathbb{R})$ aller reellen *invertierbaren* 2×2 Matrizen. Wegen $\det A = \det A^t$ gilt $\det A = \pm 1$ für alle $A \in O(2)$. Durch

$$SO(2) := \{A \in O(2) : \det A = 1\}$$

wird ein *Normalteiler von* $O(2)$, die Gruppe der *eigentlich orthogonalen* reellen 2×2 Matrizen, definiert. Es gilt, wenn $\mathscr{C}$ den in 2.5 eingeführten Unterkörper von $\mathrm{Mat}(2, \mathbb{R})$ bezeichnet:

Satz. $$SO(2) = \{A \in \mathscr{C} : \det A = 1\}.$$

Beweis. Für $A = \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ verifiziert man sofort $AA^t = (\det A)E$, hieraus folgt $\{A \in \mathscr{C} : \det A = 1\} \subset SO(2)$.

Für $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SO(2)$ gilt $A^{-1} = A^t = \left(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right)$ nach (1). Da andererseits $A^{-1} = \left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$ wegen $\det A = 1$, so folgt $d = a$, $c = -b$, das heißt, $A \in \mathscr{C}$. □

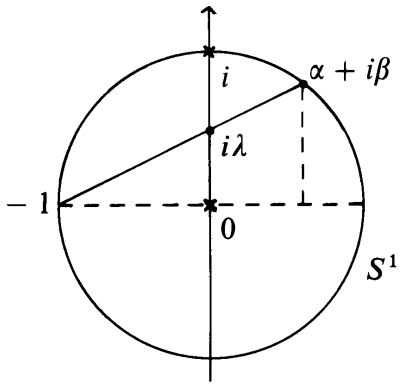
Es ergibt sich nun sofort:

Isomorphiesatz. *Vermöge* $F: \mathbb{C} \to \mathscr{C}$, $a + bi \mapsto \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$, *wird die Kreisgruppe* S^1 *isomorph auf die Gruppe* $SO(2)$ *abgebildet.*

Beweis. Klar wegen $F(S^1) = \{A = \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right) \in \mathscr{C} : \det A = a^2 + b^2 = 1\}$. □

Die orthogonalen Gruppen $SO(3)$ und $SO(4)$ werden in Kap. 6, § 3 mittels Quaternionen beschrieben.

4. Rationale Parametrisierung eigentlich orthogonaler 2×2 Matrizen. Man bildet $S^1 \setminus \{-1\}$ bijektiv auf die imaginäre Achse ab, indem man dem Punkt $\alpha + i\beta \in S^1$



den Schnittpunkt $i\lambda$ der Geraden durch -1 und $\alpha + i\beta$ mit der imaginären Achse zuordnet (vgl. Figur). Eine einfache Rechnung gibt (Strahlensatz):

$$\alpha = \frac{1 - \lambda^2}{1 + \lambda^2}, \qquad \beta = \frac{2\lambda}{1 + \lambda^2}, \qquad \lambda = \frac{\beta}{1 + \alpha}. \tag{1}$$

Es folgt $\alpha + i\beta = \dfrac{1 + i\lambda}{1 - i\lambda}$, also die *rationale Parametrisierung*

$$S^1 \setminus \{-1\} = \left\{ \frac{1 + i\lambda}{1 - i\lambda} : \lambda \in \mathbb{R} \right\}, \tag{2}$$

dabei sind Real- und Imaginärteil von $c := \dfrac{1+i\lambda}{1-i\lambda}$ *genau dann rational, das heißt aus* $\mathbb{Q}$, *wenn* λ *rational ist.*

Mittels $F(S^1) = SO(2)$ übersetzt sich das Ergebnis zu:

$$(3) \qquad SO(2) \setminus \{-E\} = \left\{ \frac{1}{1+\lambda^2} \begin{pmatrix} 1-\lambda^2 & 2\lambda \\ -2\lambda & 1+\lambda^2 \end{pmatrix} : \lambda \in \mathbb{R} \right\},$$

die Matrix ist genau dann rational, wenn λ *rational ist.*

Bemerkung. Man beseitigt die Ausnahmerolle von -1 bzw. $-E$ in den Gleichungen (2) und (3), indem man λ durch λ/κ ersetzt und mit κ erweitert. Dann gilt einschränkungslos

$$(2') \qquad S^1 = \left\{ \frac{\kappa + i\lambda}{\kappa - i\lambda} : (\kappa, \lambda) \in \mathbb{R}^2 \setminus \{0\} \right\}$$

$$= \left\{ \frac{1}{\kappa^2 + \lambda^2} [(\kappa^2 - \lambda^2) + 2\kappa\lambda i] : (\kappa, \lambda) \in \mathbb{R}^2 \setminus \{0\} \right\},$$

$$(3') \qquad SO(2) = \left\{ \frac{1}{\kappa^2 + \lambda^2} \begin{pmatrix} \kappa^2 - \lambda^2 & 2\kappa\lambda \\ -2\kappa\lambda & \kappa^2 + \lambda^2 \end{pmatrix} : (\kappa, \lambda) \in \mathbb{R}^2 \setminus \{0\} \right\}.$$

In 6.3.5 werden wir für die Gruppe $SO(3)$ die berühmte rationale Parameterdarstellung von EULER kennenlernen, welche die Darstellung (3') von $SO(2)$ als Spezialfall umfaßt.

Die Darstellung (3) orthogonaler 2×2 Matrizen A ist nichts anderes als die CAYLEYsche Darstellung

$$(*) \qquad A = (E - X)^{-1}(E + X), \qquad X \in \mathrm{Mat}(2, \mathbb{R}) \text{ schiefsymmetrisch.}$$

Da nämlich $X = \lambda \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, $\lambda \in \mathbb{R}$, *alle* schiefsymmetrischen 2×2 Matrizen sind und da $X^2 = -\lambda^2 E$, so ist $(*)$ das Analogon zur Gleichung $(1 - \lambda i)^{-1}(1 + \lambda i)$; wegen $(E - X)^{-1} = (1 + \lambda^2)^{-1}(E + X)$ gilt

$$A = (1+\lambda^2)^{-1}(E+X)^2 = (1+\lambda^2)^{-1}[(1-\lambda^2)E + 2X] = (1+\lambda^2)^{-1} \begin{pmatrix} 1-\lambda^2 & 2\lambda \\ -2\lambda & 1-\lambda^2 \end{pmatrix}. \quad \square$$

In den Gleichungen (1) für rationale Punkte auf S^1 sind die sogenannten „indischen Formeln" für pythagoräische Tripel enthalten. Ein Tripel k, l, m natürlicher Zahlen $\neq 0$ heißt *pythagoräisch*, wenn $k^2 + l^2 = m^2$. Dann ist wenigstens eine der Zahlen k, l gerade (Beweis!); wir zeigen:

Ist (k, l, m) *ein pythagoräisches Tripel, und ist* l *gerade, so gibt es natürliche Zahlen* $r, s, t \neq 0$, *so daß gilt:*

$$k = (r^2 - s^2)t, \qquad l = 2rst, \qquad m = (r^2 + s^2)t \qquad (\textit{indische Formeln}).$$

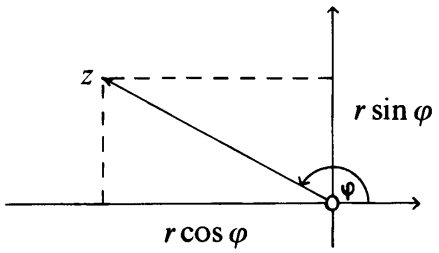
Beweis. Zu $m^{-1}k + im^{-1}l \in S^1 \setminus \{-1\}$ gibt es ein $\lambda = \frac{s}{r}, r, s \in \mathbb{N} \setminus 0$, so daß nach (1) gilt: $k = (r^2 - s^2)\dfrac{m}{r^2 + s^2}$, $l = 2rs\dfrac{m}{r^2 + s^2}$. Wählt man r, s teilerfremd, so sind auch $r^2 + s^2$, rs teilerfremd (Beweis!). Da $\frac{1}{2}l = \dfrac{rsm}{r^2 + s^2} \in \mathbb{N}$, so folgt $t := \dfrac{m}{r^2 + s^2} \in \mathbb{N}$ und somit die Behauptung.

§ 6. Polarkoordinaten und n-te Wurzeln

In der Zahlenebene führt man Polarkoordinaten ein, indem man jeden Punkt $z \in \mathbb{C} = \mathbb{R}^2$ in der Form $(r\cos\varphi, r\sin\varphi)$ schreibt (Figur). Dabei ist $r := |z|$ die Entfernung von z zum Nullpunkt, und φ ist der Winkel (im Bogenmaß) zwischen der positiven x-Achse und dem Ortsvektor von z. Jede komplexe Zahl $z \neq 0$ hat somit die Form

$$z = r(\cos\varphi + i\sin\varphi),$$

der Winkel φ ist bis auf ein ganzzahliges Vielfaches von 2π eindeutig bestimmt.



Die Präzisierung dieser intuitiv klaren Dinge ist *nicht trivial.* Man benötigt Eigenschaften der Cosinus- und Sinusfunktion, die zwar wohlbekannt sind, deren Beweise aber tiefer liegen. Wir arbeiten im folgenden vorwiegend mit der in ganz $\mathbb{C}$ definierten komplexen Exponentialfunktion

$$\exp z = \sum_0^\infty \frac{z^\nu}{\nu!}.$$

Wir setzen $e^{i\varphi} := \exp(i\varphi)$ und werden entscheidend heranziehen

Epimorphiesatz. *Die Abbildung $p: \mathbb{R} \to S^1$, $\varphi \mapsto e^{i\varphi}$ ist ein Gruppenepimorphismus der (additiven) Gruppe $\mathbb{R}$ auf die (multiplikative) Kreisgruppe S^1. Es gibt genau eine positive reelle Zahl π, so daß gilt:*

a) *Die Gruppe $2\pi\mathbb{Z}$ ist der Kern $\{r \in \mathbb{R} : p(r) = 1\}$ von p, speziell gilt:*

$$p(\varphi) = p(\psi) \Leftrightarrow \varphi - \psi \in 2\pi\mathbb{Z}; \qquad p([0, 2\pi)) = S^1.$$

b) $p(\frac{\pi}{2}) = i$.

Aus b) folgt automatisch $p(\pi) = -1, p(\frac{3}{2}\pi) = -i$. Wir nennen p den *Polarkoordinatenepimorphismus.* Der Zusammenhang zwischen p und den trigonometrischen Funktionen

$$\cos z := \sum_0^\infty \frac{(-1)^\nu}{(2\nu)!} z^{2\nu}, \qquad \sin z := \sum_0^\infty \frac{(-1)^\nu}{(2\nu+1)!} z^{2\nu+1}, \qquad z \in \mathbb{C},$$

wird durch die EULERsche Formel

$$\exp iz = \cos z + i\sin z$$

vermittelt, die offensichtlich impliziert:

c) $p(\varphi) = e^{i\varphi} = \cos\varphi + i\sin\varphi$ *für alle* $\varphi \in \mathbb{R}$.

Die EULERsche Formel und vor allem der Epimorphiesatz werden ausführlich im Kapitel 5 diskutiert, vgl. insbesondere 5.3.1 und 5.3.6.

1. Polarkoordinaten. Der Epimorphiesatz hat zur Folge:

Satz. *Jede komplexe Zahl $z \in \mathbb{C}^\times$ läßt sich eindeutig schreiben in der Form*

(1) $z = re^{i\varphi} = r(\cos\varphi + i\sin\varphi)$ *mit* $r := |z|$ *und* $\varphi \in [0, 2\pi)$.

Für jede weitere Darstellung $z = \rho e^{i\psi} = \rho(\cos\psi + i\sin\psi)$ mit $\rho, \psi \in \mathbb{R}$, $\rho > 0$, gilt $\rho = r$ und $\psi = \varphi + 2n\pi$ mit $n \in \mathbb{Z}$.

Beweis. Da $r^{-1}z \in S^1$, so gibt es ein $\varphi \in [0, 2\pi)$ mit $p(\varphi) = r^{-1}z$. Dies bedeutet $z = re^{i\varphi} = r(\cos\varphi + i\sin\varphi)$. Aus $z = \rho e^{i\psi}$ mit $\rho > 0$, $\psi \in \mathbb{R}$ folgt $|z| = \rho$ wegen $e^{i\psi} \in S^1$. Es ist nun $e^{i\varphi} = e^{i\psi}$, das heißt, $\varphi - \psi \in 2\pi\mathbb{Z}$. □

Die Gleichung (1) heißt eine *Polarkoordinatendarstellung*, die Zahlen r, φ und allgemeiner r, ψ, wo $\psi = \varphi + 2n\pi$, heißen *Polarkoordinaten von z*. Die Zahl $\varphi \in [0, 2\pi)$ heißt *Argument* oder *Amplitude von $z \in \mathbb{C}^\times$*.

Polarkoordinaten wurden bereits 1671 von NEWTON bei der Untersuchung von Spiralen in der Ebene benutzt. Die Darstellung komplexer Zahlen in Polarkoordinaten erscheint zuerst bei EULER und D'ALEMBERT; den Faktor $\cos\varphi + i\sin\varphi$ nennt CAUCHY 1821 „expression réduite" (im „Cours d'Analyse").

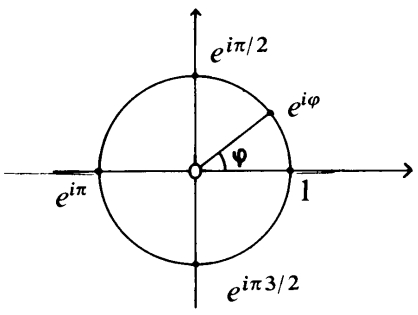
Die Zahlen $1, i, -1, -i$ haben die Polarkoordinatendarstellung

$$1 = 1 \cdot (\cos 0 + i\sin 0), \qquad i = 1 \cdot \left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right),$$

$$-1 = 1 \cdot (\cos\pi + i\sin\pi), \qquad -i = 1 \cdot \left(\cos\frac{3\pi}{2} + i\sin\frac{3\pi}{2}\right);$$

wir haben also die nachstehende klassische Figur mit den 4 Werten

$$e^{i\pi/2} = i, \qquad e^{i\pi} = -1, \qquad e^{i3\pi/2} = -i, \qquad e^{2\pi i} = 1;$$



dies sind Spezialfälle der Identität

$$i^m = (e^{i\pi/2})^m = e^{im\pi/2}, \qquad m \in \mathbb{Z}.$$

Die Darstellung konjugiert komplexer Zahlen sowie von Inversen ist einfach in Polarkoordinaten. Da $\cos(-\varphi) = \cos\varphi$ und $\sin(-\varphi) = -\sin\varphi$, so folgt:

Falls $z = |z|e^{i\varphi} = |z|(\cos\varphi + i\sin\varphi)$, *so gilt*:

(2) $\bar{z} = |z|e^{-i\varphi} = |z|(\cos\varphi - i\sin\varphi), \quad z^{-1} = |z|^{-1}e^{-i\varphi} = |z|^{-1}(\cos\varphi - i\sin\varphi).$

Die zweite Gleichung ergibt sich aus der ersten wegen $z^{-1} = |z|^{-2}\bar{z}$.

Die *reelle* Polarkoordinatenabbildung

$$\{r \in \mathbb{R} : r > 0\} \times \mathbb{R} \to \mathbb{C}^\times, \qquad (r, \varphi) \mapsto (x, y) := (r\cos\varphi, r\sin\varphi)$$

ist beliebig oft *reell differenzierbar*, für ihre Funktionaldeterminante gilt:

$$\det\begin{pmatrix} x_r & x_\varphi \\ y_r & y_\varphi \end{pmatrix} = \det\begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix} = r \neq 0,$$

daher existiert lokal überall eine reell differenzierbare Umkehrabbildung (die durch $(x, y) \mapsto \left(\sqrt{x^2 + y^2}, \arccos \frac{x}{\sqrt{x^2 + y^2}}\right)$ beschrieben wird, wenn der Zweig der Arcuscosinusfunktion richtig gewählt wird).

2. Multiplikation komplexer Zahlen in Polarkoordinaten. Da $e^{i\psi}e^{i\varphi} = e^{i(\psi+\varphi)}$ wegen der Homomorphieeigenschaft von p, so folgt direkt für $w, z \in \mathbb{C}^\times$:

Satz. *Falls*

$$w = |w|e^{i\psi} = |w|(\cos\psi + i\sin\psi), \qquad z = |z|e^{i\varphi} = |z|(\cos\varphi + i\sin\varphi),$$

so gilt:

(1) $$wz = |w|\,|z|e^{i(\psi+\varphi)} = |w|\,|z|(\cos(\psi + \varphi) + i\sin(\psi + \varphi)),$$

also auch

$$\frac{w}{z} = \frac{|w|}{|z|}e^{i(\psi-\varphi)} = \frac{|w|}{|z|}(\cos(\psi - \varphi) + i\sin(\psi - \varphi)).$$

Man erhält somit das Produkt bzw. den Quotienten zweier komplexer Zahlen, indem man ihre Beträge multipliziert bzw. dividiert und ihre Argumentwinkel addiert bzw. subtrahiert (Fig. a). Die Gleichung (1) ist fundamental und weitaus mehr als nur eine bequeme Rechenregel, die die vorteilhafte Verwendung von Polarkoordinaten bei der Multiplikation komplexer Zahlen evident macht. Sie ist eine tiefe und unerwartete Rechtfertigung für die geometrische Deutung der

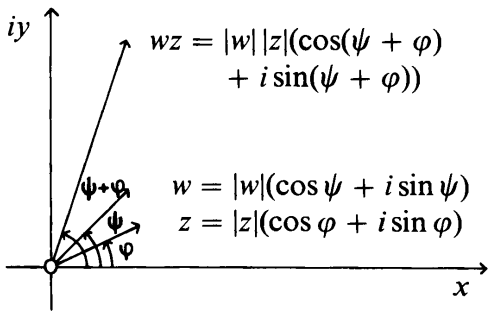


Fig. a

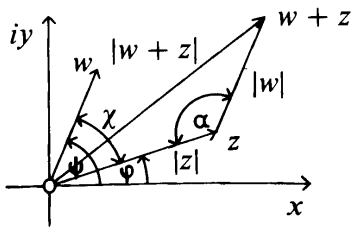


Fig. b

komplexen Zahlen in der Ebene. Die mathematische Kraft dieser Gleichung hat schon EULER gekannt.*)

Das Skalarprodukt $\langle w, z\rangle = \mathrm{Re}(w\bar{z})$ erhält, wenn man Gleichung (1) in der Form $w\bar{z} = |w|\,|z|(\cos(\psi - \varphi) + i\sin(\psi - \varphi))$ verwendet, die bekannte Form $\langle w, z\rangle = |w|\,|z|\cos\chi$, wo $\chi := \psi - \varphi$ der „Winkel zwischen den Vektoren w und z" ist (Fig. b). Jetzt wird verständlich, warum in 4.2 die Gleichung $|w + z|^2 = |w|^2 + |z|^2 + 2\,\mathrm{Re}(w\bar{z})$ *Cosinussatz* heißt; wegen $\alpha + \chi = \pi$ (vgl. Fig. b) gilt $\cos\chi = -\cos\alpha$ und also $|w + z|^2 = |w|^2 + |z|^2 - 2|w|\,|z|\cos\alpha$.

3. MOIVREsche Formel: $(\cos\varphi + i\sin\varphi)^n = \cos n\varphi + i\sin n\varphi$ *für* $n \in \mathbb{Z}$. Dies ist klar wegen $(e^{i\varphi})^n = e^{in\varphi}$ (Homomorphieeigenschaft von p); allgemeiner folgt

Satz. *Für jede komplexe Zahl* $z = re^{i\varphi} = r(\cos\varphi + i\sin\varphi) \in \mathbb{C}^\times$ *gilt* $z^n = r^n e^{in\varphi} = r^n(\cos n\varphi + i\sin n\varphi)$ *für alle* $n \in \mathbb{Z}$.

Der französische Mathematiker Abraham De MOIVRE (1667–1754, Hugenotte; emigrierte nach Aufhebung des Ediktes von Nantes 1685 nach London; 1697 Mitglied der Royal Society und später der Akademien in Paris und Berlin; veröffentlichte 1718 sein berühmtes Werk „Doctrine of Chances" zur Wahrscheinlichkeitsrechnung; entdeckte vor STIRLING die „STIRLINGsche Formel" $n! \approx \sqrt{2\pi n}(n/e)^n$; 1712 von der Royal Society bestellter Bevollmächtigter im Streit zwischen NEWTON und LEIBNIZ über die Entdeckung der Infinitesimalrechnung; NEWTON soll im fortgeschrittenen Alter gesagt haben, wenn man ihn etwas Mathematisches fragte: „Go to Mr. De MOIVRE; he knows these things better than I do.") deutete die Entdeckung seiner „magischen" Formel 1707 an Zahlenbeispielen an; 1730 scheint er die allgemeine Formel

$$\cos\varphi = \tfrac{1}{2}\sqrt[n]{\cos n\varphi + i\sin n\varphi} + \tfrac{1}{2}\sqrt[n]{\cos n\varphi - i\sin n\varphi}, \qquad n > 0,$$

zu kennen; 1738 beschreibt er (umständlich) ein Verfahren zum Auffinden von Wurzeln der Form $\sqrt[n]{a + ib}$, seine Lösungsvorschrift deckt sich inhaltlich mit der nach ihm benannten Formel. Die heutige Fassung findet sich erst bei EULER im Cap. VIII seiner „Introductio in Analysin infinitorum" von 1748, den ersten stichhaltigen Beweis für alle $n \in \mathbb{R}$ gab ebenfalls EULER 1749 mit Hilfe der Differentialrechnung. Die MOIVREsche Formel ist durch die Gleichung $(e^{i\varphi})^n = e^{in\varphi}$ entmystifiziert.

Die MOIVREsche Formel liefert eine extrem einfache Methode, um $\cos n\varphi$ und $\sin n\varphi$ für alle $n \geqslant 1$ als Polynom in $\cos\varphi$ und $\sin\varphi$ auszudrücken. So erhält man z. B. für $n = 3$, wenn man zu Real- und Imaginärteilen übergeht:

$$\cos 3\varphi = \cos^3\varphi - 3\cos\varphi\sin^2\varphi, \qquad \sin 3\varphi = 3\cos^2\varphi\sin\varphi - \sin^3\varphi.$$

Die in 3.5 angekündigte Darstellung der Lösungen der quadratischen Gleichung $z^2 + az + b = 0$ in trigonometrischer Form geschieht wie folgt: man

*) Im Cauchyschen „Cours d'Analyse" (vgl. 1.7) liest man 1821 aber aus heutiger Sicht noch Erstaunliches (S. 154): „L'équation $\cos(a + b) + \sqrt{-1}\,\sin(a + b) = (\cos a + \sqrt{-1}\sin a)(\cos b + \sqrt{-1}\sin b)$ elle-même, prise à la lettre, se trouve inexacte et n'a pas de sens".

schreibt $\frac{1}{4}(a^2 - 4b) = r(\cos\varphi + i\sin\varphi)$ und erhält die gesuchten Wurzeln in der Form

$$z_1 = -\frac{1}{2}a + \sqrt{r}\left(\cos\frac{\varphi}{2} + i\sin\frac{\varphi}{2}\right), \qquad z_2 = -\frac{1}{2}a - \sqrt{r}\left(\cos\frac{\varphi}{2} + i\sin\frac{\varphi}{2}\right).$$

4. Einheitswurzeln. Als wichtigste Anwendung der Polarkoordinaten zeigen wir

Lemma. *Es sei $n \geqslant 1$ eine natürliche Zahl. Dann gibt es genau n verschiedene komplexe Zahlen z mit $z^n = 1$, nämlich:*

$$\zeta_\nu := \exp\frac{2\pi i}{n}\nu, \qquad \nu = 0, 1, \ldots, n-1.$$

Mit $\zeta := \zeta_1$ gilt $\zeta_\nu = \zeta^\nu$.

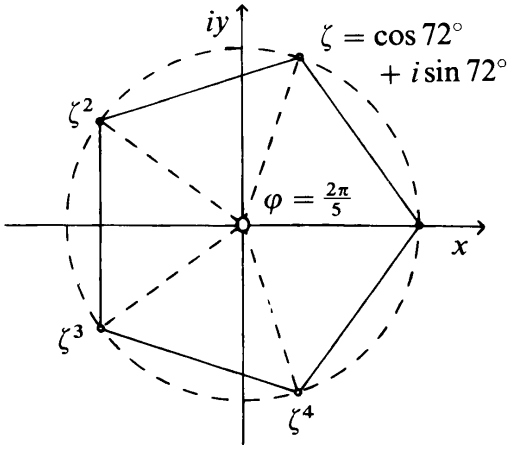
Beweis. Die Gleichungen $\zeta_\nu = \zeta^\nu$ und $\zeta_\nu^n = 1$ sind klar (MOIVRE). Da

$$\zeta_\nu \zeta_\mu^{-1} = \exp\frac{2\pi i}{n}(\nu - \mu),$$

so gilt $\zeta_\nu = \zeta_\mu$ wegen $\operatorname{Kern} p = 2\pi\mathbb{Z}$ genau dann, wenn $\frac{1}{n}(\nu - \mu) \in \mathbb{Z}$. Wegen $-n < \nu - \mu < n$ folgt $\nu = \mu$, das heißt, $\zeta_0, \zeta_1, \ldots, \zeta_{n-1}$ sind paarweise verschieden.

Für $z = |z|e^{i\varphi}$ gilt $z^n = 1$ genau dann, wenn $|z| = 1$ und $e^{in\varphi} = 1$, das heißt, wenn $\varphi = \frac{2\pi k}{n}$ mit $k \in \mathbb{Z}$. Da $0 \leqslant \varphi < 2\pi$, so folgt $k \in \{0, 1, \ldots, n-1\}$, das heißt, $z = \zeta_k$. Mithin gibt es außer $\zeta_0, \zeta_1, \ldots, \zeta_{n-1}$ keine weitere komplexe Zahl z mit $z^n = 1$. □

Die n Zahlen $1, \zeta, \zeta^2, \ldots, \zeta^{n-1}$ heißen *die n-ten Einheitswurzeln*, geometrisch bilden sie die Eckpunkte eines regulären n-Ecks (die Figur zeigt die 5-ten



Einheitswurzeln). Eine n-te Einheitswurzel, deren Potenzen sämtliche übrigen n-ten Einheitswurzeln darstellen, heißt *primitiv*; ζ ist stets primitive n-te Einheitswurzel, für $n = 5$ gilt z. B.:

$$\zeta = \frac{\sqrt{5} - 1}{4} + \frac{i}{4}\sqrt{2(5 + \sqrt{5})}.$$

Die vorangehenden Überlegungen lassen sich sofort verallgemeinern. Jede komplexe Zahl c mit $c^n = a$ heißt eine *n-te Wurzel* aus a. Setzt man

$$c := \sqrt[n]{|a|} \exp \frac{i\varphi}{n} \qquad \text{für} \qquad a = |a|e^{i\varphi} \in \mathbb{C}^\times,$$

wobei $\sqrt[n]{|a|}$ *die positive reelle* n-te Wurzel aus $|a|$ bezeichnet, so folgt:

Existenz- und Eindeutigkeitssatz für n-te Wurzeln. *Jede komplexe Zahl $a = |a|e^{i\varphi} \in \mathbb{C}^\times$ hat für jedes $n \in \mathbb{N}, n \geqslant 1$, genau n verschiedene n-te komplexe Wurzeln, nämlich $c, c\zeta, \ldots, c\zeta^{n-1}$, wobei $\zeta := \exp \frac{2\pi i}{n}$.*

Die Einsicht in die Mehrdeutigkeit von Wurzeln hat sich im 17. Jahrhundert entwickelt. Der Satz, daß n-te Wurzeln n Werte besitzen, war z. B. 1690 Michael Rolle (1652–1719; Paris, Mitglied der Académie Française) wohlvertraut; Rolle hat übrigens den nach ihm benannten Satz in der Differentialrechnung bei Untersuchungen über Wurzeln von Polynomen gefunden (zwischen benachbarten reellen Wurzeln eines reellen Polynoms liegt stets eine Wurzel der 1. Ableitung).

Der bewiesene Existenzsatz besagt insbesondere:

Jedes reine Polynom $z^n - a \in \mathbb{C}[z]$ vom Grade $n \geqslant 1$ hat eine komplexe Nullstelle.

Dies ist ein wichtiger Spezialfall des Fundamentalsatzes der Algebra.

Der britische Mathematiker Roger Cotes (1682–1716; Student und Professor in Cambridge; Freund Newtons) beschäftigte sich 1714 anläßlich seiner Untersuchungen über die Integration rationaler Funktionen mittels Partialbruchzerlegung mit der Faktorisierung der Polynome $z^n - 1$ und $z^{2n} + az^n + 1$ in reelle quadratische Faktoren, er kannte z. B. die Formel

$$z^{2n} + 1 = \prod_{\nu=1}^{n} \left(z^2 - 2z \cos \frac{2\nu - 1}{2n} \pi + 1 \right).$$

Die Resultate von Cotes wurden erst 1722 posthum unter dem Titel „Harmonia mensurarum" publiziert; der Wunsch, die Ergebnisse von Cotes zu vervollständigen, motivierte u. a. de Moivre bei seinen Überlegungen.

Kapitel 4. Fundamentalsatz der Algebra

R. Remmert

> Was beweisbar ist, soll in der Wissenschaft nicht ohne Beweis geglaubt werden (R. DEDEKIND 1887).

Wir haben in 3.3.5 gesehen, daß jedes quadratische komplexe Polynom (zwei) Nullstellen in $\mathbb{C}$ hat. Diese Aussage ist ein Spezialfall eines viel allgemeineren Satzes, den GAUSS 1849 *Grundlehrsatz* der Theorie der algebraischen Gleichungen (Werke 3, 73) genannt hat und der in die Literatur als sog. *Fundamentalsatz der Algebra* eingegangen ist.

Jedes nicht konstante, komplexe Polynom hat im Körper $\mathbb{C}$ wenigstens eine Nullstelle.

Die Bezeichnung Fundamentalsatz der Algebra stammt aus einer Zeit, wo man noch unter Algebra die Theorie der Polynome mit reellen oder komplexen Koeffizienten verstand. Dieser Existenzsatz, der ja bereits für reine Polynome $z^n - a$ nicht trivial ist (vgl. 3.6.4),wird in diesem Kapitel ausführlich diskutiert und „elementar" bewiesen; er ist äquivalent zum Satz, daß jedes reelle Polynom in reelle Linearfaktoren und reelle quadratische Faktoren zerlegbar ist.

Dem Fundamentalsatz der Algebra kommt in der Geschichte der Theorie der komplexen Zahlen eine alles andere überragende Bedeutung zu: die Möglichkeit, diesen Satz im Komplexen beweisen zu können, ist es vor allem gewesen, die der allgemeinen Anerkennung der komplexen Zahlen den Weg bereitet hat.

Die Genesis des Fundamentalsatzes wird ausführlich im Paragraphen 1 dargestellt. Im Paragraphen 2 beweisen wir den Fundamentalsatz nach einer schönen alten Idee von ARGAND, die auf D'ALEMBERT zurückgeht; dies ist wohl der einfachste Beweis. Erste Anwendungen des Fundamentalsatzes, der in den späteren Kapiteln über Algebren immer wieder herangezogen wird, geben wir im Paragraphen 3, insbesondere beweisen wir in 3.5 den erstmals 1867 von HANKEL publizierten Satz über die Einzigkeit des Körpers $\mathbb{C}$.

In einem Anhang besprechen wir noch den eleganten Beweis von LAPLACE, der „algebraischer" ist als der Beweis von ARGAND. Wegen funktionentheoretischer Beweise verweisen wir auf Grundwissen Mathematik, Bd. 5, Funktionentheorie I.

§ 1. Zur Geschichte des Fundamentalsatzes

Mit $f = a_0 + a_1 x + \cdots + a_n x^n \in \mathbb{R}[x]$ wird in diesem Paragraphen stets ein *reelles Polynom n-ten Grades* bezeichnet (also $a_\nu \in \mathbb{R}$, $a_n \neq 0$). Wir betrachten nur nicht konstante Polynome, das heißt, wir setzen $n \geqslant 1$ voraus. Unter einer *Nullstelle* oder auch *Wurzel* von f versteht man jedes Element c eines $\mathbb{R}$ umfassenden Körpers K

mit $f(c) = 0$; man nennt c auch eine *Lösung der Polynomgleichung* $f = 0$. Unter *Gleichung* verstehen wir stets eine Polynomgleichung.

Der natürlichste und direkteste Weg zu zeigen, daß reelle Gleichungen immer komplexe Lösungen haben, besteht darin, *Lösungen durch ein Rechenverfahren, das nicht aus* $\mathbb{C}$ *herausführt, explizit anzugeben.* So geschieht es bei quadratischen Gleichungen (vgl. 3.3.5); so ist CARDANO bei kubischen Gleichungen vorgegangen; so lassen sich auch biquadratische Gleichungen lösen: man hat Lösungsformeln (die „ineinander geschachtelte Radikale" enthalten mit Radikanden, die Polynome in den Koeffizienten $a_0, \ldots, a_n$ sind), aus denen man mühelos abliest, daß die konstruierten Lösungen komplexe Zahlen sind, vgl. B. L. V. D. WAERDEN, Algebra I, Berlin 1955, § 59.

Ganz anders wird die Situation bei Gleichungen fünften und höheren Grades. Hier konnte man keine Auflösungsmethode durch Radikale finden*); bis GAUSS haben alle Mathematiker an die Existenz von Lösungen im Niemandsland (heute würde man sagen: in einem unbekannten Oberkörper von $\mathbb{C}$) *geglaubt* und ideenreich zu zeigen versucht, daß diese Lösungen in Wahrheit bereits komplexe Zahlen sind.

Wir stellen im folgenden die entscheidenden Daten vom ersten mystischen Auftreten des Fundamentalsatzes bis zu seinem Selbstverständnis in unserer Zeit zusammen; neben der in 3.1 angegebenen Literatur wurde noch herangezogen:

Abrégé d'histoire des mathématiques I, sous la direction de Jean DIEUDONNÉ; Hermann, Paris 1978, insb. Chapitre IV.

1. GIRARD (1595–1632) und DESCARTES (1596–1650). Gleichungen n-ten Grades haben höchstens n Lösungen (Peter ROTH 1608); VIETA (1540–1603) konnte aufgrund seines Wurzelsatzes Gleichungen n-ten Grades anschreiben, die wirklich n Lösungen haben. Es war der heute vergessene flämische Mathematiker Albert GIRARD, der als erster behauptete, daß immer n Lösungen vorhanden sind. In seinem 1629 erschienenen Werk „L'Invention en l'algèbre" schreibt er: „Toutes les equations d'algèbre reçoivent autant de solutions, que la denomination de la plus haute quantité le demonstre...". Einen Beweis oder Andeutungen eines Beweises gibt GIRARD nicht, er erläutert den Satz nur an Beispielen, unter anderem an der Gleichung $x^4 - 4x + 3 = 0$ mit den Lösungen $1,\ 1,\ -1 + i\sqrt{2},\ -1 - i\sqrt{2}$.

GIRARD behauptet nicht, daß die Lösungen stets die Form $a + b\sqrt{-1}, a, b \in \mathbb{R}$, haben müssen; neben reelle Lösungen („solche, die > 0 sind, solche, die < 0 sind") stellt er vielmehr „autres enveloppées, comme celles qui ont des $\sqrt{-}$, comme $\sqrt{-3}$, ou autres nombres semblables"; er läßt also die Möglichkeit von Lösungen, die nicht komplex sind, offen. In heutiger Sprache besagt also die

These von GIRARD. *Zu jedem Polynom n-ten Grades* $f \in \mathbb{R}[x]$ *existiert ein Oberkörper* K *von* $\mathbb{R}$, *so daß* f *in* K *genau* n *Nullstellen hat; der Körper* K *ist möglicherweise ein echter Oberkörper von* $\mathbb{C}$.

---

*) N. H. ABEL zeigte 1826 in seiner im ersten Band des Crelleschen Journals, 65–84, publizierten Arbeit „Beweis der Unmöglichkeit, algebraische Gleichungen von höheren Graden, als den vierten, allgemein aufzulösen" (vgl. auch Œuvres complètes, 1, 66–87), daß es grundsätzlich kein Verfahren gibt, allgemeine Gleichungen von Grad $\geqslant 5$ durch Radikale zu lösen.

DESCARTES gibt 1637 im dritten und letzten Buch seiner „La Géométrie" eine knappe Darstellung der damals bekannten Dinge über Gleichungen. Er notiert den wichtigen Satz, daß ein Polynom mit der Nullstelle c stets den Faktor $x - c$ abspaltet*) ; er beschreibt ferner die nach ihm benannte *Vorzeichenregel* (vgl. hierzu O. HAUPT „Einführung in die Algebra", 2. Teil, Akad. Verl. Ges. Geest u. Portig 1954, S. 411). Zur These von GIRARD nimmt DESCARTES nur vage Stellung, vgl. 3.1.3.

2. LEIBNIZ (1646–1716). Bei seinen Bemühungen, rationale Funktionen mittels Partialbruchzerlegung zu integrieren, kam LEIBNIZ zu der Frage, ob jedes reelle Polynom als Produkt von Faktoren ersten und zweiten Grades darstellbar ist. Er vertrat 1702 in einer Arbeit in den *Acta Eruditorum* die Ansicht, daß dies nicht zutrifft: zur Bestätigung führt er an, daß in der Zerlegung

$$x^4 + a^4 = (x^2 - a^2 i)(x^2 + a^2 i) = (x + a\sqrt{i})(x - a\sqrt{i})(x + a\sqrt{-i})(x - a\sqrt{-i})$$

das Produkt aus irgend zwei Faktoren rechts niemals ein quadratisches reelles Polynom ist. LEIBNIZ scheint nicht der Gedanke gekommen zu sein, daß $\sqrt{i}$ von der Form $a + bi$ sein könnte; dann hätte er nämlich wohl sofort

$$\sqrt{i} = \tfrac{1}{2}\sqrt{2}(1 + i) \qquad \text{und} \qquad \sqrt{-i} = \tfrac{1}{2}\sqrt{2}(1 - i)$$

gesehen, den ersten mit dem dritten und den zweiten mit dem vierten Faktor multipliziert und anstelle seiner falschen Behauptung notiert:

$$x^4 + a^4 = (x^2 + a\sqrt{2}x + a^2)(x^2 - a\sqrt{2}x + a^2);$$

es ist bemerkenswert, daß er hierauf auch nicht durch den Rechentrick $x^4 + a^4 = (x^2 + a^2)^2 - 2a^2x^2$ geführt wurde.

3. EULER (1707–1783). In einem Brief an Nikolaus BERNOULLI vom 1. 11. 1742 formuliert EULER den Faktorisierungssatz für reelle Polynome genau in der Form, in der LEIBNIZ ihn für falsch hielt. Das angebliche Gegenbeispiel $x^4 - 4x^3 + 2x^2 + 4x + 4$ BERNOULLIS mit den Nullstellen

$$x_{1,2} = 1 \pm \sqrt{2 + i\sqrt{3}}, \qquad x_{3,4} = 1 \pm \sqrt{2 - i\sqrt{3}}$$

entkräftet er durch den Nachweis, daß $(x - x_1)(x - x_3)$ und $(x - x_2)(x - x_4)$ *reelle* Polynome sind, nämlich

$$x^2 - (2 + a)x + 1 + \sqrt{7} + a \qquad \text{und} \qquad x^2 - (2 - a)x + 1 + \sqrt{7} - a$$

mit $a := \sqrt{4 + 2\sqrt{7}}$.

Bald darauf, in einem Brief vom 15. 12. 1742 an seinen vertrauten Briefpartner GOLDBACH, wiederholt EULER seine Behauptung, wobei er sagt, daß er dieses Theorem nicht vollständig habe beweisen können, sondern nur „ungefähr, wie gewisse Fermatsche Sätze." In diesem Brief erwähnt er übrigens auch – was aus

---

*) Diesen Satz kannte wahrscheinlich auch schon Thomas HARRIOT (1560[?]–1621, vermaß 1585 im Auftrag von Sir Walter RALEIGH die Kolonie Virginia und war damit der erste Mathematiker, der in Nordamerika lebte).

heutiger Sicht ganz klar ist und nichts mit dem eigentlichen Problem der Existenz komplexer Nullstellen zu tun hat –, daß man die imaginären Wurzeln reeller Polynome stets so paarweise zusammenfassen kann, daß beim Ausmultiplizieren reelle Polynome zweiten Grades entstehen*). GOLDBACH bleibt selbst gegenüber dieser einfachen Behauptung skeptisch und führt zur Widerlegung das Polynom $z^4 + 72z - 20$ an, welches EULER sofort faktorisiert.

Die Eulersche Faktorisierungsbehauptung geht über die These von GIRARD, die er gekannt haben dürfte, hinaus: da quadratische Gleichungen stets komplexe Lösungen haben, so behauptet EULER nichts anderes als den

Fundamentalsatz der Algebra für reelle Polynome. *Jedes Polynom n-ten Grades* $f \in \mathbb{R}[x]$ *hat genau n Nullstellen im Oberkörper* $\mathbb{C}$.

EULER konnte diesen Satz für alle Polynome vom Grade $\leqslant 6$ streng beweisen. 1749 hat er (Recherches sur les racines imaginaires des équations, Histoire de l'Academie Royale des Sciences et Belles Lettres, Année MDCCXLIX (Berlin 1751), 222–288, vgl. auch Opera Omnia 6, 1. Ser., 78–147) den Allgemeinfall aufgegriffen: Seine Idee ist, jedes normierte Polynom P, dessen Grad eine Zweierpotenz $2^n \geqslant 4$ ist, in ein Produkt $P_1 P_2$ zweier normierter Polynome vom Grad $m := 2^{n-1}$ zu zerlegen(!). Gelingt dies, so ist sein Satz bewiesen, denn beliebige Polynome $\not\equiv 0$ lassen sich durch Multiplikation mit ax^d stets in solche Polynome überführen, und Iteration des Zerlegungsverfahrens gibt schließlich eine Zerlegung von P in reelle quadratische Polynome. EULER macht für P den Ansatz

$$P(x) = x^{2m} + Bx^{2m-2} + Cx^{2m-3} + \cdots,$$

dies ist erlaubt, da sich der Koeffizient A von x^{2m-1} durch die Translation $x \mapsto x - \frac{1}{2m}A$ stets zu 0 machen läßt (diese Reduktion war spätestens seit CARDANO, Ars Magna Kap. 17, bekannt; VIÈTA nennt diesen Prozeß „expurgatio"). Die Polynome P_1, P_2 sollen nun die Form

$$x^m + ux^{m-1} + \alpha x^{m-2} + \beta x^{m-3} + \cdots, \qquad x^m - ux^{m-1} + \lambda x^{m-2} + \mu x^{m-3} + \cdots$$

haben, denn die Koeffizienten von x^{m-1} unterscheiden sich wegen des Fehlens von x^{2m-1} in $P(x)$ nur im Vorzeichen. Ausmultiplizieren und Koeffizientenvergleich liefert Gleichungen zwischen $B, C, \ldots$ und $u, \alpha, \beta, \ldots, \lambda, \mu, \ldots$. EULER behauptet, daß $\alpha, \beta, \ldots, \lambda, \mu, \ldots$ *rationale Funktionen* in $B, C, \ldots$ und u sind, und daß sich durch Elimination der $\alpha, \beta, \ldots, \lambda, \mu, \ldots$ ein normiertes reelles Polynom in u vom Grad $\binom{2m}{m}$ einstellt, dessen konstantes Glied *negativ* ist. Dieses Polynom hat nun nach dem *Zwischenwertsatz*, den EULER klar gesehen hat, eine Nullstelle u. Es ist klar, daß damit der Satz bewiesen ist. Für $2m = 4$ führt EULER alles explizit aus (vgl. loc. cit. S. 93/94); indessen ist der Beweis im Allgemeinfall nur skizzenhaft (vgl. S. 105/106), bei vielen wesentlichen Einzelheiten hüllt EULER (wie GAUSS später bemängelte, vgl. Abschnitt 6) sich in Schweigen.

EULER hat seinen Satz auch komplex formuliert (loc. cit. S. 112):

Si une équation algébrique, de degré qu'elle soit, a des racines imaginaires, chacune sera comprise dans cette formule générale $M + N\sqrt{-1}$, *les lettres M et N marquant des quantités réelles.*

---

*) Dies hat bereits BOMBELLI um 1560 bemerkt.

4. D'ALEMBERT (1717–1783). Drei Jahre vor EULER machte 1746 Jean le Rond D'ALEMBERT (Recherches sur le calcul intégral, Histoire de l'Académie Royale des Sciences et Belles Lettres, Année MDCCXLVI (Berlin 1748), 182–224) den *ersten* ernst zu nehmenden Versuch, den Faktorisierungssatz zu beweisen; dieser Satz heißt seither in der französischen Literatur der Satz von D'ALEMBERT. Die Grundidee ist einfach: es wird (wenn auch sehr verborgen) versucht, *den Betrag des Polynoms f durch geeignete Wahl des Argumentes zu verkleinern.* D'ALEMBERT verwendet folgenden Hilfssatz, den er unbewiesen annimmt und den erst 1851 PUSIEUX (unter impliziter Verwendung des Fundamentalsatzes!) korrekt hergeleitet hat:

Zu jedem Paar (b, c) komplexer Zahlen mit $f(b) = c$ gibt es eine natürliche Zahl $q \geqslant 1$ und eine in der Umgebung von c konvergente Reihe

$$h(w) = b + \sum_{\nu=1}^{\infty} c_\nu (w - c)^{\nu/q},$$

so daß für alle Zahlen w nahe bei c gilt: $f(h(w)) = w$.

D'ALEMBERT geht nun von reellen Zahlen b, c mit $f(b) = c$ aus (in der Tat wählt er b so, daß die reelle Funktion in b ein Minimum hat) und findet, falls $c \neq 0$, auf Grund seiner Pusieuxentwicklung komplexe Zahlen z_1, w_1 mit $|w_1| < |c|$ so, daß gilt $f(z_1) = w_1$. Wiederholung des Verfahrens führt zu immer kleineren Werten der Absolutbeträge von f und, wenn man ein einfaches Kompaktheitsargument benutzt (was D'ALEMBERT nicht konnte), zu einer Nullstelle von f.

Die zeitbedingten Schwächen des d'Alembertschen Schlusses unterliegen mit Recht der Kritik durch GAUSS (siehe Abschnitt 6); allerdings sagt GAUSS auch nahezu seherisch (Werke 3, S. 11): „Aus diesen Gründen vermag ich den d'Alembertschen Beweis nicht für ausreichend zu halten. Allein das verhindert mich nicht, daß mir der wahre Nerv des Beweises trotz aller Einwände unberührt zu sein scheint; ich glaube ..., daß man auf dieselben Grundlagen einen strengen Beweis unseres Satzes aufbauen kann." Das genau hat ARGAND 1814 getan (siehe Abschnitt 8).

Durch die Arbeiten von D'ALEMBERT und EULER setzte sich die Ansicht durch, „daß es nur der Zulassung einer fingierten Grösse $\sqrt{-1}$ bedürfe, um jeder entwickelten algebraischen Gleichung die ihrer Ordnungszahl gleiche Menge von Wurzeln zu verschaffen" (GAUSS, Werke 10, 1, S. 404).

5. LAGRANGE (1736–1813) und LAPLACE (1749–1827). Bereits 1772 hat Joseph Louis LAGRANGE in seiner Abhandlung „Sur la forme des racines imaginaires des équations" (Nouveaux mémoires de l'Académie Royale des Sciences et Belles Lettres, Année MDCCLXXVII (Berlin 1774), 222–258, vgl. auch Œuvres Complètes 3, 477–516) Bedenken gegen Eulers Beweis erhoben. Er bemerkt u. a., daß die Eulersche Gleichung für u unbestimmte Koeffizienten der Form $\frac{0}{0}$ haben könne. LAGRANGE macht einen neuen Versuch, die Existenz der von EULER gesuchten Faktorisierung $P = P_1 P_2$ zu zeigen. Dank seiner Resultate über die Permutation der Wurzeln von Gleichungen gelingt es ihm weitgehend, die Eulerschen Lücken zu schließen; doch muß auch er fiktive Wurzeln heranziehen.

Im Jahre 1795 machte Pierre Simon de LAPLACE*) in seinen „Leçons de Mathématiques données à l'Ecole Normale" (Journal de l'Ecole Polytechnique (Septième et huitième cahier) Tome II, 1–278, Paris 1812, besonders 56–58; siehe auch Œuvres Complètes 14, 10–111, besonders 63–65) einen Ansatz zum Beweis des Fundamentalsatzes, der vom Euler-Lagrangeschen Ansatz völlig verschieden ist und Vorstellungen über die Diskriminante eines Polynoms verwendet. Auch LAPLACE arbeitet mit den platonischen Wurzeln des Polynoms. Sein überaus eleganter Beweis ist lange vergessen gewesen, wir reproduzieren ihn in moderner Form im Anhang dieses Kapitels.

6. Die Kritik durch GAUSS. Im Oktober 1797 schreibt GAUSS in sein Tagebuch: „Aequationes habere radices imaginarias methodo genuina demonstratum" (vgl. Math. Ann. 57, S. 18 (1903)). Den hier angekündigten Beweis des Fundamentalsatzes, der allerdings nach heutigen Maßstäben auch keineswegs streng ist, veröffentlichte er 1799 in seiner Arbeit „Demonstratio nova theorematis omnem functionem algebraicam rationalem integram unius variabilis in factores reales primi vel secundi gradus resolvi posse" (Werke 3, 1–30), mit der er in absentia an der Universität Helmstedt bei J. F. PFAFF (1765–1825) promovierte. GAUSS beginnt seine Abhandlung mit einer eingehenden Kritik aller ihm bekannten Beweisansätze. Es ist hier nicht der Platz, die Einwände des 22jährigen gegen die Beweise von D'ALEMBERT, EULER und LAGRANGE – also gegen die führenden etablierten Mathematiker der damaligen Zeit – im einzelnen zu diskutieren (vgl. hierzu etwa: TROPFKE, Bd. 1, 1980, 494–499): der Haupteinwand von GAUSS ist, daß die zu beweisende Existenz der Nullstelle stets vorausgesetzt wird. So rügt er z. B. EULERS Gebrauch der hypothetischen Wurzeln (Werke 3, S. 5, 14)**): „..., wenn man dann mit diesen unmöglichen Wurzeln so verfährt, als ob sie etwas Wirkliches seien, und beispielsweise sagt, die Summe aller Wurzeln der Gleichung $x^m + AX^{m-1} + \cdots = 0$ sei $= -A$, obschon unmögliche unter ihnen sind (das heißt eigentlich: *wiewohl einige fehlen*), so kann ich dies durchaus nicht billigen."

Der verbesserte Beweis von LAGRANGE wird ebenfalls verbannt, man liest (Werke 3, S. 20)**) „Dieser große Mathematiker bemühte sich vor Allem, die Lücken in Eulers erstem Beweise auszufüllen, und wirklich hat er das, was oben § 8 den zweiten und den vierten Einwurf ausmacht, so tief durchforscht, daß nichts Weiteres zu wünschen übrig bleibt .... Den dritten Einwurf dagegen berührt er überhaupt nicht; ja auch seine ganze Untersuchung ist auf der Voraussetzung aufgebaut, jede Gleichung m-ten Grades habe wirklich m Wurzeln." Und 1815 (Werke 3, S. 106) spricht er in diesem Zusammenhang gar von einer „wahren petitio principii".

Den Beweis von LAPLACE kannte GAUSS 1799 noch nicht. Doch später findet auch dieser Ansatz keine Gnade vor seinen Augen; er bemerkt dazu 1815 in den

*) 1799 von NAPOLEON zum Minister des Innern ernannt, nach 6 Wochen amtsenthoben, da er „den Geist des unendlich Kleinen bis in die Verwaltung hineingetragen" habe. Nach der Restauration durch die Bourbonen Marquis und Pair von Frankreich.

**) Zitate nach der deutschen Übersetzung in Ostwald's Klassikern der Exakten Wissenschaften, Nr. 14. „Die vier Gaußschen Beweise für die Zerlegung ganzer algebraischer Funktionen in reelle Faktoren ersten und zweiten Grades (1799–1849)", herausgegeben 1890 von E. NETTO.

Göttingischen gelehrten Anzeigen (Werke 3, S. 105), daß „die scharfsinnige Art, wie später LAPLACE diesen Gegenstand behandelt hat, gerade von dem Hauptvorwurfe, welcher alle jene versuchten Beweise trifft, nicht freigesprochen werden" könne.

Wir wollen uns die Situation noch einmal aus heutiger Sicht vor Augen führen. In allen Beweisversuchen vor GAUSS wird von vornherein nicht nach der *Existenz* der Wurzeln einer Gleichung, sondern nur nach der *Form* derselben, und ob diese die Gestalt $a + b\sqrt{-1}$ haben, gefragt. Die These von GIRARD wird stillschweigend als Axiom angenommen, es werden keinerlei Rechtfertigungsgründe dafür angeführt. Man hat sogar lange geglaubt, daß es eine Hierarchie imaginärer Größen – von GAUSS in seiner Dissertation (Werke 3, S. 14) „vera umbrae umbra" (= wahre Schatten von Schatten) genannt – gebe, unter denen die komplexen Zahlen $a + b\sqrt{-1}$, $a, b \in \mathbb{R}$, die einfachsten seien. Erst als sich im 18. Jahrhundert die Auffassung durchsetzte, daß Lösungen durch „algebraisch-analytische Methoden, die nicht aus $\mathbb{C}$ herausführen", beschreibbar sind, stellte man ernsthaft folgendes Problem, das bei Kenntnis der Hintergründe nun auch keineswegs mehr so paradox erscheint:

„*Man zeige, daß jede imaginäre Größe die Form* $a + b\sqrt{-1}$ *hat.*"

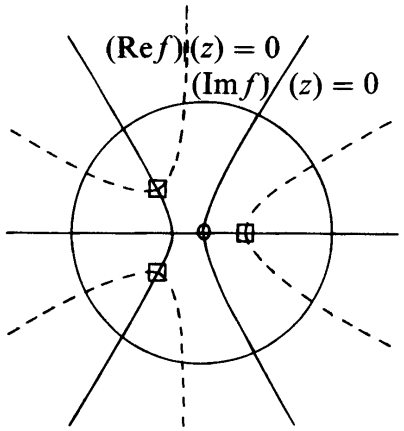
Dies ist bei wohlwollender Interpretation nichts anderes als die Aussage, daß $\mathbb{C}$ *vollständig* ist und *keine algebraischen Körpererweiterungen zuläßt*; in der im Abschnitt 3 zitierten Arbeit „Recherches sur les racines . . ." von EULER liest sich das so (S. 147): „Puisque donc toutes ces quantités imaginaires, qui sont formées par des opérations transcendantes, sont aussi comprises dans la forme générale $M + N\sqrt{-1}$, nous pourrons soutenir sans balancer, que généralement toutes les quantités imaginaires, quelques compliqués qu'elles puissent être, sont toujours réductibles à la forme $M + N\sqrt{-1}$."

Der Gaußsche Einwand gegen die Ansätze von EULER-LAGRANGE und LAPLACE war in dem Augenblick, als die Algebra für jedes Polynom die Existenz eines Zerfällungskörpers sicherstellte, entkräftet; von da an waren die Ansätze, wie Adolf KNESER bereits 1888 bemerkte (Crelles Journ. 102, S. 21), vollgültige Beweise. 1907 sagte FROBENIUS (Ges. Abhandl. 3, S. 733) anläßlich des Festaktes der Universität Basel zur Feier des zweihundertsten Geburtstages Leonhard Eulers: „Für die Existenz der Wurzeln einer Gleichung führt er jenen am meisten algebraischen Beweis, der darauf fußt, daß jede reelle Gleichung unpaaren Grades eine reelle Wurzel besitzt. Ich halte es für unrecht, diesen Beweis ausschließlich GAUSS zuzuschreiben, der doch nur die letzte Feile daran gelegt hat."

7. Die vier Beweise von GAUSS. Das grundsätzlich Neue am GAUSSschen Beweis von 1799 ist, daß er nicht daran geht, eine Wurzel zu berechnen, sondern ihre *Existenz zu beweisen*. Dazu bedurfte es, um mit HANKEL zu sprechen (S. 97), „einen eminenten Aufwand von Schärfe des Gedankens und Productionskraft, wie beides in Gauß wunderbar vereinigt war." GAUSS erhebt aber in seiner Doktorarbeit nicht den Anspruch, den ersten korrekten Beweis des Fundamentalsatzes zu schaffen, wie bereits das Wort „nova" im Titel andeutet und wie auch seine Bemerkung zum Beweisversuch von D'ALEMBERT bezeugt (vgl. Abschnitt 4). GAUSS lieferte insgesamt *vier* Beweise für den Fundamentalsatz der Algebra, den vierten Beweis publizierte

er 1849 im Jahr seines Goldenen Doktorjubiläums (vgl. Ostwald's Klassiker, Nr. 14).

Der erste Beweis von 1799 ist topologisch, hat aber nach heutigem Verständnis wesentliche Lücken; wir gehen auf die Problematik näher ein: Die komplexen Nullstellen des reellen Polynoms f vom Grad n sind die *Schnittpunkte* der beiden *reell-algebraischen Kurven* $(\mathrm{Re}\, f)(z) = 0$ und $(\mathrm{Im}\, f)(z) = 0$. Ist R hinreichend groß, so liegen auf jeder Kreislinie $|z| = r$, $r > R$, genau $2n$ Punkte jeder Kurve. Diese Punkte lassen sich außerhalb der Kreisscheibe $\{z \in \mathbb{C}\colon |z| \leqslant R\}$ zu je $2n$ *stetigen* Kurvenzügen A_ν und B_ν, $1 \leqslant \nu \leqslant 2n$, zusammenfügen, die sich ins Unendliche erstrecken, und zwar so, daß zwischen je zwei aufeinander folgenden „Zweigen" der Kurve $(\mathrm{Re}\, f)(z) = 0$ ein Zweig der Kurve $(\mathrm{Im}\, f)(z) = 0$ liegt und umgekehrt (vgl. Figur mit dem Beispiel $f(z) := z^3 + z^2 - 2$ und den Nullstellen $1, -1 \pm i$). Gauss



sagt (Artikel 21): „Nun läßt sich aus der gegenseitigen Lage der in die Kreisscheibe eintretenden Zweige der Schluss, dass innerhalb des Kreises ein Schnitt eines Zweiges der ersten mit einem Zweige der zweiten Linie vorhanden sein müsse, auf so viele Arten ziehen, daß ich fast nicht weiss, welche Methode an erster Stelle vor den übrigen zu bevorzugen sei." Bei der dann folgenden geometrischen Begründung benutzt er Sätze aus der höheren Geometrie, insbesondere: „..., wenn ein [nicht kompakter] Zweig einer algebraischen Curve in einen begrenzten Raum [Kreisscheibe] eintritt, er nothwendig aus demselben wieder heraustreten muß." Dieser Satz, der über ein Jahrhundert lang nicht angezweifelt wurde, ist das Herzstück des Beweises, Topologen können ihn bis heute nur auf subtile Weise zeigen. Gauss bemerkt erläuternd in einer Fußnote (Werke 3, S. 27; Ostwald's Klassiker Nr. 14, S. 33): „Wie mir scheint, ist es wohl hinreichend sicher bewiesen, daß eine algebraische Curve weder plötzlich irgendwo abbricht, noch sich nach unendlich vielen Umläufen gewissermaßen in einem Punkt verlieren kann (wie die logarithmische Spirale)".

Eine ausgewogene Kritik und eine Ergänzung des ersten Gaußschen Beweises wurde erst 1920 von A. Ostrowski gegeben („Über den ersten und vierten Gaußschen Beweis des Fundamentalsatzes der Algebra", Gauss Werke 10.2, Abh. 3); Ostrowski beginnt seine Arbeit mit den Sätzen: „Während die im ersten Teil der Gaußschen Dissertation enthaltene Besprechung der früheren Beweisversuche des Fundamentalsatzes der Algebra sich durch ganz außerordentliche Sorgfalt auszeichnet, fällt daneben der im zweiten Teil entwickelte Beweis dieses Satzes etwas ab. Nicht etwa, weil dieser Beweis in geometrischer Einkleidung vorgetragen wird,

sondern, weil bei ihm Eigenschaften der algebraischen Kurven verwendet werden, die weder in der *Dissertation* selbst, noch in der vorgaußschen Literatur bewiesen sind."

GAUSS gab 1816 einen zweiten Beweis des Fundamentalsatzes, der sehr algebraisch ist. Aus der Analysis wird lediglich benutzt, daß *reelle Polynome ungeraden Grades stets reelle Nullstellen haben.* GAUSS greift die algebraische Grundidee von EULER mit einer 1759 von DE FONCENEX vorgeschlagenen Vereinfachung auf und benutzt, obgleich er nicht über den allgemeinen Körperbegriff verfügt, das echt algebraische Hilfsmittel der Unbestimmten; er führt Rechnungen, die seine Vorläufer mit den *unzulässigerweise* angenommenen Wurzeln anstellten, aus, indem er eben diese Größen *zulässigerweise* als Unbestimmte auffaßt. Solche Überlegungen liegen noch heute dem üblichen Beweis der Existenz eines Zerfällungskörpers zugrunde; der zweite Gaußsche Beweis ist – auch mit heutigen Maßstäben – absolut korrekt.

Der dritte Beweis von GAUSS stammt ebenfalls aus dem Jahre 1816; er ist wieder topologisch: diesmal werden mittels Doppelintegralen die Umläufe gezählt, die die Bildpunkte $f(z)$ um den Nullpunkt beschreiben, wenn z eine geschlossene Kurve um 0 durchläuft; die Grundidee dieses Beweises findet man heute noch in funktionentheoretischen Beweisen, die das Integral $(1/2\pi i)\int (f'(z)/f(z))\,dz$ heranziehen (Satz von ROUCHÉ).

Bis 1849 werden in allen Beweisen – auch in den inzwischen von CAUCHY, ABEL, JACOBI u. a. gefundenen – nur reelle Polynome betrachtet. Erst in seinem vierten Beweis, der eine Variante des ersten ist, läßt GAUSS 1849, als die Zeit reif war, beliebige komplexe Polynome zu. Allerdings ist dies keine echte Verallgemeinerung, da man von einem *komplexen* Polynom $f\in\mathbb{C}[z]$ sofort vermöge $g(z):=\overline{f(\bar{z})}\,f(z)\in\mathbb{R}[z]$ zu einem *reellen* Polynom übergehen kann (!): ist dann c Nullstelle von g, so ist c oder $\bar{c}$ Nullstelle von f. Aus heutiger Sicht ist indessen der Beweis für reelle Polynome um nichts einfacher als der Beweis für komplexe Polynome.

8. ARGAND (1768–1822) und CAUCHY (1789–1857). Der wohl einfachste Beweis des Fundamentalsatzes der Algebra wurde 1814 von R. ARGAND veröffentlicht in der Arbeit „Réflexions sur la nouvelle théorie des imaginaires, suivies d'une application à la démonstration d'un théorème d'Analyse", Annales de Mathématiques 5, 197–209. ARGAND, der bereits 1806 in seinem Essay über die Darstellung komplexer Zahlen seine Beweisidee skizzierte, vereinfacht in verblüffender Weise die Anwendung der d'Alembertschen Grundidee; er benutzt den allgemeinen Satz vom Vorhandensein des kleinsten Wertes einer (stetigen) Funktion und gewinnt so einen ganz neuartigen Beweis. Da ARGAND nichts zur Existenz des Minimums sagt, wurde sein elementarer Beweis zunächst nicht akzeptiert. CAUCHY gab 1820 in „Sur les racines imaginaires des équations" (Œuvres 1, 2. Ser., 258–263) im wesentlichen denselben Beweis, aber in zugänglicherer Gestalt; er hat damit sehr zur Verbreitung der Argandschen Ideen beigetragen.

Auch bei CAUCHY wird nicht sauber begründet, daß $|f(z)|$ irgendwo einen kleinsten Wert annimmt; das wurde erst möglich, nachdem der Allgemeinbegriff

der unteren Grenze eingeführt worden war. CAUCHY widmet dem Fundamentalsatz der Algebra in seinem „Cours d'Analyse" ein ganzes Kapitel (Chapitre X), ARGAND wird nicht zitiert.

Im 19. Jahrhundert fand die Argandsche Schlußweise Aufnahme in Lehrbücher, so z. B. 1877 im „Lehrbuch der Analysis, Bd. 1" von R. LIPSCHITZ, ferner 1886 im von G. CHRYSTAL veröffentlichten Werk „Algebra, An Elementary Text-Book For Higher Classes of Secondary Schools And For Colleges". CHRYSTAL, dessen Lehrbuch im englischen Sprachraum ungewöhnlich großen Einfluß hatte (vgl. hierzu die Besprechung der Chrystalschen Algebra von S. ABHYANKAR in „The Mathematical Intelligencer" 1, S. 37 (1978)), nennt Argands Beweis „both ingenious and profound" (S. 248).

Der Argandsche Beweis ist in neuerer Zeit infolge der funktionentheoretischen Beweise etwas in Vergessenheit geraten. Ende der zwanziger Jahre hat O. SCHREIER diesen Beweis noch wiederholt in seinen Hamburger Vorlesungen über „Analytische Geometrie und Algebra" vorgetragen; man findet ihn z. B. im ersten Band der *ersten* Auflage des Buches von SCHREIER und SPERNER (Teubner Verlag 1931, S. 221 ff.) Auch E. LANDAU hat 1934 in seiner „Einführung in die Differentialrechnung und Integralrechnung" den Argandschen Beweis in der für LANDAU charakteristischen Weise dargestellt (S. 233 ff.); weiter steht der Argandsche Beweis im 2. Band der „Einführung in die Höhere Mathematik" von H. v. MANGOLDT und K. KNOPP (11. Aufl. Hirzel Verlag Stuttgart 1958, S. 546 ff.). – In diesem Kapitel wird der Argandsche Beweis reproduziert.

9. Fundamentalsatz der Algebra: einst und jetzt. Man kann nur spekulieren, wie sich die Mathematiker bis zum Beginn des 19. Jahrhunderts die Lösungen von Gleichungen vorgestellt haben. Es ist schwer verständlich, warum man bis GAUSS unerschütterlich an eine „außerirdische" Existenz von Lösungen im „irgendwo" glaubte und dann versuchte zu zeigen, daß diese Lösungen bereits komplexe Zahlen sind; noch weniger begreift man indessen, daß in Lehrbüchern der Algebra bis weit ins 19. Jahrhundert hinein häufig „nicht einmal ein Enoncé dieses Fundamentalsatzes gegeben wird [sondern] derselbe auf wunderliche Weise hinein escamotirt wird" (HANKEL 1867, S. 98). Eine rühmliche Ausnahme macht hier 1767 der Göttinger Mathematiker, Physiker und (vor GAUSS) Leiter der Sternwarte Abraham Gotthelf KÄSTNER (1719–1800, schrieb auch Sinngedichte, Aphorismen, satirische Epigramme sowie pointierte Glossen auf literarische Neuheiten, befreundet mit GOTTSCHED), der im Artikel 210 seiner „Anfangsgründe der endlichen Analysis" den Fundamentalsatz ganz ausdrücklich als *Axiom postuliert.*

Heute gehört der Fundamentalsatz der Algebra zu den etablierten Sätzen der Algebra bzw. Funktionentheorie, den Studierende ohne Protest akzeptieren. Alle Beweise benutzen letzten Endes *nicht-algebraische* (*analytische, transzendente*) Hilfsmittel. Entweder verkleinert man – wie D'ALEMBERT, ARGAND und CAUCHY – durch geschickte Wahl des Argumentes sukzessive den Betrag des Polynoms: dann muß man reine Gleichungen lösen und einen Minimumsatz zur Verfügung haben; oder man spaltet – wie EULER, LAGRANGE und LAPLACE – Faktoren ab: dann lassen sich die Hilfsmittel der Analysis stärker zurückdrängen, man benutzt „nur" das Vorhandensein von Quadratwurzeln aus komplexen Zahlen und den Satz, daß reelle Polynome ungeraden Grades reelle Nullstellen haben.

Besonders beliebt sind Beweise, die Resultate der Cauchyschen Funktionentheorie heranziehen: etwa das *Maximumprinzip* oder den *Offenheitssatz* oder den Satz von LIOUVILLE, nach dem jede in ganz $\mathbb{C}$ holomorphe und beschränkte Funktion konstant ist (vgl. Grundwissen Mathematik, Bd. 5, Funktionentheorie I, wo sich vier solche Beweise finden). Viele Mathematiker glauben, daß es keinen rein algebraischen Beweis geben kann, denn der Körper $\mathbb{R}$, und damit sein Oberkörper $\mathbb{C}$, ist ein Erzeugnis der Analysis.

10. Kurzbiographie von Carl Friedrich GAUSS: geb. 30. April 1777 in Braunschweig, Mathematiker, Astronom, Geodät und Physiker; vermutete bereits 1792 mit 15 Jahren den erst 100 Jahre später bewiesenen *Primzahlsatz* durch Abzählungen in ihm geschenkten Primzahl- und Logarithmentafeln; als Stipendiat des Herzogs von Braunschweig von 1795–1798 Student zu Göttingen; 1796 Konstruktion des regulären 17-Ecks mit Zirkel und Lineal; 1799 Promotion in absentia bei PFAFF an der zu Braunschweig gehörenden Universität Helmstedt; 1801 Veröffentlichung der „Disquisitiones Arithmeticae", des grundlegenden Werkes über Zahlentheorie mit Ewigkeitswert; 1801 korr. Mitglied der Petersburger Akademie; 1801 Berechnung der Ceres-Bahn mittels Ausgleichsrechnung aus wenigen Beobachtungsdaten; 1807 Professor der Astronomie und Direktor der Sternwarte in Göttingen; 1810 Ablehnung eines Rufes nach Berlin; 1818 Beginn der Arbeiten zur Vermessung des Königreichs Hannover; 1820 Erfindung des Heliotropen; 1821–1825 Leitung der geodätischen Arbeiten im Gelände; 1828 Gast in Berlin bei Alexander von HUMBOLDT, Bekanntschaft mit Wilhelm WEBER; 1841 Studium der russischen Sprache, um die Arbeiten von LOBATSCHEWSKI zur ihm seit langem vertrauten nichteuklidischen Geometrie zu lesen; 1842 Gründungsmitglied*) des Ordens „Pour Le Mérite für Wissenschaften und Künste"; 1845–1850 langwierige Berechnungen zur Reorganisation der Göttinger Professoren-Witwenkasse; gest. 23. Februar 1855 in Göttingen. – Große Teile des Gaußschen mathematischen Wissens wurden erst durch die Veröffentlichung seines Nachlasses bekannt, sein Motto war: Pauca sed matura. Nach seinem Tod wurden im Königreich Hannover auf Veranlassung des Königs Münzen geprägt, auf denen er wie bereits zu Lebzeiten „Princeps mathematicorum" genannt wird. Durch Lektüre auswärtiger Zeitungen in einer Göttinger Lesestube und systematische Auswertung der Börsennachrichten erwarb GAUSS mit erfolgreichen Börsenspekulationen ein beträchtliches Privatvermögen. – Bereits 1856 erschien ein Nachruf „Gauß zum Gedächtnis" von Sartorius v. WALTERSHAUSEN. Sehr anregend und kritisch ist das 1981 im Springer-Verlag publizierte Buch von W. K. BÜHLER „GAUSS, A Bibliographical Study".

---

*) Weitere Gründungsmitglieder der 1842 vom König Friedrich Wilhelm IV. von Preußen gestifteten Friedensklasse des Pour Le Mérite Ordens waren u. a.: J. I. BERZELIUS (Chemiker), F. W. BESSEL (Astronom), J. DAGUERRE (Maler, Erfinder der Lichtbilder), J. L. GAY-LUSSAC (Chemiker und Physiker), J. GRIMM (Germanist), F. H. A. v. HUMBOLDT (Naturforscher und Geograph, erster Ordenskanzler), C. G. J. JACOBI (Mathematiker), F. LISZT (Tonkünstler), J. L. F. MENDELSSOHN-BARTHOLDY (Komponist), F. RÜCKERT (Dichter und Orientalist), A. W. v. SCHLEGEL (Dichter) und L. TIECK (Dichter). – Alle Angaben nach „Orden Pour Le Mérite Für Wissenschaften Und Künste" Gebr. Mann Verlag, Berlin 1975.

§ 2. Beweis des Fundamentalsatzes nach ARGAND

Der ARGANDsche Beweis benutzt drei Hilfsmittel:

0) *Jedes komplexe Polynom ist eine stetige Funktion in* $\mathbb{C}$.

1) *Jede stetige Funktion* $f\colon K \to \mathbb{R}$ *auf einem Kompaktum* K *in* $\mathbb{R}^2$ *nimmt in* K *ein Minimum an.*

2) *Jede komplexe Zahl* $\neq 0$ *hat* k*-te Wurzeln,* $1 \leqslant k < \infty$.

Die Aussagen 0) und 1) gehören zu den Grundlagen der Analysis; die Aussage 2) wurde in 3.6.4 bewiesen.

Wir führen den Beweis in drei Schritten: zunächst wird in 2.1 mittels einer Wachstumsüberlegung gezeigt, daß die reelle Betragsfunktion $|f(z)|$ eines jeden komplexen Polynoms $f(z)$ in $\mathbb{C}$ stets ein *Minimum annimmt*, dies ist der sog. Minimumsatz von CAUCHY. Der Satz von D'ALEMBERT-GAUSS besagt nun, daß für nichtkonstante Polynome dieses Minimum immer Null ist; der Nachweis dafür gelingt in 2.2 in drei Zeilen mittels der Argandschen Ungleichung, die eine Abschätzung für Werte komplexer Polynome gibt. Diese Ungleichung ist der Kern des Argandschen Argumentes; sie wird in 2.3 hergeleitet und beruht auf einer einfachen allgemeinen Ungleichung für Polynome des Typs $a + z^k g(z)$.

1. Der CAUCHYsche Minimumsatz. *Zu jedem Polynom* $f(z) = a_0 + a_1 z + \cdots + a_n z^n \in \mathbb{C}[z]$ *gibt es ein* $c \in \mathbb{C}$, *so daß gilt*: $|f(c)| = \inf |f(\mathbb{C})|$.

Beweis. Man darf $a_n \neq 0$ mit $n \geqslant 1$ annehmen. Wir benötigen eine *Wachstumsaussage*:

(*): *Es gibt ein* $r \in \mathbb{R}$, *so daß* $|f(z)| > |f(0)|$ *für alle* $z \in \mathbb{C}$ *mit* $|z| > r$.

Für $z \neq 0$ gilt $|f(z)| = |z|^n |a_n + h(z^{-1})|$ mit $h(w) := a_{n-1} w + \cdots + a_0 w^n \in \mathbb{C}[w]$. Da h stetig in 0 ist, gibt es ein $\delta > 0$, so daß $|h(w)| \leqslant \frac{1}{2}|a_n|$, falls $|w| < \delta$. Es folgt $|f(z)| \geqslant |z|^n (|a_n| - |h(z^{-1})|) \geqslant \frac{1}{2}|a_n|\,|z|^n$, falls $|z| > \delta^{-1}$. Es genügt somit, $r > \delta^{-1}$ so zu wählen, daß gilt: $|a_n| r^n > 2|a_0|$.

Nach dieser Vorbereitung ist der eigentliche Beweis des Minimumsatzes schnell erbracht: Da mit $f(z)$ auch die Funktion $|f(z)|$ stetig in $\mathbb{C}$ ist, so nimmt $|f(z)|$ im kompakten Kreis $K := \{z \in \mathbb{C}\colon |z| \leqslant r\}$ ein Minimum an (Aussage 1) der Einleitung): es gibt also ein $c \in K$ mit $|f(c)| = \inf |f(K)|$. Da $|f(c)| \leqslant |f(0)| \leqslant \inf |f(\mathbb{C} \setminus K)|$ nach (*), so folgt: $|f(c)| = \inf |f(\mathbb{C})|$. □

CAUCHY hat die Existenz des Minimums 1821 in seinem „Cours d'Analyse" ebenfalls zu einem Beweis des Satzes von D'ALEMBERT-GAUSS herangezogen (Chapitre X). Die Existenz von Minima in kompakten Mengen, die wir aus der reellen Analysis ohne Beweis übernommen haben, wird natürlich bei CAUCHY noch nicht bewiesen.

Auch in den funktionentheoretischen Beweisen des Fundamentalsatzes benötigt man eine Aussage über das Wachstum von Polynomen, die hier durch (*) wiedergegeben wird.

2. Beweis des Fundamentalsatzes. Neben dem Minimumsatz benötigen wir die

Argandsche Ungleichung: *Es sei $f(z)$ ein nichtkonstantes Polynom. Dann gibt es zu jedem Punkt $c \in \mathbb{C}$ mit $f(c) \neq 0$ einen Punkt $c' \in \mathbb{C}$ mit*

$$|f(c')| < |f(c)|.$$

Der Beweis dieser Ungleichung wird im nächsten Abschnitt geführt, indem man k-te Wurzeln zieht. Es folgt nun schnell, daß jedes nichtkonstante komplexe Polynom $f(z) \in \mathbb{C}[z]$ eine Nullstelle c in $\mathbb{C}$ hat. Nach dem Minimumsatz existiert nämlich ein $c \in \mathbb{C}$, so daß $|f(c)| \leqslant |f(z)|$ für alle $z \in \mathbb{C}$ gilt. Wäre $f(c) \neq 0$, dann gäbe es nach der Argandschen Ungleichung ein $c' \in \mathbb{C}$ mit $|f(c')| < |f(c)|$, was absurd ist. □

In der Algebra nennt man einen Körper *K algebraisch abgeschlossen*, wenn jedes Polynom $f \in K[X] \setminus K$ eine Nullstelle in K besitzt. Der Fundamentalsatz läßt sich dann auch so aussprechen:

Der Körper $\mathbb{C}$ der komplexen Zahlen ist algebraisch abgeschlossen.

3. Beweis der Argandschen Ungleichung. Entscheidend für den Beweis ist folgendes

Lemma: *Es sei k eine natürliche Zahl $\geqslant 1$, und es sei g ein komplexes Polynom mit $g(0) \neq 0$. Dann gibt es zu jedem $a \in \mathbb{C}^\times$ ein $d \in \mathbb{C}$ mit*

$$|a + d^k g(d)| < |a|.$$

Beweis. Da g als Polynom stetig in $\mathbb{C}$ ist (Aussage 0) der Einleitung), gibt es ein $\delta > 0$, so daß $|g(z) - g(0)| < \frac{1}{2}|g(0)|$ gilt, falls $|z| < \delta$. Damit gewinnen wir für $a + z^k g(z) = a + g(0)z^k + z^k(g(z) - g(0))$ die Abschätzung

$$(*) \quad |a + z^k g(z)| \leqslant |a + g(0)z^k| + \tfrac{1}{2}|g(0)|\,|z^k| \quad \text{für alle} \quad z \in \mathbb{C} \quad \text{mit} \quad |z| < \delta.$$

Wir wählen nun $d \in \mathbb{C}$ so, daß $g(0)d^k$ „entgegengesetzt zu a gerichtet" ist: genauer soll gelten

$$g(0)d^k = -ta \qquad \text{mit} \qquad 0 < t < 1,$$

wobei über t unten näher verfügt wird; solche Zahlen d existieren, da es k-te Wurzeln zu $-tag(0)^{-1} \neq 0$ gibt (Aussage 2) der Einleitung). Dann folgt

$$|a + g(0)d^k| = (1 - t)|a| \qquad \text{und} \qquad \tfrac{1}{2}|g(0)|\,|d^k| = \tfrac{1}{2}t|a|.$$

Um dies in $(*)$ eintragen zu können, muß $|d| < \delta$ sichergestellt sein: das führt wegen $t = -g(0)a^{-1}d^k$ zu $t < |g(0)a^{-1}|\delta^k$. Unterwirft man t zusätzlich dieser Ungleichung, so erhält man für ein zugehöriges d aus $(*)$:

$$|a + d^k g(d)| \leqslant (1 - t)|a| + \tfrac{1}{2}t|a| = (1 - \tfrac{1}{2}t)|a| < |a|. \qquad \square$$

Der Leser wird bemerken, daß von $g: \mathbb{C} \to \mathbb{C}$ nur die Stetigkeit im Nullpunkt benutzt wurde, das Lemma gilt also für alle solchen Funktionen g.

Nunmehr ist die Argandsche Ungleichung schnell bewiesen: mit $f(z)$ ist auch $f(c + z)$ ein nichtkonstantes Polynom. Es gilt also eine Gleichung

$$f(c + z) = f(c) + b_k z^k + b_{k+1} z^{k+1} + \cdots + b_n z^n \qquad \text{mit} \qquad b_k \neq 0.$$

Dann ist $g(z) := b_k + b_{k+1}z + \cdots + b_n z^{n-k}$ ein Polynom mit $g(0) \neq 0$ und $f(c+z) = f(c) + z^k g(z)$. Wegen $f(c) \neq 0$ gibt es dann nach dem Lemma ein $d \in \mathbb{C}$, so daß für $c' := c + d$ gilt: $|f(c')| = |f(c+d)| < |f(c)|$.

In der Funktionentheorie ist die Argandsche Ungleichung ein Spezialfall des allgemeinen „Offenheitssatzes", nach dem nicht konstante holomorphe Funktionen offene Mengen stets auf offene Mengen abbilden (vgl. Grundwissen Mathematik, Bd. 5, Funktionentheorie I).

4. Konstruktive Beweise des Fundamentalsatzes. Der Argand-Cauchysche Beweis ist ein Existenzbeweis und *nicht konstruktiv*. Bereits WEIERSTRASS hat 1859 in seiner Note „Neuer Beweis des Fundamentalsatzes der Algebra" (Math. Werke 1, 247–256) folgenden Beweisansatz gemacht: bei vorgegebenem Polynom $f(z)$ wird $z_0 := c \in \mathbb{C}$ beliebig gewählt und alsdann eine Folge $z_n := z_{n-1} - f(z_{n-1})$ induktiv definiert. WEIERSTRASS sagt (S. 247): „... es läßt sich zeigen, daß unter bestimmten Bedingungen, wenn n ohne Ende wächst, z_n einer bestimmten Grenze sich nähert, welche für z gesetzt die Gleichung $f(z) = 0$ befriedigt." Mehr als 30 Jahre später (1891, Math. Werke 3, 251–269) diskutiert WEIERSTRASS nochmals ausführlich das Problem eines *konstruktiven* Beweises, welches sich wie folgt präzisieren läßt:

„Läßt sich für jedes Polynom $f \in \mathbb{C}[z]$ auf effektive Weise eine konvergente Folge komplexer Zahlen z_n herstellen, bei der $|f(z_n)|$ gegenüber $|f(z_{n-1})|$ so wirksam verkleinert ist, daß z_n gegen eine Nullstelle von f konvergiert?" H. KNESER hat 1940 in seiner Arbeit „Der Fundamentalsatz der Algebra und der Intuitionismus", Math. Zeitschr. 46, 287–302, ein solches Herstellungsverfahren beschrieben, das eine konstruktive Variante des Beweises von ARGAND-CAUCHY ist und das auch der Kritik eines Intuitionisten standhält. M. KNESER hat 1981 das Verfahren seines Vaters nochmals vereinfacht: Ergänzung zu einer Arbeit von Hellmuth KNESER über den Fundamentalsatz der Algebra, Math. Zeitschr. 177, 285–287.

1979 haben M. W. HIRSCH und St. SMALE einen sogenannten „sure fire algorithm" angegeben, der *für jedes* nicht konstante Polynom $f(z) \in \mathbb{C}[z]$ bei *beliebigem* Ausgangspunkt $c \in \mathbb{C}$ eine Folge z_n mit $z_0 := c$ produziert, die gegen eine Nullstelle von f konvergiert; genauer wird gezeigt:

$$|f(z_n)| \leqslant K^n |f(c)|, \qquad n = 0, 1, 2, \ldots$$

mit einer nur vom Polynomgrad (nicht aber von f selbst!) abhängenden positiven reellen Konstanten $K < 1$. Wegen Einzelheiten sei auf die Arbeit „On Algorithms for Solving $f(x) = 0$" in Comm. Pure Appl. Math. 32, 281–312 verwiesen, insb. S. 303 ff.

§ 3. Anwendungen des Fundamentalsatzes

Die Existenz einer *einzigen* Nullstelle für jedes nichtkonstante komplexe Polynom impliziert bereits, daß komplexe Polynome in Linearfaktoren und reelle Polynome in lineare und quadratische Faktoren *zerfallen*. Diese Folgerungen aus dem Fundamentalsatz sind absolut elementar, sie resultieren aus der simplen Tatsache, daß ein Polynom mit der Nullstelle c stets den Faktor $z - c$ abspaltet.

1. Lemma über die Abspaltung von Nullstellen. Ist $c \in \mathbb{C}$ *eine Nullstelle eines Polynoms* $f \in \mathbb{C}[z]$ *vom Grade* n, *so gibt es genau ein Polynom* $g \in \mathbb{C}[z]$ *vom Grade* $n - 1$, *so daß gilt*: $f(z) = (z - c)g(z)$.

Beweis. Sei $f = a_0 + a_1 z + \cdots + a_n z^n$, $a_n \neq 0$. Wegen $z^\nu - c^\nu = (z-c)q_\nu(z)$ mit $q_\nu(z) := z^{\nu-1} + z^{\nu-2}c + \cdots + c^{\nu-1}$ folgt

$$f(z) = f(z) - f(c) = \sum_1^n a_\nu (z^\nu - c^\nu) = (z-c)g(z), \quad \text{wobei} \quad g(z) := \sum_1^n a_\nu q_\nu(z).$$

Es ist klar, daß g den Grad $n-1$ hat; wegen $g(z) = (z-c)^{-1} f(z)$, $z \neq c$, ist g durch f und c eindeutig bestimmt. □

Das Abspaltungslemma gilt für alle kommutativen Ringe, wenn man auf die Eindeutigkeit von g verzichtet. Durch Induktion nach n folgt sofort das

Korollar. *Ein Polynom $f \in \mathbb{C}[z]$ vom Grad n hat höchstens n Nullstellen.*

2. Faktorisierung komplexer Polynome. *Jedes komplexe Polynom $f \in \mathbb{C}[z]$ vom Grad $n \geqslant 1$ ist (bis auf die Reihenfolge der Faktoren) eindeutig darstellbar in der Form*

$$(1) \qquad f(z) = a(z-c_1)^{n_1}(z-c_2)^{n_2} \cdot \ldots \cdot (z-c_r)^{n_r},$$

wobei $a \in \mathbb{C}^\times$; $r \in \mathbb{N}$, $c_1, \ldots, c_r \in \mathbb{C}$ paarweise verschieden, $n_1, \ldots, n_r \in \mathbb{N} \setminus \{0\}$ mit $n_1 + n_2 + \cdots + n_r = n$.

Beweis. Durch Induktion nach n, der Fall $n = 1$ ist klar. Sei $n > 1$. Nach dem Fundamentalsatz der Algebra existiert eine Nullstelle $c_1 \in \mathbb{C}$ von f. Nach Lemma 1 gilt $f(z) = (z-c_1)g(z)$, wobei $g(z) \in \mathbb{C}[z]$ vom Grad $n-1$ ist. Nach Induktionsannahme gilt eine eindeutige Faktorisierung

$$g(z) = a(z-c_1)^{n_1-1}(z-c_2)^{n_2} \cdot \ldots \cdot (z-c_r)^{n_r}$$

mit $n_1 \geqslant 1, \ldots, n_r \geqslant 1$, $n_1 - 1 + n_2 + \cdots + n_r = n-1$; $c_1, \ldots, c_r \in \mathbb{C}$ paarweise verschieden, $a \in \mathbb{C}^\times$. Damit folgt (1). □

Der eben bewiesene Satz wird häufig so ausgesprochen (wobei die Nullstellen c_j mit ihrer Vielfachheit n_j gezählt werden):

Jedes komplexe Polynom n-ten Grades hat genau n Nullstellen.

3. Faktorisierung reeller Polynome. Jedes *reelle* Polynom $f = \sum a_\nu z^\nu$ ist ein komplexes Polynom, dabei gilt zusätzlich

$$\overline{f(z)} = f(\bar{z}) \qquad \text{für alle} \qquad z \in \mathbb{C},$$

denn wegen $\bar{a}_\nu = a_\nu$ gilt: $\overline{\sum a_\nu z^\nu} = \sum a_\nu \bar{z}^\nu$. *Speziell ist mit c auch stets $\bar{c}$ eine Nullstelle von $f \in \mathbb{R}[x]$.* Hieraus ergibt sich leicht:

Satz. *Jedes reelle Polynom $f \in \mathbb{R}[x]$ vom Grad $n \geqslant 1$ ist (bis auf die Reihenfolge der Faktoren) eindeutig darstellbar in der Form*

$$(1) \qquad f(x) = a(x-c_1)^{m_1} \cdot \ldots \cdot (x-c_s)^{m_s} q_1(x)^{n_1} \cdot \ldots \cdot q_t(x)^{n_t},$$

wobei gilt:

(a) $a \in \mathbb{R}$, $a \neq 0$; $s, t \in \mathbb{N}$; $c_1, \ldots, c_s \in \mathbb{R}$ *paarweise verschieden*; $m_1, \ldots, m_s$, $n_1, \ldots, n_t \in \mathbb{N} \setminus \{0\}$ *mit* $m_1 + \cdots + m_s + 2n_1 + \cdots + 2n_t = n$.

(b) $q_j(x) = x^2 - b_j x - a_j$ *mit* $b_j^2 + 4a_j < 0$ *für* $j = 1, \ldots, t$; $q_1, \ldots, q_t$ *sind paarweise verschieden.*

Beweis. Wir fassen f als komplexes Polynom auf und faktorisieren gemäß Satz 2. Wir bezeichnen mit $c_1, \ldots, c_s$ die *reellen* Nullstellen. Die übrigen echt komplexen Nullstellen fassen wir zu reellen quadratischen Polynomen $q(x) = (x - c)(x - \bar{c})$ $= x^2 - (c + \bar{c})x + c\bar{c} \in \mathbb{R}[x]$ zusammen. Mit $b := c + \bar{c}$, $a := -c\bar{c}$ gilt $b^2 + 4a$ < 0, da sonst $q(x) = (x - \frac{1}{2}b)^2 - \frac{1}{4}(b^2 + 4a)$ eine reelle Nullstelle hätte. Die Aussage des Satzes folgt nun unmittelbar. □

In der Formulierung des vorangehenden Satzes kommen keine komplexen Zahlen mehr vor! Im Beweis spielen sie aber die wesentliche Rolle (deus ex machina). GAUSS selbst hat übrigens in seiner Dissertation, wie schon deren Titel sagt (vgl. 1.6), den Fundamentalsatz der Algebra als Faktorisierungssatz für reelle Polynome formuliert. Dieser Satz wird u. a. in der reellen Integralrechnung verwendet, um unbestimmte Integrale rationaler Funktionen mittels Partialbruchzerlegung zu berechnen, siehe zum Beispiel A. OSTROWSKI: Vorlesungen über Differential- und Integralrechnung I, Birkhäuser, Basel 1952, S. 275f.

4. Primpolynome in $\mathbb{C}[z]$ und $\mathbb{R}[x]$. Wir ordnen die Ergebnisse der Abschnitte 2 und 3 in einen größeren Zusammenhang ein. Ist K irgendein (kommutativer) Körper, so heißt ein Polynom $p \in K[x] \setminus K$ mit höchstem Koeffizienten 1 ein *normiertes Primpolynom*, wenn p nicht als Produkt zweier Polynome $g, h \in K[x] \setminus K$ darstellbar ist. Alle Polynome $x - c$, $c \in K$, sind normierte Primpolynome. Aus der Algebra entnehmen wir:

Der Polynomring $K[x]$ ist faktoriell, das heißt jedes Polynom $f \in K[x] \setminus \{0\}$ ist (bis auf die Reihenfolge der Faktoren) eindeutig darstellbar in der Form

$$f = a p_1^{m_1} p_2^{m_2} \cdot \ldots \cdot p_r^{m_r} \qquad \text{mit} \qquad r \in \mathbb{N};\ m_1, \ldots, m_r \in \mathbb{N} \setminus \{0\},$$

wo $a \in K \setminus \{0\}$ und $p_1, p_2, \ldots, p_r \in K[x]$ paarweise verschiedene normierte Primpolynome sind.

In den Fällen $K = \mathbb{C}$ bzw. $\mathbb{R}$ wird diese Primfaktorzerlegung von Polynomen durch die Sätze 2 und 3 präzisiert. Insbesondere folgt unmittelbar

Im Polynomring $\mathbb{C}[z]$ ist jedes normierte Primpolynom p linear, das heißt, $p(z) = z - c$, $c \in \mathbb{C}$.

Im Polynomring $\mathbb{R}[x]$ ist jedes normierte Primpolynom p linear oder quadratisch: $p(x) = x - c$, $c \in \mathbb{R}$, oder $p(x) = x^2 - bx - a$ mit $b^2 + 4a < 0$.

Jede dieser letzten beiden Aussagen ist äquivalent zum Fundamentalsatz der Algebra. Bei beliebigem Grundkörper K gibt es im allgemeinen Primpolynome beliebig hohen Grades in $K[x]$: so ist z. B. in $\mathbb{Q}[x]$ für jedes $n \geqslant 1$ das Polynom $x^n - 2$ ein normiertes Primpolynom.

5. Einzigkeit von $\mathbb{C}$. Die Wahl des Körpers $\mathbb{C}$ der komplexen Zahlen ist weder willkürlich noch zufällig. Wir haben bereits in 3.2.3 eine Eindeutigkeitsaussage für $\mathbb{C}$ kennengelernt. Wir zeigen nun mit Hilfe des Fundamentalsatzes der Algebra einen allgemeineren

Einzigkeitssatz für $\mathbb{C}$. *Es sei $\mathbb{K}$ ein kommutativer, nullteilerfreier Oberring von $\mathbb{R}$ mit Eins* 1, *derart, daß jedes Element aus $\mathbb{K}$ algebraisch über $\mathbb{R}$, das heißt Nullstelle eines reellen Polynoms $\neq 0$, ist.*

Dann ist $\mathbb{K}$ isomorph zu $\mathbb{R}$ oder zu $\mathbb{C}$.

Den Beweis dieses Satzes stützen wir auf folgende einfache (aber auf dem Fundamentalsatz basierende)

Hilfsaussage. *Unter den Voraussetzungen des Einzigkeitssatzes besteht für jedes Element $v \in \mathbb{K} \setminus \mathbb{R}$ eine Gleichung: $v^2 = a + bv$ mit $a, b \in \mathbb{R}$.*

Beweis. Laut Annahme gibt es ein reelles Polynom $f \neq 0$ mit $f(v) = 0$. Da $\mathbb{K}$ nullteilerfrei ist, annulliert v dann nach Satz 3 auch ein Polynom p vom Grad 1 oder 2. Da p wegen $v \notin \mathbb{R}$ nicht linear sein kann, folgt $p(x) = x^2 - bx - a$, das heißt, $v^2 = a + bv$. □

Wir kommen nun zum eigentlichen Beweis des Einzigkeitssatzes. Es sei $\mathbb{K} \neq \mathbb{R}$. Wir wählen ein Element $v \in \mathbb{K} \setminus \mathbb{R}$ und betrachten den von 1 und v erzeugten, 2-dimensionalen reellen Vektorraum $V = \mathbb{R} + \mathbb{R}v$. Da v nach der Hilfsaussage einer Gleichung $v^2 = a + bv$ mit $a, b \in \mathbb{R}$ genügt, so folgt für beliebige Elemente $x_1 + y_1 v$, $x_2 + y_2 v \in V$:

$$(x_1 + y_1 v)(x_2 + y_2 v) = (x_1 x_2 + y_1 y_2 a) + (x_1 y_2 + y_1 x_2 + y_1 y_2 b)v \in V.$$

Mithin ist V ein kommutativer, nullteilerfreier, 2-dimensionaler Oberring von $\mathbb{R}$ mit Eins und also nach Satz 3.2.3 zu $\mathbb{C}$ isomorph.

Es bleibt zu zeigen: $\mathbb{K} = V$. Sei $u \in \mathbb{K} \setminus \mathbb{R}$ beliebig. Es gibt ein reelles Polynom $f \neq 0$ mit $f(u) = 0$. Über $\mathbb{C} \simeq V \subset \mathbb{K}$ zerfällt f in Linearfaktoren $x - c$, $c \in V$. Da $\mathbb{K}$ nullteilerfrei ist, annulliert u einen solchen Faktor, das heißt $u = c \in V$. Damit ist $\mathbb{K} = V \simeq \mathbb{C}$ verifiziert. □

Die Voraussetzung des Einzigkeitssatzes, daß jedes Element $w \in \mathbb{K}$ *algebraisch* ist, ist immer dann erfüllt, wenn $\mathbb{K}$ ein endlich-dimensionaler Vektorraum über $\mathbb{R}$ ist: dann sind nämlich die Potenzen $1, w, w^2, \ldots, w^n, \ldots$ linear abhängig, das heißt, es gilt eine Gleichung $a_0 + a_1 w + \cdots + a_n w^n = 0$, wo nicht alle a_ν verschwinden.

6. Ausblick auf „höhere komplexe Zahlen". Der Einzigkeitssatz besagt speziell:

Der Körper $\mathbb{C}$ ist bis auf Isomorphie der einzige echte kommutative algebraische Erweiterungskörper des Körpers $\mathbb{R}$, insbesondere gibt es keinen algebraischen kommutativen Oberkörper $\neq \mathbb{C}$ von $\mathbb{C}$.

Diesen Satz hat WEIERSTRASS ab 1863 in seinen Vorlesungen zu Berlin vorgetragen; er wurde erstmals 1867 von H. HANKEL in seinem Buch „Theorie der complexen Zahlensysteme" veröffentlicht. Bei HANKEL liest es sich so (S. 107):

„*Ein höheres complexes Zahlensystem*, dessen formale Rechenoperationen nach den Bedingungen des § 28 bestimmt sind*), und dessen Einheitsprodukte in's

*) Die Bedingungen des § 28 besagen, daß ein kommutativer Ring mit 1 vorliegt, der ein endlich-dimensionaler Vektorraum über $\mathbb{R}$ ist.

Besondere lineare Functionen der ursprünglichen Einheiten sind, und in welchem kein Product verschwinden kann, ohne dass einer seiner Factoren Null würde, enthält also in sich einen Widerspruch und *kann nicht existieren.*"

HANKEL erklärt dazu stolz (S. 107): „Damit ist die Frage beantwortet, deren Lösung 1831 GAUSS (Werke 2, S. 178) versprochen, aber nicht gegeben hat, ‚*warum die Relationen zwischen Dingen, die eine Mannigfaltigkeit von mehr als zwei Dimensionen darbieten, nicht noch andere in der allgemeinen Arithmetik zulässige Arten von Größen liefern können*'".

Im Einzigkeitssatz ist die Voraussetzung der Kommutativität wesentlich. Bekanntlich bildet das von HAMILTON im Jahre 1843 beschriebene *hyperkomplexe System der Quaternionen* einen 4-*dimensionalen, nicht kommutativen* Oberkörper von $\mathbb{R}$. Darüber hinaus gibt es das 8-*dimensionale, weder kommutative noch assoziative, aber noch nullteilerfreie hyperkomplexe System der* CAYLEY-*Zahlen über* $\mathbb{R}$. Wir werden diese Algebren in den Kapiteln 6–8 dieses Bandes eingehend diskutieren.

Auch die Voraussetzung der Nullteilerfreiheit ist für die Gültigkeit des Einzigkeitssatzes unabdingbar. So ist z. B. $\mathbb{R} \times \mathbb{R}$ mit der „ringdirekten Multiplikation"

$$(a, b)(c, d) := (ac, bd)$$

eine 2-dimensionale kommutative Ringerweiterung von $\mathbb{R}$ mit Einselement $e := (1, 1)$, die Nullteiler, z. B. $(1, 0)$, hat und also nicht zu $\mathbb{C}$ isomorph ist. WEIERSTRASS (1884) und DEDEKIND (1885) zeigten, daß dieses Beispiel signifikant ist: *jeder endlich-dimensionale, kommutative Oberring von* $\mathbb{R}$ *mit Eins, der keine nilpotenten Elemente hat, ist isomorph zu einer ringdirekten Summe aus Exemplaren von* $\mathbb{R}$ *und* $\mathbb{C}$ (dabei heißt $x \neq 0$ *nilpotent*, wenn es einen Exponenten $n \geqslant 2$ mit $x^n = 0$ gibt).

Anhang: Beweis des Fundamentalsatzes nach LAPLACE

Wir besprechen hier den schönen *algebraischen* Beweis, den LAPLACE 1795 skizzierte, und der etwas anders und vielleicht einfacher ist als der zweite Beweis, den GAUSS 1816 gab. Dieser Beweis findet sich bei N. BOURBAKI in *Algèbre*, Chap. VI, 1952, S. 40/41; in der *Note Historique* schreibt BOURBAKI den Beweis GAUSS zu (S. 150). In der „Einführung in die Algebra" von G. FISCHER und R. SACHER, Teubner Verlag, Stuttgart 1974, wird der Beweis mit Hilfe von Galoistheorie geführt und LAGRANGE zugesprochen. Unsere Quelle ist ein 1939 in der Deutschen Mathematik 4, 318–322, publizierter Artikel von Hellmuth KNESER: „Laplace, Gauß und der Fundamentalsatz der Algebra".

1. Hilfsmittel. Wir werden benutzen

1) *Jedes reelle Polynom ungeraden Grades hat wenigstens eine reelle Nullstelle* (*Folgerung aus dem Zwischenwertsatz*).

2) *Zu jedem nichtkonstanten reellen Polynom f gibt es einen Oberkörper K des Körpers* $\mathbb{R}$, *so daß f in $K[x]$ in Linearfaktoren zerfällt* (*Existenz des Zerfällungskörpers*).

3) *Es sei K ein Oberkörper von $\mathbb{R}$, es seien $\zeta_1, \ldots, \zeta_n$ Elemente aus K, und es seien*

$$\eta_k := \sum_{1 \leqslant \nu_1 < \cdots < \nu_k \leqslant n} \zeta_{\nu_1} \cdot \ldots \cdot \zeta_{\nu_k}$$

die „elementarsymmetrischen Funktionen" in $\zeta_1, \ldots, \zeta_n$ (also $\eta_1 = \zeta_1 + \cdots + \zeta_n$, $\ldots$, $\eta_n = \zeta_1 \cdot \ldots \cdot \zeta_n$). Dann gilt (mit x als Unbestimmte)

$$\prod_{\nu=1}^{n} (x - \zeta_\nu) = x^n - \eta_1 x + \eta_2 x^{n-2} - \cdots + (-1)^n \eta_n;$$

jedes in $\zeta_1, \ldots, \zeta_n$ symmetrische Polynom) aus $\mathbb{R}[\zeta_1, \ldots, \zeta_n]$ ist ein reelles Polynom in $\eta_1, \ldots, \eta_n$ (Hauptsatz über symmetrische Funktionen).*

4) *Jedes quadratische komplexe Polynom zerfällt in $\mathbb{C}[z]$ in Linearfaktoren.*

Von diesen vier Aussagen gehört lediglich der Hauptsatz über symmetrische Funktionen, den NEWTON 1673 bewiesen hat, nicht unbedingt zur mathematischen Allgemeinbildung.

2. Beweis. Um die Aussage 1.3) mühelos anwenden zu können, schreiben wir die Koeffizienten des gegebenen Polynoms mit alternierenden Vorzeichen. Der Fundamentalsatz der Algebra ist bewiesen, sobald folgendes gezeigt ist:

Jedes Polynom $h = x^n - b_1 x^{n-1} + b_2 x^{n-2} - \cdots + (-1)^n b_n \in \mathbb{R}[x]$, $n \geqslant 1$, hat eine Nullstelle $c \in \mathbb{C}$.

Beweis (nach LAPLACE). Wir schreiben n in der Form $2^k q$, wo $q \in \mathbb{N}$ eine ungerade Zahl ist, und führen Induktion nach k. Der Induktionsbeginn $k = 0$ ist klar nach 1). Sei $k \geqslant 1$. Nach 2) gibt es einen Körper $K \supset \mathbb{R}$ und Elemente $\zeta_1, \ldots, \zeta_n \in K$, so daß gilt $h = (x - \zeta_1)(x - \zeta_2) \cdot \ldots \cdot (x - \zeta_n) \in K[x]$. Wir bilden nun für jede reelle Zahl t das Polynom

$$L_t := \prod_{1 \leqslant \mu < \nu \leqslant n} (x - \zeta_\mu - \zeta_\nu - t\zeta_\mu\zeta_\nu) \in K[x] \qquad \text{(Laplacescher Kunstgriff).}$$

Entwickelt man nach Potenzen von x, so sind alle Koeffizienten reelle *symmetrische* (!) Polynome in $\zeta_1, \ldots, \zeta_n$, denn L_t is per definitionem invariant, wenn man $\zeta_1, \ldots, \zeta_n$ irgendwie permutiert. Nach 3) sind diese Koeffizienten reelle Polynome in den elementar-symmetrischen Funktionen der $\zeta_1, \ldots, \zeta_n$, das heißt in den reellen Zahlen $b_1, \ldots, b_n$. Es folgt $L_t \in \mathbb{R}[x]$. Da L_t den Grad $\frac{1}{2}n(n-1) = 2^{k-1} q(2^k q - 1)$ hat, und da mit q auch $q(2^k q - 1)$ wegen $k \geqslant 1$ ungerade ist, so hat L_t nach Induktionsannahme eine Nullstelle in $\mathbb{C}$. Auf Grund der Produktform von L_t gibt es also zu jedem $t \in \mathbb{R}$ Indices $\mu < \nu$, so daß $\zeta_\mu + \zeta_\nu + t\zeta_\mu\zeta_\nu$ in $\mathbb{C}$ liegt. Da es nur $\frac{1}{2}n(n-1)$ Indexpaare (μ, ν) mit $1 \leqslant \mu < \nu \leqslant n$ und unendlich viele reelle Zahlen gibt, lassen sich $r, s \in \mathbb{R}$ mit $r \neq s$ und κ, λ mit $1 \leqslant \kappa < \lambda \leqslant n$ so finden, daß gilt:

$$\zeta_\kappa + \zeta_\lambda + r\zeta_\kappa\zeta_\lambda \in \mathbb{C}, \qquad \zeta_\kappa + \zeta_\lambda + s\zeta_\kappa\zeta_\lambda \in \mathbb{C}.$$

---

*) Ein Polynom $p(\zeta_1, \ldots, \zeta_n)$ heißt *symmetrisch*, wenn es bezüglich jeder Permutation der Indices $1, \ldots, n$ invariant ist.

Wegen $r \neq s$ folgt hieraus $u := \zeta_\kappa + \zeta_\lambda \in \mathbb{C}$, $v := \zeta_\kappa \zeta_\lambda \in \mathbb{C}$. Daher sind ζ_κ und ζ_λ die Wurzeln des Polynoms $z^2 - uz + v \in \mathbb{C}[z]$; nach 4) gilt folglich: $\zeta_\kappa, \zeta_\lambda \in \mathbb{C}$.

3. Historisches. LAGRANGE sagte 1797/98 vom Laplaceschen Beweis, daß er „ne laisse rien à désirer comme simple démonstration"; er hielt ihm aber entgegen, daß die wirkliche Durchführung der Rechnungen „comme impossible" sein würde (De la résolution des équations numériques de tous les degrés, Paris, An VI, 1797/98, S. 200–201). In der 1808 erschienenen 2. Auflage dieses Lagrangeschen Werkes wird übrigens Gaußens erster Beweis von 1799 nicht erwähnt, was seinen Grund in dessen geringer Verbreitung haben dürfte. H. KNESER notiert noch: „Merkwürdiger ist es, daß sich hierzu auch bei der dritten Auflage – 1828, nach LAGRANGES Tod von POINSOT veranstaltet – nichts geändert hat: POINSOT weiß nicht nur nichts von Gaußens zweitem und drittem Beweis, die 1816 in den Göttinger Commentationes erschienen, sondern er äußert sogar seine volle Befriedigung über das von LAGRANGE und LAPLACE Erreichte. Gaußens kritischer Gedanke war also in beinahe 30 Jahren bzw. in 12 Friedensjahren nicht bis Paris gedrungen."

Kapitel 5. Was ist π?

R. Remmert

> Und er machte ein Meer, gegossen, zehn Ellen weit, von einem Rande zum andern, rund umher, und fünf Ellen hoch, und eine Schnur dreyßig Ellen lang war das Maaß rings um
> (Könige 1, Kap. 7, Vers 23).

Es gibt viele Möglichkeiten, die Kreiszahl π einzuführen. Wir gewinnen π aus der *komplexen* Exponentialfunktion

$$\exp z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \cdots .$$

Es gibt eine (eindeutig bestimmte) reelle Zahl $\pi > 0$, *so daß genau die Zahlen* $2n\pi i$, $n \in \mathbb{Z}$, *vermöge* $\exp z$ *auf* 1 *abgebildet werden*:

$$\{w \in \mathbb{C} : \exp w = 1\} = 2\pi i\mathbb{Z}. \tag{1}$$

Wir nehmen (1) als Definition von π und leiten hieraus alle wohlbekannten Eigenschaften her; im einzelnen gehen wir – nachdem im Paragraphen 1 die Geschichte der Zahl π geschildert ist – wie folgt vor: Zunächst wird die Theorie der Exponentialfunktion im Komplexen so weit wie nötig entwickelt. Wir setzen beim Leser eine gewisse Vertrautheit mit den grundlegenden Begriffen der reellen Analysis voraus. Absolut konvergente Reihen sind wie im Reellen definiert. Der Körper $\mathbb{C}$ erbt die Vollständigkeit des Körpers $\mathbb{R}$; daher gilt auch für absolut konvergente Reihen komplexer Zahlen der Cauchysche Multiplikationssatz. Wir verwenden diese elementaren Dinge hier ohne weitere Begründung auch im Komplexen, wobei wir auch nichts zum allgemeinen Grenzwertbegriff für Reihen von Funktionen sagen*). Das zentrale Ergebnis des Paragraphen 2 ist der Epimorphiesatz 2.3, der die Exponentialfunktion als *Homomorphismus* $\exp: \mathbb{C} \to \mathbb{C}^\times$ der *additiven* Gruppe $\mathbb{C}$ **auf** die *multiplikative* Gruppe $\mathbb{C}^\times$ beschreibt. Dies folgt schnell, wenn man weiß, daß die Bildmenge $\exp(\mathbb{C})$ eine Umgebung des Punktes 1 enthält. Wir geben hierfür zwei Beweise: einen ganz kurzen mittels Differentiation und einen völlig elementaren, der ohne Differentialrechnung auskommt und lediglich vom Zwischenwertsatz für reelle stetige Funktionen Gebrauch macht, vgl. 2.3 und Anhang zum § 2.

Sobald der Epimorphiesatz zur Verfügung steht, ist es leicht, die Gleichung (1) zu verifizieren. Alsdann ergibt sich auch schnell die Existenz des für die Einführung von Polarkoordinaten unentbehrlichen Polarkoordinatenepimorphismus

---

*) Wir berufen uns bei diesem unsystematischen Vorgehen auf einen Hauptsatz der angewandten Didaktik, den Schiller im Briefwechsel mit Goethe am 5.2.1796 so aussprach: „Wo es die Sache leidet, halte ich es immer für besser, nicht mit dem Anfang anzufangen, der immer das Schwerste ist." Dieses Theorem, dem Mathematiker in Vorlesungen und Lehrbüchern kaum gerecht werden können, fand sich übrigens in Riemanns Nachlaß mitten unter Rechnungen.

$p\colon \mathbb{R} \to S^1$, $\varphi \mapsto e^{i\varphi}$ mit Kern $p = 2\pi\mathbb{Z}$; allerdings ist der Nachweis, daß $p(\frac{\pi}{2}) = i$, ohne Zwischenwertsatz nicht möglich (vgl. 3.5 und 3.6).

„Nach ... den Exponentialgrößen müssen ... der Sinus und der Cosinus betrachtet werden, weil sie aus ... den Exponentialgrößen selbst entspringen, sobald dieselben imaginäre Zahlgrößen enthalten" (EULER 1748 im § 126 seiner Introductio in Analysin infinitorum). Getreu diesem Satz führen wir im Paragraphen 3 die trigonometrischen Funktionen mittels der Exponentialfunktion ein. Die berühmten EULERschen Formeln

$$\cos z = \frac{1}{2}(e^{iz} + e^{-iz}), \qquad \sin z = \frac{1}{2i}(e^{iz} - e^{-iz})$$

werden zur Definition erhoben. Diese EULERsche Entdeckung des Zusammenhanges der trigonometrischen Funktionen mit der Exponentialfunktion hat die gesamte Analysis von Grund auf umgestaltet, alle Aussagen der elementaren Theorie der Kreisfunktionen folgen nun nahezu von selbst; insbesondere die BALTZER-LANDAUsche Beschreibung von π (vgl. 1.5 und 1.6). Im Paragraphen 4 diskutieren wir klassische Formeln für π; wir referieren dort auch über Irrationalität und Transzendenz; der Schlüssel für die Lösung der Frage nach der Quadratur des Zirkels liegt in der fundamentalen Relation $e^{2\pi i} = 1$.

Ein Kapitel über die Zahl π und ihre Einführung mittels der komplexen Exponentialfunktion gehört nicht unbedingt in einen Band über Zahlen. Wir haben es hier u. a. deshalb aufgenommen, um den ARGANDschen Beweis des Fundamentalsatzes der Algebra, der wesentlich die Existenz primitiver n-ter Einheitswurzeln und also letzten Endes die Gleichung (1) heranzieht (vgl. 3.6.4), lückenlos zu machen.

§ 1. Zur Geschichte der Zahl π

Wir stellen wichtige historische Daten zusammen, unsere Quellen sind:

TROPFKE, J.: Geschichte der Elementar-Mathematik, Bd. 4, Ebene Geometrie, 3. Aufl., S. 260 ff., De Gruyter, Berlin 1940

JUSCHKEWITSCH, A. P.: Geschichte der Mathematik im Mittelalter, Teubner-Verlag, Leipzig 1964

RUDIO, F.: ARCHIMEDES, HUYGENS, LAMBERT, LEGENDRE. Vier Abhandlungen über die Kreismessung. Deutsch herausgegeben und mit einer Übersicht über die Geschichte des Problems von der Quadratur des Zirkels, von den ältesten Zeiten bis auf unsere Tage, Teubner Verlag, Leipzig 1892. Nachdruck Dr. Martin Sändig OHG 1971

Nach Fertigstellung dieses Manuskriptes wurde ich noch auf ein Buch von

BECKMANN, P.: A history of π (Pi), The Golem Press, Boulder Colorado, 4. Aufl. 1977

aufmerksam gemacht.

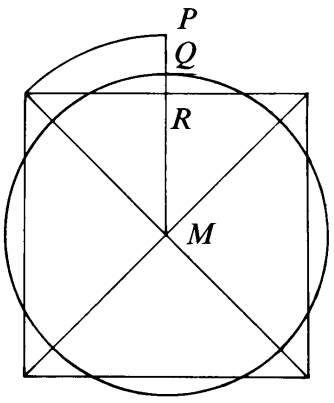
1. Definition mittels Kreismessung. Bei allen Kreisen ist das Verhältnis des Umfangs U zum Durchmesser und das Verhältnis des Flächeninhalts F zum Quadrat des Radius konstant. ARCHIMEDES (287–212 v. Chr.) erkannte, daß es sich in beiden Fällen um dieselbe Konstante handelt. Seit EULER (1737) bezeichnet man diese Konstante mit π; es gilt also, wenn r den Radius bezeichnet:

$$U = 2r\pi, \qquad F = r^2\pi.$$

Der Buchstabe π kommt erstmals bei W. OUGHTRED (1575–1660, englischer Mathematiker, Lehrer von J. WALLIS) vor in „Theorematum in libris Archimedis de Sphaera et Cylindro Declaratio", Oxoniae 1663. Ob EULER die Arbeit von OUGHTRED gekannt hat, scheint schwer feststellbar zu sein. Aber er wird wohl wie OUGHTRED bei π auch an den Anfangsbuchstaben von Peripherie ($\pi\varepsilon\rho\iota\varphi\acute{\varepsilon}\rho\varepsilon\iota\alpha$) gedacht haben. Bis 1735 schrieb EULER noch p statt π.

2. Näherungswerte aus der Praxis. Für die Architekten beim Bau des ehernen Meeres im Vorhof des Tempels von König SALOMON ist π nach dem Buch der Könige gleich 3. Dieser Wert wurde auch vorwiegend von den Babyloniern benutzt. Erstaunlich ist die gute Näherung, die sich im ägyptischen Rechenbuch des AHMES (ca. 1900 v. Chr.) findet: Der Flächeninhalt eines Kreises vom Durchmesser d ist $(d - \frac{d}{9})^2$. Das entspricht $\pi \approx (\frac{16}{9})^2 \approx 3{,}16$. Wie dieser Wert gefunden wurde, ist nicht überliefert.

In den indischen Śulbasūtras (das heißt „Schnurregeln", nämlich Regeln zur Konstruktion von Altären bestimmter Formen mit Hilfe von Schnüren) finden sich zwei Vorschriften: 1) Will man zu einem Kreis das flächengleiche Quadrat finden, so ziehe man vom Durchmesser $\frac{2}{15}$ ab, was zu $\pi \approx (\frac{26}{15})^2 \approx 3{,}0044$ führt. 2) Will man zu einem Quadrat den flächengleichen Kreis finden, so nehme man in der nachstehenden Figur als Radius MQ mit $RQ = \frac{1}{3}RP$, das führt zu $\pi \approx 3{,}088$. Die Śulbasūtras wurden um 500 v. Chr. aufgeschrieben; wie lange ihr Inhalt vorher mündlich überliefert wurde, ist unbekannt.



Albrecht DÜRER (1471–1528, Nürnberg) gibt für die zweite Aufgabe folgende Lösung an*): Teile die Diagonale des Quadrats in 10 Teile und nimm 8 davon als Durchmesser des Kreises. Das besagt $1 \approx (\frac{2}{5}\sqrt{2})^2\pi$, das heißt, $\pi \approx 3\frac{1}{8}$. DÜRER nimmt

*) Underweysung der Messung mit dem Zirckel und Richtscheyt in Linien, Ebnen, und gantzen Corporen, Nuremberg 1525, ^{2}1528; Ende des 2. Buches, Figur 34 (Faksimile ed. A. Jaeggli und Chr. Papesch, Zürich 1966).

also nicht den damals allgemein bekannten Wert $3\frac{1}{7}$, was wohl dadurch zu erklären ist, daß er nicht rechnen, sondern zeichnen möchte, und daß keine rationale Teilungskonstruktion zu $3\frac{1}{7}$ führt (Beweis!).

Nach K. R. POPPER (*The open society and its enemies*, Vol. I. *The spell of* PLATO, 5. revid. Aufl., Routledge and Kegan Paul, London and Henley 1966) war bereits PLATON (427–348/47) eine überraschend gute Näherung für π bekannt: er soll $\sqrt{2} + \sqrt{3} \approx 3{,}14626$ angegeben haben, der Fehler ist weniger als 1,5 Promille.

3. Methodische Approximation. ARCHIMEDES hat erstmals den Wert von π in zwei Grenzen eingeschlossen. Er vergleicht den Kreisumfang mit den Seitenlängen ein- und umbeschriebener regulärer n-Ecke; für $n = 96$ erhält er die Ungleichung $3\frac{10}{71} < \pi < 3\frac{1}{7}$. Die Abschätzung $\pi > 3$ folgt trivial, da das einbeschriebene reguläre 6-Eck den Umfang $6r$ hat. Der Wert $\pi \approx 3\frac{1}{7} \approx 3{,}14$ wird noch heute in der Praxis als ausreichender Näherungswert benutzt.

Mit dem Archimedischen Verfahren wurde es möglich, den Wert von π immer genauer zu bestimmen. Schon APOLLONIOS, der etwa 25 Jahre jünger war als ARCHIMEDES, hat bessere Werte berechnet. Das berichtet EUTOKIOS im Kommentar zu ARCHIMEDES' „Kreismessung" [ARCHIMEDES, Opera, Bd. III, Leipzig, Teubner 1915, Nachdruck 1972, S. 258/9]; leider gibt er keine Zahlen an. PTOLEMAIOS (um 150 n. Chr.) wählte den Mittelwert aus den beiden Werten des ARCHIMEDES, nämlich $\pi \approx 3\frac{17}{120} \approx 3{,}14166\ldots$ (Handbuch der Astronomie, Deutsch von K. MANITIUS, 2. Aufl. Leipzig 1963, S. 384/5).

Seitdem haben die Astronomen aller Völker sich um verbesserte Werte von π bemüht. Die Chinesen kannten solche schon seit dem 1. Jh. n. Chr. So arbeitet der Astronom und Philosoph ZHANG HENG (78–139) mit dem Wert $\sqrt{10} \approx 3{,}162$; der gelehrte Heerführer WANG FAN (gest. 267) kennt den besseren Näherungsbruch $\frac{142}{45} \approx 3{,}155$. LIU HUI berechnete (ca. 263) aus dem 192-Eck: $3{,}14\frac{64}{625} < \pi < 3{,}14\frac{169}{625}$ und später aus dem 3072-Eck einen Näherungswert, der dem Dezimalbruch 3,14159 entspricht. Von ZU CHONG-ZHI (430–501) schließlich stammt die Approximation $\pi \approx \frac{355}{113}$, die bis auf sechs Dezimalen nach dem Komma genau ist; bekanntlich ist $\frac{355}{113}$ ein Näherungsbruch der Kettenbruchentwicklung von π (vgl. 5.6); diesen Bruch fand erneut der Holländer Valentin OTHO gegen Ende des 16. Jh. – Ob die Chinesen etwas von den Erkenntnissen von ARCHIMEDES oder PTOLEMAIOS erfahren haben, ist unbekannt; immerhin wurde chinesische Seide damals bis nach Rom verkauft.

In dem indischen astronomischen Werk Sūryasiddhānta (ca. 400 n. Chr.) wird $\sqrt{10}$ benutzt, ĀRAYBHATA gibt $\frac{62832}{20000}$ (498 n. Chr.) an. Dieser Wert tritt auch bei al-HWĀRIZMĪ (Bagdad, Anfang des 9. Jh. n. Chr.) auf. Den Höhepunkt der Berechnungen der islamischen Astronomen bildet die – allerdings sehr späte – des al-KĀŠĪ (1427). Al-KĀŠĪ war Astronom an der von ULUG BEG errichteten Sternwarte in Samarkand. Er berechnete mittels des $3 \cdot 2^{28}$-Ecks den Umfang des Kreises vom Radius 1, also 2π, sexagesimal zu 6; 16, 59, 28, 1, 34, 51, 46, 14, 50 mit einem Fehler von weniger als $\frac{1}{4}$ Einheit der letzten Stelle. Diesen Wert rechnete er in den Dezimalbruch um: 6,283 185 307 179 586 5. (Das ist eines der ältesten Vorkommen von Dezimalbrüchen.)

Vorschriften zur Kreisberechnung, die auf dem Wert $3\frac{1}{7}$ beruhen, scheinen sich auf dem Weg über die römischen Agrimensoren und BOETIUS (ca. 480–524 n. Chr.)

im Abendland verbreitet zu haben. LEONARDO von PISA (ca. 1170–1240 [?]), der sich auf Reisen in den Orient das dortige mathematische Wissen angeeignet hatte, berechnete aus dem 96-Eck $\pi \approx \frac{864}{275} \approx 3{,}141818\ldots$ (La pratica di geometria. In: Scritti di Leonardo Pisano, Ed. B. Boncompagni. Bd. 2. Rom 1862, S. 90 ff.); LUDOLPH VAN CEULEN (1540–1610, Leiden) gab den Wert auf 35 Stellen hinter dem Komma genau an, nach ihm nennt man π auch oft die LUDOLPHsche Zahl. Die ersten 20 korrekten Dezimalen sind

$$\pi = 3{,}14159\,26535\,89793\,23846\ldots.$$

Das House of representatives (Landtag) des US-Bundesstaates Indiana hat 1897 ein Gesetz „for an act introducing a new mathematical truth" einstimmig verabschiedet, das zwei Werte für π, und zwar 4 bzw. 3,2, vorschlägt. Der Senat von Indiana hat die Verabschiedung dieses Gesetzes „indefinitely" zurückgestellt. „Fortunately for the people of Indiana, the 'indefinitely' still continues!" (vgl. A. E. HALLERBERG: Indiana's Squared Circle, Math. Magazine 50, 136–140 (1977)).

4. Analytische Formeln. Die erste analytische Darstellung für π fand VIETA 1579 in Form des unendlichen Produktes

$$\frac{2}{\pi} = \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\cdot\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}} \cdot \ldots,$$

dies ist wohl das erste unendliche Produkt in der Geschichte der Mathematik überhaupt. 1655 entdeckte J. WALLIS bei Integrationsbetrachtungen sein berühmtes Produkt

$$\frac{\pi}{2} = \frac{2\cdot 2}{1\cdot 3}\cdot\frac{4\cdot 4}{3\cdot 5}\cdot\frac{6\cdot 6}{5\cdot 7}\cdot\ldots\cdot\frac{2n\cdot 2n}{(2n-1)\cdot(2n+1)}\cdot\ldots;$$

es ist bemerkenswert, daß diese ersten Formeln für π nicht unendliche Reihen sind.

Weitere große Fortschritte zum Verständnis der Zahl π wurden erst mittels der Infinitesimalrechnung und der Theorie der unendlichen Reihen erzielt. 1671 gibt James GREGORY die klassische Reihendarstellung

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots,$$

sie wurde 1674 von LEIBNIZ wiedergefunden, ist aber wie das Produkt von WALLIS wegen ihrer langsamen Konvergenz für numerische Rechnungen nicht geeignet. NEWTON gewinnt um 1665 aus der arcsin-Reihe

$$\arcsin z = z + \frac{1}{2}\frac{z^3}{3} + \frac{1}{2}\cdot\frac{3}{4}\frac{z^5}{5} + \cdots + \frac{1\cdot 3\cdot\ldots\cdot(2n-1)}{2\cdot 4\cdot\ldots\cdot 2n}\frac{z^{2n+1}}{2n+1} + \cdots$$

durch Einsetzen von $z := \frac{1}{2}$ die Darstellung

$$\frac{\pi}{6} = \frac{1}{2} + \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{8} + \frac{1}{2}\cdot\frac{3}{4}\cdot\frac{1}{5}\cdot\frac{1}{32} + \frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}\cdot\frac{1}{7}\cdot\frac{1}{128} + \cdots,$$

die ihm mühelos die ersten 14 Dezimalstellen von π liefert.

5. Die Definition von BALTZER. Will man die geometrischen Definitionen von π des Abschnittes 1 analytisch fassen, so benötigt man Integrale. Der Einheitskreis wird durch $x^2 + y^2 = 1$ beschrieben; für Flächeninhalt bzw. Bogenlänge seiner oberen Hälfte gilt mithin

$$\int_{-1}^{1} y\,dx = \int_{-1}^{1} \sqrt{1-x^2}\,dx = \frac{\pi}{2} \text{ bzw. } \int_{-1}^{1} \sqrt{1+y'^2}\,dx = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\,dx = \pi.$$

Man kann diese Gleichungen zur Definition von π erheben; es sei hier bereits erwähnt, daß K. WEIERSTRASS 1841 bei seinem funktionentheoretischen Beweis des heute nach LAURENT benannten Entwicklungssatzes die Zahl π gar durch das uneigentliche Integral

$$\pi := \int_{-\infty}^{\infty} \frac{dx}{1+x^2}$$

einführte (Math. Werke 1, S. 53).

In Vorlesungen und Büchern zur Infinitesimalrechnung benutzt man zur Definition von π üblicherweise keine Integrale, da man die Integralrechnung in der Regel erst nach der Differentialrechnung behandelt und π bzw. $\frac{1}{2}\pi$ schon sehr früh als Nullstellen des Sinus bzw. Cosinus benötigt werden. Man definiert $\frac{1}{2}\pi$ vielmehr als die *kleinste positive* Nullstelle der durch ihre Potenzreihe erklärten Cosinusfunktion, wobei man die Existenz positiver Nullstellen von $\cos x$ mittels des Zwischenwertsatzes beweist. Diesen Zugang zu π findet man bereits im vergangenen Jahrhundert bei Richard BALTZER (1818–1887, seit 1869 o. Prof. in Giessen, Freund KRONECKERS). Im ersten Band seiner „Elemente der Mathematik" liest man (vgl. z. B. 5. Aufl. 1875, S. 195): „Während x den realen Weg von 1 bis 2 zurücklegt, geht $\cos x$ ohne Unterbrechung der Continuität aus dem Positiven ins Negative:

$$\cos 1 = 1 - \frac{1}{2} + \frac{1}{4!}\left(1 - \frac{1}{5\cdot 6}\right) + \cdots > 0,$$

$$\cos 2 = -\frac{1}{3} - \frac{2^6}{6!}\left(1 - \frac{2^2}{7\cdot 8}\right) - \cdots < 0$$

also giebt es zwischen 1 und 2 einen realen Werth x, bei welchem $\cos x$ null ist. Dieser Werth ... wird durch $\frac{1}{2}\pi$ bezeichnet."

6. LANDAU und die zeitgenössische Kritik. Die BALTZERsche Einführung von π ist nicht geometrisch, aber wohl der bequemste Weg, im Reellen schnell zu π zu gelangen. Edmund LANDAU (1877–1938, Schüler von FROBENIUS; 1909 o. Professor der Mathematik in Göttingen als Nachfolger von MINKOWSKI; 1933 aus rassischen Gründen amtsenthoben; Nachruf von K. KNOPP in Jahresber. DMV 54, 55–62 (1951)) hat diesen Zugang in seinen Göttinger Vorlesungen propagiert und 1934 in seiner „Einführung in die Differentialrechnung und Integralrechnung" (Verlag Noordhoff, Groningen) in dem für ihn charakteristischen Telegrammstil dargestellt; auf S. 193 heißt es: *Die Weltkonstante aus* Satz 262 *werde dauernd mit* π *bezeichnet.*

Die Definition von $\frac{1}{2}\pi$ als kleinste positive Nullstelle von $\cos x$ ist heute gang und gäbe. Um so unverständlicher mutet es an, daß u. a. gerade diese *Art* der Einführung von π in Deutschland 1934 eine Polemik auslöste, die mit dem Wort „beschämend" nicht annähernd charakterisiert wird. Ein hoch angesehener Kollege aus Berlin griff damals LANDAU scharf an; es seien hier nur zwei Sätze von ihm zitiert: „Uns Deutsche läßt eine solche Rumpftheorie unbefriedigt" (Sonderausg. Sitz. Ber. Preuss. Akad. Wiss., Phys.-Math. Kl. XX, p. 6); und weitaus deutlicher: „So ist ... die mannhafte Ablehnung, die ein großer Mathematiker, Edmund LANDAU, bei der Göttinger Studentenschaft gefunden hat, letzten Endes darin begründet, daß der undeutsche Stil dieses Mannes in Forschung und Lehre deutschem Empfinden unerträglich ist. Ein Volk, das eingesehen hat, ... wie Volksfremde daran arbeiten, ihm fremde Art aufzuzwingen, muß Lehrer von einem ihm fremden Typus ablehnen." (Persönlichkeitsstruktur und mathematisches Schaffen, Forsch. u. Fortschr., 10. Jahrg. Nr. 18, 1934, S. 236).

Derartige abstruse Ungeheuerlichkeiten hat der britische Mathematiker G. H. HARDY sofort im August 1934 in seiner Note: „The *J*-type and the *S*-type among mathematicians" (Collected Papers 7, 610–611) scharf zurückgewiesen, er urteilt: „There are many of us, many Englishmen and many Germans, who said things during the War which we scarcely meant and are sorry to remember now. Anxiety for one's own position, dread of falling behind the rising torrent of folly, determination at all costs not to be outdone, may be natural if not particularly heroic excuses. Prof. Bieberbach's reputation excludes such explanations of his utterances; and I find myself driven to the more uncharitable conclusion that he really believes them true.".

§ 2. Der Exponentialhomomorphismus exp: $\mathbb{C} \to \mathbb{C}^\times$

Die zuerst von NEWTON (Brief an LEIBNIZ vom 24. Octob. 1676, vgl. Math. Schriften, ed. GERHARDT, Bd. 1, S. 138) für reelle Argumente aufgestellte *Exponentialreihe*

$$\exp z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots + \frac{z^n}{n!} + \cdots = \sum_0^\infty \frac{z^\nu}{\nu!}$$

ist für alle $z \in \mathbb{C}$ *absolut konvergent*; dies beweist man genauso wie im Reellen. Damit ist in ganz $\mathbb{C}$ eine komplexe Funktion $\exp: \mathbb{C} \to \mathbb{C}$ eingeführt, sie heißt die (*komplexe*) *Exponentialfunktion* und ist die natürliche Fortsetzung der reellen Exponentialfunktion ins Komplexe. Diese Funktion spielt seit EULER unter den sogenannten „elementaren transzendenten Funktionen" die dominierende Rolle. Das für die Theorie der Funktion $\exp z$ grundlegende Additionstheorem gewinnen wir mittels des CAUCHYschen Reihenproduktsatzes:

Es seien $\sum_0^\infty a_\mu$, $\sum_0^\infty b_\nu$ *absolut konvergente Reihen. Dann ist auch ihr „Cauchyprodukt"* $\sum_0^\infty p_\lambda$, *wobei* $p_\lambda := \sum_{\mu+\nu=\lambda} a_\mu b_\nu$, *absolut konvergent, es gilt:*

$$\left(\sum_0^\infty a_\mu\right)\left(\sum_0^\infty b_\nu\right) = \sum_0^\infty p_\lambda.$$

Beweise dieses Satzes findet der Leser z. B. im berühmten, 1821 in Paris erschienenen CAUCHYschen Buch „*Cours d'Analyse*" auf S. 237 (Œuvres 3, 2. Ser.)

sowie im klassischen Buch von K. KNOPP „*Theorie und Anwendung der unendlichen Reihen*", Springer-Verlag. Man vergleiche auch Grundwissen Mathematik 3 (Analysis I) und 5 (Funktionentheorie I).

Das Additionstheorem besagt, daß die Abbildung

$$\exp: \mathbb{C} \to \mathbb{C}^\times, \qquad z \mapsto \exp z$$

ein *Homomorphismus* der *additiven Gruppe* $\mathbb{C}$ in die *multiplikative Gruppe* $\mathbb{C}^\times$ ist. Wann immer Mathematiker Gruppenhomomorphismen $\sigma: G \to H$ sehen, fragen sie nach der Bildgruppe $\sigma(G)$ und der Kerngruppe $\operatorname{Kern}\sigma := \{g \in G: \sigma(g) = \text{neutrales Element von } H\}$. Für den Exponentialhomomorphismus werden wir zeigen

$$\exp(\mathbb{C}) = \mathbb{C}^\times, \qquad \operatorname{Kern}(\exp) = 2\pi i\mathbb{Z},$$

wobei π eine positive reelle Zahl ist. Zum Beweis von $\exp(\mathbb{C}) = \mathbb{C}^\times$ benutzen wir ein simples

Konvergenzlemma. *Zu jeder Zahl $w \in \mathbb{C}^\times$ gibt es eine Folge $w_1, w_2, \ldots$ in $\mathbb{C}^\times$ mit $w_n^{2^n} = w$ und $\lim w_n = 1$.*

Wir wollen dies hier sofort beweisen. Wir schreiben $w = |w|c$ mit $c \in S^1$. Es gibt ein $c_1 = a_1 + ib_1 \in S^1$ mit $a_1 \geqslant 0$ und $c_1^2 = c$. Aufgrund der Schlußbemerkung in 3.3.5 finden wir nun sukzessive Zahlen $c_n = a_n + ib_n \in S^1$, so daß $c_n^2 = c_{n-1}$, $a_n \geqslant 0$, $|b_n| \leqslant \frac{1}{\sqrt{2}}|b_{n-1}|$. Dann gilt $c_n^{2^n} = c$ und $|b_n| \leqslant \left(\frac{1}{\sqrt{2}}\right)^{n-1}|b_1| \leqslant \left(\frac{1}{\sqrt{2}}\right)^{n-1}$. Wegen $\sqrt{2} > 1$ folgt $\lim b_n = 0$, also $\lim a_n^2 = \lim(1 - b_n^2) = 1$, also $\lim a_n = 1$ wegen $a_n \geqslant 0$. Damit ist klar: $\lim c_n = \lim(a_n + ib_n) = 1$. Für die Folge $w_n := \sqrt[2^n]{|w|}c_n$ gilt nun $w_n^{2^n} = w$ und $\lim w_n = 1$, da für alle $r > 0$ bekanntlich gilt $\lim \sqrt[n]{r} = 1$. □

Neben dem Konvergenzlemma werden noch zwei Fakten aus den Elementen der Funktionentheorie herangezogen (vgl. hierzu Grundwissen Mathematik 5, Funktionentheorie I.

1) *Konvergente Potenzreihen $f(z) = \sum_0^\infty a_\nu z^\nu$ sind in ihren Konvergenzkreisscheiben holomorph; dort gilt $f'(z) = \sum_1^\infty \nu a_\nu z^{\nu-1}$.*

2) *Ist f holomorph in einer Kreisscheibe und gilt dort $f' \equiv 0$, so ist f konstant.*

1. Additionstheorem. $(\exp w)(\exp z) = \exp(w + z)$. Zum Beweis schreibt man

$$(\exp w)(\exp z) = \sum_{\lambda=0}^{\infty} p_\lambda \quad \text{mit} \quad p_\lambda := \sum_{\mu+\nu=\lambda} \frac{w^\mu}{\mu!}\frac{z^\nu}{\nu!} = \sum_{\nu=0}^{\lambda} \frac{1}{(\lambda-\nu)!\nu!} w^{\lambda-\nu}z^\nu$$

nach dem Reihenmultiplikationssatz von CAUCHY. Nun ist $\frac{1}{(\lambda-\nu)!}\frac{1}{\nu!} = \frac{1}{\lambda!}\binom{\lambda}{\nu}$, daher folgt nach der binomischen Formel $p_\lambda = \frac{1}{\lambda!}\sum_{\nu=0}^{\lambda}\binom{\lambda}{\nu}w^{\lambda-\nu}z^\nu = \frac{1}{\lambda!}(w+z)^\lambda$, das heißt, $(\exp w)(\exp z) = \sum_{\lambda=0}^{\infty}\frac{(w+z)^\lambda}{\lambda!} = \exp(w+z)$. □

Das Additionstheorem ist eine „Potenzregel". Um dies besonders deutlich hervortreten zu lassen, schreibt man gern

$$e^z := \exp z, \qquad \text{wobei} \qquad e := 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots \qquad \text{(EULER 1728)}.$$

Man spricht von der komplexen e-Funktion; benutzt man diese nicht ungefährliche Schreibweise, so liest sich das Additionstheorem suggestiv als

Potenzregel. $e^w e^z = e^{w+z}$ *für alle* $w, z \in \mathbb{C}$.

Setzt man $w := -z$ im Additionstheorem, so folgt wegen $\exp 0 = 1$:

$$(\exp z)^{-1} = \exp(-z) \qquad \text{für alle} \qquad z \in \mathbb{C};$$

speziell ist die Funktion $\exp z$ nullstellenfrei und also eine Abbildung von $\mathbb{C}$ nach $\mathbb{C}^\times$. Das Additionstheorem besagt nun:

Die Abbildung $\exp: \mathbb{C} \to \mathbb{C}^\times$ *ist ein Homomorphismus der additiven Gruppe* $\mathbb{C}$ *in die multiplikative Gruppe* $\mathbb{C}^\times$.

2. Elementare Folgerungen. Das Konjugieren konvergenter Folgen ist mit der Limesbildung verträglich. Daher gilt $\overline{\exp z} = \exp \bar{z}$, womit folgt:

(1) $$|\exp z| = \exp(\operatorname{Re} z) \qquad \textit{für alle} \qquad z \in \mathbb{C}.$$

Beweis. Wegen $z + \bar{z} = 2 \operatorname{Re} z$ gilt aufgrund des Additionstheorems

$$|\exp z| = |\exp \tfrac{1}{2}z|^2 = (\exp \tfrac{1}{2}z)\overline{(\exp \tfrac{1}{2}z)} = (\exp \tfrac{1}{2}z)(\exp \tfrac{1}{2}\bar{z})$$
$$= \exp[\tfrac{1}{2}(z + \bar{z})] = \exp(\operatorname{Re} z). \qquad \square$$

Da $\exp x > 1$ für $x > 0$ aufgrund der Gestalt der Exponentialreihe, so folgt $\exp x = (\exp(-x))^{-1} < 1$ für $x < 0$. In (1) ist daher enthalten:

(2) $$|\exp z| = 1 \Leftrightarrow z \in \mathbb{R}i;$$

speziell ist $y \mapsto \exp(iy)$ eine Abbildung von $\mathbb{R}$ in die Kreislinie S^1. Über das Werteverhalten dieser Funktion zeigen wir

(3) $$\operatorname{Im}(\exp(iy)) > 0 \qquad \textit{für} \qquad 0 < y < \sqrt{6}.$$

Beweis. Da $\exp(iy) = \sum_0^\infty \frac{1}{\nu!}(iy)^\nu$ und da stets $(iy)^{2n} \in \mathbb{R}$, so gilt

$$\operatorname{Im}(\exp(iy)) = y - \frac{1}{3!}y^3 + \cdots + \frac{(-1)^n}{(2n+1)!}y^{2n+1} + \cdots \qquad \text{(Sinusreihe, vgl. 3.1(2))}.$$

Hieraus liest man durch Klammern die Behauptung ab:

$$\operatorname{Im}(\exp(iy)) = y\left(1 - \frac{1}{6}y^2\right) + \frac{1}{5!}y^5\left(1 - \frac{1}{6 \cdot 7}y^2\right) + \cdots > 0 \text{ für } 0 < y < \sqrt{6}. \qquad \square$$

Da $\exp(-iy) = (\exp(iy))^{-1}$, so folgt aus (3) direkt

Lemma. *Die Funktion* $\mathbb{R} \to S^1$, $y \mapsto \exp(iy)$ *hat im offenen Intervall* $(-1, 1)$ *nur im Nullpunkt einen reellen Wert.*

3. Epimorphiesatz. *Der Exponentialhomomorphismus* $\exp: \mathbb{C} \to \mathbb{C}^\times$ *ist ein Epimorphismus, das heißt surjektiv.*

Den Beweis stützen wir auf folgenden

Hilfssatz. *Es gibt eine Umgebung* U *des Punktes* $1 \in \mathbb{C}$ *mit* $U \subset \exp(\mathbb{C})$.

Beweis. Die logarithmische Reihe $\lambda(z) := z - \frac{z^2}{2} + \frac{z^3}{3} - + \cdots$ ist für $|z| < 1$ konvergent, also holomorph mit $\lambda'(z) = 1 - z + z^2 - z^3 + - \cdots = (1+z)^{-1}$ (Aussage 1) der Einleitung), ebenso ist $\exp z$ in $\mathbb{C}$ holomorph mit $(\exp z)' = \exp z$. Also ist auch $f(z) := (1+z)\exp(-\lambda(z))$ im Einheitskreis holomorph, nach der Kettenregel folgt $f'(z) \equiv 0$ für $|z| < 1$. Daher ist f konstant (Aussage 2) der Einleitung), wegen $f(0) = 1$ gilt also $\exp\lambda(z) = 1 + z$, falls $|z| < 1$. Es folgt nun unmittelbar, daß die Kreisscheibe $U := \{z \in \mathbb{C} : |z - 1| < 1\}$ in $\exp(\mathbb{C})$ liegt, denn für jedes $a \in \mathbb{C}$ mit $|a - 1| < 1$ ist $b := \lambda(a - 1)$ wohldefiniert, und es gilt: $\exp b = a$. □

Nunmehr ist der Beweis des Epimorphiesatzes schnell erbracht. Nach dem Konvergenzlemma der Einleitung gibt es zu jedem $w \in \mathbb{C}^\times$ eine Folge $w_n \in \mathbb{C}$ mit $w_n^{2^n} = w$ und $\lim w_n = 1$. Aufgrund des Hilfssatzes existieren also ein Index $m \geqslant 1$ und ein $\hat{z} \in \mathbb{C}$ mit $w_m = \exp\hat{z}$. Für $z := 2^m\hat{z}$ gilt dann aufgrund des Additionstheorems $\exp z = (\exp\hat{z})^{2^m} = w$. Wir sehen $\exp(\mathbb{C}) = \mathbb{C}^\times$. □

Wir skizzieren einen zweiten Beweis des Epimorphiesatzes, der ohne die Folge w_n auskommt. Für jedes $w \in \mathbb{C}^\times$ ist $W := \{wz : z \in U\}$ eine Umgebung von w in $\mathbb{C}^\times$. Falls $w \in \exp(\mathbb{C})$, so folgt $W \subset \exp(\mathbb{C})$ aufgrund des Hilfssatzes und der Gruppeneigenschaft von $\exp(\mathbb{C})$. Damit ist $\exp(\mathbb{C})$ eine offene Untergruppe der *zusammenhängenden* Gruppe $\mathbb{C}^\times$. Nach einem elementaren Satz der allgemeinen Theorie topologischer Gruppen hat aber eine zusammenhängende Gruppe G keine offenen Untergruppen $\neq G$.

Im Beweis des obigen Hilfssatzes ist die Identität $\exp\lambda(z) = 1 + z$ entscheidend. Es scheint nicht möglich, diese Identität *elementar*, z. B. ohne Benutzung von Differentialrechnung, herzuleiten. Im Anhang dieses Paragraphen geben wir einen elementaren Beweis des Hilfssatzes, der ohne komplexe Differentialrechnung auskommt.

4. Der Kern des Exponentialhomomorphismus. Definition von π. Mit Hilfe der Gleichung $\exp(\mathbb{C}) = \mathbb{C}^\times$ bestimmen wir nun mühelos die additive Untergruppe $\mathrm{Kern}(\exp) = \{w \in \mathbb{C} : \exp w = 1\}$ von $\mathbb{C}$.

Satz. *Es gibt eine eindeutig bestimmte, positive reelle Zahl* π, *so daß gilt*:

$$\mathrm{Kern}(\exp) = 2\pi i\mathbb{Z}.$$

Beweis. Wir setzen $K := \mathrm{Kern}(\exp)$. Für jedes $c \in K$ gilt $1 = |\exp c|$, woraus nach 2.(2) folgt: $K \subset \mathbb{R}i$. Nach dem Epimorphiesatz gibt es ein $a \in \mathbb{C}$ mit $\exp a = -1$. Es gilt $a \neq 0$, da $\exp 0 = 1$. Für $c := 2a \neq 0$ folgt $\exp c = (\exp a)^2 = 1$. Wir sehen $K \neq (0)$. Da $K \subset \mathbb{R}i$, so gibt es also Zahlen $s > 0$ mit $si \in K$ (beachte, daß mit c auch stets $-c$ in K liegt). Da nach Lemma 2 keine Zahl $iy \neq 0$ mit $y \in (-1, 1)$ zu K gehört, gibt es wegen der Stetigkeit von $\exp z$ eine *kleinste positive reelle* Zahl π mit $2\pi i \in K$. Dann ist $2\pi i\mathbb{Z} \subset K$ trivial. Hat man umgekehrt ein $r \in \mathbb{R}$ mit $ri \in K$, so gibt es wegen $\pi > 0$ ein $n \in \mathbb{Z}$, so daß gilt: $2n\pi \leqslant r < 2(n+1)\pi$. Da K eine additive Untergruppe von $\mathbb{C}$ ist, folgt $i(r - 2n\pi) \in K$. Da $0 \leqslant r - 2n\pi < 2\pi$, so folgt $r = 2n\pi$ wegen der minimalen Wahl von π. Damit ist $K = 2\pi i\mathbb{Z}$ gezeigt. Die Eindeutigkeit von π ist klar. □

Wir verwenden im folgenden die Aussage des Satzes als Definition von π.

Wegen $e^{2\pi i} = 1$ und $\pi i \notin \mathrm{Ker}(\exp)$ gilt $e^{i\pi} = -1$. In der einen Gleichung

$$0 = 1 + e^{i\pi}$$

sind die fünf Fundamentalzahlen 0, 1, i, e, π „auf gar wundersame Art miteinander verwoben", diese Relation gibt gelegentlich zu metaphysischen Spekulationen Anlaß.

Anhang: Elementarer Beweis von Hilfssatz 3. Wir benutzen Lemma 2 und folgende reelle Hilfsmittel:

1) *Ist I ein kompaktes Intervall in $\mathbb{R}$, und ist $f\colon I \to \mathbb{R}$ stetig, so ist das Bild $f(I)$ ein kompaktes Intervall (Zwischenwertsatz).*

2) *Die Exponentialfunktion ist stetig; es gilt* $\exp(\mathbb{R}) = \{r \in \mathbb{R} : r > 0\}$.

Die Aussage 1) übernehmen wir aus der Analysis ohne Beweis; Aussage 2) verifiziert man wie folgt: mit $q := 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots$ gilt

$$|\exp w - 1| \leqslant |w| \left| 1 + \frac{w}{2!} + \frac{w^2}{3!} + \cdots \right| \leqslant q|w| \quad \text{für alle} \quad w \in \mathbb{C} \quad \text{mit} \quad |w| \leqslant 1.$$

Damit folgt für jedes $c \in \mathbb{C}$ und alle $z \in \mathbb{C}$ mit $|z - c| \leqslant 1$:

$$|\exp z - \exp c| = |\exp c|\,|\exp(z - c) - 1| \leqslant q|\exp c|\,|z - c|;$$

somit gilt $|\exp z - \exp c| \leqslant \varepsilon$, falls $|z - c| \leqslant \delta := \min(1, |q \exp c|^{-1}\varepsilon)$. Aus der Stetigkeit von $\exp x$ folgt $\exp(\mathbb{R}) = \{r \in \mathbb{R} : r > 0\}$ mit Hilfe des Zwischenwertsatzes, wenn man $\exp s > 1 + s$ für $s > 0$ und $\exp(-x) = (\exp x)^{-1}$ beachtet. □

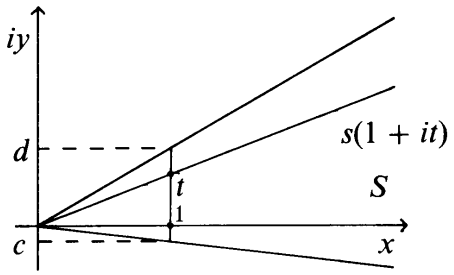
Zum Beweis des Hilfssatzes setzen wir nun $u(y) := \mathrm{Re}\exp(iy)$ und $v(y) := \mathrm{Im}\exp(iy)$. Da

$$(*) \qquad \exp z = e^x u(y) + ie^x v(y) \qquad \text{für} \qquad z = x + iy,$$

und da stetige komplexe Funktionen stetige Realteil-, Imaginärteil- und Betragsfunktionen haben, so ist nach 2) die Funktion $u(y) = \dfrac{\mathrm{Re}\exp z}{|\exp z|}$ in ganz $\mathbb{R}$ und

die Funktion $h(y) := \dfrac{\operatorname{Im} \exp z}{\operatorname{Re} \exp z} = \dfrac{v(y)}{u(y)}$ überall dort, wo $u(y)$ nicht verschwindet, stetig (natürlich gilt $u(y) = \cos y$, $v(y) = \sin y$ und $h(y) = \tan y$, doch ist dies für unsere Überlegungen unwichtig). Wegen $u(0) = 1$ und der Stetigkeit von u gibt es ein $\varepsilon > 0$, so daß u in $I := [-\varepsilon, \varepsilon]$ positiv ist. Dann ist h auf I wohldefiniert und stetig; wir behaupten:

Das Bild $h(I)$ ist ein Intervall $[c, d]$ mit $c < d$; das Bild $\exp(\mathbb{R} \times I)$ *ist der „Sektor"* $S := \{s(1 + it) : s > 0,\ t \in h(I)\}$ (vgl. Figur).



Beweis. Nach 1) ist $h(I)$ ein kompaktes Intervall $[c, d]$ in $\mathbb{R}$. Wegen $v(0) = 0$ gilt $0 = h(0) \in I$. Wäre $c = d$, so würde h und also v in I identisch verschwinden. Wegen $(*)$ würde dies $e^{iy} \in \mathbb{R}$ für alle $y \in I$ bedeuten, was nach Lemma 2 unmöglich ist. Also gilt $c < d$.

Für $z = (x, y) \in \mathbb{R} \times I$ gilt: $\exp z = e^x u(y)[1 + ih(y)]$. Da $u(y) > 0$ für jedes $y \in I$, und da e^x alle positiven reellen Zahlen durchläuft, folgt $\exp(\mathbb{R} \times I) = S$. Da $1 \in \mathbb{C}$ das Bild von $(0, 0) \in \mathbb{R} \times I$ ist, so ist $S \subset \exp(\mathbb{C})$ eine gesuchte Umgebung des Punktes 1.

§ 3. Klassische Charakterisierungen von π

In diesem Paragraphen zeigen wir, daß die in 2.4 definierte Zahl π alle Eigenschaften hat, die man üblicherweise in der reellen Analysis kennenlernt. Die Charakterisierung von π bzw. $\frac{1}{2}\pi$ als kleinste positive Nullstelle der Sinus- bzw. Cosinusfunktion ist mit den Hilfsmitteln des vorangehenden Paragraphen elementar, wenn man den im Reellen unsichtbaren, von EULER entdeckten Zusammenhang

$$e^{iz} = \cos z + i \sin z$$

zwischen der Exponentialfunktion und den trigonometrischen Funktionen heranzieht. Um Umfang bzw. Inhalt eines Kreises mittels π zu bestimmen, übernehmen wir die grundlegenden Definitionen aus der Analysis.

1. Definition von $\cos z$ und $\sin z$. Wir erklären auf ganz $\mathbb{C}$ durch

$$\cos z := \frac{e^{iz} + e^{-iz}}{2}, \qquad \sin z := \frac{e^{iz} - e^{-iz}}{2i}, \qquad z \in \mathbb{C}, \tag{1}$$

die *komplexe Cosinus-* bzw. *Sinusfunktion* und bemerken sofort, daß dies die

bekannten, üblicherweise durch ihre Potenzreihen erklärten trigonometrischen Funktionen

$$(2) \qquad \cos z = \sum_0^\infty \frac{(-1)^\nu}{(2\nu)!} z^{2\nu}, \qquad \sin z = \sum_0^\infty \frac{(-1)^\nu}{(2\nu+1)!} z^{2\nu+1}, \qquad z \in \mathbb{C},$$

sind. Setzt man nämlich $s_n(z) := \sum_0^n z^\nu/\nu!$, so gilt für alle $m \in \mathbb{N}$

$$s_{2m+1}(\pm iz) = \sum_{\mu=0}^{m} (-1)^\mu \frac{z^{2\mu}}{(2\mu)!} \pm i \sum_{\mu=0}^{m} (-1)^\mu \frac{z^{2\mu+1}}{(2\mu+1)!}, \qquad z \in \mathbb{C}.$$

Hieraus folgen durch Addition bzw. Subtraktion wegen $e^z = \lim_{n\to\infty} s_n(z)$ die Gleichungen (2). □

Aus (1) erhält man durch Addition die klassische EULERsche Formel:

$$\exp(iz) = \cos z + i \sin z, \qquad z \in \mathbb{C}.$$

Für reelle Argumente $z = x$ gilt $\cos x, \sin x \in \mathbb{R}$; daher ist

$$\exp(ix) = \cos x + i \sin x, \qquad x \in \mathbb{R},$$

die Zerlegung von $\exp(ix)$ *in Real- und Imaginärteil*; diese Darstellung wurde in Kap. 3 § 6 dauernd benutzt. Es folgt jetzt z. B. wegen $e^{2\pi i} = 1$ und $e^{i\pi} = -1$:

$$\cos 2\pi = 1, \qquad \sin 2\pi = 0; \qquad \cos \pi = -1, \qquad \sin \pi = 0.$$

Aus (1) folgt unmittelbar, daß die Cosinusfunktion *gerade* und die Sinusfunktion *ungerade* ist: $\cos(-z) = \cos z$, $\sin(-z) = -\sin z$.

2. Additionstheoreme. *Für alle* $w, z \in \mathbb{C}$ *gilt*:

$$(1) \qquad \begin{aligned} \cos(w+z) &= \cos w \cos z - \sin w \sin z, \\ \sin(w+z) &= \sin w \cos z + \cos w \sin z. \end{aligned}$$

Beweis. Man geht aus von der Identität

$$\begin{aligned} e^{i(w+z)} = e^{iw} \cdot e^{iz} &= (\cos w + i \sin w)(\cos z + i \sin z) \\ &= \cos w \cos z - \sin w \sin z + i(\sin w \cos z + \cos w \sin z). \end{aligned}$$

Schreibt man $-w$ und $-z$ anstelle von w und z, so erhält man:

$$e^{-i(w+z)} = \cos w \cos z - \sin w \sin z - i(\sin w \cos z + \cos w \sin z).$$

Addition bzw. Subtraktion liefert die Gleichungen (1). □

Aus den Additionstheoremen ergeben sich wie im Reellen unzählige weitere Formeln, z. B. $\cos^2 z + \sin^2 z = 1$ und die „Halbierungsformeln"

$$\sin z = 2 \sin \tfrac{1}{2} z \cos \tfrac{1}{2} z, \qquad \cos^2 \tfrac{1}{2} z = \tfrac{1}{2}(1 + \cos z).$$

Im Abschnitt 4 werden wir wesentlich benutzen:

$$\begin{aligned} \cos w - \cos z &= -2\sin\tfrac{1}{2}(w+z)\sin\tfrac{1}{2}(w-z), \\ \sin w - \sin z &= 2\cos\tfrac{1}{2}(w+z)\sin\tfrac{1}{2}(w-z). \end{aligned} \tag{2}$$

Beweis. Aus den Gleichungen (1) folgt durch Subtraktion

$$\cos(w+z) - \cos(w-z) = -2\sin w \sin z,$$

$$\sin(w+z) - \sin(w-z) = 2\cos w \sin z.$$

Schreibt man $\frac{1}{2}(w+z)$ bzw. $\frac{1}{2}(w-z)$ statt w bzw. z, so folgt (2).

3. Die Zahl π und die Nullstellen von $\cos z$ und $\sin z$. Im Gegensatz zu $\exp z$ haben $\cos z$ und $\sin z$ Nullstellen.

Nullstellensatz. *Genau die reellen Zahlen $n\pi$, $n \in \mathbb{Z}$ sind alle (komplexen) Nullstellen von* $\sin z$.

Genau die reellen Zahlen $\frac{1}{2}\pi + n\pi$, $n \in \mathbb{Z}$, sind alle (komplexen) Nullstellen von $\cos z$.

Beweis. Es gilt, wenn man $e^{i\pi} = -1$ beachtet:

$$2i\sin z = e^{-iz}(e^{2iz} - 1), \qquad 2\cos z = e^{i(\pi - z)}(e^{2i(z - \frac{1}{2}\pi)} - 1).$$

Hieraus liest man aufgrund von Satz 2.4 ab:

$$\begin{aligned} \sin w = 0 &\Leftrightarrow 2iw \in \operatorname{Kern}(\exp) = 2\pi i\mathbb{Z} &&\Leftrightarrow w = n\pi, \qquad n \in \mathbb{Z}, \\ \cos w = 0 &\Leftrightarrow 2i(w - \tfrac{1}{2}\pi) \in 2\pi i\mathbb{Z} &&\Leftrightarrow w = \tfrac{1}{2}\pi + n\pi, \qquad n \in \mathbb{Z}. \end{aligned}$$ □

Wir sehen, daß π bzw. $\frac{1}{2}\pi$ in der Tat die kleinste positive Nullstelle von sin bzw. cos ist. Selbst wenn man aus der reellen Theorie bereits alle reellen Nullstellen von cos und sin kennt, muß man zeigen, daß bei Erweiterung des Argumentbereichs auf komplexe Zahlen keine neuen echt komplexen Nullstellen hinzukommen.

4. Die Zahl π und die Perioden von $\exp z$, $\cos z$ und $\sin z$. Eine Funktion $f\colon \mathbb{C} \to \mathbb{C}$ heißt *periodisch*, wenn es eine komplexe Zahl $\omega \neq 0$ gibt, so daß für alle $z \in \mathbb{C}$ gilt: $f(z+\omega) = f(z)$; die Zahl ω heißt alsdann eine *Periode von f*. Ist f periodisch, so ist die Menge

$$\operatorname{Per}(f) := \{\omega \in \mathbb{C} \colon \omega \text{ ist Periode von } f\} \cup \{0\}$$

aller Perioden von f einschließlich 0 *eine additive Untergruppe von* $\mathbb{C}$.

Periodizitätssatz. *Die Funktionen* exp, cos *und* sin *sind periodisch, es gilt*

$$\operatorname{Per}(\exp) = \operatorname{Kern}(\exp) = 2\pi i\mathbb{Z}, \qquad \operatorname{Per}(\cos) = \operatorname{Per}(\sin) = 2\pi\mathbb{Z}.$$

Beweis. Für eine Zahl $\omega \in \mathbb{C}$ stimmt $\exp(z+\omega) = \exp z \exp\omega$ genau dann für alle $z \in \mathbb{C}$ mit $\exp z$ überein, wenn gilt: $\exp\omega = 1$. Dies beweist $\operatorname{Per}(\exp) = \operatorname{Kern}(\exp) = 2\pi i\mathbb{Z}$ wegen Satz 2.4.

Wegen $\cos(z+\omega)-\cos z=-2\sin(z+\frac{\omega}{2})\sin\frac{\omega}{2}$ gilt $\omega\in\mathrm{Per}(\cos)$ genau dann, wenn $\sin\frac{\omega}{2}=0$, das heißt, wenn $\omega\in 2\pi\mathbb{Z}$. Ebenso folgt die Behauptung für die Sinusfunktion wegen $\sin(z+\omega)-\sin z=2\cos(z+\frac{\omega}{2})\sin\frac{\omega}{2}$. □

Wir erkennen einen wesentlichen Unterschied im Verhalten der Exponentialfunktion im Reellen und Komplexen: Im Reellen nimmt sie wegen $\mathrm{Kern}(\exp)\cap\mathbb{R}=\{0\}$ *jede positive reelle Zahl genau einmal als Wert an*, im Komplexen hingegen besitzt sie die rein imaginäre (reell unsichtbare) „Minimalperiode" $2\pi i$ und nimmt *jeden* Wert $c\neq 0$ – auch negative reelle Werte – *abzählbar unendlich oft an.*

Wir sehen weiter, daß die Zahl 2π auch als die kleinste positive Periode von cos und sin charakterisiert werden kann. Selbst wenn man weiß, daß cos und sin im Reellen die Minimalperiode 2π haben, hat man beim Übergang zum Komplexen noch zu zeigen, daß 2π Periode bleibt und daß zu den reellen Perioden keine neuen echt komplexen Perioden hinzukommen.

5. Die Ungleichung $\sin y>0$ für $0<y<\pi$ und die Gleichung $e^{i\frac{\pi}{2}}=i$. Im Anschluß an die Gleichung $e^{i\pi}=-1$ stellt sich naturgemäß die Frage nach den Werten $\xi:=e^{i\frac{\pi}{2}}$, $\eta:=e^{i\frac{\pi}{4}}$ usf. Wegen $\xi^2=e^{i\pi}=-1$, $\eta^2=\xi$ gibt es die zwei bzw. vier Möglichkeiten $\xi=\pm i$ bzw. $\eta=\pm\frac{1}{2}\sqrt{2}(1\pm i)$. Um die Vorzeichen zu bestimmen, bemerken wir

(1) $$\sin y>0 \qquad \textit{für} \qquad 0<y<\pi.$$

Beweis. Die Sinusfunktion ist in $\mathbb{R}$ stetig und nach 2.2(3) im Intervall $(0,\sqrt{6})$ positiv. Wäre $\sin y$ irgendwo in $(0,\pi)$ negativ, so gäbe es *aufgrund des Zwischenwertsatzes* eine Nullstelle r, $0<r<\pi$ im Widerspruch zum Nullstellensatz 3. □

Aus (1) folgt, daß oben für ξ und η nur die Pluszeichen gelten, das heißt

(2) $$e^{i\frac{\pi}{2}}=i, \qquad e^{i\frac{\pi}{4}}=\tfrac{1}{2}\sqrt{2}(1+i).$$

Die erste Gleichung hier kannte Johann Bernoulli bereits 1702; in der Form

$$\ln i=i\frac{\pi}{2} \qquad \text{oder} \qquad \pi=2\frac{\ln i}{i}$$

spielte sie eine wichtige Rolle in der Kontroverse zwischen Leibniz und Bernoulli über die wahren Werte des natürlichen Logarithmus für -1 und i. □

Es muß deutlich gesagt werden, daß sich ohne Heranziehung des Zwischenwertsatzes in der Formel $e^{i\frac{\pi}{2}}=\pm i$ das Minuszeichen nicht ausschließen läßt. Alle bisherigen Schlüsse sind nämlich auch für die Funktion $\exp(-z)$ richtig, die für $i\frac{\pi}{2}$ den Wert $-i$ hat (Beweis!). Um die wichtige Gleichung $e^{i\frac{\pi}{2}}=i$ zu erhalten, ist also die erneute Bemühung des Zwischenwertsatzes unerläßlich. □

Die Gleichungen (2) besagen, wenn man sie reell schreibt:

$$\cos\frac{\pi}{2} = 0, \qquad \sin\frac{\pi}{2} = 1, \qquad \cos\frac{\pi}{4} = \sin\frac{\pi}{4} = \frac{1}{2}\sqrt{2}.$$

Mit Hilfe des Zwischenwertsatzes und $\cos 0 = 1$ zeigt man wie oben:

(1′) $$\cos y > 0 \qquad \text{für} \qquad -\tfrac{1}{2}\pi < y < \tfrac{1}{2}\pi.$$

6. Der Polarkoordinatenepimorphismus p: $\mathbb{R} \to S^1$. In Kap. 3, § 6 wurden Polarkoordinaten eingeführt. Die dort ohne Beweis herangezogenen Aussagen sind jetzt klar. Aus dem Epimorphiesatz 2.3 und der Tatsache, daß $|\exp z| = 1$ genau für die Zahlen $z \in \mathbb{R}i$ gilt (vgl. 2.2(2)), erhält man $\exp(i\mathbb{R}) = S^1$, wobei S^1 wieder die (multiplikative) Kreisgruppe bezeichnet (vgl. 3.3.4). Damit folgt der in 3.6.1 entscheidend benutzte

Epimorphiesatz. *Die Abbildung $p\colon \mathbb{R} \to S^1$, $\varphi \mapsto e^{i\varphi}$ ist ein Gruppenepimorphismus mit der Gruppe $2\pi\mathbb{Z}$ als Kern, es gilt: $p(\frac{\pi}{2}) = i$.*

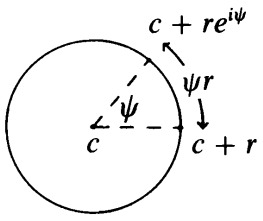
Beweis. Für $\varphi, \psi \in \mathbb{R}$ gilt $p(\varphi + \psi) = \exp(i\varphi + i\psi) = (\exp i\varphi)(\exp i\psi) = p(\varphi)p(\psi)$; daher ist p wegen $p(\mathbb{R}) = \exp(i\mathbb{R}) = S^1$ ein Epimorphismus. Wegen $\operatorname{Ker}(\exp) = 2\pi i\mathbb{Z}$ folgt weiter: $\operatorname{Kern} p = \{t \in \mathbb{R}\colon it \in \operatorname{Kern}(\exp)\} = \{t \in \mathbb{R}\colon t \in 2\pi\mathbb{Z}\}$. Die letzte Behauptung wurde im vorigen Abschnitt bewiesen.

7. Die Zahl π und Umfang und Inhalt eines Kreises. Eine Abbildung $\gamma\colon I \to \mathbb{C}$, $t \mapsto z(t) = x(t) + iy(t)$ eines abgeschlossenen Intervalles $I = [a, b]$ von $\mathbb{R}$ in $\mathbb{C}$ heißt ein *stetig differenzierbarer Weg* in $\mathbb{C}$, wenn die Funktionen $x(t)$ und $y(t)$ in I stetig differenzierbar sind. Für solche Wege γ existiert das Integral

$$L(\gamma) := \int_a^b |z'(t)|\,dt \qquad \text{mit} \qquad z'(t) := x'(t) + iy'(t);$$

wegen $|z'(t)| = \sqrt{x'(t)^2 + y'(t)^2}$ ist dies, wie in der Analysis gezeigt wird, die (euklidische) Länge $L(\gamma)$ des Weges γ.

Ist $c \in \mathbb{C}$ ein Punkt und $r > 0$, so ist der stetig differenzierbare Weg $\gamma_\psi\colon [0, \psi] \to \mathbb{C}$, $\varphi \mapsto z(\varphi) := c + re^{i\varphi}$, wobei $0 < \psi \leqslant 2\pi$, ein *Kreisbogen* mit Mittelpunkt c und Radius r, der von $c + r$ nach $c + re^{i\psi}$ läuft (vgl. Figur). Da



$z'(\varphi) = ire^{i\varphi}$, so gilt $|z'(\varphi)| = r$ und folglich: $L(\gamma_\psi) = \int_0^\psi |z'(\varphi)|\,d\varphi = \psi r$. Die Länge des Kreisbogens ist also ψr. Da $\gamma_{2\pi}$ die volle Kreislinie ist, so folgt speziell:

Der Umfang eines Kreises vom Radius r ist $2\pi r$. □

Ist $f:[a,b]\to\mathbb{R}$ stetig, so existiert das Integral $\int_a^b f(x)\,dx$; es mißt den *Flächeninhalt* unter dem Graphen von f. Da der Halbkreis vom Radius $r>0$ um 0 durch die Funktion $\sqrt{r^2-x^2}$, $x\in[-r,r]$, gegeben wird, so ist

$$I:=2\int_{-r}^{r}\sqrt{r^2-x^2}\,dx$$

der Inhalt des Kreises vom Radius r. Substituiert man $x=r\cos\varphi$, so folgt wegen $\sin^2\varphi=\frac{1}{2}(1-\cos 2\varphi)$:

$$I=2r\int_{-\pi}^{0}\sin\varphi(-r\sin\varphi)d\varphi=2r^2\int_0^{\pi}\sin^2\varphi\,d\varphi=r^2(\varphi-\tfrac{1}{2}\sin 2\varphi)\Big|_0^{\pi}=\pi r^2.$$

Der Inhalt eines Kreises vom Radius r ist $r^2\pi$.

§ 4. Klassische Formeln für π

Aus der schier unüberschaubaren Menge von Formeln für π stellen wir die Formeln von LEIBNIZ, VIETA, WALLIS und EULER heraus, da ihnen eine besondere historische Bedeutung zukommt.

Im Abschnitt 5 geben wir eine Darstellung von π durch ein Integral, die WEIERSTRASS 1841 in seiner Jugendarbeit „Darstellung einer analytischen Function einer complexen Veränderlichen, deren absoluter Betrag zwischen zwei gegebenen Grenzen liegt" (Math. Werke 1, 51–66) geradezu als Definition von π benutzt hat. Abschließend diskutieren wir noch Kettenbruchformeln und das Transzendenzproblem für π.

1. Die LEIBNIZsche Reihe für π. Die Tangensfunktion ist im Intervall $(-\frac{1}{2}\pi,\frac{1}{2}\pi)$ wegen $\tan'(x)=\frac{1}{\cos^2 x}>0$ streng monoton wachsend und nimmt alle reellen Werte an. Es gibt also eine Umkehrfunktion arctan: $\mathbb{R}\to(-\frac{1}{2}\pi,\frac{1}{2}\pi)$, für deren Ableitung bekanntlich gilt

$$\arctan'(x)=\frac{1}{\tan'(\arctan x)}=\cos^2(\arctan x)=\frac{1}{1+x^2};$$

letzteres folgt, wenn man $y:=\arctan x$ setzt und $x^2=\tan^2 y=(1/\cos^2 y)-1$ beachtet. Die für $|t|<1$ kompakt konvergente geometrische Reihe $(1+t^2)^{-1}=\sum_0^{\infty}(-1)^\nu t^{2\nu}$ liefert nun, wenn man legitim Integration und Summation vertauscht, die *Arcustangensreihe*

$$\arctan x=\int_0^x\frac{dt}{1+t^2}=\sum_0^{\infty}(-1)^\nu\int_0^x t^{2\nu}\,dt \tag{1}$$

$$=x-\frac{x^3}{3}+\frac{x^5}{5}-\frac{x^7}{7}+\cdots,\qquad |x|<1.$$

Aufgrund des ABELschen Grenzwertsatzes (vgl. K. KNOPP: Theorie und Anwendung der unendlichen Reihen, Springer-Verlag, 4. Aufl. 1947, S. 179) ist (1)

auch noch richtig für $x := 1$; da $\arctan 1 = \frac{\pi}{4}$, so entsteht die „eben noch" konvergente Reihe

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + - \cdots = \sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{2\nu+1}.$$

Dies ist die LEIBNIZsche Reihe für π, sie wurde von LEIBNIZ durch geometrische Überlegungen gefunden und „liefert sozusagen die Zahl π der rein arithmetischen Behandlung aus. Es ist, als ob der Schleier, der über dieser seltsamen Zahl lag, durch diese Darstellung fortgezogen sei" (so steht es bei KNOPP, a. a. O., S. 220).

Die LEIBNIZsche Reihe, die übrigens auch schon in Indien um 1500 bekannt war, ist für die praktische Berechnung von π völlig ungeeignet: will man π mit einer Genauigkeit von 10^{-k} berechnen, so muß man ungefähr $\frac{1}{4}10^k$ Glieder berücksichtigen; es gilt nämlich für $n \geqslant 1$:

$$\frac{1}{4}\frac{1}{2n+1} < \frac{\pi}{4} - \sum_{\nu=0}^{2n-1} \frac{(-1)^\nu}{2\nu+1} < \frac{1}{4}\frac{1}{2n-1}.$$

2. Das VIETAsche Produkt für π. Aus der „Halbierungsformel" $\sin z = 2\sin\frac{z}{2}\cos\frac{z}{2}$ der Sinusfunktion erhält man durch Induktion

$$\sin z = 2^n \sin\frac{z}{2^n} \prod_{\nu=1}^{n} \cos\frac{z}{2^\nu}, \qquad z \in \mathbb{C}, \qquad n = 1, 2, \ldots.$$

Da $\lim_{n\to\infty} 2^n \sin\frac{z}{2^n} = z$, so entstehen die unendlichen Produkte

$$\sin z = z \prod_{\nu=1}^{\infty} \cos\frac{z}{2^\nu}, \qquad z \in \mathbb{C},$$

$$(*) \qquad \frac{2}{\pi} = \cos\frac{\pi}{4} \cdot \cos\frac{\pi}{8} \cdot \cos\frac{\pi}{16} \cdot \ldots \cdot \cos\frac{\pi}{2^{\nu+1}} \cdot \ldots \qquad (\text{für } z := \tfrac{1}{2}\pi).$$

Dies ist fast schon die VIETAsche Formel: da $\cos^2\frac{z}{2} = \frac{1}{2}(1+\cos z)$ nach 3.2, und da $\cos x \geqslant 0$ für $x \in [0, \frac{1}{2}\pi]$ nach 3.5(2′), so gilt $\cos\frac{x}{2} = \sqrt{\frac{1}{2} + \frac{1}{2}\cos x}$ für solche x; speziell folgt:

$$\cos\frac{\pi}{4} = \sqrt{\frac{1}{2}}, \quad \cos\frac{\pi}{8} = \sqrt{\frac{1}{2} + \frac{1}{2}\cdot\sqrt{\frac{1}{2}}}, \quad \cos\frac{\pi}{16} = \sqrt{\frac{1}{2} + \frac{1}{2}\cdot\sqrt{\frac{1}{2} + \frac{1}{2}\cdot\sqrt{\frac{1}{2}}}}, \ldots,$$

womit (*) zur VIETAschen Formel wird:

$$\frac{2}{\pi} = \sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\cdot\sqrt{\frac{1}{2}}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\cdot\sqrt{\frac{1}{2} + \frac{1}{2}\cdot\sqrt{\frac{1}{2}}}} \cdots.$$

Die „VIETAsche Folge" $v_n := \prod_{\nu=1}^{n} \cos\frac{\pi}{2^{\nu+1}} = (2^n \sin\frac{\pi}{2^{n+1}})^{-1}$ konvergiert rasch:

$$(1) \qquad 0 < v_n - \frac{2}{\pi} < \frac{1}{48}\sqrt{2}\,\pi^2 \frac{1}{4^n} < \frac{3}{10}\frac{1}{4^n}.$$

Beweis. Da v_n monoton fällt, gilt $v_n > 2\pi^{-1}$. Da $\sin x > x - \frac{x^3}{3!}$ für $0 < x < \sqrt{42}$ (Abschätzung der reellen Taylorreihe!), so folgt $2\pi^{-1}v_n^{-1} > 1 - \frac{\pi^2}{24}\frac{1}{4^n}$ und hieraus (1) durch Multiplikation mit v_n wegen $v_n \leqslant v_1 = \frac{1}{2}\sqrt{2}$, $\pi^2 < 10$, $\sqrt{2} < 1{,}44$. □

Folgende Zahlenbeispiele verdeutlichen die gute Konvergenz:

n	v_n	$2v_n^{-1}$
5	0,6368755077217...	3,140331156954...
15	0,6366197726114...	3,141592652386...
21	0,6366197723676...	3,141592653589...;

der letzte Wert ist bereits bis zur 12. Stelle hinter dem Komma korrekt.

Die Abschätzung (1) läßt sich im übrigen leicht verbessern zur Gleichung

$$\lim 4^n\left(v_n - \frac{2}{\pi}\right) = \frac{\pi}{12}.$$

3. Das EULERsche Sinusprodukt und das WALLISsche Produkt für π. Der französische Mathematiker J. HADAMARD (1865–1963) soll gesagt haben: „*Le plus court chemin entre deux énoncés réels passe par le complexe.*" Als Beispiel für dieses „Prinzip des kürzesten Weges durchs Komplexe" leiten wir hier die EULERsche Produktformel für die Sinusfunktion und damit die WALLISsche Formel für π her.

Aus der MOIVREschen Formel $(\cos t + i\sin t)^k = \cos kt + i\sin kt$, $t \in \mathbb{R}$, entsteht, wenn man links Real- und Imaginärteil trennt:

$$\sin kt = \sin t\left[k\cos^{k-1}t - \binom{k}{3}\cos^{k-3}t\sin^2 t + \cdots\right], \qquad k \in \mathbb{N}.$$

Hieraus entnimmt man, da $\cos^{2k}t = (1 - 2\sin^2 t)^k$:

Die Funktion $\sin kt$ *ist für ungerades* $k = 2n + 1$ *ein rationales Polynom* $p(\sin t)$ *in* $\sin t$ *vom Grad* k.

Von jetzt an verläuft alles in $\mathbb{R}$: Da $p(\sin t) = \sin kt$ die k verschiedenen reellen Nullstellen $\sin\frac{\nu\pi}{k}$, $\nu = 0, \pm 1, \ldots, \pm n$ hat, so folgt:

$$\sin kt = C\prod_{\nu=-n}^{n}\left(\sin t - \sin\frac{\nu\pi}{k}\right),$$

wobei die Konstante C, wenn man durch t dividiert und den Grenzübergang $\lim_{t\to 0}$ durchführt, durch $k = C\prod'^{n}_{\nu=-n}(-\sin\frac{\nu\pi}{k})$ bestimmt ist (der Strich am Produktzeichen bedeutet, daß der Index $\nu = 0$ übergangen wird). Schreibt man nun x statt kt, so erhält man

$$\sin x = k\sin\frac{x}{k}\prod_{\nu=-n}^{n}{}'\left(1 - \frac{\sin\frac{x}{k}}{\sin\frac{\nu\pi}{k}}\right) = k\sin\frac{x}{k}\prod_{\nu=1}^{n}\left(1 - \frac{\sin^2\frac{x}{k}}{\sin^2\frac{\nu\pi}{k}}\right),$$

wobei $n = \frac{1}{2}(k-1)$. Da

$$\lim_{k\to\infty} k\sin\frac{x}{k} = x, \qquad \lim_{k\to\infty}\frac{\sin\frac{x}{k}}{\sin\frac{\nu\pi}{k}} = \frac{x}{\nu\pi},$$

so ergibt sich, wenn man den Grenzübergang naiv ausführt, die

EULERsche Produktformel: $\sin x = x \prod_1^\infty \left(1 - \frac{x^2}{\nu^2 \pi^2}\right)$;

natürlich muß und kann dieser letzte Konvergenzschritt streng begründet werden. □

Setzt man nun $x := \frac{\pi}{2}$, so erhält man nach einfachem Umformen die

WALLISsche Formel: $$\frac{\pi}{2} = \lim_{n\to\infty} \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdot\ldots\cdot\frac{2n}{2n-1}\cdot\frac{2n}{2n+1}\cdot\ldots.$$

Hieraus folgt z. B.

$$\sqrt{\pi} = \lim_{n\to\infty} \frac{2}{1}\cdot\frac{4}{3}\cdot\ldots\cdot\frac{2n}{2n-1}\cdot\frac{1}{\sqrt{n}};$$

dies läßt sich auch als *asymptotische* Gleichung $\binom{2n}{n} \cong \frac{2^{2n}}{\sqrt{n\pi}}$ für den Binomialkoeffizienten $\binom{2n}{n}$ schreiben.

Die monoton wachsende „WALLISsche Folge" $w_n := \frac{2^2 4^2 \cdot\ldots\cdot (2n)^2}{3^2 5^2 \cdot\ldots\cdot (2n-1)^2}\cdot\frac{1}{2n+1}$ konvergiert sehr schlecht; eine elementare Rechnung zeigt:

$$\frac{1}{3}\frac{1}{n+1} < \frac{\pi}{2} - w_n < \frac{1}{2}\frac{1}{n}, \qquad \lim_{n\to\infty} n\left(\frac{\pi}{2} - w_n\right) = \frac{\pi}{8}.$$

Mit etwas mehr Aufwand läßt sich zeigen, daß gilt

$$\frac{\pi}{2} - w_n = \frac{1}{4} w_n \left(\frac{1}{n} - \frac{3}{4}\frac{1}{n^2} + \frac{M_n}{n^3}\right),$$

wobei M_n eine *beschränkte* Folge ist. Für die modifizierte WALLISfolge $\hat{w}_n := w_n(1 + \frac{1}{4n})$ gilt daher

$$\lim_{n\to\infty} n^2\left(\hat{w}_n - \frac{\pi}{2}\right) = \frac{3}{32}\pi.$$

Die Konvergenz von w_n bzw. $\hat{w}_n$ gegen $\frac{\pi}{2}$ erfolgt also mit Fehlergliedern $\simeq \frac{1}{n}$ bzw. $\simeq \frac{1}{n^2}$. Die folgende Tabelle verdeutlicht die schlechte Konvergenz:

n	$2w_n$	$2\hat{w}_n$
10	3,067703807...	3,144396403...
10^2	3,133787491...	3,141621960...
10^3	3,140807746...	3,141592948...
10^4	3,141514119...	3,141592658...
10^5	3,141584800...	3,141592655...;

die letzten Werte sind also nur bis zur 4. bzw. 8. Stelle hinter dem Komma korrekt.

4. Die EULERschen Reihen für $\pi^2, \pi^4, \ldots$. EULER gewann 1734 in seiner Arbeit „De Summis Serierum Reciprocarum" (Opera Omnia Ser. 1, XIV, 73–86) aus seiner Produktformel für den Sinus die berühmten Gleichungen

$$(*) \qquad \frac{\pi^2}{6} = \sum_1 \frac{1}{\nu^2}, \qquad \frac{\pi^4}{90} = \sum_1 \frac{1}{\nu^4}, \qquad \frac{\pi^6}{945} = \sum_1 \frac{1}{\nu^6}, \qquad \frac{\pi^8}{9450} = \sum_1 \frac{1}{\nu^8}, \ldots;$$

Jakob und Johann BERNOULLI hatten sich lange vergeblich bemüht, den Wert der Summe $1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots$ zu finden. EULER erhält die Formeln (*) aus der Identität

$$(1) \qquad 1 - \frac{1}{3!}\pi^2 x^2 + \frac{1}{5!}\pi^4 x^4 - \cdots = \frac{\sin \pi x}{\pi x} = \prod_1^\infty \left(1 - \frac{x^2}{\nu^2}\right),$$

die wegen 3.1(2) und seiner Produktformel für alle $x \in \mathbb{R}$, $x \neq 0$, gilt, durch Koeffizientenvergleich, indem er das Produkt rechts ausmultipliziert. In seiner „Einleitung in die Analysis des Unendlichen" beschreibt er das Verfahren wie folgt (Kapitel 10, § 165):

„Wenn $1 + Az + Bz^2 + Cz^3 + Dz^4 + \cdots = (1+\alpha z)(1+\beta z)(1+\gamma z)(1+\delta z)\ldots$ ist, so müssen *diese Factoren*, mag deren Anzahl eine *endliche* oder *unendliche* sein, eben *jenen Ausdruck* $1 + Az + Bz^2 + Cz^3 + Dz^4 + \cdots$ *wieder hervorbringen*, wenn man sie wirklich mit einander multipliziert." Dies gibt ihm

$$A = \alpha + \beta + \gamma + \delta + \cdots, \qquad B = \alpha\beta + \alpha\gamma + \alpha\delta + \beta\gamma + \beta\delta + \gamma\delta + \cdots, \quad \text{usf.}$$

Anwendung auf (1) liefert sofort $\frac{1}{3!}\pi^2 = \sum_1 \frac{1}{\nu^2}$ und hiermit weiter

$$\frac{1}{5!}\pi^4 = \frac{1}{2}\sum_{\mu \neq \nu}^\infty \frac{1}{\mu^2}\cdot\frac{1}{\nu^2} = \frac{1}{2}\sum_{\nu=1}^\infty \frac{1}{\nu^2}\left(\sum_{\mu=1}^\infty \frac{1}{\mu^2} - \frac{1}{\nu^2}\right) = \frac{1}{2}\sum_{\nu=1}^\infty \frac{1}{\nu^2}\left(\frac{\pi^2}{6} - \frac{1}{\nu^2}\right),$$

was zur zweiten Formel aus (*) führt. EULER zeigt mit seiner Methode, daß jede Summe $\sum_1^\infty \frac{1}{\nu^{2k}}$ ein rationales Vielfaches von π^{2k} ist; genauer gilt

$$\sum_1^\infty \frac{1}{\nu^{2k}} = (-1)^{k-1}\frac{(2\pi)^{2k}}{2(2k)!}B_{2k}, \qquad k = 1, 2, \ldots,$$

wo $B_2, B_4, B_6, \ldots$ die BERNOULLIschen Zahlen sind. Beweise für diese allgemeine Formel findet der Leser in Grundwissen Mathematik 3 (Analysis I) und 5 (Funktionentheorie I).

Die Konvergenzgüte der Reihen $\sum_1 \frac{1}{\nu^k}$ ist schlecht: ungefähr 10 Millionen Glieder werden benötigt, um $\frac{1}{6}\pi^2$ mittels der ersten EULERschen Reihe bis auf 7 Stellen hinter dem Komma genau anzugeben.

5. Die WEIERSTRASSsche Definition von π. Die Integralformel

$$(1) \qquad \int_{S^1} \frac{dz}{z} = 2\pi i$$

ist fundamental für die Funktionentheorie; man erhält sie sofort, wenn man S^1 durch $z(\varphi) := e^{i\varphi}$, $0 \leqslant \varphi \leqslant 2\pi$, beschreibt und $z'(\varphi) = iz(\varphi)$ beachtet:

$$\int_{S^1} \frac{dz}{z} = \int_0^{2\pi} \frac{z'(\varphi)}{z(\varphi)}\,d\varphi = \int_0^{2\pi} i\,d\varphi = 2\pi i.$$

WEIERSTRASS hat 1841 in seinem Beweis des LAURENTschen Entwicklungssatzes das Integral (1) wie folgt ausgerechnet (Math. Werke 1, 52–53): er beschreibt S^1 durch

$$z(\lambda) := \frac{1 + i\lambda}{1 - i\lambda}, \qquad -\infty < \lambda < \infty;$$

dies ist die in 3.5.4 betrachtete rationale Parametrisierung von S^1. Da $z'(\lambda) = \frac{2i}{(1 - i\lambda)^2}$, so gilt $\frac{z'(\lambda)}{z(\lambda)} = \frac{2i}{1 + \lambda^2}$ und folglich, wenn man noch $\int_1^\infty \frac{d\lambda}{1 + \lambda^2} = \int_0^1 \frac{d\tau}{1 + \tau^2}$ beachtet (man substituiere $\lambda := \tau^{-1}$):

$$\int_{S^1} \frac{dz}{z} = \int_{-\infty}^{\infty} \frac{z'(\lambda)}{z(\lambda)} d\lambda = 2i \int_{-\infty}^{\infty} \frac{d\lambda}{1 + \lambda^2} = 4i \int_0^\infty \frac{d\lambda}{1 + \lambda^2} = 8i \int_0^1 \frac{d\lambda}{1 + \lambda^2}.$$

Wenn Formel (1) zur Verfügung steht, so weiß man also:

$$\pi = \int_{-\infty}^{\infty} \frac{d\lambda}{1 + \lambda^2} = 4 \int_0^1 \frac{d\lambda}{1 + \lambda^2}. \tag{2}$$

Diese Identität wird bei WEIERSTRASS als mögliche Definition für π genannt.

Allgemein gilt $\int_0^x \frac{d\lambda}{1+\lambda^2} = \arctan x$, so daß (2) nichts anderes besagt als $\arctan 1 = \frac{\pi}{4}$, das heißt, $\tan \frac{\pi}{4} = 1$, was klar ist, wenn man $\sin \frac{\pi}{4} = \cos \frac{\pi}{4}$ $(= \frac{1}{2}\sqrt{2})$ weiß.

6. Irrationalität von π und Kettenbruchentwicklung. Bereits ARISTOTELES hat behauptet, daß Umfang und Durchmesser eines Kreises inkommensurabel sind. Den ersten Beweis der Irrationalität von π gab 1766 Johann Heinrich LAMBERT (1728–1777) in seiner in sehr origineller Sprache abgefaßten Arbeit „Vorläufige Kenntnisse für die, so die Quadratur und Rectificątion des Circuls suchen" (Werke 1, 194–212) mit Hilfe der Theorie der Kettenbrüche. Er fand (vgl. hierzu den Band „Zahlentheorie" dieser Lehrbuchreihe) für die Tangensfunktion den unendlichen Kettenbruch

$$\tan z = \cfrac{z}{1 - \cfrac{z^2}{3 - \cfrac{z^2}{5 - \cfrac{z^2}{7 - \cdots}}}}$$

$$\ddots$$

und folgerte hieraus die Irrationalität von $\tan z$ für alle reellen rationalen Argumente $z \neq 0$, insbesondere erhielt er $\pi \notin \mathbb{Q}$ wegen $\tan \frac{1}{4}\pi = 1$. Dem LAMBERTschen Beweis fehlt allerdings zur völligen Strenge ein Hilfssatz über die Irrationalität gewisser (besonders gut konvergierender) unendlicher Kettenbrüche. Diesen Hilfssatz bewies 1806 Adrien-Marie LEGENDRE (1752–1833) in der 6. Auflage seiner „Éléments de Géometrie, Note IV"; dort zeigt LEGENDRE auch:

π^2 ist irrational.

Der LAMBERTsche Kettenbruch von $\sqrt{q}\tan\sqrt{q}$ ist nämlich für jedes $q\in\mathbb{Q}$, $q\neq 0$, nach LEGENDRES Hilfssatz irrational, daher ist $\pi=\sqrt{q}$, $q\in\mathbb{Q}$, wegen $\tan\pi=0$ unmöglich. □

Die Abhandlungen von LAMBERT und LEGENDRE sind gut zugänglich im RUDIOschen Artikel (vgl. Einleitung zum Paragraphen 1 dieses Kapitels). – Der heute wohl einfachste Irrationalitätsbeweis für π^2 verläuft wie folgt: man führt die Polynome $p_n(x):=\frac{1}{n!}x^n(1-x)^n$, $n\geqslant 1$, ein und bemerkt zunächst:

1) $0<p_n(x)<\frac{1}{n!}$ für $0<x<1$; $p_n^{(\nu)}(0)$, $p_n^{(\nu)}(1)\in\mathbb{Z}$ für alle ν.
2) Für $P_n(x):=b^n\{\pi^{2n}p_n(x)-\pi^{2n-2}p_n''(x)+\pi^{2n-4}p_n^{(4)}(x)-\cdots+(-1)^n p_n^{(2n)}(x)\}$ gilt

$$\frac{d}{dx}(P_n'(x)\sin\pi x-\pi P_n(x)\cos\pi x)=b^n\pi^{2n+2}p_n(x)\sin\pi x,\qquad b\in\mathbb{R}.$$

Die Ungleichung in 1) ist trivial, die Ganzzahligkeitsaussage in 1) folgt durch Induktion nach n unter Beachtung der Gleichung $p_n'(x)=(1-2x)p_{n-1}(x)$; die Behauptung 2) folgt, wenn man zunächst bemerkt, daß die Ableitung links gerade $(P_n''(x)+\pi^2P_n(x))\sin\pi x$ ist.

Wäre nun π^2 rational, etwa $\pi^2=a/b$ mit natürlichen Zahlen $a, b\geqslant 1$, so wären die mit diesem b gebildeten Werte $P_n(0)$ und $P_n(1)$ wegen 1) ganze Zahlen; daher würde nach 2) wegen $b^n\pi^{2n+2}=a^n\pi^2$ folgen:

$$\pi a^n\int_0^1 p_n(x)\sin\pi x\,dx=[\pi^{-1}P_n'(x)\sin\pi x-P_n(x)\cos\pi x]_0^1=P_n(0)+P_n(1)\in\mathbb{Z}.$$

Wegen 1) gilt indessen, da $0<\sin\pi x<1$ für $0<x<1$:

$$0<\pi a^n\int_0^1 p_n(x)\sin\pi x\,dx<\pi\frac{a^n}{n!}<1 \text{ für große } n,$$

denn $\lim_{n\to\infty}\frac{a^n}{n!}=0$ für jedes $a\in\mathbb{R}$ wegen der Konvergenz der Exponentialreihe. Damit hat man für große n

$$0<P_n(0)+P_n(1)<1 \qquad \text{im Widerspruch zu} \qquad P_n(0)+P_n(1)\in\mathbb{Z}.$$ □

Dieser Beweis beruht auf einer Idee von I. NIVEN: *A Simple Proof That π Is Irrational*, Bull. Amer. Math. Soc. 53, 509 (1947), die Ausdehnung auf π^2 stammt von Y. IWAMOTO: *A Proof That π^2 Is Irrational*, Journ. Osaka Inst. Sci. Tech. 1, 147–148 (1949). Der Leser vergleiche auch das Buch von G. H. HARDY und E. M. WRIGHT: *An Introduction to the Theory of Numbers*, 3. Aufl., Oxford, At the Clarendon Press 1954, insb. S. 47. □

Für π gibt es u. a. die beiden Kettenbrüche

$$\frac{\pi}{4}=\cfrac{1}{1+\cfrac{1^2}{2+\cfrac{3^2}{2+\cfrac{5^2}{2+\cfrac{}{2+7^2}}}}}, \qquad \pi=3+\cfrac{1}{7+\cfrac{1}{15+\cfrac{1}{1+\cfrac{1}{292+1}}}}.$$

Die Entwicklung links fand 1656 Lord W. BROUNCKER (1620–1684, erster Präsident der Royal Society) durch Umwandlung des WALLISschen Produktes, EULER benutzt in seiner Introductio, § 369, die LEIBNIZsche Reihe; die Entwicklung rechts ist der sogenannte *regelmäßige* Kettenbruch für π. Jede positive reelle Zahl besitzt eine eindeutige regelmäßige Kettenbruchdarstellung: in ihr treten nur natürliche Zahlen auf, und alle „Zähler in den Nennern" sind stets 1. Ein Bildungsgesetz für den regelmäßigen Kettenbruch von π ist unbekannt, die angegebenen und weitere Werte erhält man algorithmisch aus der Dezimalbruchentwicklung von π.

Der BROUNCKERsche Kettenbruch konvergiert sehr schlecht. Regelmäßige Kettenbrüche konvergieren sehr gut; für π hat man z. B. die Näherungsbrüche $3, \frac{22}{7}, \frac{333}{106}, \frac{355}{113}, \frac{103933}{33102}$; die in 1.3 angegebene Approximation von π durch ZU CHONG-ZHI ist also die vierte Näherung. Wegen weiterer Einzelheiten zum Zusammenhang zwischen π und Kettenbrüchen verweisen wir auf die zwei Bände von O. PERRON: Die Lehre von den Kettenbrüchen, Teubner Verlag, Stuttgart, 3. Aufl. 1954/1957.

7. Transzendenz von π. Das Problem, einen Kreis mittels Zirkel und Lineal in ein flächengleiches Quadrat zu verwandeln, hat bereits die Griechen beschäftigt (Quadratur des Zirkels). In der Algebra zeigt man, daß eine reelle Zahl genau dann mit Zirkel und Lineal konstruierbar ist, wenn sie in einer endlichen Körpererweiterung von $\mathbb{Q}$ liegt, die durch sukzessive Adjunktion von Quadratwurzeln entsteht; insbesondere sind also höchstens solche Zahlen mit Zirkel und Lineal konstruierbar, die *algebraisch* (über $\mathbb{Q}$) sind, das heißt, die ein Polynom $p \in \mathbb{Q}[x] \setminus \{0\}$ annullieren.

Das Problem der Kreisquadratur ist äquivalent mit der Frage, ob π mit Zirkel und Lineal konstruierbar ist. Nach den vorangehenden Bemerkungen müßte π dazu notwendig eine algebraische Zahl sein. Schon EULER, LAMBERT und LEGENDRE glaubten, daß dies nicht zutrifft, so sagt LEGENDRE am Schluß seiner Abhandlung über die Irrationalität von π^2 ganz deutlich (vgl. RUDIO, S. 59): „Es ist wahrscheinlich, daß die Zahl π nicht einmal unter den algebraischen Irrationalitäten enthalten ist, das heißt, daß sie nicht Wurzel sein kann einer algebraischen Gleichung mit einer endlichen Anzahl von Gliedern, deren Koeffizienten rational sind. Aber es scheint sehr schwer zu sein, diesen Satz strenge zu beweisen."

Nicht algebraische Zahlen heißen *transzendent* (omnem rationem transcendunt). LEGENDRE vermutet also 1806, daß π transzendent ist. Dies war außerordentlich kühn, denn zu jener Zeit wußte man nicht, ob es überhaupt transzendente Zahlen gibt (während irrationale Zahlen, z. B. $\sqrt{2}$, seit den Griechen bekannt sind). Erst 1844 zeigte Joseph LIOUVILLE (1809–1882), daß alle Zahlen, die „sehr gut" durch rationale Zahlen approximierbar sind, wie z. B.

$$10^{-1!} + 10^{-2!} + 10^{-3!} + \cdots = 0{,}11000\,10000\cdots,$$

transzendent sind; 1874 gab Georg CANTOR (1845–1918) den sensationellen Existenzbeweis mittels seines Abzählungsarguments, daß es überabzählbar viele transzendente und nur abzählbar viele algebraische Zahlen gibt (vgl. hierzu etwa O. PERRON: Irrationalzahlen, de Gruyter, Berlin 1960, S. 174–181).

Den großen Durchbruch in der Theorie der transzendenten Zahlen erzielte 1873 der französische Mathematiker Charles HERMITE (1822–1901). Er entwickelte Methoden, mit denen er zeigen konnte: *Die Zahl e ist transzendent.* Durch

Verfeinerung der HERMITEschen Schlußweisen bewies 1882 der deutsche Mathematiker Carl Louis Ferdinand von LINDEMANN (1852–1939; Lehrer von HILBERT und HURWITZ in Königsberg, ab 1893 in München) in einer kurzen Arbeit „Über die Zahl π", Math. Ann. 20, 213–225, sein berühmtes Theorem:

π ist transzendent.

Hierdurch wird die Jahrtausende alte Frage nach der Quadratur des Zirkels negativ beantwortet. Dessenungeachtet beschäftigen sich nach wie vor auch heute noch Laienmathematiker mit diesem Problem; sie finden oft gute Näherungsverfahren und sind meistens schwer davon zu überzeugen, daß ihre „Lösung" der Transzendenz von π nicht widerspricht. □

LINDEMANN scheint damals selbst überrascht gewesen zu sein, ein Jahrtausendproblem lösen zu können; so liest man in der Einleitung seiner Arbeit (S. 213): „Man wird sonach die Unmöglichkeit der Quadratur des Kreises darthun, wenn man nachweist, dass *die Zahl π überhaupt nicht Wurzel einer algebraischen Gleichung irgend welchen Grades mit rationalen Coefficienten sein kann.* Den dafür nöthigen Beweis zu erbringen, ist im Folgenden versucht worden." Die Sätze von HERMITE und LINDEMANN sind enthalten im allgemeinen

Satz von LINDEMANN und WEIERSTRASS (*vgl.* K. WEIERSTRASS: *Zu Lindemann's Abhandlung*: „*Über die Ludolph'sche Zahl*", Math. Werke 2, 341–462, insb. S. 360/61). *Sind $c_1, \ldots, c_n \in \mathbb{C}$ paarweise verschiedene algebraische Zahlen, so gibt es keine Gleichung $a_1 e^{c_1} + \cdots + a_n e^{c_n} = 0$ mit nicht sämtlich verschwindenden algebraischen Zahlen $a_1, \ldots, a_n$.*

Nimmt man hier $n := 2$, $a_1 := -1$, $a_2 := a$, $c_1 := c$, $c_2 := 0$, so folgt:

Für jede algebraische Zahl $c \in \mathbb{C}^\times$ ist $a := e^c$ transzendent.

Für $c := 1$ gibt dies die Transzendenz von e. Da $1 = e^{2\pi i}$, so folgt auch die Transzendenz von π. □

Man weiß inzwischen auch, daß $e^\pi = i^{-2i}$ transzendent ist (A. GELFOND 1929). Über die Zahl π^e weiß man nichts, insgesamt sind die Kenntnisse über transzendente Zahlen immer noch sehr begrenzt: Da e transzendent ist, können nicht beide Zahlen $e\pi$ und $e + \pi$ algebraisch sein; es ist jedoch unbekannt, ob $e\pi$ oder $e + \pi$ irrational ist.

Als Literatur zur Theorie der transzendenten Zahlen sei angegeben:

SCHNEIDER, Th.: Einführung in die transzendenten Zahlen, Grundl. Math. Wiss. Springer-Verlag, Heidelberg 1957

SIEGEL, C. L.: Transzendente Zahlen, BI Hochschultaschenbuch 137*, Mannheim 1967

Zur Geschichte der transzendenten Zahlen vgl. man auch:

Abrégé d'histoire des mathématiques I, sous la direction de Jean DIEUDONNÉ, Hermann, Paris 1978, insb. S. 283 ff.

Teil B

Reelle Divisionsalgebren

Einleitung

M. Koecher, R. Remmert

> Erst durch die Behandlung der gewöhnlichen imaginären Zahlen ... in Gemeinschaft mit den höheren complexen Zahlen kann ihre wahre Bedeutung in das volle Licht gesetzt werden (HANKEL 1867).

1. GAUSS war 1831 davon überzeugt, daß es außer dem System der komplexen Zahlen keine „hyperkomplexen" Zahlensysteme gibt, für welche die grundlegenden Eigenschaften der komplexen Zahlen erhalten bleiben; er äußerte sich allerdings recht sibyllinisch (vgl. 4.3.6). Der in 4.3.5 bewiesene Einzigkeitssatz für den Körper $\mathbb{C}$ ist ein überzeugendes Indiz für die GAUSSsche These. Um die Auslegung des GAUSSschen Ausspruchs kam es noch in den 80er Jahren des 19. Jahrhunderts zu einer freundschaftlich geführten Kontroverse zwischen WEIERSTRASS und DEDEKIND; es ging (in heutiger Sprache) darum, alle endlich-dimensionalen, kommutativen und assoziativen $\mathbb{R}$-Algebren mit Einselement zu charakterisieren, wobei *Nullteiler* zugelassen sind.

Im Jahre 1843 erfand HAMILTON seine Quaternionen, kurz darauf konstruierten GRAVES und CAYLEY ihre Oktaven. Diese neuen hyperkomplexen Zahlensysteme sind keine Körper mehr – bei Quaternionen ist das Kommutativgesetz und bei Oktaven zusätzlich das Assoziativgesetz der Multiplikation verletzt –; es gibt aber noch zu jedem Element $\neq 0$ ein Inverses. *Die Division bleibt eindeutig ausführbar*: diese Eigenschaft der gewöhnlichen (rationalen) Zahlen war für die Väter der Theorie unabdingbar. Nullteiler oder gar nilpotente Elemente, die uns heute vom ersten Semester an bei Matrizen begegnen, wurden nicht zugelassen*). Überhaupt war die Beschäftigung mit hyperkomplexen Systemen keineswegs unumstritten. Noch 1890 sagt E. STUDY in seinem Artikel „Über Systeme complexer Zahlen und ihre Anwendungen in der Theorie der Transformationsgruppen" (Monatsh. Math. u. Phys. 1, 283–355, vgl. S. 341/42): „In weiten Kreisen, namentlich in Deutschland, ist die Ansicht verbreitet, dass die Systeme von complexen Zahlen oder ähnliche Algorithmen überhaupt gar keinen Nutzen hätten, ausgenommen allein die gewöhnlichen complexen Zahlen; und man begründet dies damit, dass durch sie nichts geleistet werden könnte, was nicht ‚ebenso gut' auch ohne sie zu leisten wäre."

2. Hyperkomplexe Systeme von Zahlen heißen seit Beginn des 20. Jahrhunderts (unverbindlicher und vor allem kürzer) *reelle Algebren*. Ist die Division eindeutig ausführbar, so spricht man von *Divisionsalgebren*. Wir orientieren uns am historischen Vorbild und stellen Divisionsalgebren in den Mittelpunkt der Diskus-

---

*) Erst WEIERSTRASS führt 1883 in seiner Arbeit „Zur Theorie der aus n Haupteinheiten gebildeten komplexen Größen" (Math. Werke 2, 311–339) den Begriff „Theiler der Null" ein (S. 314); er kämpft auch bereits mit nilpotenten Elementen (S. 319); die Bedeutung der Eigenschaft der Nullteilerfreiheit hat HANKEL 1867 klar gesehen (vgl. 4.3.6).

sion. Die klassische Divisionsalgebra ist die vierdimensionale Quaternionenalgebra; wir behandeln die Theorie der Quaternionen ausführlich im Kapitel 6. Den berühmten FROBENIUSschen Satz über die Einzigkeit der Quaternionen beweisen wir im Kapitel 7, dort wird auch der schöne Satz von HOPF hergeleitet, nach dem jede endlich-dimensionale kommutative Divisionsalgebra $\neq \mathbb{R}$ mit Einselement zu $\mathbb{C}$ isomorph ist.

Die achtdimensionale Divisionsalgebra der CAYLEY-Zahlen wird im Kapitel 8 studiert; es wird der Zornsche Satz über die Einzigkeit der CAYLEY-Zahlen bewiesen. Im Kapitel 9 werden Kompositionsalgebren betrachtet und deren Charakterisierung durch HURWITZ besprochen; damit ist der Weg bereitet, der im Kapitel 10 mit topologischen Methoden zu dem tiefliegenden Ergebnis von KERVAIRE und MILNOR führt: Nur in den Dimensionen 1, 2, 4 und 8 sind Divisionsalgebren möglich.

Um präzise und bequem formulieren zu können, stellen wir vorab weitgehend glossarisch grundlegende Begriffe und Tatsachen aus der allgemeinen Theorie der Algebren zusammen, spätere Verweise auf dieses Repertorium werden mit (R.*) zitiert.

Repertorium. Grundbegriffe aus der Theorie der Algebren

M. Koecher, R. Remmert

> Die größten und fruchtbarsten Fortschritte in der Mathematik sind vorzugsweise durch die Schöpfung neuer Begriffe gemacht, nachdem die häufige Wiederkehr zusammengesetzter Erscheinungen dazu gedrängt hat (R. DEDEKIND, Was sind und was sollen die Zahlen? 1888).

Wir legen den Körper $\mathbb{R}$ zugrunde, anstelle von $\mathbb{R}$ darf man hier auch jeden kommutativen Körper K wählen. Reelle Zahlen werden in den Kapiteln 6 bis 10 immer mit kleinen griechischen Buchstaben bezeichnet. Jeder n-dimensionale $\mathbb{R}$-Vektorraum ist isomorph zum Zahlenraum $\mathbb{R}^n$ der n-Tupel $x = (\xi_1, \ldots, \xi_n)$.

1. Reelle Algebren. Ein Vektorraum V über $\mathbb{R}$ mit einer „*Produktabbildung*" (oder *Multiplikation*) $V \times V \to V$, $(x, y) \mapsto xy$ heißt eine *Algebra über* $\mathbb{R}$ oder eine $\mathbb{R}$-*Algebra* oder (*reelle*) *Algebra*, wenn die beiden *Distributivgesetze*

$$(\alpha x + \beta y)z = \alpha \cdot xz + \beta \cdot yz, \qquad x(\alpha y + \beta z) = \alpha \cdot xy + \beta \cdot xz$$

für alle $\alpha, \beta \in \mathbb{R}$ und $x, y, z \in V$ erfüllt sind (*Bilinearität des Produktes*). Speziell gilt stets $\alpha(xy) = (\alpha x)y = x(\alpha y)$. Gilt das *Assoziativgesetz* $x(yz) = (xy)z$ für alle $x, y, z \in V$, so heißt die Algebra *assoziativ*; gilt das *Kommutativgesetz* $xy = yx$ für alle $x, y \in V$, dann spricht man von einer *kommutativen* Algebra. Nach diesen Definitionen ist eine $\mathbb{R}$-Algebra im allgemeinen also *weder* assoziativ *noch* kommutativ.

Ein Element $e \in V$ heißt *Einselement* der Algebra, wenn $ex = xe = x$ für alle $x \in V$; man sieht sofort, daß jede Algebra höchstens ein Einselement besitzt.

Um zwischen verschiedenen auf V definierten Algebren zu unterscheiden, nimmt man häufig die Multiplikation in die Bezeichnung auf und schreibt $\mathscr{A} := (V, \cdot)$. Die Dimension des $\mathbb{R}$-Vektorraumes V heißt die *Dimension der Algebra*; $\dim \mathscr{A} := \dim V$.

In jeder Algebra definiert man Potenzen induktiv durch $x^m := x \cdot x^{m-1}$. Beim Rechnen ist äußerste Vorsicht geboten, so gilt im allgemeinen $x \cdot x^2 \neq x^2 \cdot x$; selbst im kommutativen Fall läßt sich nicht zeigen, daß stets $x^4 = (x^2)^2$ gilt. Man nennt eine Algebra $\mathscr{A}$ *potenz-assoziativ*, wenn gilt

Potenzregel. $x^m x^n = x^{m+n}$ *für alle* $x \in \mathscr{A}$ *und alle* $m \geqslant 1$, $n \geqslant 1$.

Jede assoziative Algebra ist potenz-assoziativ.

Ein Element x einer Algebra $\mathscr{A}$ heißt ein *Nullteiler in* $\mathscr{A}$, wenn es ein Element $y \neq 0$ in $\mathscr{A}$ gibt, so daß $xy = 0$ oder $yx = 0$ gilt. Eine Algebra heißt *nullteilerfrei*, wenn sie keine Nullteiler $\neq 0$ hat: dies trifft genau dann zu, wenn aus $xy = 0$ stets $x = 0$ oder $y = 0$ folgt.

2. Beispiele reeller Algebren. Wir geben sieben instruktive Beispiele.

0) Die Körper $\mathbb{R}$ bzw. $\mathbb{C}$ sind assoziative und kommutative $\mathbb{R}$-Algebren mit Einselement der Dimension 1 bzw. 2.

1) Der $\mathbb{R}$-Vektorraum $\mathrm{Mat}(n, \mathbb{R})$ aller reellen $n \times n$ Matrizen ist bezüglich der Matrizenmultiplikation eine n^2-dimensionale, assoziative $\mathbb{R}$-Algebra mit Einselement (= Einheitsmatrix).

2) Der $\mathbb{R}$-Vektorraum $\mathrm{Mat}(n, \mathbb{C})$ aller komplexen $n \times n$ Matrizen ist bezüglich der Matrizenmultiplikation eine $2n^2$-dimensionale, assoziative $\mathbb{R}$-Algebra mit Einselement. Die Algebren $\mathrm{Mat}(n, \mathbb{R})$ und $\mathrm{Mat}(n, \mathbb{C})$ sind *nicht kommutativ*, falls $n > 1$.

3) Für zwei Vektoren $a = (\alpha_1, \alpha_2, \alpha_3)$, $b = (\beta_1, \beta_2, \beta_3) \in \mathbb{R}^3$ erklärt man vermöge

$$a \times b := (\alpha_2\beta_3 - \alpha_3\beta_2, \alpha_3\beta_1 - \alpha_1\beta_3, \alpha_1\beta_2 - \alpha_2\beta_1) \in \mathbb{R}^3$$

das *äußere Produkt* (= *Vektorprodukt*); vgl. Grundwissen Mathematik 2, Lineare Algebra und analytische Geometrie, 7.1.1. Der Vektorraum $\mathbb{R}^3$ wird so zu einer dreidimensionalen $\mathbb{R}$-Algebra, die *nicht assoziativ* und *antikommutativ* ist. Diese Algebra ist das einfachste, nicht-triviale Beispiel einer LIE-*Algebra*. Solche Algebren spielen in vielen Teilgebieten der modernen Mathematik eine wichtige Rolle, vgl. hierzu auch 6.1.4.

4) Der $\mathbb{R}$-Vektorraum $\mathrm{Sym}(n, \mathbb{R})$ aller reellen symmetrischen $n \times n$ Matrizen ist bezüglich des symmetrisierten Matrizen-Produkts, $(A, B) \mapsto \frac{1}{2}(AB + BA)$, eine kommutative Algebra, die im Falle $n > 1$ nicht assoziativ ist.

5) Jeder $\mathbb{R}$-Vektorraum $V \neq 0$ läßt sich zu einer *assoziativen und kommutativen* $\mathbb{R}$-*Algebra* $\mathscr{A}$ *mit Einselement* machen: man fixiert ein Element $e \in V$, $e \neq 0$, wählt zur Geraden $\mathbb{R}e \subset V$ irgendwie einen Supplementärraum U und erklärt für beliebige Vektoren $x = \alpha e + u$, $x' = \alpha' e + u' \in \mathbb{R}e \oplus U$ eine Multiplikation durch $xx' := (\alpha\alpha')e + \alpha u' + \alpha' u$. Dann ist e Einselement, es gilt $uu' = 0$ für alle $u, u' \in U$. In dieser Algebra sind also *alle* Elemente $\neq 0$ aus U Nullteiler.

6) Sind $\mathscr{A}_1 = (V_1, \cdot), \ldots, \mathscr{A}_s = (V_s, \cdot)$ reelle Algebren, so erklärt man auf der direkten Summe $V := V_1 \oplus \cdots \oplus V_s$ der zugehörigen Vektorräume ein Produkt, indem man für $x = x_1 + \cdots + x_s$, $y = y_1 + \cdots + y_s \in V$, wobei $x_i, y_i \in V_i$, setzt:

$$xy := x_1y_1 + \cdots + x_sy_s \qquad (\textit{komponentenweise Multiplikation}).$$

Die so gewonnene Algebra $\mathscr{A} := (V, \cdot)$ heißt *die direkte Summe* der Algebren $\mathscr{A}_1, \ldots, \mathscr{A}_s$, man schreibt: $\mathscr{A} = \mathscr{A}_1 \oplus \cdots \oplus \mathscr{A}_s$. Sind $\mathscr{A}_1, \ldots, \mathscr{A}_s$ sämtlich kommutativ bzw. assoziativ, so ist auch $\mathscr{A}$ kommutativ bzw. assoziativ; im Fall $s > 1$ hat $\mathscr{A}$ stets Nullteiler. Ist e_i Einselement von $\mathscr{A}_i$, $1 \leqslant i \leqslant s$, so ist $e := e_1 + \cdots + e_s$ Einselement von $\mathscr{A}$.

Alle Algebren in 0)–5) sind potenz-assoziativ.

3. Unteralgebren und Algebra-Homomorphismen. Ein reeller Untervektorraum U einer $\mathbb{R}$-Algebra $\mathscr{A} = (V, \cdot)$ heißt eine $\mathbb{R}$-*Unteralgebra* von $\mathscr{A}$, wenn $xy \in U$ für alle $x, y \in U$ gilt.

Beispiele. 1) Die Menge $\{\left(\begin{smallmatrix} \alpha & -\beta \\ \beta & \alpha \end{smallmatrix}\right) : \alpha, \beta \in \mathbb{R}\}$ ist eine $\mathbb{R}$-Unteralgebra von $\mathrm{Mat}(2, \mathbb{R})$ (vgl. Kapitel 3.2.5).

2) Die Mengen der oberen Dreiecksmatrizen bilden jeweils $\mathbb{R}$-Unteralgebren von $\mathrm{Mat}(n, \mathbb{R})$ bzw. $\mathrm{Mat}(n, \mathbb{C})$ der Dimension $\frac{n(n+1)}{2}$ bzw. $n(n+1)$.

Sind $\mathscr{A} = (V, \cdot)$ und $\mathscr{B} = (W, \cdot)$ zwei Algebren, so heißt eine $\mathbb{R}$-*lineare Abbildung* $f: V \to W$ ein $\mathbb{R}$-*Algebra-Homomorphismus*, wenn gilt:

$$f(xy) = f(x)f(y) \qquad \text{für alle} \qquad x, y \in V.$$

Man spricht von einem *Mono-*, *Epi-*, *Iso-*, *Endo-* bzw. *Automorphismus*, wenn die $\mathbb{R}$-lineare Abbildung $f: V \to W$ so beschaffen ist.

Beispiel. Die Abbildung $f: \mathbb{C} \to \mathrm{Mat}(2, \mathbb{R})$, $\alpha + \beta i \mapsto \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$ ist ein Algebra-Monomorphismus.

Bemerkung. Ist $\mathscr{A}$ eine Algebra mit Einselement e, so ist $f: \mathbb{R} \to \mathscr{A}$, $\alpha \mapsto \alpha e$, ein Algebra-Monomorphismus. Speziell ist jede eindimensionale reelle Algebra mit Einselement isomorph zu $\mathbb{R}$.

4. Bestimmung aller eindimensionalen Algebren. Jeder reelle Vektorraum V wird trivial zu einer Algebra, wenn man als Multiplikation $V \times V \to V$ die Nullabbildung $(x, y) \mapsto 0$ wählt. Wir zeigen, daß im eindimensionalen Fall diese Pathologie die einzige Ausnahme ist.

Satz. *Jede eindimensionale Algebra $\mathscr{A}$, deren Multiplikation nicht die Nullabbildung ist, ist zur Algebra $\mathbb{R}$ isomorph.*

Beweis. Aufgrund der Bemerkung 3 genügt es zu zeigen, daß $\mathscr{A}$ ein Einselement hat. Es gilt $\mathscr{A} = \mathbb{R}a$ mit $a \in \mathscr{A} \setminus \{0\}$. Da nicht stets $xy = 0$ gilt, folgt $a^2 \neq 0$, also auch $\mathscr{A} = \mathbb{R}a^2$. Daher besteht eine Gleichung $a = \varepsilon a^2$ mit $\varepsilon \in \mathbb{R}$. Dann ist $e := \varepsilon a$ Einselement von $\mathscr{A}$.

5. Divisionsalgebren. Seit HAMILTON spielen die (endlich-dimensionalen) Divisionsalgebren eine zentrale Rolle. Eine Algebra $\mathscr{A} \neq 0$ heißt eine *Divisionsalgebra*, wenn für jedes $a \in V$, $a \neq 0$, die beiden Gleichungen $ax = b$ und $ya = b$ für alle $b \in V$ *eindeutig in $\mathscr{A}$ lösbar sind.*

Die Körper $\mathbb{R}$ bzw. $\mathbb{C}$ sind assoziative und kommutative Divisionsalgebren der Dimension 1 bzw. 2. Die Matrixalgebren $\mathrm{Mat}(n, \mathbb{R})$ und $\mathrm{Mat}(n, \mathbb{C})$ sind im Falle $n > 1$ keine Divisionsalgebren.

Lemma. *Ist $\mathscr{A}$ eine assoziative Divisionsalgebra, so ist $G := \mathscr{A} \setminus \{0\}$ bezüglich der Multiplikation in $\mathscr{A}$ eine Gruppe. Das neutrale Element von G ist das Einselement von $\mathscr{A}$.*

Beweis. Da in G jede Gleichung $ax = b$ und $ya = b$ eindeutig lösbar ist, so ist G eine Gruppe. □

Jede Divisionsalgebra ist nullteilerfrei; als Umkehrung gilt nur das

Kriterium. *Folgende Aussagen über eine endlich-dimensionale Algebra $\mathscr{A}$ sind äquivalent:*

i) *$\mathscr{A}$ ist Divisionsalgebra.*
ii) *$\mathscr{A}$ ist nullteilerfrei.*

Beweis. Es ist nur ii) $\Rightarrow$ i) zu zeigen. Sei $a \in \mathscr{A} \setminus \{0\}$. Die Abbildung $\mathscr{A} \to \mathscr{A}$, $x \mapsto ax$ ist ein Vektorraumendomorphismus, nach Voraussetzung injektiv und dann wegen $\dim \mathscr{A} < \infty$ sogar bijektiv. (Siehe Grundwissen Mathematik 2, Lineare Algebra und analytische Geometrie, 1.6.5.) Damit ist jede Gleichung $ax = b$ eindeutig lösbar. Die eindeutige Lösbarkeit von $ya = b$ ergibt sich analog durch Betrachtung der Abbildung $\mathscr{A} \to \mathscr{A}$, $y \mapsto ya$. □

Es ist nicht trivial, reelle Divisionsalgebren $\neq \mathbb{R}, \mathbb{C}$ anzugeben. Die einfachste solche Algebra ist die Hamiltonsche Algebra $\mathbb{H}$ der Quaternionen, die im nächsten Kapitel besprochen wird.

6. Konstruktion von Algebren mittels Basen. Es gibt ein einfaches Verfahren, einen n-dimensionalen, reellen Vektorraum V zu einer Algebra $\mathscr{A} = (V, \cdot)$ zu machen. Man fixiert eine Basis $e_1, e_2, \ldots, e_n$ in V. Sind dann $x = \sum_1^n \alpha_\mu e_\mu$, $y = \sum_1^n \beta_\nu e_\nu \in V$ beliebig, so gilt für *jedes Produkt* $(x, y) \mapsto xy$ in V aufgrund der Distributivgesetze

$$xy = \sum_{\mu,\nu=1}^{n} (\alpha_\mu \beta_\nu) e_\mu e_\nu.$$

Eine Multiplikation in V ist also bereits vollständig bestimmt durch die n^2 Einzelprodukte $e_\mu e_\nu$. Deren Werte lassen sich *beliebig* in V wählen; man erhält so alle $\mathbb{R}$-Algebren auf V. Die meisten dieser Algebren sind völlig uninteressant.

Will man Algebren mit Einselement konstruieren, so postuliert man gern, daß e_1 das Einselement sein soll. Dann muß gelten $e_1 e_\nu = e_\nu e_1 = e_\nu$ für alle $\nu = 1, \ldots, n$; daher kann man jetzt nur noch über die $(n-1)^2$ Einzelprodukte $e_\mu e_\nu$, $2 \leqslant \mu, \nu \leqslant n$, frei verfügen.

Ist neben $\mathscr{A} = (V, \cdot)$ eine weitere $\mathbb{R}$-Algebra $\mathscr{B} = (W, \cdot)$ gegeben, so ist eine $\mathbb{R}$-lineare Abbildung $f\colon V \to W$ genau dann ein $\mathbb{R}$-Algebrahomomorphismus, wenn gilt: $f(e_\mu e_\nu) = f(e_\mu) f(e_\nu)$ für alle $\mu, \nu = 1, 2, \ldots, n$.

Man hat ein evidentes Assoziativitäts- und Kommutativitätskriterium:

Die Algebra $\mathscr{A} = (V, \cdot)$ ist genau dann assoziativ, wenn $(e_\lambda e_\mu) e_\nu = e_\lambda (e_\mu e_\nu)$ für alle $\lambda, \mu, \nu = 1, 2, \ldots, n$; sie ist genau dann kommutativ, wenn $e_\mu e_\nu = e_\nu e_\mu$ für alle $\mu, \nu = 1, 2, \ldots, n$.

Die Verifikation der n^3 Assoziativitätsbedingungen ist für praktische Rechnungen mühsam; selbst im Falle, wo e_1 Einselement ist, bleiben $(n-1)^3$ Gleichungen zu testen. Aus diesem Grunde werden heute kaum noch Algebren durch Vorgabe der Basisprodukte $e_\mu e_\nu$ beschrieben. Klassisch ist HAMILTON bei seinen Quaternionen aber genau so verfahren; auch DEDEKIND und WEIERSTRASS haben im kommutativen und assoziativen Fall Basen (sogenannte Haupteinheiten) benutzt.

Kapitel 6. HAMILTONsche Quaternionen

M. Koecher, R. Remmert

Love of fame moves and cheers great mathematicians (W. R. HAMILTON).

Einleitung

1. Sir William Rowan HAMILTON (geb. 1805 in Dublin; liest mit 5 Jahren Lateinisch, Griechisch und Hebräisch; 1823 Student am Trinity College in Dublin; 1827 als Undergraduate Student Berufung zum Professor der Astronomie an der Universität Dublin und zum Direktor der Sternwarte von Dunsink mit dem Titel: Royal Astronomer of Ireland; entwickelt 1827 die geometrische Optik aus Extremalprinzipien; 1834/35 Übertragung der Extremalprinzipien auf die Dynamik, Hamiltonsches Prinzip der kleinsten Wirkung, Hamiltonsche Wirkungsfunktion, kanonische Bewegungsgleichungen; 1835 Ritterschlag; 1837–1845 Präsident der Royal Irish Academy; 1843 Erfindung der Quaternionen; gest. 1865 in Dunsink) hat 1835 das Rechnen mit komplexen Zahlen $x + iy$ logisch gerechtfertigt als ein Operieren mit geordneten reellen Zahlenpaaren (x, y) nach postulierten Rechenregeln (vgl. 3.1.8). Dies war der Ausgangspunkt für HAMILTONS Interesse an der Frage, ob die geometrische Deutung des Addierens und vor allem des Multiplizierens von komplexen Zahlen in der Ebene $\mathbb{R}^2$ nicht irgendwie – durch Schaffung *hyperkomplexer* Zahlen – auch im Anschauungsraum $\mathbb{R}^3$ ein Analogon haben könne.

HAMILTON hat jahrelang gehofft, eine Multiplikation für reelle Tripel mit guten Eigenschaften zu finden. An seinen Sohn schreibt er 1865 kurz vor seinem Tode (Math. Papers 3, p. XV): „Every morning, on my coming down to breakfast, you used to ask me: ‚Well, Papa, can you multiply triplets?' Whereto I was always obliged to reply, with a sad shake of the head: ‚No, I can only add and subtract them'."

Heute ist es leicht, sich klar zu machen, daß es keine Multiplikation im $\mathbb{R}^3$ aller reellen Tripel (α, β, γ) geben kann, welche die Multiplikation von $\mathbb{C} = \mathbb{R}^2 \subset \mathbb{R}^3$ der Paare (α, β) fortsetzt. Bezeichnet nämlich $e := (1,0,0), i := (0,1,0), j := (0,0,1)$ die kanonische Basis des $\mathbb{R}^3$, so müßte gelten: $ij = \rho e + \sigma i + \tau j$. Daraus würde folgen, wenn man $i^2 = -e$ und $i(ij) = (ii)j = -j$ unterstellt:

$$-j = \rho i - \sigma e + \tau ij = \rho i - \sigma e + \tau(\rho e + \sigma i + \tau j) = (\tau\rho - \sigma)e + (\tau\sigma + \rho)i + \tau^2 j,$$

also (wegen der linearen Unabhängigkeit von e, i, j): $\tau^2 = -1$, das heißt $\tau \notin \mathbb{R}$.*)

*) Es gilt ein besserer

Satz. *Jede reelle Divisionsalgebra $\mathscr{A}$ ungerader Dimension mit Einselement e ist isomorph zu $\mathbb{R}$, hat also die Dimension* 1.

Beweis. Sei $a \in \mathscr{A}$. Die „Linksmultiplikation" $L_a : \mathscr{A} \to \mathscr{A}$, $x \mapsto ax$, ist ein Vektorraum-Endomorphismus. Da $\dim \mathscr{A}$ ungerade ist, so hat L_a einen reellen Eigenwert λ (Zwischenwertsatz). Ist $v \neq 0$ ein zugehöriger Eigenvektor, so gilt also $av = \lambda v$, das heißt, $(a - \lambda e)v = 0$. Da $\mathscr{A}$ Divisionsalgebra ist, folgt $a = \lambda e$, das heißt, $a \in \mathbb{R}e$. Wir sehen $\mathscr{A} = \mathbb{R}e$.

2. Hamiltons Anstrengungen sind zunächst erfolglos: er sucht nach einer Multiplikation für Tripel, so daß wie bei Zahlenpaaren (vgl. 3.2.1) die üblichen Regeln weiterhin gelten (das heißt, er unterstellt ein Permanenzprinzip). Er macht den Ansatz

$$\alpha + \beta i + \gamma j \qquad \text{mit} \qquad i^2 = j^2 = -1,$$

(worin die Existenz eines neutralen Elementes enthalten ist) und betrachtet, indem er „kommutativ rechnet", den einfachsten Fall

$$(*) \qquad (\alpha + \beta i + \gamma j)^2 = \alpha^2 - \beta^2 - \gamma^2 + 2i\alpha\beta + 2j\alpha\gamma + 2ij\beta\gamma.$$

Prüfstein für den Wert der Multiplikation wird wie im Fall $\mathbb{C}$ die Produktregel (law of moduli), nach der die Länge von Produktvektoren mit dem Produkt der Längen übereinstimmen soll (wobei $\alpha + \beta i + \gamma j$ die euklidische Länge $\sqrt{\alpha^2 + \beta^2 + \gamma^2}$ hat). Die Berechnung der Summe der Quadrate der Koeffizienten von $1, i$ und j auf der rechten Seite von $(*)$ gibt

$$(\alpha^2 - \beta^2 - \gamma^2)^2 + (2\alpha\beta)^2 + (2\alpha\gamma)^2 = (\alpha^2 + \beta^2 + \gamma^2)^2;$$

damit stellt Hamilton fest, daß die Produktregel sicher dann gilt, wenn man $ij = 0$ setzt. Doch das gefällt ihm nicht. Und nun bemerkt er, daß rechts in $(*)$ ja gar nicht $2ij$ steht, sondern (!) $ij + ji$. Das muß verschwinden: $ji = -ij$; so wird er dazu geführt, das Kommutativgesetz zu opfern. Man kann dies sehr schön nachlesen in einem Brief von Hamilton an John Graves vom 17. Oktober 1843 (Math. Papers 3, 106–110): „Behold me therefore tempted for a moment to fancy that $ij = 0$. But this seemed odd and uncomfortable, and I perceived that the same suppression of the term which was *de trop* might be attained by assuming what seemed to me less harsh, namely that $ji = -ij$. I made therefore $ij = k, ji = -k$, reserving to myself to inquire whether k was 0 or not."

Und nun hat Hamilton den genialen Einfall, der dem ganzen Problem die entscheidende Wendung gibt: er „springt mit k in eine vierte Dimension", das heißt, er nimmt k als von $1, i$ und j linear unabhängig an. In seinem Brief an Graves schreibt er (loc. cit.): „And there dawned on me the notion that we must admit, in some sense, a *fourth dimension* of space for the purpose of calculating with triplets."

Hamilton untersucht nun vorsichtig, was k^2 ist. Bei Verwendung des Assoziativgesetzes wäre sofort klar:

$$k^2 = (ij)(ij) = i(ji)j = -i(ij)j = -i^2j^2 = -1;$$

doch er geht nicht so vor, da er noch nicht sicher ist, ob seine Multiplikation assoziativ wird (seine diesbezüglichen Notizen findet man in Math. Papers 3, 103–105).

Später hat er die Gültigkeit des Assoziativgesetzes klar herausgestellt, so schreibt er (Math. Papers 3, S. 114): „... the commutative character is lost .... However it will be found that another important property of the old multiplication is preserved, or extended to the new, namely, that which may be called the *associative* character of the operation ...". Das dürfte die erste Einführung der Bezeichnung „assoziativ" in die Mathematik sein.

3. Der Durchbruch gelang Hamilton am 16. Oktober 1843 auf dem Wege zur Sitzung der Royal Irish Academy; noch auf jener Sitzung kündigt er seine

Erfindung der Quaternionen an. Sein weiteres Leben widmet er ausschließlich der Erforschung der Quaternionen. Den Augenblick der Entdeckung hat er selbst 1858 wie folgt beschrieben (North. British Review 14, 1858): „... Tomorrow will be the fifteenth birthday of the Quaternions. They started into life, or light, full grown, on the 16th of October, 1843, as I was walking with Lady Hamilton to Dublin, and came up to Brougham Bridge. That is to say, I then and there felt the galvanic circuit of thought closed, and the sparks which fell from it were the fundamental equations between i, j, k *exactly such* as I have used them ever since. I pulled out, on the spot, a pocketbook, which still exists, and made an entry, on which, *at the very moment*, I felt that it might be worth my while to expend the labour of at least ten (or it might be fifteen) years to come. But then it is fair to say that this was because I felt a *problem* to have been at that moment *solved*, an intellectual *want relieved*, which had *haunted* me for at least *fifteen years* before ..."

Und im bereits erwähnten Brief an seinen Sohn sagt er noch zu jenem denkwürdigen Oktobertag: „Nor could I resist the impulse – unphilosophical as it may have been – to cut with a knife on a stone of Brougham Bridge the fundamental formula with the symbols i, j, k:

$$i^2 = j^2 = k^2 = ijk = -1."$$

Mit großer Freude verifiziert HAMILTON die Gültigkeit der Produktregel für seine Quaternionenmultiplikation, er schreibt (Math. Papers 3, S. 108): „But I considered it essential to try whether [my] equations were consistent with the law of moduli, ..., without which consistence being verified, I should have regarded the whole speculation as a failure."

Weder HAMILTON noch andere wußten damals, daß EULER bereits 1748 die typischen Rechengesetze für Quaternionen besaß. In einem Brief an GOLDBACH vom 4. Mai gibt er die Produktregel als „Vier-Quadrate-Satz" an (vgl. hierzu 2.3). Auch GAUSS hat die Rechenregeln für Quaternionen gekannt, er schrieb 1819 eine (damals nicht publizierte) kurze Note über *Mutationen des Raumes*, wo die Quaternionenformeln auftreten (Werke 8, 357–362).

4. HAMILTON betrachtete seine Schöpfung der Quaternionen als ebenbürtig mit der Schöpfung der Infinitesimalrechnung. An zeitgenössischen Mathematikern erkannte er außer GAUSS und GRASSMANN niemanden an. „Graßmann teilt sich mit Gauß in die Ehre, daß Hamilton ihm zutraut, er könne die Quaternionen gefunden haben, und sich immer von Neuem freut, daß es allem Anscheine nach doch nicht der Fall ist" (so F. ENGEL auf S. 208 in seinem lesenswerten Band „Graßmanns Leben", Teubner Verlag, Leipzig 1911).

HAMILTON glaubte, daß seinen Quaternionen eine Schlüsselstellung für die mathematische Physik zukomme. Mit missionarischem Eifer hat er ihre Verbreitung in der mathematischen Welt betrieben. So wurden in Dublin Quaternionen ein offizielles Examensfach; man sprach ihnen eine kosmische Signifikanz zu. Sehr hart ist das Urteil von Felix KLEIN in seinen „Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert"; er sagt (Bd. 1, S. 184): „Hamilton selbst gestaltete sie [=Quaternionen] für sich zu einer Art orthodoxer Lehre des mathematischen Credo, in die er alle seine geometrischen und sonstigen

Interessen hineinzwang, je mehr sich gegen Ende seines Lebens sein Geist vereinseitigte und ...".

5. HAMILTON wurde in Irland und England zur Galionsfigur einer Schule von Quaternionisten, „die ihren Meister an Starrheit und Intoleranz noch übertraf". Im Mittelpunkt stand ein mit Scheu und Verehrung gehandhabter Formalismus. Man träumte von einer quaternionistischen Funktionentheorie und erhoffte sich ganz neue und tiefe Einsichten in die gesamte Mathematik. Zur Förderung dieser utopischen Ziele wurde gar 1895 an der Yale University in New Haven, Conn., eine „International Association for Promoting the Study of Quaternions" gegründet. Die große Zeit der Quaternionisten hat in Irland bis in die Gegenwart hinein ausgestrahlt, so kam noch Eamon de VALERA, Präsident der Republik Irland von 1959–1973, während seiner Amtszeit gelegentlich zum mathematischen Colloquium in Dublin, wenn die Vortragsankündigung das Wort „Quaternionen" enthielt.

Die Geschichte der Algebra hat gezeigt, daß im letzten Jahrhundert die Bedeutung der Quaternionen weit überschätzt wurde. Heute ist klar, daß die Quaternionenalgebra nur eine spezielle Algebra komplexer 2×2 Matrizen ist (vgl. § 1). Nicht die Erfindung der Quaternionen war die große Leistung, sondern vielmehr die dadurch gewonnene Erkenntnis der großen Freiheiten, die man hat, hyperkomplexe Systeme zu konstruieren. Sehr scharf ist das Urteil von Lord KELVIN (1824–1907, brit. Physiker, Thermodynamik): „Quaternions came from Hamilton after his really good work had been done; and though beautifully ingenious, have been an unmixed evil to those who have touched them in any way."

Demgegenüber steht ein berühmter Satz von Thomas HILL (Schüler von B. PEIRCE, 1862 Präsident von Harvard): „In the great mathematical birth of 1843, the Quaternions of HAMILTON, there is as much real promise of benefit to mankind as in any event of Victoria's reign."

Leser, die an weiteren historischen Einzelheiten interessiert sind, seien verwiesen auf:

CROWE, M. J.: A History of Vector Analysis, University of Notre Dame Press, Notre Dame, London 1967

ROTHE, H.: Die Hamiltonschen Quaternionen und ihre Verallgemeinerungen, Encykl. Math. Wiss. III. 1.2, 1300–1423, Teubner Verlag, Leipzig 1914–1931

VAN DER WAERDEN, B. L: Hamiltons Entdeckung der Quaternionen, Veröffentlichungen der Joachim Jungius Gesellschaft der Wissenschaften, Vandenhoeck u. Ruprecht, Göttingen 1973, 14 Seiten

§ 1. Die Quaternionenalgebra $\mathbb{H}$

Die Einführung der Quaternionen geschieht im Abschnitt 1 nach HAMILTONschem Vorbild durch Angabe einer Multiplikationstabelle für die natürliche Basis. Im Abschnitt 2 werden Quaternionen als spezielle komplexe 2×2 Matrizen realisiert: es werden eine Unteralgebra $\mathscr{H}$ von $\mathrm{Mat}(2, \mathbb{C})$ und ein natürlicher Isomorphismus

$F: \mathbb{H} \to \mathscr{H}$ der Quaternionenalgebra $\mathbb{H}$ auf $\mathscr{H}$ konstruiert, den CAYLEY bereits 1858 kannte; durch diese Isomorphie wird u. a. evident, daß $\mathbb{H}$ eine assoziative Divisionsalgebra über $\mathbb{R}$ ist. HAMILTON mußte die Assoziativität von $\mathbb{H}$ direkt nachweisen, da im Entdeckungsjahr 1843 Matrizen noch unbekannt waren; erst 1858 hat CAYLEY in „A Memoir on the Theory of Matrices" (Math. Papers 2, 475–496) Matrizen und den Matrizenkalkül eingeführt, der den Quaternionenkalkül als Spezialfall umfaßt. Die Algebra $\mathscr{H}$ und der Isomorphismus F können in diesem ganzen Kapitel nutzbringend verwendet werden.

In den Abschnitten 3 bis 7 dieses Paragraphen werden grundlegende algebraische Eigenschaften der Quaternionen diskutiert.

1. Die Algebra $\mathbb{H}$ der Quaternionen. Im vierdimensionalen $\mathbb{R}$-Vektorraum $\mathbb{R}^4$ der geordneten reellen Quadrupel wird die Standardbasis

$$e_1 := (1,0,0,0), \quad e_2 := (0,1,0,0), \quad e_3 := (0,0,1,0), \quad e_4 := (0,0,0,1)$$

ausgezeichnet. Wir führen die sogenannte *Hamiltonsche Multiplikation* ein. Es sei e_1 das Einselement. Dann sind gemäß R.6 noch die neun Produkte $e_\mu e_\nu$, $2 \leqslant \mu, \nu \leqslant 4$, festzulegen. Wir setzen:

$$\left.\begin{array}{lll} e_2e_2 := -e_1, & e_2e_3 := e_4, & e_2e_4 := -e_3 \\ e_3e_2 := -e_4, & e_3e_3 := -e_1, & e_3e_4 := e_2 \\ e_4e_2 := e_3, & e_4e_3 := -e_2, & e_4e_4 := -e_1 \end{array}\right\} \text{(HAMILTONsche Bedingungen).}$$

Man schreibt dies häufig in Form der Multiplikationstabelle

	e_2	e_3	e_4
e_2	$-e_1$	e_4	$-e_3$
e_3	$-e_4$	$-e_1$	e_2
e_4	e_3	$-e_2$	$-e_1$

.

Die so konstruierte vierdimensionale $\mathbb{R}$-Algebra nennt man *die Quaternionen-Algebra* und bezeichnet sie mit $\mathbb{H}$. Nach HAMILTON werden die Elemente von $\mathbb{H}$ *Quaternionen**) genannt. Wegen $e_2e_3 \neq e_3e_2$ ist klar: *Die Quaternionen-Algebra* $\mathbb{H}$ *ist nicht kommutativ.*

Durch Verifikation der 27 Gleichungen $(e_\lambda e_\mu)e_\nu = e_\lambda(e_\mu e_\nu)$, $2 \leqslant \lambda, \mu, \nu \leqslant 4$, läßt sich direkt einsehen, daß die Quaternionenalgebra assoziativ ist. Wir verzichten auf die Durchführung der Rechnungen, da sich Assoziativität und mehr im nächsten Abschnitt eleganter ergeben. Traditionell schreibt man $e := e_1$, $i := e_2$, $j := e_3$, $k := e_4$ und entsprechend

$$i^2 = j^2 = k^2 = ijk = -e, \qquad ij = -ji = k.$$

---

*) Bezeichnung aus der Vulgata für die vier Rotten von je vier Kriegsknechten des Herodes, die Petrus im Gefängnis bewachten: „he put him in prison, and delivered him to four quaternions of soldiers to keep him" (Apost. Gesch. 12.4); vgl. hierzu G. TEMPLE: *100 Years of Mathematics*, Duckworth, London 1981, S. 46.

Die weiteren Produkte entstehen hieraus durch zyklische Vertauschung von i,j,k. Durch distributives Ausrechnen erhält man damit die

Produktformel. $(\alpha e + \beta i + \gamma j + \delta k)(\alpha' e + \beta' i + \gamma' j + \delta' k)$

$$= (\alpha\alpha' - \beta\beta' - \gamma\gamma' - \delta\delta')e + (\alpha\beta' + \beta\alpha' + \gamma\delta' - \delta\gamma')i$$
$$+ (\alpha\gamma' - \beta\delta' + \gamma\alpha' + \delta\beta')j + (\alpha\delta' + \beta\gamma' - \gamma\beta' + \delta\alpha')k.$$

Die klassische Schreibweise der Quaternionen mittels i, j, k birgt Gefahren, z. B. gewiß dann, wenn man Quaternionen mit komplexen Zahlen statt reellen Zahlen als Koeffizienten betrachtet.

$\mathbb{R}e$ ist eine $\mathbb{R}$-Unteralgebra von $\mathbb{H}$. Im Gegensatz zu $\mathbb{C}$ identifizieren wir aber nicht $\mathbb{R}e$ mit $\mathbb{R}$; wir schreiben dementsprechend auch konsequent e und nicht 1 für das Einselement von $\mathbb{H}$.

2. Die Matrixalgebra $\mathscr{H}$ und der Isomorphismus $F: \mathbb{H} \to \mathscr{H}$. Die Menge $\mathscr{C}$ aller reellen 2×2 Matrizen $\left(\begin{smallmatrix} \alpha & -\beta \\ \beta & \alpha \end{smallmatrix}\right)$, $\alpha, \beta \in \mathbb{R}$, ist eine $\mathbb{R}$-Unteralgebra von $\mathrm{Mat}(2, \mathbb{R})$, die Abbildung $\alpha + \beta i \mapsto \left(\begin{smallmatrix} \alpha & -\beta \\ \beta & \alpha \end{smallmatrix}\right)$ ist ein $\mathbb{R}$-Algebra-Isomorphismus $\mathbb{C} \to \mathscr{C}$ (vgl. 3.2.5). In Analogie hierzu gilt

Satz. *Die Menge* $\mathscr{H} := \{\left(\begin{smallmatrix} w & -z \\ \bar{z} & \bar{w} \end{smallmatrix}\right): w, z \in \mathbb{C}\}$ *ist eine* $\mathbb{R}$*-Unteralgebra von* $\mathrm{Mat}(2, \mathbb{C})$ *mit Einselement* $E := \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. *Jede Matrix* $A = \left(\begin{smallmatrix} w & -z \\ \bar{z} & \bar{w} \end{smallmatrix}\right) \in \mathscr{H}$ *genügt über* $\mathbb{R}$ *der quadratischen Gleichung*

(1) $A^2 - (\mathrm{Spur}\, A)A + (\det A)E = 0$ *mit* $\mathrm{Spur}\, A = 2 \operatorname{Re} w$, $\det A = |w|^2 + |z|^2$.

$\mathscr{H}$ *ist eine* 4*-dimensionale, assoziative Divisionsalgebra.*

Beweis. 1) Durch direktes Nachrechnen verifiziert man, daß $\mathscr{H}$ ein vierdimensionaler $\mathbb{R}$-Untervektorraum von $\mathrm{Mat}(2, \mathbb{C})$ ist, der abgeschlossen bezüglich Matrizenmultiplikation ist. Die Gleichung $A^2 - (\mathrm{Spur}\, A)A + (\det A)E = 0$ bestätigt man ebenfalls durch Nachrechnen.

2) Die Algebra $\mathscr{H}$ ist assoziativ, da $\mathrm{Mat}(2, \mathbb{C})$ assoziativ ist. Um einzusehen, daß $\mathscr{H}$ eine Divisionsalgebra ist, benutzen wir das Kriterium R.5. Seien also $A, B \in \mathscr{H}$ und $AB = 0$. Dann folgt $\det A \cdot \det B = 0$ und also $\det A = 0$ oder $\det B = 0$. Da $\det\left(\begin{smallmatrix} w & -z \\ \bar{z} & \bar{w} \end{smallmatrix}\right) = |w|^2 + |z|^2$ nur verschwindet, wenn $w = z = 0$ gilt, so folgt die Behauptung. □

Gleichung (1) ist die Aussage des sogenannten Satzes von CAYLEY bzw. HAMILTON-CAYLEY für den Spezialfall von 2×2 Matrizen, vgl. Grundwissen Mathematik 2, Lineare Algebra und analytische Geometrie, 2.5.7.

Lemma. *Die Abbildung*

$$F: \mathbb{H} \to \mathscr{H}, \qquad (\alpha, \beta, \gamma, \delta) \mapsto \begin{pmatrix} \alpha + \beta i & -\gamma - \delta i \\ \gamma - \delta i & \alpha - \beta i \end{pmatrix},$$

ist ein $\mathbb{R}$*-Algebra-Isomorphismus; es gilt:*

$$F(e_1) = E, \qquad F(e_2) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} =: I,$$

$$F(e_3) = \begin{pmatrix} 0 & -1 \\ +1 & 0 \end{pmatrix} =: J, \qquad F(e_4) = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} =: K.$$

Beweis. Die Abbildung F ist offensichtlich $\mathbb{R}$-linear und bijektiv. Es bleibt zu zeigen (vgl. R.6): $F(e_\mu)F(e_\nu) = F(e_\mu e_\nu)$ für $\mu, \nu = 1, 2, 3, 4$. Dies aber ist klar, denn die Matrizen E, I, J, K sind die F-Bilder von e_1, e_2, e_3, e_4 und genügen denselben Multiplikationsregeln wie e_1, e_2, e_3, e_4: Man verifiziert $I^2 = J^2 = -E$, $IJ = -JI = K$ und erhält die fehlenden Beziehungen aus dem Assoziativgesetz, z. B. $K^2 = (IJ)(-JI) = -IJ^2I = I^2 = -E$.

Korollar. *Die Hamiltonsche Algebra* $\mathbb{H}$ *ist eine assoziative Divisionsalgebra.*

Aufgrund von Lemma R.5 ist $\mathscr{H} \setminus \{0\}$ bezüglich der Multiplikation eine Gruppe. Man verifiziert sofort:

Die Menge $\{E, -E, I, -I, J, -J, K, -K\}$ *ist eine nichtkommutative Untergruppe von* $\mathscr{H} \setminus \{0\}$; *jedes ihrer Elemente* $\neq \pm E$ *hat die Ordnung* 4.

In der Literatur nennt man diese Gruppe und jede zu ihr isomorphe Gruppe *die* (*endliche*) *Quaternionengruppe.*

Die hier benutzte Darstellung von Quaternionen durch komplexe 2×2 Matrizen war CAYLEY bereits 1858 geläufig. In seinem berühmten „Memoir on the Theory of Matrices" schreibt er (Math. Papers 2, S. 491): „It may be noticed in passing, that if L, M are skew convertible matrices of the order 2, and if these matrices are also such that $L^2 = -1$, $M^2 = -1$, then putting $N = LM = -ML$, we obtain

$$L^2 = -1, \qquad M^2 = -1, \qquad N^2 = -1,$$

$$L = MN = -NM, \qquad M = NL = -NL\,[\text{sic}], \qquad N = LM = -ML,$$

which is a system of relations precisely similar to that in the theory of quaternions." Explizite Beispiele für L, M gibt CAYLEY aber nicht an.

Da sich mit komplexen Matrizen eleganter als mit Quaternionen rechnen läßt, beweist man häufig – analog wie oben – Sätze für $\mathbb{H}$ dadurch, daß man sie für die Algebra $\mathscr{H}$ herleitet und dann mittels des Isomorphismus $F: \mathbb{H} \to \mathscr{H}$ nach $\mathbb{H}$ „liftet". Dieses Prinzip wird auch weiterhin benutzt. □

Wie früher bei komplexen Zahlen gibt es viele Möglichkeiten, die Quaternionenalgebra $\mathbb{H}$ als eine $\mathbb{R}$-Unteralgebra von $\mathrm{Mat}(2, \mathbb{C})$ zu realisieren. Man wähle irgendwie drei Matrizen $I_2, I_3, I_4 \in \mathrm{Mat}(2, \mathbb{C})$, so daß die neun HAMILTONschen Bedingungen erfüllt sind. Dann ist die Abbildung

$$\mathbb{H} \to \mathrm{Mat}(2, \mathbb{C}), \qquad (\alpha, \beta, \gamma, \delta) \to \alpha E + \beta I_2 + \gamma I_3 + \delta I_4$$

ein $\mathbb{R}$*-Algebra-Monomorphismus.* Es läßt sich in Verallgemeinerung des Satzes 3.2.6 zeigen:

Ist $g: \mathbb{H} \to \mathrm{Mat}(2, \mathbb{C})$ ein $\mathbb{R}$-*Algebra-Monomorphismus, so gibt es eine invertierbare Matrix* $W \in \mathrm{Mat}(2, \mathbb{C})$, *so daß für den zugehörigen „inneren Automorphismus"* $\iota_W : \mathrm{Mat}(2, \mathbb{C}) \to \mathrm{Mat}(2, \mathbb{C})$, $A \mapsto W^{-1}AW$ *gilt*: $g = \iota_W \circ F$.

3. Der Imaginärraum von $\mathbb{H}$. Wir benutzen die HAMILTONsche Standardbasis e, i, j, k. Der dreidimensionale Untervektorraum

$$\mathrm{Im}\,\mathbb{H} := \mathbb{R}i + \mathbb{R}j + \mathbb{R}k \tag{1}$$

von $\mathbb{H}$ heißt – in Analogie zu den komplexen Zahlen – der *Imaginärraum von* $\mathbb{H}$, seine Elemente heißen „rein-imaginär". Es gilt:

$$\mathbb{H} = \mathbb{R}e \oplus \mathrm{Im}\,\mathbb{H} \qquad \text{(direkte Summe von Vektorräumen).} \tag{2}$$

Hier ist die Gerade $\mathbb{R}e$ durch das Einselement e invariant definiert. Die Definition von $\mathrm{Im}\,\mathbb{H}$ ist zunächst basisabhängig. Um auch $\mathrm{Im}\,\mathbb{H}$ invariant zu charakterisieren, beachten wir, daß die Quaternion $x = \alpha e + \beta i + \gamma j + \delta k$ aufgrund von Satz 2 der quadratischen Gleichung

$$x^2 = 2\alpha x - (\alpha^2 + \beta^2 + \gamma^2 + \delta^2)e \tag{3}$$

genügt. Da $x \in \mathrm{Im}\,\mathbb{H}$ genau dann gilt, wenn $\alpha = 0$, so gewinnen wir die basisfreie Darstellung

$$\mathrm{Im}\,\mathbb{H} = \{x \in \mathbb{H} : x^2 \in \mathbb{R}e \text{ und } x \notin \mathbb{R}e \setminus \{0\}\}. \tag{4}$$

$\mathrm{Im}\,\mathbb{H}$ ist *keine* $\mathbb{R}$-Unteralgebra von $\mathbb{H}$. Wir notieren:

Für rein-imaginäre Quaternionen u, v *gilt* $u^2 = -\omega e$ *mit* $\omega \geqslant 0$ *und* $uv + vu \in \mathbb{R}e$.

Beweis. Für $u = \beta i + \gamma j + \delta k$ gilt $u^2 = -(\beta^2 + \gamma^2 + \delta^2)e$ mit $\beta^2 + \gamma^2 + \delta^2 \geqslant 0$. Da $u, v, u + v \in \mathrm{Im}\,\mathbb{H}$, so folgt $uv + vu = (u + v)^2 - u^2 - v^2 \in \mathbb{R}e$. □

Insbesondere gibt es zu jedem $u \in \mathrm{Im}\,\mathbb{H}$, $u \neq 0$, einen Skalar ρ (nämlich $\rho := \sqrt{\omega}$) mit $(\rho u)^2 = -e$ (Normierung).

Der Imaginärraum $\mathrm{Im}\,\mathbb{H}$ spielt in der Theorie der Quaternionen eine hervorragende Rolle, seine Elemente heißen auch *vektorielle* Quaternionen. Der Ausdruck „Vektor" erscheint bei HAMILTON zuerst 1845, Quaterly Journal 1, S. 56; den langandauernden Widerstand gegen die Vektorrechnung drückt Lord KELVIN noch 1896 so aus: „Vector is a useless survival, or offshoot from quaternions, and has never been of the slightest use to any creature."

Nach (2) läßt sich jede Quaternion x eindeutig schreiben

$$x = \alpha e + u \qquad \text{mit} \qquad \alpha \in \mathbb{R} \qquad \text{und} \qquad u \in \mathrm{Im}\,\mathbb{H}; \tag{5}$$

hier heißt αe manchmal der *skalare Anteil* (oder *Realteil*) und u der *vektorielle Anteil* (oder *Imaginärteil*) von x.

Für jedes Element $u \in \mathrm{Im}\,\mathbb{H}$ mit $u^2 = -e$ ist $\mathbb{R}e + \mathbb{R}u$ eine zu $\mathbb{C}$ isomorphe Unteralgebra von $\mathbb{H}$. Es ist aber grundsätzlich unmöglich, $\mathbb{H}$ „irgendwie" zu einer $\mathbb{C}$-Algebra zu machen*).

---

*) Es gilt ein besserer

Satz. *Jede endlich-dimensionale, komplexe Divisionsalgebra* $\mathcal{A}$ *mit Einselement ist zu* $\mathbb{C}$ *isomorph.*

Das beweist man analog wie den Satz in der Fußnote auf Seite 131: die Linksmultiplikation L_a hat jetzt einen *komplexen* Eigenwert λ (Fundamentalsatz der Algebra).

4. Quaternionenprodukt, Vektorprodukt und Skalarprodukt. Für vektorielle Quaternionen $u = \beta i + \gamma j + \delta k$, $v = \rho i + \sigma j + \tau k$ gilt:

$$uv = -(\beta\rho + \gamma\sigma + \delta\tau)e + (\gamma\tau - \delta\sigma)i + (\delta\rho - \beta\tau)j + (\beta\sigma - \gamma\rho)k.$$

Hier ist der „skalare Teil" bis auf das Vorzeichen das kanonische euklidische *Skalarprodukt* $\langle u, v\rangle$ der Vektoren $u = (\alpha, \beta, \gamma)$, $v = (\rho, \sigma, \tau) \in \mathbb{R}^3$; der „vektorielle Teil" von uv ist das *antikommutative vektorielle Produkt* $u \times v$ dieser Vektoren. Man gewinnt so die ästhetische Darstellung

$$uv = -\langle u, v\rangle e + u \times v \qquad \text{mit} \qquad u \times v \in \operatorname{Im}\mathbb{H}, \qquad u \times v = -v \times u.$$

Eine Verifikation ergibt unmittelbar

$$u \times v = \tfrac{1}{2}(uv - vu),$$

wodurch evident wird, daß der $\mathbb{R}^3$ zusammen mit dem Vektorprodukt eine LIE-Algebra ist (vgl. R.2, 3).

Historische Bemerkung. Die vektorielle Multiplikation wurde von H. GRASSMANN 1844 (ein Jahr nach HAMILTONS Erfindung der Quaternionen) als Spezialfall viel allgemeinerer sogenannter „äußerer Produkte" erfunden. Die Algebra der Vektoren des $\mathbb{R}^3$ wurde aber erst in den achtziger Jahren des vergangenen Jahrhunderts durch die Arbeiten des amerikanischen Physikers und Mathematikers Josiah Williard GIBBS (1839–1903, Professor an der Yale University) populär. GIBBS argumentierte u. a., – was für uns heute selbstverständlich ist – daß dem Skalarprodukt $\langle u, v\rangle$ und dem Vektorprodukt $u \times v$ eine selbständige Bedeutung zukomme, und daß das Quaternionenprodukt uv, welches diese beiden Produkte miteinander verknüpft, für viele Probleme unwesentlich ist. GIBBS war ein Gegner der Quaternionisten; aufgrund dieser Kontroverse gründete ein Kollege von GIBBS in Yale 1895 den in der Einleitung bereits erwähnten Weltbund zur Förderung der Quaternionen.

5. Zur Nichtkommutativität von $\mathbb{H}$. Zentrum. Die Tatsache, daß $\mathbb{H}$ nicht kommutativ ist, hat viele ungewohnte Konsequenzen. So können Polynome mehr Nullstellen haben als ihr Grad angibt: z. B. hat das quadratische Polynom $X^2 + e$ *alle* rein-imaginären Quaternionen $u = \beta i + \gamma j + \delta k$, deren „Länge" $\beta^2 + \gamma^2 + \delta^2$ eins ist, als Nullstellen, diese Quaternionen bilden die Oberfläche der Einheitskugel im 3-dimensionalen Raum $\mathbb{R}^3$ der Tripel (β, γ, δ).*)

Wir behaupten weiter:

Es gibt kubische Polynome über $\mathbb{H}$, *z. B.* $X^2iXi + iX^2iX - iXiX^2 - XiX^2i$, *die von allen Quaternionen annulliert werden.*

Da jede Quaternion einer Gleichung $X^2 = \alpha X + \beta e$ genügt, folgt die Behauptung durch Einsetzen von $\alpha X + \beta e$ anstelle von X^2 im obigen Polynom. □

*) Der Grund für dieses Phänomen ist, daß sich Polynome über $\mathbb{H}$ nicht mehr in der gewohnten Art faktorisieren: z. B. gilt $(X - x)(X - y) = X^2 - xX - Xy + xy$, und hier darf man die linearen Glieder nicht zu $-(x + y)X$ zusammenfassen!

Wegen der Nichtkommutativität lassen sich über $\mathbb{H}$ auch nicht mehr Determinanten naiv definieren. So ist z. B. weder

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} := ad - bc \qquad \text{noch} \qquad \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} := ad - cb$$

sinnvoll: im ersten bzw. zweiten Fall würde gelten

$$\det\begin{pmatrix} i & j \\ i & j \end{pmatrix} = ij - ji = 2k \neq 0 \qquad \text{bzw.} \qquad \det\begin{pmatrix} i & i \\ j & j \end{pmatrix} = ij - ji \neq 0,$$

was man nicht haben möchte, da Zeilen bzw. Spalten gleich sind.

Um für eine Algebra $\mathscr{A}$ die Abweichung von der Kommutativität zu messen, betrachtet man ihr *Zentrum*

$$Z(\mathscr{A}) := \{z \in \mathscr{A} : zx = xz \text{ für alle } x \in \mathscr{A}\}.$$

Ist $\mathscr{A}$ assoziativ, so ist $Z(\mathscr{A})$ eine *Unteralgebra* von $\mathscr{A}$; es gilt $Z(\mathscr{A}) = \mathscr{A}$ genau dann, wenn $\mathscr{A}$ kommutativ ist. Für Algebren mit Einselement e gilt $\mathbb{R}e \subset Z(\mathscr{A})$. Der Extremfall $Z(\mathscr{A}) = \mathbb{R}e$ ist möglich.

Für die Algebra $\mathbb{H}$ *gilt*: $Z(\mathbb{H}) = \mathbb{R}e = \{x \in \mathbb{H} : xu = ux \text{ für alle } u \in \operatorname{Im}\mathbb{H}\}$.

Dies ist enthalten in folgender Aussage:

Für alle $u \in \mathbb{H} \setminus \mathbb{R}e$ *gilt* $\{x \in \mathbb{H} : xu = ux\} = \mathbb{R}e + \mathbb{R}u$.

Beweis. Da $\{x \in \mathbb{H} : xu = ux\} = \{x \in \mathbb{H} : xv = vx\}$ für $v \in \operatorname{Im}\mathbb{H}$ mit $u - v \in \mathbb{R}e$, so darf man $u \in \operatorname{Im}\mathbb{H}$, $u \neq 0$, voraussetzen. Wir dürfen sogar $u^2 = -e$ und $x^2 = -e$ annehmen (man gehe von x zu $x - \alpha e$ über und normiere!). Dann folgt $(x - u)(x + u) = x^2 - ux + xu - u^2 = 0$, also $x = \pm u$. □

Die Nichtkommutativität von $\mathbb{H}$ ist auch der Grund dafür, daß $\mathbb{H}$ viele $\mathbb{R}$-Algebra-Automorphismen hat: jedes $a \in \mathbb{H}$, $a \neq 0$, induziert vermöge

$$h_a : \mathbb{H} \to \mathbb{H}, \qquad x \mapsto axa^{-1}$$

einen sogenannten *inneren Automorphismus*. Wegen $Z(\mathscr{A}) = \mathbb{R}e$ gilt $h_a = h_b$ genau dann, wenn $ab^{-1} \in \mathbb{R}e$. Wir werden in 3.2 zeigen, daß die $\mathbb{R}$-Algebra $\mathbb{H}$ keine weiteren Automorphismen hat.

Aufgabe. Zeigen Sie, daß für zwei Elemente $a, b \in \mathbb{H}$ die Gleichung $ab = ba$ genau dann gilt, wenn e, a, b linear abhängig sind.

6. Die Endomorphismen des $\mathbb{R}$-Vektorraumes $\mathbb{H}$. Für je zwei Quaternionen a, b ist $\mathbb{H} \to \mathbb{H}$, $x \mapsto axb$ eine $\mathbb{R}$-lineare Abbildung von $\mathbb{H}$ in sich (*Endomorphismus*). Wir bezeichnen mit $\operatorname{End}\mathbb{H}$ den $\mathbb{R}$-Vektorraum aller Endomorphismen von $\mathbb{H}$ und behaupten

Satz. *Ist* $a_1, \ldots, a_4$ *eine Basis von* $\mathbb{H}$, *so ist die Abbildung*

$$\mathbb{H}^4 \to \operatorname{End}\mathbb{H}, \qquad (b_1, b_2, b_3, b_4) \mapsto f \in \operatorname{End}\mathbb{H} \qquad \textit{mit} \qquad f(x) := \sum_1^4 a_\nu x b_\nu$$

$\mathbb{R}$-*linear und bijektiv.*

Beweis. Die $\mathbb{R}$-Linearität ist trivial. Da $\dim \mathbb{H}^4 = \dim(\mathrm{End}\,\mathbb{H}) = 16$ wegen $\dim \mathbb{H} = 4$, so ist nur die Injektivität der in Rede stehenden Abbildung zu zeigen. Das ist der Fall $n = 4$ folgender Hilfsaussage:

Sei $n = 1, 2, 3, 4$, *sei* $\sum_1^n a_\nu x b_\nu = 0$ *für alle* $x \in \mathbb{H}$. *Dann gilt*: $b_1 = \cdots = b_n = 0$.

Man schließt induktiv, der Fall $n = 1$ ist klar. Sei $n > 1$. Wäre $b_1 \neq 0$, so wäre

$$(*) \qquad a_1 x + \sum_2^n a_\nu x q_\nu = 0 \qquad \text{mit} \qquad q_\nu := b_\nu b_1^{-1}.$$

Multipliziert man hier zum einen von rechts mit y und setzt man zum anderen xy statt x, so entsteht durch Subtraktion

$$\sum_2^n \alpha_\nu x(q_\nu y - y q_\nu) = 0, \qquad \text{also} \qquad q_\nu y = y q_\nu \qquad \text{für alle} \qquad y \in \mathbb{H}$$

nach Induktionsannahme. Wegen $Z(\mathbb{H}) = \mathbb{R}e$ folgt $q_\nu = \alpha_\nu e$, $\alpha_\nu \in \mathbb{R}$. Aus $(*)$ wird jetzt

$$\left(a_1 + \sum_2^n \alpha_\nu a_\nu\right) x = 0, \qquad \text{das heißt,} \qquad a_1 + \sum_2^n \alpha_\nu a_\nu = 0,$$

das heißt, $a_1, \ldots, a_4$ wären linear abhängig. Somit gilt $b_1 = 0$, analog folgt $b_2 = b_3 = b_4 = 0$. □

Beispiel. Die Konjugierung $x \to \bar{x}$ (vgl. § 2, 1) gehört zu $\mathrm{End}\,\mathbb{H}$, bezüglich der Basis $1, i, j, k$ gilt:

$$\bar{x} = -\tfrac{1}{2}(x + ixi + jxj + kxk).$$

Der hier bewiesene Satz findet sich in HAMILTONS Werk „Elements of Quaternions", das 1866 von seinem Sohn herausgegeben wurde (vgl. z. B. S. 733 im Band 1 der 1882 erschienenen deutschen Übersetzung). Für den Körper $\mathbb{C}$ gilt das Analogon des Satzes nicht, hier haben (vgl. 3.3.1) die $\mathbb{R}$-linearen Abbildungen $\mathbb{C} \to \mathbb{C}$ die Form $z \mapsto az + b\bar{z}$. Daß man im Falle $\mathbb{H}$ ohne konjugierte Quaternionen auskommt, liegt daran, daß $\mathbb{H}$ ein 1-dimensionales Zentrum hat (was für $\mathbb{C}$ falsch ist).

7. Quaternionenmultiplikation und Vektoranalysis. HAMILTON hat die Quaternionenmultiplikation verwendet, um wichtige Formeln der Vektoranalysis elegant herzuleiten. Er führte den „Nabla"-Operator

$$\nabla := \frac{\partial}{\partial x} i + \frac{\partial}{\partial y} j + \frac{\partial}{\partial z} k$$

ein (die Bezeichnung „Nabla" wählte er wegen der Ähnlichkeit des Zeichens ∇ mit einem altjüdischen Musikinstrument dieses Namens). Anwendung von ∇ auf eine differenzierbare Funktion $f(x, y, z)$ dreier reeller Variabler gibt den *Gradienten* von f

$$\nabla f := \frac{\partial f}{\partial x} i + \frac{\partial f}{\partial y} j + \frac{\partial f}{\partial z} k = \mathrm{grad} f.$$

Anwendung von ∇ auf ein „differenzierbares Quaternionenfeld" $F(x, y, z) = u(x, y, z)i + v(x, y, z)j + w(x, y, z)k$ gibt, wenn man ∇F formal ausmultipliziert:

$$\nabla F = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right) + \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z}\right) i + \left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x}\right) j + \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) k;$$

der Realteil ist bis aufs Vorzeichen die *Divergenz* $\operatorname{div} F$, der Imaginärteil ist die *Rotation* $\operatorname{rot} F$ des Feldes F:

$$\nabla F = -\operatorname{div} F + \operatorname{rot} F.$$

Zweimalige Anwendung von ∇ auf eine Funktion f führt zum LAPLACE-*Operator* Δ der Potentialtheorie, genauer:

$$\nabla^2 f = -\Delta f = -\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}\right).$$

Das alles wirkt verblüffend. Felix KLEIN sagt dazu im 1. Band seiner „Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert" (S. 188): „Die Leichtigkeit und Eleganz ist in der Tat überraschend, und es läßt sich wohl von hier aus die alles andere ablehnende Begeisterung der Quaternionisten für ihr System begreifen, die bald über vernünftige Grenzen hinauswuchs, in einer weder der Mathematik als Ganzem noch der Quaternionentheorie selbst förderlichen Weise."

§ 2. Die Algebra $\mathbb{H}$ als euklidischer Vektorraum

Ist V ein *reeller* Vektorraum, so heißt eine *Bilinearform* $V \times V \to \mathbb{R}$, $(x, y) \mapsto \langle x, y\rangle$, ein *Skalarprodukt*, wenn sie *symmetrisch* und *positiv-definit* ist:

$$\langle x, y\rangle = \langle y, x\rangle \qquad \text{und} \qquad \langle x, x\rangle > 0 \qquad \text{für} \qquad x \neq 0.$$

V zusammen mit einem Skalarprodukt heißt ein *euklidischer* Vektorraum. Die Zahl $|x| := +\sqrt{\langle x, x\rangle} \geqslant 0$ heißt die (*euklidische*) *Länge* oder auch die *Norm* von $x \in V$; zwei Vektoren $x, y \in V$ heißen *orthogonal* (*stehen senkrecht aufeinander*), wenn $\langle x, y\rangle = 0$.

Ziel dieses Paragraphen ist, in der Quaternionen-Algebra $\mathbb{H}$ ein Skalarprodukt einzuführen, das gut mit der Multiplikation in $\mathbb{H}$ harmoniert. In $\mathbb{C} = \mathbb{R}^2$ ist $\langle w, z\rangle = \operatorname{Re}(w\bar{z})$ ein optimal mit der Multiplikation in $\mathbb{C}$ verträgliches Skalarprodukt, wie u. a. die Produktregel $|wz| = |w|\,|z|$ zeigt (vgl. 3.3.4). Wir werden sehen, daß analoges für $\mathbb{H} = \mathbb{R}^4$ gilt, wenn man für Quaternionen $x = \alpha e + \beta i + \gamma j + \delta k$, $x' = \alpha' e + \beta' i + \gamma' j + \delta' k \in \mathbb{H}$, das *kanonische Skalarprodukt*

$$\langle x, x'\rangle := \alpha\alpha' + \beta\beta' + \gamma\gamma' + \delta\delta' \in \mathbb{R} \tag{1}$$

einführt. Dann ist trivial, daß e, i, j, k eine *Orthonormalbasis von* $\mathbb{H}$ bilden. Wegen (1) ist die Länge $|x|$ von x gegeben durch

$$|x|^2 := \langle x, x\rangle = \alpha^2 + \beta^2 + \gamma^2 + \delta^2. \tag{2}$$

1. Konjugierung und Linearform Re. Nach 1.3(5) hat jede Quaternion x die basisinvariante Darstellung $x = \alpha e + u$, $u \in \operatorname{Im}\mathbb{H}$. Wir betrachten – in Analogie zu $\mathbb{C}$ – die $\mathbb{R}$-*lineare Konjugierung*

$$\mathbb{H} \to \mathbb{H}, \qquad x \mapsto \bar{x} := \alpha e - u. \tag{1}$$

Dann gilt

$$\bar{\bar{x}} = x, \qquad \operatorname{Im}\mathbb{H} = \{x \in \mathbb{H} : \bar{x} = -x\}, \tag{2}$$

und die *Fixpunktmenge* ist die *Gerade* $\mathbb{R}e$; weiter ist klar:

$$|\bar{x}| = |x|, \qquad x \in \mathbb{H} \qquad (\textit{Längentreue}). \tag{3}$$

Immer wieder benutzt wird die Multiplikationsregel

$$\overline{xy} = \bar{y}\bar{x}; \tag{4}$$

sie folgt z. B. aus der Produktformel 1.1, indessen braucht man (4) nur für die Basis e, i, j, k zu verifizieren; hieraus folgt schon die Allgemeingültigkeit, da die Abbildung $(x, y) \mapsto \overline{xy} - \bar{y}\bar{x}$ bilinear ist. Wegen (2) und (4) heißt die Abbildung $x \mapsto \bar{x}$ eine *Involution* der Quaternionen-Algebra $\mathbb{H}$.

Wir simulieren weiter die *Realteilbildung in* $\mathbb{C}$ und führen vermöge

$$\operatorname{Re} : \mathbb{H} \to \mathbb{R},\ x \mapsto \operatorname{Re}(x) := \alpha, \qquad \text{falls} \qquad x = \alpha e + u, \qquad u \in \operatorname{Im}\mathbb{H}, \tag{5}$$

eine $\mathbb{R}$-*Linearform* ein; ersichtlich ist Re durch die Eigenschaften

$$\operatorname{Re}(e) = 1 \qquad \text{und} \qquad \operatorname{Kern}\operatorname{Re} = \operatorname{Im}\mathbb{H}$$

charakterisiert. Per definitionem ist klar (Analogon zu 3.3.1(1)):

$$x + \bar{x} = 2\operatorname{Re}(x)e \qquad \text{und} \qquad \operatorname{Re}(\bar{x}) = \operatorname{Re}(x). \tag{6}$$

Die wichtige quadratische Gleichung (3) aus 1.3 schreibt sich nun als

$$x^2 = 2\operatorname{Re}(x)x - |x|^2 e. \tag{7}$$

Da $(x, y) \mapsto \operatorname{Re}(xy) - \operatorname{Re}(yx)$ bilinear ist, so gilt allgemein

$$\operatorname{Re}(xy) = \operatorname{Re}(yx), \tag{8}$$

da dies für e, i, j, k zutrifft. Es sei nebenbei erwähnt, daß $\operatorname{Re}(xy)$ die Bilinearform der *Lorentzmetrik* im $\mathbb{R}^4$ ist, denn es gilt (z. B. nach der Produktformel 1.1)

$$\operatorname{Re}(xy) = \alpha\alpha' - \beta\beta' - \gamma\gamma' - \delta\delta'$$

für

$$x = \alpha e + \beta i + \gamma j + \delta k, \qquad x' = \alpha' e + \beta' i + \gamma' j + \delta' k \in \mathbb{H}.$$

Aufgabe. 1) Zeigen Sie, daß für jedes Paar $u, v \in \operatorname{Im}\mathbb{H}$ mit $|u| = 1$ die Gleichung $xu = v$ durch $x := \langle u, v\rangle + u \times v$ gelöst wird.

2) Seien $a, b \in \mathbb{H}$, sei $\operatorname{Re}(a) \neq 0$. Zeigen Sie, daß $x := \dfrac{b + \bar{a}ba^{-1}}{4\operatorname{Re}(a)} \in \mathbb{H}$ eine Lösung der Gleichung $ax + xa = b$ ist.

Bemerkung. Die Beweise der Regeln (4) und (8) werden besser verstanden, wenn man den in 1.2 eingeführten Algebra-Isomorphismus

$$F : \mathbb{H} \to \mathscr{H}, \quad x = (\alpha, \beta, \gamma, \delta) \mapsto F(x) = \begin{pmatrix} w & -z \\ \bar{z} & \bar{w} \end{pmatrix}, \quad w := \alpha + i\beta, \quad z := \gamma + i\delta \in \mathbb{C},$$

heranzieht und in der Matrixalgebra $\mathscr{H}$ arbeitet. Es gilt nämlich (wenn A^t die zu A transponierte Matrix bedeutet):

$$F(\bar{x}) = \overline{F(x)}^t, \qquad \operatorname{Re} x = \tfrac{1}{2}\operatorname{Spur} F(x);$$

wohlvertraute Rechenregeln für Matrizen liefern nun

$$F(\overline{xy}) = \overline{F(xy)}^t = (\overline{F(x)F(y)})^t = (\overline{F(x)}\ \overline{F(y)})^t = \overline{F(y)}^t\,\overline{F(x)}^t = F(\bar{y})F(\bar{x}) = F(\bar{y}\bar{x}),$$

also $\overline{xy} = \bar{y}\bar{x}$ wegen der Injektivität von F. Wegen $\operatorname{Re}(xy) = \frac{1}{2}\operatorname{Spur}(F(x)F(y))$ folgt (8) unmittelbar aus der Kommutativität der Spur: $\operatorname{Spur}(AB) = \operatorname{Spur}(BA)$.

2. Eigenschaften des Skalarproduktes. In der Einleitung wurde das Skalarprodukt $\langle x, x'\rangle$ durch (1) bezüglich der Basis e, i, j, k von $\mathbb{H}$ definiert. Es ist leicht, basisunabhängige Beschreibungen mittels der Konjugierung zu geben. Man verifiziert zunächst die „Inversenregel"

(1) $\quad x\bar{x} = \bar{x}x = \langle x, x\rangle e, \qquad$ *speziell* $\qquad x^{-1} = |x|^{-2}\bar{x} \qquad$ *für* $\qquad x \neq 0.$

Schreibt man hier $x + y$ anstelle von x, so folgt

(2) $$\langle x, y\rangle e = \tfrac{1}{2}(x\bar{y} + y\bar{x})$$

durch Auflösen der Klammern (Verfahren der Linearisierung). Aus (2) folgt direkt:

Orthogonalitätskriterium: $\quad \langle x, y\rangle = 0 \Leftrightarrow x\bar{y} = -y\bar{x} \Leftrightarrow x\bar{y} \in \operatorname{Im}\mathbb{H}.$

Das Skalarprodukt in $\mathbb{C}$ wird durch $\operatorname{Re}(w\bar{z})$ gegeben. Die gleiche Formel gilt für $\mathbb{H}$:

(3) $\quad \langle x, y\rangle = \operatorname{Re}(x\bar{y}) = \operatorname{Re}(\bar{x}y), \qquad$ speziell $\qquad \langle x, e\rangle = \operatorname{Re}(x).$

Will man (3) nicht einfach aus der Produktformel in 1.1 ablesen, so schließe man wie folgt: Da $x\bar{y} + \overline{x\bar{y}} = 2\operatorname{Re}(x\bar{y})e$ nach 1.(6), und da $\overline{x\bar{y}} = y\bar{x}$ nach 1.(4), so folgt (3) aus (2). – Ein zweiter Beweis für (3) besteht darin, daß man die Abbildung $\mathbb{H} \times \mathbb{H} \to \mathbb{R}$, $(x, y) \mapsto \operatorname{Re}(x\bar{y})$, als bilinear, symmetrisch und positiv definit nachweist und bemerkt, daß e, i, j, k eine Orthonormalbasis bilden. □

Fundamental ist

(4) $$|xy| = |x|\,|y| \qquad (\textit{Produktregel}).$$

Beweis. Mit (1), 1.(4), dem Assoziativgesetz und erneuter zweimaliger Anwendung von (1) hat man:

$$|xy|^2 e = \langle xy, xy\rangle e = (\overline{xy})(xy) = \bar{y}(\bar{x}x)y = \langle x, x\rangle \bar{y}y = \langle x, x\rangle\langle y, y\rangle e = |x|^2|y|^2 e.$$

□

Wir beweisen schließlich noch eine Formel, die in 3.2 hilfreich sein wird, und die in überraschender Weise dreifache Produkte der Form yxy als Linearkombination von y und $\bar{x}$ ausdrückt:

(5) $\quad yxy = 2\langle \bar{x}, y\rangle y - \langle y, y\rangle \bar{x}, \qquad x, y \in \mathbb{H} \qquad$ (*Dreier-Identität*).

Beweis. Aufgrund von (2) gilt: $2\langle y, \bar{x}\rangle e = \bar{x}\bar{y} + yx$. Rechtsmultiplikation mit y gibt wegen $\bar{y}y = \langle y, y\rangle e$ die Behauptung. □

Bemerkung. Betrachtet man in der Algebra $\mathscr{H}$ die Abbildung

$$\mathscr{H} \times \mathscr{H} \to \mathbb{R}, \qquad (A, B) \mapsto \langle A, B\rangle := \tfrac{1}{2}\mathrm{Spur}(A\bar{B}^t),$$

so ist aufgrund der Bemerkung im Abschnitt 1 klar, daß für den Algebra-Isomorphismus $F: \mathbb{H} \to \mathscr{H}$ gilt: $\langle F(x), F(y)\rangle = \mathrm{Re}(x\bar{y})$. Formel (3) besagt daher $\langle F(x), F(y)\rangle = \langle x, y\rangle$; dies bedeutet, daß $\langle A, B\rangle$ ein Skalarprodukt in $\mathscr{H}$ ist (was sich natürlich auch direkt verifizieren läßt) und daß $F: \mathbb{H} \to \mathscr{H}$ eine orthogonale Abbildung ist (zu diesem Begriff siehe 3.1). Da (!) $\mathrm{Spur}(\bar{A}A^t) = 2 \det A$, so gilt $\det F(x) = |x|^2$; damit übersetzt sich die Produktregel (4) in die Produktregel für Determinanten.

3. Der „Vier-Quadrate-Satz". Aus der Produktregel für $\mathbb{C}$ wurde in 3.3.4 der „Zwei-Quadrate-Satz" abgeleitet. Aus der Produktregel für $\mathbb{H}$ gewinnen wir analog den berühmten

Vier-Quadrate-Satz. *Für alle* $\alpha, \beta, \gamma, \delta, \alpha', \beta', \gamma', \delta' \in \mathbb{R}$ *gilt*:

$$\begin{aligned}(\alpha^2 + \beta^2 + \gamma^2 + \delta^2)(\alpha'^2 + \beta'^2 + \gamma'^2 + \delta'^2)\\ = (\alpha\alpha' - \beta\beta' - \gamma\gamma' - \delta\delta')^2 + (\alpha\beta' + \beta\alpha' + \gamma\delta' - \delta\gamma')^2\\ + (\alpha\gamma' + \gamma\alpha' + \delta\beta' - \beta\delta')^2 + (\alpha\delta' + \delta\alpha' + \beta\gamma' - \gamma\beta')^2.\end{aligned}$$

Beweis. Man benutze die Produktregel 2.(4) und die Produktformel aus 1.1. □

Der „Vier-Quadrate-Satz" wurde 1748 von EULER entdeckt (Brief an GOLDBACH vom 4. Mai; vgl. Correspondance entre LEONHARD EULER et Chr. Goldbach 1729–1763, in „Correspondance Mathématique et Physique de quelques célèbres géomètres du XVIII$^{\text{ème}}$ siècle" ed. P.-H. FUSS, St.-Pétersbourg 1843, Bd. 1, S. 452). EULER bemühte sich, den bereits 1659 von P. FERMAT ausgesprochenen Satz, daß jede natürliche Zahl Summe von vier Quadraten natürlicher Zahlen ist, zu beweisen; durch seinen Vier-Quadrate-Satz reduzierte er diesen Satz auf Primzahlen. Den ersten vollständigen Beweis des Satzes von FERMAT gab 1770 J. L. LAGRANGE (näheres hierzu findet man im Band „Von Fermat bis Minkowski" von W. SCHARLAU und H. OPOLKA, Springer-Verlag 1980). □

GAUSS bemerkte (nachgelassenes Manuskript, Werke 3, 383/4), daß bei Benutzung komplexer Zahlen der „Vier-Quadrate-Satz" in der Identität

$$(|u|^2 + |v|^2)(|w|^2 + |z|^2) = |uw + vz|^2 + |u\bar{z} - v\bar{w}|^2, \qquad u, v, w, z \in \mathbb{C},$$

enthalten ist; dies ist nichts anderes als der Determinantenproduktsatz für die Matrizen $\begin{pmatrix} u & v \\ -\bar{v} & \bar{u} \end{pmatrix}$ und $\begin{pmatrix} w & -\bar{z} \\ z & \bar{w} \end{pmatrix}$ aus $\mathscr{H}$.

HAMILTON hat, wie in der Einleitung dieses Kapitels herausgestellt wurde, den „Vier-Quadrate-Satz" zum Prüfstein für den Wert seiner Quaternionen erhoben. Kennt man erst einmal die Vier-Quadrate-Formel, so ist klar, daß sie wie die Zwei-Quadrate-Formel in *jedem kommutativen* Ring richtig ist.

4. Konjugierungs- und Längentreue von Automorphismen. Das bereits deutlich gewordene gute Zusammenspiel zwischen Skalarprodukt, Konjugierung und Multiplikation in $\mathbb{H}$ wird noch unterstrichen durch folgenden

Satz. *Jeder $\mathbb{R}$-Algebra-Automorphismus $h: \mathbb{H} \to \mathbb{H}$ ist „konjugierungs- und längentreu", das heißt, es gilt:*

$$h(\bar{x}) = \overline{h(x)}, \qquad |h(x)| = |x|, \qquad x \in \mathbb{H}. \tag{1}$$

Beweis. Wegen $h(e) = e$ und $\operatorname{Im}\mathbb{H} = \{x \in \mathbb{H} : x^2 = -\omega e \text{ mit } \omega \geqslant 0\}$ gilt $h(\operatorname{Im}\mathbb{H}) \subset \operatorname{Im}\mathbb{H}$. Für $x = \alpha e + u \in \mathbb{H}$, $\alpha \in \mathbb{R}$, $u \in \operatorname{Im}\mathbb{H}$, folgt also $h(x) = \alpha e + h(u)$ mit $h(u) \in \operatorname{Im}\mathbb{H}$. Dies impliziert: $\overline{h(x)} = \alpha e - h(u) = h(\alpha e - u) = h(\bar{x})$. Aus $h(x)\overline{h(x)}e = |h(x)|^2 e$ folgt nun $|h(x)| = |x|$ wegen $h(x)\overline{h(x)} = h(x\bar{x}) = |x|^2 e$. □

Im eben bewiesenen Satz wird die Bijektivität von h nirgends benutzt. Die Aussage gilt in der Tat für alle $\mathbb{R}$-Algebra-Endomorphismen $h \neq 0$ von $\mathbb{H}$, denn dann gilt stets $\operatorname{Ker} h = 0$, da $\mathbb{H}$ Divisionsalgebra ist, und weiter $h(e) = e$. – Obiger Satz wird in 3.2 benutzt zum Nachweis, daß alle Automorphismen von $\mathbb{H}$ die Form axa^{-1}, $a \neq 0$, haben. Für die $\mathbb{R}$-Algebra $\mathbb{C}$ ist die analoge Aussage des Satzes trivial, da $\mathbb{C}$ nur die Identität und die Konjugierung als $\mathbb{R}$-Automorphismen hat (vgl. 3.3.2).

5. Die Gruppe S^3 der Quaternionen der Länge 1. Wie bei komplexen Zahlen (vgl. 3.3.4) liefert die Produktregel unmittelbar

Die Menge $S^3 := \{x \in \mathbb{H} : |x| = 1\}$ aller Quaternionen der Länge 1 ist bezüglich der Multiplikation in $\mathbb{H}$ eine Untergruppe von $(\mathbb{H} \setminus \{0\}, \cdot)$.

Es gilt $e, i, j, k \in S^3$; daher ist die Gruppe S^3 *nicht abelsch*. Im $\mathbb{R}^4 \simeq \mathbb{H}$ ist S^3 die „Oberfläche der Einheitskugel" um den Nullpunkt. S^3 ist *kompakt*, man nennt S^3 auch die *dreidimensionale Sphäre*; topologisch entsteht S^3 aus dem Anschauungsraum $\mathbb{R}^3$ durch Kompaktifizierung mittels *eines unendlich fernen Punktes*. Die Gruppe S^3 wird im nächsten Paragraphen beim Studium der orthogonalen Abbildungen der Räume $\mathbb{H}$ und $\operatorname{Im}\mathbb{H}$ eine zentrale Rolle spielen.

Im Raum $\mathbb{R}^{n+1}$ der $(n+1)$-Tupel $x = (\xi_0, \ldots, \xi_n)$, $y = (\eta_0, \ldots, \eta_n)$ mit dem Skalarprodukt $\langle x, y\rangle = \sum_0^n \xi_\nu \eta_\nu$ definiert man die „n-dimensionale Sphäre" durch $S^n := \{x \in \mathbb{R}^{n+1} : |x| = 1\}$; ein nichttrivialer Satz besagt, daß S^1 und S^3 die einzigen Sphären mit einer „stetigen" Gruppenstruktur sind.

Für jede Quaternion $x \neq 0$ gilt $x\bar{x}^{-1} \in S^3$. Man verifiziert direkt:

Die Abbildung $h: \mathbb{H} \setminus \{0\} \to S^3$, $x \mapsto x\bar{x}^{-1} = |x|^{-2}x^2$, ist surjektiv:

$$h\left(e + \frac{b}{1+\alpha}\right) = \alpha e + b, \quad \textit{falls} \quad \alpha e + b \in S^3 \setminus \{-e\}, \quad \alpha \in \mathbb{R}, \quad b \in \operatorname{Im}\mathbb{H}; \tag{1}$$

$$h(i) = -e. \qquad \square$$

Setzt man $x = \kappa e + b$, $b \in \operatorname{Im}\mathbb{H}$, so gilt $b^2 = -|b|^2 e$ und also

$$h(x) = \frac{\kappa^2 - |b|^2}{\kappa^2 + |b|^2} e + \frac{2\kappa}{\kappa^2 + |b|^2} b; \tag{2}$$

wir haben somit für die Gruppe S^3 folgende *Parameterdarstellung*

$$(3) \qquad S^3 = \left\{ \frac{1}{\kappa^2 + |b|^2} [(\kappa^2 - |b|^2)e + 2\kappa b] : (\kappa, b) \in (\mathbb{R} \times \operatorname{Im} \mathbb{H}) \setminus \{0\} \right\}$$

als Verallgemeinerung der Parameterdarstellung 3.5.4(2′) der Kreisgruppe S^1. Die Gleichungen (1) und (2) liefern überdies (Beweis!):

Jede „rationale" Quaternion $\alpha e + \beta_1 i + \beta_2 j + \beta_3 k \in S^3 \setminus \{e\}$, $\alpha, \beta_\nu \in \mathbb{Q}$, *hat die Form*

$$(4) \quad \alpha = \frac{1 - q^2}{1 + q^2}, \qquad \beta_\nu = \frac{2q_\nu}{1 + q^2} \qquad \text{mit} \qquad q_\nu := \frac{\beta_\nu}{1 + \alpha} \in \mathbb{Q}, \qquad 1 \leqslant \nu \leqslant 3,$$

$$q^2 := q_1^2 + q_2^2 + q_3^2 \in \mathbb{Q}.$$

Diese Darstellung kann man benutzen, um in Analogie zu 3.5.4 *pythagoräische Quintupel* – das sind 5-tupel (k, l, m, n, p) natürlicher Zahlen $\neq 0$ mit $k^2 + l^2 + m^2 + n^2 = p^2$ – zu parametrisieren. Der interessierte Leser möge die einfachen Rechnungen durchführen.

6. Die spezielle unitäre Gruppe $SU(2)$ und der Isomorphismus $S^3 \to SU(2)$. Die Menge

$$(1) \qquad U(2) := \{U \in GL(2, \mathbb{C}) : U\bar{U}^t = E\}$$

aller *unitären* 2×2 Matrizen ist eine wichtige Untergruppe der Gruppe $GL(2, \mathbb{C})$ aller komplexen, *invertierbaren* 2×2 Matrizen. Wegen $\det \bar{A}^t = \overline{\det A}$ gilt $|\det U| = 1$ für alle $U \in U(2)$. Durch

$$SU(2) := \{U \in U(2) : \det U = 1\}$$

wird ein *Normalteiler von* $U(2)$, die sogenannte *spezielle unitäre Gruppe*, definiert. Mit der in 1.2 definierten Unteralgebra $\mathscr{H}$ von $\operatorname{Mat}(2, \mathbb{C})$ gilt der

Satz. $SU(2) = \{A \in \mathscr{H} : \det A = 1\}$, *speziell* $SU(2) \subset \mathscr{H}$.

Beweis. Für $A = \begin{pmatrix} w & -z \\ \bar{z} & \bar{w} \end{pmatrix} \in \mathscr{H}$ verifiziert man sofort $A\bar{A}^t = (\det A) \cdot E$; hieraus folgt die Inklusion $\{A \in \mathscr{H} : \det A = 1\} \subset SU(2)$.

Für $U = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SU(2)$ gilt $U^{-1} = \bar{U}^t = \begin{pmatrix} \bar{a} & \bar{c} \\ \bar{b} & \bar{d} \end{pmatrix}$ nach (1). Da andererseits $U^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$, so folgt $d = \bar{a}$, $c = -\bar{b}$, das heißt, $U \in \mathscr{H}$. □

Es ergibt sich nun sofort

Isomorphiesatz. *Der Algebra-Isomorphismus* $F : \mathbb{H} \to \mathscr{H}$ *bildet die Gruppe* S^3 *aller Quaternionen der Länge* 1 *isomorph auf die spezielle unitäre Gruppe* $SU(2)$ *ab.*

Beweis. Wegen $|x|^2 = \det F(x)$ gilt $F(S^3) = \{A \in \mathscr{H} : \det A = 1\} = SU(2)$. □

Aus 5.(3) folgt nun eine „EULERsche Parameterdarstellung" der Gruppe

$$SU(2) = \left\{ \frac{1}{\kappa^2 + \lambda^2 + \mu^2 + \nu^2} \begin{pmatrix} \kappa^2 - \lambda^2 - \mu^2 - \nu^2 + 2\kappa\lambda i & 2\kappa\mu + 2\kappa\nu i \\ -2\kappa\mu + 2\kappa\nu i & \kappa^2 - \lambda^2 - \mu^2 - \nu^2 - 2\kappa\lambda i \end{pmatrix} \right\},$$

hier durchlaufen $(\kappa, \lambda, \mu, \nu)$ alle reellen Quadrupel $\neq 0$.

Der Leser vergleiche die Betrachtungen dieses Abschnittes mit den Überlegungen aus 3.5.3 und 3.5.4.

§ 3. Die orthogonalen Gruppen $O(3)$, $O(4)$ und Quaternionen

HAMILTON hat viele Jahre versucht, eine algebraische Struktur auf dem Anschauungsraum zu finden, mit deren Hilfe man die euklidische Geometrie des $\mathbb{R}^3$ besser verstehen kann. Wir haben gesehen, daß die Struktur einer Divisionsalgebra erst im umfassenden Raum $\mathbb{R}^4$ realisiert werden kann und daß zwischen Quaternionenmultiplikation und natürlichem Skalarprodukt im $\mathbb{R}^4$ interessante Zusammenhänge bestehen. Es zeigt sich nun, daß man mit den „rein-imaginären Quaternionen" auch die Drehungen des $\mathbb{R}^3$ elegant als Quaternionenmultiplikation beschreiben kann: schon 1844, also ein Jahr nach der Erfindung der Quaternionen, wußten HAMILTON und CAYLEY, daß *jede eigentlich orthogonale Abbildung von* $\mathbb{R}^3 = \operatorname{Im}\mathbb{H}$ die Gestalt

$$\operatorname{Im}\mathbb{H} \to \operatorname{Im}\mathbb{H}, \qquad u \mapsto aua^{-1},$$

hat, wo a alle Quaternionen $\neq 0$ durchläuft (vgl. HAMILTON: *Quaternions: Applications in Geometry*, Math. Papers 3, 353–362, insbesondere Formel (i') in der Fußnote auf S. 361; und CAYLEY: *On certain results relating to quaternions*, Math. Papers 1, 123–126). CAYLEY selbst schreibt die Priorität HAMILTON zu: „the discovery of the formula $q(ix + jy + kz)q^{-1} = ix' + jy' + kz'$, as expressing a rotation, was made by Sir W. R. HAMILTON some months previous to the date of this paper" (vgl. Math. Papers 1, S. 586).

1855 bemerkte CAYLEY in einer im Crelleschen Journal Bd. 50 erschienenen Arbeit (S. 312, Math. Papers 2, S. 214), daß *jede eigentlich orthogonale Abbildung von* $\mathbb{R}^4 = \mathbb{H}$ von der Form

$$\mathbb{H} \to \mathbb{H}, \qquad x \mapsto \frac{axb}{|a|\,|b|},$$

ist, wo a, b unabhängig voneinander alle Quaternionen $\neq 0$ durchlaufen. Im folgenden werden diese Sätze von HAMILTON und CAYLEY eingehend diskutiert. Wir behandeln – abweichend vom vielfach üblichen Vorgehen – zunächst die Situation im $\mathbb{R}^4 = \mathbb{H}$ und erhalten dann den vielleicht mehr interessierenden Fall des $\mathbb{R}^3$ geschenkt durch die natürliche *Einbettung* von $\mathbb{R}^3 = \operatorname{Im}\mathbb{H}$ in $\mathbb{H}$.

1. Orthogonale Gruppen. Es bezeichnet V einen *endlich-dimensionalen euklidischen* Vektorraum. Eine lineare Abbildung $f: V \to V$ heißt *orthogonal, wenn*

$$\langle f(x), f(y)\rangle = \langle x, y\rangle \qquad \text{für} \qquad x, y \in V;$$

dies trifft genau dann zu, wenn f *längentreu* ist: $|f(x)| = |x|$ für $x \in V$. Jede orthogonale Abbildung ist bijektiv, ihre Umkehrabbildung ist wieder orthogonal. Die orthogonalen Abbildungen von V bilden bei Komposition eine Gruppe $O(V)$; man nennt $O(V)$ *die orthogonale Gruppe des euklidischen Vektorraumes* V.

Jeder Endomorphismus $f: V \to V$ besitzt eine Determinante. Es gilt

$$\det f = \pm 1 \qquad \text{für} \qquad f \in O(V).$$

Man definiert die *Untergruppe der eigentlich orthogonalen Abbildungen durch*

$$SO(V) = O^+(V) := \{f \in O(V) : \det f = 1\};$$

die *Nebenklasse der Spiegelungen* ist erklärt durch

$$O^-(V) := \{f \in O(V) : \det f = -1\}.$$

Es gilt $O(V) = O^+(V) \cup O^-(V)$.

Die Gruppen $O(\mathbb{R}^n)$ und $SO(\mathbb{R}^n)$ des euklidischen Zahlenraumes $\mathbb{R}^n$ werden traditionell mit $O(n)$ und $SO(n)$ bezeichnet und meistens mit den Matrixgruppen $\{A \in GL(n, \mathbb{R}) : A^t A = E\}$ bzw. $\{A \in GL(n, \mathbb{R}) : A^t A = E \text{ und } \det A = 1\}$ identifiziert.

Eine besondere Rolle spielen die Abbildungen

$$s_a : V \to V, \qquad x \mapsto x - 2\langle a, x\rangle a, \qquad a \in V, \qquad |a| = 1.$$

s_a ist stets orthogonal: nämlich *die Spiegelung an der auf der Geraden* $\mathbb{R}a$ *senkrecht stehenden Hyperebene* $\{x \in V : \langle a, x\rangle = 0\}$. Es gilt

1) $s_a \in O^-(V)$, $s_a^2 = \mathrm{id}$, $f \circ s_a \in O^+(V)$ *für* $f \in O^-(V)$.
2) $f \circ s_a = s_{f(a)} \circ f$ *für* $f \in O(V)$.

Aus Grundwissen Mathematik 2, Lineare Algebra und Analytische Geometrie, 6.1.5, entnehmen wir den

Erzeugungssatz für orthogonale Gruppen. *Die Gruppe* $O(V)$ *wird von ihren Spiegelungen erzeugt. Die Abbildungen* $f \in SO(V)$ *sind (genau die) Produkte aus einer geraden Anzahl* k *von Spiegelungen, wobei* $k \leqslant \dim V$.

2. Die Gruppe $O(\mathbb{H})$. Satz von CAYLEY. Jede Abbildung

$$\mathbb{H} \to \mathbb{H}, \quad x \mapsto axb, \qquad \mathbb{H} \to \mathbb{H}, \quad x \mapsto a\bar{x}b, \qquad a, b \in S^3,$$

ist aufgrund der Produktregel 2.2.(4) orthogonal. Um zu zeigen, daß dies bereits alle orthogonalen Abbildungen von $\mathbb{H}$ sind, ziehen wir die Spiegelungen $s_a : \mathbb{H} \to \mathbb{H}$ heran. Wegen $2\langle e, x\rangle e = x + \bar{x}$ folgt

$$s_e(x) = -\bar{x}. \tag{1}$$

Wir bezeichnen mit p_a die spezielle Abbildung $x \mapsto axa$, $a \in S^3$. Es gilt

$$p_a = s_a \circ s_e, \tag{2}$$

da $s_a \circ s_e(x) = -s_a(\bar{x}) = -\bar{x} + 2\langle a, \bar{x}\rangle a = axa$ nach der Dreieridentität 2.2.(5). Wir zeigen weiter:

$$s_a \circ s_b = p_a \circ p_{-b} \qquad \textit{für alle} \qquad a, b \in S^3. \tag{3}$$

Beweis. Da $f \circ s_b = s_{f(b)} \circ f$ für $f \in O(\mathbb{H})$, so folgt wegen $s_e^2 = \mathrm{id}$:

$$s_a \circ s_b = (s_a \circ s_e) \circ (s_e \circ s_b) = (s_a \circ s_e) \circ (s_c \circ s_e) \qquad \text{mit} \qquad c := s_e(b) = -\bar{b}.$$

Dies ist wegen (2) die Behauptung. □

Aus (1)–(3) ergibt sich aufgrund des Erzeugungssatzes 1 sofort

Erzeugungssatz für $O(\mathbb{H})$. *Jede orthogonale Abbildung $f \in SO(\mathbb{H})$ ist ein Produkt aus höchstens vier Abbildungen p_a, $a \in S^3$.*

Die Gruppe $O(\mathbb{H})$ wird von den Abbildungen $x \mapsto axa$, $a \in S^3$, und der Konjugierung $x \mapsto \bar{x}$ erzeugt.

Hieraus folgt unmittelbar:

Satz (CAYLEY). *Zu jeder orthogonalen Abbildung $f: \mathbb{H} \to \mathbb{H}$ gibt es zwei Quaternionen $a, b \in S^3$ mit folgender Eigenschaft:*

a) *Es gilt $f(x) = axb$, falls $f \in O^+(\mathbb{H})$.*
b) *Es gilt $f(x) = a\bar{x}b$, falls $f \in O^-(\mathbb{H})$.*

Beweis. a) Für $f \in O^+(\mathbb{H})$ gilt $f = p_{a_1} \circ \cdots \circ p_{a_4}$ mit $a_1, \ldots, a_4 \in S^3$. Setzt man $a := a_1 a_2 a_3 a_4$, $b := a_4 a_3 a_2 a_1$, so folgt $a, b \in S^3$ und $f(x) = axb$.

b) Für $f \in O^-(\mathbb{H})$ gilt $f \circ s_e \in O^+(\mathbb{H})$, also $f(-\bar{x}) = f \circ s_e(x) = cxb$ mit $b, c \in S^3$ nach a). Wir sehen $f(x) = a\bar{x}b$ mit $a := -c$. □

Aus dem Satz von CAYLEY folgt der bereits in 1.5 angekündigte

Satz. *Jeder $\mathbb{R}$-Algebra-Automorphismus $h: \mathbb{H} \to \mathbb{H}$ hat die Form $h(x) = axa^{-1}$, $a \in S^3$.*

Beweis. Nach Satz 2.4 gilt $h \in O(\mathbb{H})$. Da $h(e) = e$, so folgt also

$$h(x) = axa^{-1} \qquad \text{oder} \qquad h(x) = a\bar{x}a^{-1} \qquad \text{mit} \qquad a \in S^3.$$

Der zweite Fall ist nicht möglich, da dann $h(xy) = \overline{axya^{-1}} = a\bar{y}a^{-1}a\bar{x}a^{-1} = h(y)h(x)$.

3. Die Gruppe $O(\mathrm{Im}\,\mathbb{H})$. Satz von HAMILTON. Jede orthogonale Abbildung $\mathbb{H} \to \mathbb{H}$, $x \mapsto \pm axa$, $a \in S^3$, bildet den Unterraum $\mathrm{Im}\,\mathbb{H} = \{u \in \mathbb{H} : \bar{u} = -u\}$ aller rein imaginären Quaternionen wegen $\overline{aua} = a\bar{u}a = -aua$ in sich ab und induziert somit eine orthogonale Abbildung $\mathrm{Im}\,\mathbb{H} \to \mathrm{Im}\,\mathbb{H}$, $u \mapsto \pm aua$. Wir behaupten, daß alle orthogonalen Abbildungen von $\mathrm{Im}\,\mathbb{H}$ so gewonnen werden: da der Raum $\mathrm{Im}\,\mathbb{H}$ senkrecht auf der Geraden $\mathbb{R}e$ steht, so wird jede orthogonale Abbildung f von $\mathrm{Im}\,\mathbb{H}$ durch die Festlegung

$$\hat{f} := \mathrm{id} \quad \text{auf} \quad \mathbb{R}e, \qquad \hat{f} := f \quad \text{auf} \quad \mathrm{Im}\,\mathbb{H}$$

eindeutig zu einer orthogonalen Abbildung $\hat{f}: \mathbb{H} \to \mathbb{H}$ fortgesetzt. In Matrix-Schreibweise gehört zu $\hat{f}$ die Matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & B \end{smallmatrix}\right)$, falls die 3×3 Matrix B zu f gehört.

Damit ist klar:

$$\det \hat{f} = \det f, \qquad \textit{speziell:} \qquad f \in O^+(\operatorname{Im}\mathbb{H}) \Leftrightarrow \hat{f} \in O^+(\mathbb{H}).$$

Aus dem Satz von CAYLEY folgt nun mühelos:

Satz (HAMILTON). *Zu jeder orthogonalen Abbildung* $f: \operatorname{Im}\mathbb{H} \to \operatorname{Im}\mathbb{H}$ *gibt es eine Quaternion* $a \in S^3$ *mit folgender Eigenschaft:*

a) *Es gilt* $f(u) = au\bar{a}$, *falls* $f \in O^+(\operatorname{Im}\mathbb{H})$.
b) *Es gilt* $f(u) = -au\bar{a}$, *falls* $f \in O^-(\operatorname{Im}\mathbb{H})$.

Beweis. a) Sei $f \in O^+(\operatorname{Im}\mathbb{H})$. Dann gilt $\hat{f} \in O^+(\mathbb{H})$, so daß aus Satz 2, a) folgt: $\hat{f}(x) = axb$ mit $a, b \in S^3$. Aus $\hat{f}(e) = e$ folgt $ab = e$, das heißt, $b = a^{-1} = \bar{a}$.

b) Klar nach a), denn aus $f \in O^-(\operatorname{Im}\mathbb{H})$ folgt $-f \in O^+(\operatorname{Im}\mathbb{H})$, da $\operatorname{Im}\mathbb{H}$ die Dimension 3 hat. □

Aufgrund der Dreier-Identität 2.2.(5) gilt

$$bub = u - 2\langle u, b\rangle b = s_b(u), \qquad b \in S^3, \qquad u \in \operatorname{Im}\mathbb{H}.$$

Da jede eigentlich orthogonale Abbildung des 3-dimensionalen Raumes $\operatorname{Im}\mathbb{H}$ Produkt von zwei Spiegelungen s_b, s_c, $b, c \in \operatorname{Im}\mathbb{H} \cap S^3$ ist, so gibt es also zu jedem $a \in S^3$ Elemente $b, c \in \operatorname{Im}\mathbb{H} \cap S^3$, so daß gilt:

$$au\bar{a} = s_c(s_b(u)) = cbubc, \qquad u \in \operatorname{Im}\mathbb{H}.$$

Hieraus folgt (wenn man die Gleichung $\operatorname{Kern}\varphi = \{\pm e\}$ des nächsten Abschnitts benutzt):

Jede Quaternion $a \in \mathbb{H}$ *ist (auf unendlich viele Weisen) darstellbar als Produkt* $a = bc$ *zweier rein-imaginärer Quaternionen.*

4. Die Epimorphismen $S^3 \to SO(3)$ und $S^3 \times S^3 \to SO(4)$. Die Sätze von HAMILTON und CAYLEY liefern wichtige Informationen über die klassischen Gruppen $SO(3)$ und $SO(4)$. Wir ordnen jedem $a \in S^3$ bzw. jedem Paar $(a, b) \in S^3 \times S^3$ die orthogonale Abbildung

$$\varphi(a): \operatorname{Im}\mathbb{H} \to \operatorname{Im}\mathbb{H}, \quad u \mapsto au\bar{a}, \qquad \text{bzw.} \qquad \psi(a,b): \mathbb{H} \to \mathbb{H}, \quad x \mapsto ax\bar{b},$$

zu und betrachten die Abbildungen $\varphi: S^3 \to O(\operatorname{Im}\mathbb{H})$, $\psi: S^3 \times S^3 \to O(\mathbb{H})$. Mit S^3 ist auch das kartesische Produkt $S^3 \times S^3$ bezüglich der Verknüpfung $(a, b) \cdot (c, d) = (ac, bd)$ eine *nicht-abelsche, kompakte Gruppe.* Wir behaupten:

Satz. *Die Abbildungen* $\varphi: S^3 \to O(\operatorname{Im}\mathbb{H})$ *und* $\psi: S^3 \times S^3 \to O(\mathbb{H})$ *sind Gruppenhomomorphismen*). Die Kerngruppen haben 2 Elemente:* $\operatorname{Kern}\varphi = \{\pm e\}$, $\operatorname{Kern}\psi = \{\pm(e, e)\}$; *für die Bildgruppen gilt:* $\varphi(S^3) = SO(\operatorname{Im}\mathbb{H})$, $\psi(S^3 \times S^3) = SO(\mathbb{H})$.

Beweis. Die Homomorphieeigenschaft ergibt sich in beiden Fällen direkt, z. B. gilt $\psi((a, b) \cdot (c, d)) = \psi(a, b) \circ \psi(c, d)$ wegen

$$[\psi(a,b) \circ \psi(c,d)](x) = \psi(a,b)(cx\bar{d}) = acx\bar{d}\bar{b} = (ac)x(\overline{bd}) = \psi(ac, bd)(x), \; x \in \mathbb{H}^{*)}.$$

---

*) Da die Multiplikation in S^3 und $S^3 \times S^3$ *nicht abelsch* ist, würde man keinen Homomorphismus erhalten, wenn man $a \in S^3$ bzw. $(a, b) \in S^3 \times S^3$ die Isometrie $u \mapsto aua$ bzw. $x \mapsto axb$ zuordnet. Man beachte, daß $\bar{a} = a^{-1}$ für $a \in S^3$.

Sei $a \in \operatorname{Kern}\varphi$, also $u = aua\bar{}$ für alle $u \in \operatorname{Im}\mathbb{H}$. Nach 1.5 trifft dies genau im Fall $a \in \mathbb{R}e$ zu. Wegen $|a| = 1$ folgt $a = \pm e$, also $\operatorname{Kern}\varphi = \{\pm e\}$. Sei weiter $(a,b) \in \operatorname{Kern}\psi$, also $axb\bar{} = x$ für alle $x \in \mathbb{H}$. Für $x := e$ folgt $a = b$ und also $a \in \operatorname{Kern}\psi$, das heißt, $a = \pm e$, das heißt, $\operatorname{Kern}\varphi = \{\pm (e,e)\}$.

Die Sätze 3 und 2 liefern die nichttrivialen Inklusionen $\varphi(S^3) \supset SO(\operatorname{Im}\mathbb{H})$, $\psi(S^3 \times S^3) \supset SO(\mathbb{H})$. Hier besteht in beiden Fällen Gleichheit: dies folgt unmittelbar „aus Stetigkeitsgründen (mittels der Determinante)" oder auch direkt: gäbe es z. B. ein $\psi(a,b) \in O^-(\mathbb{H})$, so gäbe es nach Satz 2,b) Elemente $c, d \in S^3$, so daß $axb\bar{} = c\bar{x}d$ für alle $x \in \mathbb{H}$. Es wäre stets $\bar{x} = pxq^{-1}$ mit $p := c^{-1}a$, $q := db$. Für $x := e$ folgt $p = q$, für $x := p$ folgt weiter $\bar{p} = p$, also $p \in \mathbb{R}e$. Dies führt zur Absurdität $\bar{x} = x$. □

Nach dem Vorangehenden gibt es natürliche Gruppenepimorphismen $S^3 \to SO(3)$, $S^3 \times S^3 \to SO(4)$, deren Kerne jeweils 2 Elemente haben. Da S^3 nach 2.6 zu $SU(2)$ isomorph ist, hat man entsprechend auch Epimorphismen $SU(2) \to SO(3)$, $SU(2) \times SU(2) \to SO(4)$, deren Kerne zwei Elemente haben.

Da S^3 bzw. $S^3 \times S^3$ drei- bzw. sechsdimensional ist, so folgt u. a.:

Die Gruppe $SO(3)$ ist 3-dimensional, die Gruppe $SO(4)$ ist 6-dimensional (allgemein gilt $\dim SO(n) = \frac{1}{2}n(n-1)$, vgl. Abschnitt 6).

Die Mengen $G := \psi(S^3 \times e)$, $G' := \psi(e \times S^3)$ sind Normalteiler in $SO(4)$, die vermöge $a \mapsto \psi(a,e)$ bzw. $b \mapsto \psi(e,b)$ zur Gruppe S^3 isomorph sind. Es gilt $G \cdot G' = SO(4)$, aber $G \cap G' = \pm \mathrm{id}$ (Beweis!). Wir sehen speziell:

Die Gruppe $SO(4)$ enthält zur Gruppe S^3 isomorphe Normalteiler und ist also keine „einfache" LIEsche Gruppe (dagegen sind alle Gruppen $SO(n)$, $n > 4$, einfach, das heißt, sie haben keine nichttrivialen zusammenhängenden Normalteiler: die Gruppen $SO(2n+1)$ haben überhaupt keine echten Normalteiler $\neq \{e\}$; die Gruppen $SO(2n)$ haben genau den einen nichttrivialen Normalteiler $\{\pm e\}$).

5. Drehachse und Drehwinkel. Für $a \in S^3$ sei $f_a : \operatorname{Im}\mathbb{H} \to \operatorname{Im}\mathbb{H}$ definiert durch $f_a(u) := aua\bar{}$. Es gilt $f_a \in O(\operatorname{Im}\mathbb{H})$, weiter ist f_a genau dann die Identität, wenn $a = \pm e$ (man beachte, daß $f_a = \varphi(a)$ gemäß Abschnitt 4).

Falls $f_a \neq \mathrm{id}$, so gilt $0 \neq a - \bar{a} \in \operatorname{Im}\mathbb{H}$ und $f_a(a - \bar{a}) = a - \bar{a}$; die von $a - \bar{a}$ erzeugte Gerade ist also Fixgerade von f_a.

Beweis. Da $a \neq \pm e$, so folgt $0 \neq a - \bar{a} \in \operatorname{Im}\mathbb{H}$ und weiter $f_a(a - \bar{a}) = a(a - \bar{a})\bar{a}$ $= a(a\bar{a}) - (a\bar{a})\bar{a} = a - \bar{a}$. □

Um die Abbildung f_a anders zu beschreiben, benutzen wir folgendes

Lemma. *Jede Quaternion $a \in S^3 \setminus \{\pm e\}$ hat die eindeutige Darstellung*

(1) $a = \cos\frac{1}{2}\omega \cdot e + \sin\frac{1}{2}\omega \cdot q$ *mit* $q \in \operatorname{Im}\mathbb{H}$, $|q| = 1$, *und* $0 < \omega < 2\pi$.

Beweis. Man schreibt $a = \alpha e + \beta q$ mit $q \in \operatorname{Im}\mathbb{H}$, $|q| = 1$ und $\beta > 0$. Wegen $\alpha^2 + \beta^2 = 1$ gibt es genau ein $\omega \in (0, 2\pi)$ mit $\alpha = \cos\frac{1}{2}\omega$, $\beta = \sin\frac{1}{2}\omega$. □

Wir behaupten nun, daß f_a eine Drehung um $\mathbb{R}q$ mit dem „Drehwinkel" ω ist, das heißt, daß die zu $\mathbb{R}q$ senkrechte Ebene in $\operatorname{Im}\mathbb{H}$ um den Winkel ω gedreht wird. Dies und mehr impliziert folgender

Satz. *Wählt man zu $a \in S^3 \setminus \{\pm e\}$ die Größen ω und q gemäß* (1), *so gilt*

$$f_a(u) = \cos\omega \cdot u + \sin\omega \cdot q \times u + (1 - \cos\omega)\langle q, u\rangle \qquad \textit{für alle} \qquad u \in \operatorname{Im}\mathbb{H}.$$

Beweis. Mit den Abkürzungen $\alpha := \cos\frac{1}{2}\omega, \beta := \sin\frac{1}{2}\omega$ hat man

$$au\bar{a} = (\alpha e + \beta q)u(\alpha e - \beta q) = \alpha^2 u + \beta\alpha qu - \alpha\beta uq - \beta^2 quq.$$

Nach Definition des Vektorproduktes (vgl. 1.4) gilt $2q \times u = qu - uq$. Da $\bar{u} = -u$ und $\langle q, q\rangle = 1$, so folgt $quq = u - 2\langle q, u\rangle q$ aufgrund der Dreier-Identität 2.2.(5), also:

$$f_a(u) = (\alpha^2 - \beta^2)u + 2\alpha\beta q \times u + 2\beta^2\langle q, u\rangle q, \qquad u \in \operatorname{Im}\mathbb{H}.$$

Laut Definition von α, β gilt nach elementaren Formeln $\alpha^2 - \beta^2 = \cos\omega$, $2\alpha\beta = \sin\omega$, $2\beta^2 = 1 - (\alpha^2 - \beta^2) = 1 - \cos\omega$. □

Korollar. $f_a(q) = q$ *und* $\langle f_a(u), u\rangle = \cos\omega$ *für alle* $u \in \operatorname{Im}\mathbb{H}$ *mit* $|u| = 1$ *und* $\langle u, q\rangle = 0$.

Hieraus entnimmt man, daß f_a eine Drehung um $\mathbb{R}q$ mit dem Drehwinkel ω ist; es gilt übrigens: $\cos\omega = \operatorname{Re}(a^2)$ (Beweis!). Ist a rein imaginär, so gilt $\omega = \pi$ und $f_a = -s_a$ ist eine Drehung um 180° um die Achse $\mathbb{R}a$. □

Bemerkung. Nach allgemeiner Theorie (vgl. Grundwissen Mathematik 2, Lineare Algebra und analytische Geometrie, 7.3.6) ist jede eigentlich orthogonale Abbildung $\neq \mathrm{id}$ des $\mathbb{R}^3$ eine Drehung um eine eindeutig bestimmte Achse. Jedes $f \in SO(\operatorname{Im}\mathbb{H}) \setminus \{\mathrm{id}\}$ ist daher eine Drehung um eine Drehachse $\mathbb{R}q$, $q \in \operatorname{Im}\mathbb{H}$, $|q| = 1$, um einen Winkel ω, $0 < \omega < 2\pi$. Bestimmt man nun $a \in S^3$ gemäß (1), so gilt $a \neq \pm e$, und $f_a \in SO(\operatorname{Im}\mathbb{H})$ ist nach dem Satz eine Drehung um den Winkel ω mit Drehachse $\mathbb{R}q$. Damit hat man erneut die Aussage a) des Satzes 3 von HAMILTON bewiesen: jedes $f \in SO(\operatorname{Im}\mathbb{H})$ hat die Form f_a mit $a \in S^3$.

6. EULERsche Parameterdarstellung der $SO(3)$. Die Abbildung

$$\mathbb{H} \setminus \{0\} \to SO(\operatorname{Im}\mathbb{H}), \quad a \mapsto h_a \quad \text{mit} \quad h_a : \operatorname{Im}\mathbb{H} \to \operatorname{Im}\mathbb{H}, \quad u \mapsto \frac{1}{|a|^2} au\bar{a} = aua^{-1},$$

ist aufgrund von Satz 4 ein *Epimorphismus* der multiplikativen Gruppe $\mathbb{H} \setminus \{0\}$ auf die Gruppe $SO(\operatorname{Im}\mathbb{H})$ mit $\mathbb{R}e \setminus \{0\}$ als Kern. Setzt man $a := \kappa e + \lambda i + \mu j + \nu k$ und schreibt man $u := xi + yj + zk$ als Spaltenvektor, so gilt

$$h_a(u) = A \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

mit einer eigentlich orthogonalen Matrix $A \in SO(3)$. Diese Matrix wird bestimmt, indem man $|a|^{-2} a u \bar{a}$ bezüglich der Basis i, j, k von $\operatorname{Im} \mathbb{H}$ ausrechnet. Man gewinnt die von EULER 1770 (Opera Omnia 6, 1. Ser., 287–315) entdeckte

Rationale Parameterdarstellung orthogonaler 3 × 3 Matrizen. *Für jedes Quadrupel* $(\kappa, \lambda, \mu, \nu) \in \mathbb{R}^4 \setminus \{0\}$ *ist die* 3×3 *Matrix*

$$(1) \qquad \frac{1}{\kappa^2 + \lambda^2 + \mu^2 + \nu^2} \begin{pmatrix} \kappa^2 + \lambda^2 - \mu^2 - \nu^2 & -2\kappa\nu + 2\lambda\mu & 2\kappa\mu + 2\lambda\nu \\ 2\kappa\nu + 2\lambda\mu & \kappa^2 - \lambda^2 + \mu^2 - \nu^2 & -2\kappa\lambda + 2\mu\nu \\ -2\kappa\mu + 2\lambda\nu & 2\kappa\lambda + 2\mu\nu & \kappa^2 - \lambda^2 - \mu^2 + \nu^2 \end{pmatrix}$$

eigentlich orthogonal; man erhält so alle eigentlich orthogonalen 3×3 *Matrizen.*

Beweis. Setzt man $a := \kappa e + b \in (\mathbb{R} \times \operatorname{Im} \mathbb{H}) \setminus \{0\}$, so hat man $a u \bar{a} = (\kappa e + b) u (\kappa e - b) = \kappa^2 u + 2\kappa b \times u - bub$. Da $bub = |b|^2 u - 2\langle b, u \rangle b$ nach der Dreieridentität 2.2.(5) wegen $\bar{u} = -u$, so sieht man $a u \bar{a} = (\kappa^2 - |b|^2) u + 2\kappa b \times u + 2\langle b, u \rangle b$. Hieraus folgt (1) direkt, wenn $b := \lambda i + \mu j + \nu k$. □

Die in 3.5.4 angegebene rationale Parameterdarstellung für eigentlich orthogonale 2×2 Matrizen entsteht aus (1), wenn man im rechten unteren Kästchen $\mu = \nu = 0$ setzt (und $-\lambda$ statt λ schreibt). Einen anderen Zugang zur EULERschen Parameterdarstellung findet der Leser in Grundwissen Mathematik 2, Lineare Algebra und analytische Geometrie, 7.3.7.

Da der Epimorphismus $\mathbb{H} \setminus \{0\} \to SO(\operatorname{Im} \mathbb{H})$, $a \mapsto h_a$, die Gruppe $\mathbb{R}e \setminus \{0\}$ zum Kern hat, so gehört zu zwei Quadrupeln $a, a' \in \mathbb{R}^4 \setminus \{0\}$ genau dann dieselbe Matrix A gemäß (1), wenn $a' = \alpha a$ mit $\alpha \neq 0$, das heißt, wenn a und a' denselben Punkt im 3-dimensionalen reellen projektiven Raum $\mathbb{P}^3(\mathbb{R})$ mit den homogenen Koordinaten κ, λ, μ, ν bestimmen. Daher kann der Satz von EULER auch wie folgt ausgesprochen werden:

Die durch (1) *definierte Abbildung* $\mathbb{P}^3(\mathbb{R}) \to SO(3)$ *ist bijektiv, insbesondere ist* $SO(3)$ *eine rationale Mannigfaltigkeit.*

Diese Aussage wurde 1846 von CAYLEY (Math. Papers 1, 332–336) verallgemeinert:

Die Gruppe $SO(n)$ *ist eine* $\frac{1}{2}n(n-1)$*-dimensionale rationale Mannigfaltigkeit; vermöge der „CAYLEY-Abbildung“*

$$X \mapsto (\kappa E - X)^{-1}(\kappa E + X), \qquad X \in \operatorname{Mat}(n, \mathbb{R}), \qquad X \text{ schiefsymmetrisch},$$

wird der $\frac{1}{2}n(n-1)$*-dimensionale reell projektive Raum „birational“ auf* $SO(n)$ *abgebildet.*

Im Fall $n = 3$ ist diese CAYLEYsche Darstellung nichts anderes als die EULERsche Parameterdarstellung.

Kapitel 7. Isomorphiesätze von Frobenius und Hopf

M. Koecher, R. Remmert

Einleitung

1. In der zweiten Hälfte des 19. Jahrhunderts wurden neben den Quaternionen viele weitere hyperkomplexe Systeme entdeckt und erforscht. Vor allem in England stand diese Kunst in hohem Ansehen. Kurz nach Entdeckung der Quaternionen und *vor* Einführung von Matrizen erfanden John T. Graves und Arthur Cayley die nichtassoziative Divisionsalgebra der *Oktaven*. Hamilton führte 1853 in seinen „Lectures on Quaternions" *Biquaternionen*, das sind Quaternionen mit komplexen Koeffizienten, ein und bemerkte, daß sie keine Divisionsalgebra bilden. William Kingdon Clifford (1845–1879) schuf 1878 die nach ihm benannten assoziativen Algebren.

Die Flut neuer hyperkomplexer Systeme überschwemmte bald die gesamte Algebra. Die wichtige Frage, wie viel Freiheit in der Fülle aller Beispiele wirklich vorhanden ist, rückte nur langsam in den Vordergrund des Interesses. War Gauss noch 1831 davon überzeugt gewesen, daß keine hyperkomplexen Zahlensysteme existieren, für welche die grundlegenden Eigenschaften der komplexen Zahlen erhalten bleiben (vgl. 4.3.6), so glaubte man nach Entdeckung der Quaternionen und Oktaven zunächst, daß immerzu neue interessante hyperkomplexe Systeme erfunden werden könnten. Es ist allerdings bezeichnend, daß Hamilton nicht beweisen konnte, daß 3-dimensionale, kommutative und assoziative Divisionsalgebren (das sind Körper) über $\mathbb{R}$ nicht existieren. Auch Grassmann hat dazu nichts gesagt. 1871 veröffentlichte Benjamin Peirce (1809–80, Professor der Mathematik in Harvard) einen Artikel „Linear Associative Algebras", wo er alle bis dahin bekannten solchen Algebren zusammenstellte (abgedruckt im Amer. Journ. Math. 4, 97–229, 1881).

2. Die Einsicht, daß es weitaus weniger interessante $\mathbb{R}$-Algebren gibt, als man erhofft hatte, wurde erst von der folgenden Mathematikergeneration gewonnen. Einen ersten präzisen Einzigkeitssatz bewies 1877 Ferdinand Georg Frobenius (geb. 1849 in Berlin; Schüler von Weierstrass, 1875 Professor am Polytechnikum in Zürich, ab 1892 Professor an der Universität in Berlin; förderte die abstrakte Betrachtungsweise in der Algebra, wichtige Anwendungen seiner Darstellungstheorie ergaben sich in der Quantentheorie; gest. 1917 in Charlottenburg). In seiner im Crelleschen Journal veröffentlichten Arbeit „Über lineare Substitutionen und bilineare Formen" (Ges. Abhandl. 1, 343–405) zeigte er, daß es bis auf Isomorphie nur *drei* reelle endlich-dimensionale, assoziative Divisionsalgebren gibt: $\mathbb{R}$ selbst, $\mathbb{C}$ und $\mathbb{H}$. Dieser berühmte Satz, der 1881 unabhängig von dem amerikanischen

Mathematiker Charles Sanders PEIRCE (1839–1914, Sohn von Benjamin PEIRCE) bewiesen wurde (Appendix zur Arbeit des Vaters im Amer. Journ. Math. 4), zeigte den Algebraikern erstmals Grenzen ihrer für allmächtig gehaltenen Konstruktionsverfahren. Hätte HAMILTON den Satz von FROBENIUS gekannt, so wären ihm Jahre harter Arbeit bei seiner vergeblichen Suche nach drei-dimensionalen, assoziativen Divisionsalgebren erspart geblieben.

In den ersten beiden Paragraphen dieses Kapitels wird der Satz von FROBENIUS bewiesen. Das zentrale Resultat ist ein Existenzsatz für HAMILTONsche Tripel, aus dem das FROBENIUSsche Resultat folgt. Um spätere Wiederholungen zu vermeiden, setzen wir *nicht* von vornherein alle vorkommenden Algebren als assoziativ voraus; vielmehr stellen wir jeweils die benötigten (schwächeren) Voraussetzungen, wie *potenz-assoziativ* oder *alternativ* oder *quadratisch*, bewußt heraus. Dieser abstrakte Standpunkt dürfte heute, nach der BOURBAKI-Zeit, keinen Leser mehr abschrecken; wir verstoßen hier nicht gegen die neuerdings wieder zu Ehren kommende These der Didaktik: „Premature abstraction falls on deaf ears whether they belong to mathematicians or to students."

3. Im Jahre 1940 hat Heinz HOPF (schweizerischer Mathematiker deutscher Herkunft; geb 1894 in Grätschen (Schlesien); Studium in Berlin, Heidelberg und Göttingen, hier 1925 Bekanntschaft mit Paul ALEXANDROFF und Emmy NOETHER; 1931 Nachfolger von Hermann WEYL an der Eidgenössischen Technischen Hochschule Zürich, gest. 1971 in seiner Wahlheimat Zollikon, Kanton Zürich. Bahnbrechende Arbeiten zur Topologie der Mannigfaltigkeiten und ihrer Abbildungen sowie zur Differentialgeometrie. Meister der echten Didaktik. „Hopf hat stets gleichzeitig die Lösung des Einzelproblems gegeben und die Methode zu seiner Bezwingung geschaffen, aus der die leitende Idee, der tiefere Grund, die weiteren Möglichkeiten klar wurden") das Problem gestellt, alle endlich-dimensionalen, reellen, *kommutativen* Divisionsalgebren (die nicht mehr notwendig assoziativ sind) anzugeben. Diese auf den ersten Blick gekünstelt erscheinende Problemstellung führt zu überraschenden und unerwarteten Einsichten. HOPF hat in seiner berühmt gewordenen Arbeit „Systeme symmetrischer Bilinearformen und euklidische Modelle der projektiven Räume" (Vierteljahreszeitschrift der Naturforschenden Gesellschaft in Zürich, LXXXV (1940), Beibl. Nr. 32, Festschrift Rudolf FUETER; siehe auch H. HOPF: Selecta, Springer-Verlag 1964) gezeigt, daß eine reelle, kommutative Divisionsalgebra endlicher Dimension *höchstens 2-dimensional ist*. Das Bemerkenswerte am HOPFschen Problem ist, daß eine einfach zu formulierende algebraische Frage eine einfache algebraische Antwort findet, daß indessen zur Lösung *nichttriviale Methoden der Topologie* erforderlich sind; hier erscheint zum ersten Mal der „topologische Stachel im Fleisch der Algebra", der bis auf den heutigen Tag von vielen Algebraikern als so schmerzhaft empfunden wird.

§ 1. Hamiltonsche Tripel in alternativen Algebren

Das Multiplikationsverhalten der HAMILTONschen Basisquaternionen i, j, k erweist sich für die allgemeine Theorie der Algebren als so wichtig, daß wir es durch eine Definition herausheben.

In einer Algebra $\mathscr{A}$ mit Einselement e nennt man drei Elemente u, v, w ein *Hamiltonsches Tripel*, wenn die *neun* HAMILTONschen Bedingungen erfüllt sind:

$$u^2 = v^2 = w^2 = -e, \qquad w = uv = -vu, \qquad u = vw = -wv, \qquad v = wu = -uw.$$

Es ist das Ziel dieses vorbereitenden Paragraphen, Bedingungen für die Existenz HAMILTONscher Tripel zu finden.

1. Die rein-imaginären Elemente einer Algebra. In den $\mathbb{R}$-Algebren $\mathbb{C}$ und $\mathbb{H}$ besteht der *Imaginärraum* aus allen Elementen $x \notin \mathbb{R}e \setminus \{0\}$, deren Quadrat „reell" ist: $x^2 \in \mathbb{R}e$. Wir führen für jede Algebra $\mathscr{A}$ mit Einselement e die *Menge*

$$\operatorname{Im}\mathscr{A} := \{x \in \mathscr{A} : x^2 \in \mathbb{R}e \text{ und } x \notin \mathbb{R}e \setminus \{0\}\}$$

der „*rein-imaginären*" Elemente ein. Dann ist trivial:

$$\mathbb{R}e \cap \operatorname{Im}\mathscr{A} = 0, \qquad u \in \operatorname{Im}\mathscr{A} \Rightarrow \alpha u \in \operatorname{Im}\mathscr{A} \qquad \textit{für jedes} \qquad \alpha \in \mathbb{R}.$$

Es ist keineswegs klar, daß $\operatorname{Im}\mathscr{A}$ ein Untervektorraum von $\mathscr{A}$ ist, das heißt, aus $u, v \in \operatorname{Im}\mathscr{A}$ folgt *nicht* ohne weiteres $u + v \in \operatorname{Im}\mathscr{A}$ (vgl. hierzu 2.1). Ganz allgemein gilt (also insbesondere für $\mathbb{H}$) das

Unabhängigkeitslemma. *Sind $u, v \in \operatorname{Im}\mathscr{A}$ linear unabhängig, so sind auch e, u, v linear unabhängig.*

Beweis. Wäre das nicht der Fall, so dürfte man ohne Einschränkung $v = \alpha e + \beta u$ mit $\alpha, \beta \in \mathbb{R}$ annehmen. Hieraus würde folgen $2\alpha\beta u = v^2 - \alpha^2 e - \beta^2 u^2 \in \mathbb{R}e$, also $\alpha\beta = 0$. Da $\alpha = 0$ bzw. $\beta = 0$ wegen der linearen Unabhängigkeit von u, v bzw. wegen $v \in \operatorname{Im}\mathscr{A}$ nicht möglich ist, hat man einen Widerspruch. □

Wegen $uv + vu = (u + v)^2 - u^2 - v^2$ ist weiter klar:

$$u, v, u + v \in \operatorname{Im}\mathscr{A} \Rightarrow uv + vu \in \mathbb{R}e.$$

Wir zeigen nun:

Ist $\mathscr{A}$ nullteilerfrei, so gilt $u^2 = -\omega e$ mit $\omega > 0$ für $u \in \operatorname{Im}\mathscr{A}$, $u \neq 0$.

Beweis. Es gilt $u^2 = \alpha e$ mit $\alpha \in \mathbb{R}$. Im Falle $\alpha \geqslant 0$ könnte man $\alpha = \beta^2$ mit $\beta \in \mathbb{R}$ schreiben, und man hätte $(u - \beta e)(u + \beta e) = u^2 - \beta^2 e = u^2 - \alpha e = 0$. Da $\mathscr{A}$ nullteilerfrei ist, müßte einer der Faktoren links verschwinden, das heißt, es würde $u \in \mathbb{R}e$ folgen. Widerspruch! □

Bei nullteilerfreien Algebren kann man also insbesondere (wie bei $\mathbb{C}$ und $\mathbb{H}$) von jedem rein-imaginären Element $u' \neq 0$ durch Skalarenmultiplikation zu einem Element $u = \gamma u'$ mit $u^2 = -e$ übergehen (Normierung).

2. Hamiltonsche Tripel. Jedes Element eines HAMILTONschen Tripels ist rein-imaginär. Wir zeigen darüber hinaus

Satz. *Ist u, v, w ein Hamiltonsches Tripel in $\mathscr{A}$, so gilt:*

1) *die Abbildung $f\colon \mathbb{H} \to \mathscr{A}$, $(\alpha, \beta, \gamma, \delta) \mapsto \alpha e + \beta u + \gamma v + \delta w$ ist ein Algebra-Monomorphismus,*

2) $\mathbb{R}u + \mathbb{R}v + \mathbb{R}w \subset \operatorname{Im}\mathscr{A}$, *speziell enthält* $\operatorname{Im}\mathscr{A}$ *einen 3-dimensionalen Untervektorraum.*

Beweis. 1) Da $f(e) = e, f(i) = u, f(j) = v, f(k) = w$, so ist f ein Algebra-Homomorphismus. Die Injektivität von f ist äquivalent damit, daß e, u, v, $w \in \mathscr{A}$ linear unabhängig sind. Zunächst ist klar, daß u und v linear unabhängig sind, denn sonst wäre $v \in \mathbb{R}u$ und also $vu = uv$, das heißt, $-w = w$, das heißt, $w = 0$ im Widerspruch zu $w^2 = -e$. Nach dem Unabhängigkeitslemma 1 folgt nun, daß e, u, v linear unabhängig sind. Wären jetzt e, u, v, w linear abhängig, so gäbe es *eindeutig* bestimmte Zahlen $\rho, \sigma, \tau \in \mathbb{R}$, so daß $w = uv = \rho e + \sigma u + \tau v$. Linksmultiplikation mit u ergibt $-v = \rho u - \sigma e + \tau w$. Die Eindeutigkeit der Darstellung erzwingt $\tau^2 = -1$. Widerspruch!

2) Man verifiziert $(\beta u + \gamma v + \delta w)^2 \in \mathbb{R}e$ durch Nachrechnen. □

Eine wichtige Vorstufe zur Konstruktion Hamiltonscher Tripel besteht darin, zu einem Vektor $p \in \operatorname{Im}\mathscr{A}$ einen linear unabhängigen Vektor $q \in \operatorname{Im}\mathscr{A}$ zu finden, so daß die *eine* Hamiltonsche Bedingung $pq + qp = 0$ erfüllt ist. Wir zeigen:

Lemma. *Es sei U ein 2-dimensionaler Untervektorraum von* $\operatorname{Im}\mathscr{A}$. *Dann gibt es zu jedem $p \in U$ ein $q \in U \setminus \mathbb{R}p$, so daß gilt: $pq + qp = 0$.*

Beweis. Wir dürfen $p \neq 0$ annehmen, also $p^2 = \alpha e$ mit $\alpha \neq 0$. Man wähle $x \in U$, so daß p und x linear unabhängig sind. Es gilt $px + xp = \beta e$, $\beta \in \mathbb{R}$ (vgl. Abschnitt 1). Nun leistet $q := x + \xi p$ mit $\xi := -\beta(2\alpha)^{-1}$ das Gewünschte.

3. Existenz Hamiltonscher Tripel in alternativen Algebren. Das weitere Vorgehen ist durch Lemma 2 vorgezeichnet: hat man zwei Vektoren gemäß dieses Lemmas, so geht man im Falle einer nullteilerfreien Algebra (durch Multiplikation mit Skalaren, vgl. Abschnitt 1) sofort zu zwei Vektoren $u, v \in \operatorname{Im}\mathscr{A}$ mit $u^2 = v^2 = -e$ und $uv = -vu$ über. Dann sind $u, v, w := uv$ Kandidaten für ein Hamiltonsches Tripel. Man kann nun aber nicht „ohne weiteres" zeigen, daß z. B. wirklich $vw = u$ gilt, denn in $v(uv) = -v(vu)$ darf man nicht bedenkenlos umklammern. Man macht die Not zur Tugend und *postuliert* diese *schwache Assoziativität.*

Eine Algebra $\mathscr{A}$ heißt alternativ, wenn für alle $x, y \in \mathscr{A}$ gilt:

$$x(xy) = x^2 y, \qquad (xy)y = xy^2.$$

Jede assoziative Algebra ist alternativ*). Ist $\mathscr{A}$ alternativ, so gilt auch

$$(xy)x = x(yx) \qquad \text{für alle} \qquad x, y \in \mathscr{A};$$

zum Beweis ersetze man in $(xy)y = xy^2$ das Element y durch $x + y$ und rechne.

---

*) Das Wort „alternativ" soll darauf hinweisen, daß der *Assoziator* $(xy)z - x(yz)$ das Vorzeichen wechselt, wenn man zwei Argumente vertauscht. Alternative Algebren sind also keineswegs, wie moderne Schlagworte vermuten lassen könnten, „Gegenalgebren".

Aufgabe. Man zeige: je zwei der drei Identitäten $x(xy) = x^2y$, $(xy)y = xy^2$, $(xy)x = x(yx)$ implizieren die dritte.

Existenzsatz für Hamiltonsche Tripel. *Es sei $\mathscr{A}$ eine alternative, nullteilerfreie Algebra mit Einselement e, und es sei U ein 2-dimensionaler Untervektorraum von* Im $\mathscr{A}$. *Dann gibt es zu jedem Element $u \in U$ mit $u^2 = -e$ ein $v \in U$, so daß u, v, uv ein Hamiltonsches Tripel bilden.*

Beweis. Aufgrund der vorangehenden Überlegungen gibt es ein $v \in U$ mit $v^2 = -e$ und $uv = -vu$. Da $\mathscr{A}$ alternativ ist, sind für $u, v, w := uv$ die HAMILTONschen Bedingungen erfüllt; so folgt zunächst $vw = v(uv) = -v(vu) = -v^2u = u$ und $wv = (uv)v = uv^2 = -u$. Ebenso zeigt man $wu = v = -uw$. Es bleibt $w^2 = -e$ zu zeigen. Da $vw^2 = (vw)w = uw = -v$, so folgt $v(w^2 + e) = 0$ und also $w^2 = -e$ wegen der Nullteilerfreiheit.

4. Alternative Algebren. Durch den Existenzsatz 3 gewinnen die alternativen Algebren eine besondere Bedeutung. Wir notieren für sie zwei Aussagen, die im nächsten Paragraphen nützlich sind:

Jede alternative Algebra $\mathscr{A}$ ist potenz-assoziativ.

Dazu ist die Potenzregel $x^m x^n = x^{m+n}$, $x \in \mathscr{A}$, zu verifizieren. Das geschieht durch Induktion mittels der Alternativregeln, z. B. gilt, wenn $x^{m-1}x^n = x^{m-1+n}$ schon für alle n bekannt ist: $x^m x = (x^{m-1}x)x = x^{m-1}x^2 = x^{m+1}$. Wir überlassen dem Leser die detaillierte Ausführung des Induktionsbeweises. □

Die im Existenzsatz 3 gemachte Voraussetzung, daß $\mathscr{A}$ ein Einselement hat, ist für alternative Divisionsalgebren von selbst erfüllt:

Jede alternative Divisionsalgebra $\mathscr{A}$ hat ein Einselement.

Beweis. Wähle $a \in \mathscr{A}$, $a \neq 0$. Da $\mathscr{A}$ Divisionsalgebra ist, gibt es ein $e \in \mathscr{A}$ mit $ea = a$. Es gilt $e \neq 0$ wegen $a \neq 0$. Weiter folgt: $e(ea) = ea$, also wegen der Alternativität: $e^2a = ea$, das heißt $(e^2 - e)a = 0$ und also $e^2 = e$. Nunmehr folgt $e(ex - x) = e(ex) - ex = e^2x - ex = 0$, das heißt, $ex = x$ für alle $x \in \mathscr{A}$. Analog sieht man $xe = x$. □

Ohne Beweis sei noch erwähnt:

Satz von E. ARTIN. *Eine Algebra $\mathscr{A}$ ist genau dann alternativ, wenn je zwei Elemente $x, y \in \mathscr{A}$ eine assoziative Unteralgebra von $\mathscr{A}$ erzeugen.*

§ 2. Satz von FROBENIUS

> Wir sind also zu dem Resultate gelangt, dass ausser den reellen Zahlen, den imaginären Zahlen und den Quaternionen keine andern complexen Zahlen in dem oben definirten Sinne existiren (G. FROBENIUS 1877).

Um den Existenzsatz 1.3 für HAMILTONsche Tripel anwenden zu können, benötigt man vorab einen 2-dimensionalen Untervektorraum des Imaginärraumes $\operatorname{Im}\mathscr{A}$. In diesem Paragraphen zeigen wir, daß in wichtigen Fällen, die weit über $\mathbb{C}$ und $\mathbb{H}$ hinausgehen, die Menge $\operatorname{Im}\mathscr{A}$ selbst ein Untervektorraum von $\mathscr{A}$ ist: dies ist nämlich für alle sogenannten *quadratischen Algebren* richtig; dabei heißt eine Algebra $\mathscr{A}$ mit Einselement e *quadratisch*, wenn jedes Element $x\in\mathscr{A}$ einer quadratischen Gleichung $x^2=\alpha e+\beta x$ mit $\alpha,\beta\in\mathbb{R}$ genügt. Die Algebren $\mathbb{C}$ und $\mathbb{H}$ sind quadratisch (letztere aufgrund von Satz 6.1.2). Quadratische Algebren spielen in der allgemeinen Theorie der Algebren eine wesentliche Rolle; hier und im nächsten Kapitel benutzen wir sie aber lediglich als Hilfsbegriff*), um bequem formulieren zu können; in den finalen Fassungen der Hauptresultate treten sie nicht mehr auf. Jede endlich-dimensionale, alternative Divisionsalgebra erweist sich als quadratisch.

Das Hauptresultat dieses Paragraphen ist das Quaternionen-Lemma 3, das unmittelbar zum eigentlichen Satz von FROBENIUS führt.

1. Lemma von FROBENIUS. *Ist $\mathscr{A}$ eine quadratische Algebra, so ist $\operatorname{Im}\mathscr{A}$ ein Untervektorraum von $\mathscr{A}$, und es gilt: $\mathscr{A}=\mathbb{R}e\oplus\operatorname{Im}\mathscr{A}$.*

Beweis. 1) Wir zeigen: $u,v\in\operatorname{Im}\mathscr{A}\Rightarrow u+v\in\operatorname{Im}\mathscr{A}$. Sind u,v linear abhängig, etwa $v=\alpha u$, so ist $u+v=(1+\alpha)u\in\operatorname{Im}\mathscr{A}$ trivial. Seien also u,v linear unabhängig. Da $\mathscr{A}$ quadratisch ist, bestehen Gleichungen

$$(u+v)^2=\alpha_1 e+\beta_1(u+v),\qquad (u-v)^2=\alpha_2 e+\beta_2(u-v)$$

mit $\alpha_1,\alpha_2,\beta_1,\beta_2\in\mathbb{R}$. Ausmultiplizieren und Addieren liefert:

$$(\beta_1+\beta_2)u+(\beta_1-\beta_2)v=2u^2+2v^2-(\alpha_1+\alpha_2)e\in\mathbb{R}e$$

wegen $u,v\in\operatorname{Im}\mathscr{A}$. Aus dem Unabhängigkeitslemma 1.1 folgt $\beta_1+\beta_2=\beta_1-\beta_2=0$, also $\beta_1=\beta_2=0$ und somit $(u+v)^2=\alpha_1 e\in\mathbb{R}e$. Da $u+v\notin\mathbb{R}e$ (Unabhängigkeitslemma), so folgt $u+v\in\operatorname{Im}\mathscr{A}$.

2) Sei $x\in\mathscr{A}, x\notin\mathbb{R}e$ beliebig. Es gilt $x^2=\alpha e+2\beta x$ mit $\alpha,\beta\in\mathbb{R}$. Man erhält $(x-\beta e)^2=x^2-2\beta x+\beta^2 e=(\alpha+\beta^2)e$. Da $x-\beta e\notin\mathbb{R}e$, so folgt $x=\beta e+u$ mit $u:=x-\beta e\in\operatorname{Im}\mathscr{A}$. Damit ist $\mathscr{A}=\mathbb{R}e+\operatorname{Im}\mathscr{A}$ gezeigt. Da $\mathbb{R}e\cap\operatorname{Im}\mathscr{A}=\{0\}$, so ergibt sich $\mathscr{A}=\mathbb{R}e\oplus\operatorname{Im}\mathscr{A}$. □

---

*) Der Leser diskutiere, um sich mit quadratischen Algebren etwas vertraut zu machen, folgenden

Satz. *Ist $\mathscr{A}$ eine quadratische Algebra, so ist für jedes $x\in\mathscr{A}$ der Untervektorraum $\mathbb{R}e+\mathbb{R}x$ eine kommutative und assoziative $\mathbb{R}$-Unteralgebra von $\mathscr{A}$; speziell ist $\mathscr{A}$ potenz-assoziativ.*

Bemerkung. Der im ersten Teil des Beweises benutzte Trick, die Gleichungen für $(u+v)^2$ und $(u-v)^2$ heranzuziehen und sie zu addieren, findet sich bereits bei FROBENIUS (Ges. Abhandl. 1, S. 403).

2. Beispiele quadratischer Algebren. In potenz-assoziativen Algebren gilt die *Potenzregel*: $x^m x^n = x^{m+n}$ (R.1). Eine unmittelbare Verallgemeinerung ist die

Substitutionsregel. *Es sei $\mathfrak{A}$ eine potenz-assoziative Algebra mit Einselement e; für jedes reelle Polynom $f = \alpha_0 + \alpha_1 X + \cdots + \alpha_n X^n \in \mathbb{R}[X]$ setze man $f(x) := \alpha_0 e + \alpha_1 x + \cdots + \alpha_n x^n \in \mathfrak{A}$ (Substitution von $x \in \mathfrak{A}$ für X). Dann gilt:*

$$(f \cdot g)(x) = f(x)g(x) \quad \textit{für alle Polynome} \quad f, g \in \mathbb{R}[X] \quad \textit{und alle} \quad x \in \mathfrak{A}.$$

Der einfache Beweis wird in der Algebra geführt. – Man kann die Substitutionsregel auch wie folgt aussprechen:

Ist $\mathfrak{A}$ potenz-assoziativ und hat $\mathfrak{A}$ ein Einselement, so definiert jedes Element $x \in \mathfrak{A}$ vermöge $\mathbb{R}[X] \to \mathfrak{A}$, $f \mapsto f(x)$, einen Algebra-Homomorphismus (Substitutionshomomorphismus zu x).

Mittels der Substitutionsregel ergibt sich schnell:

Satz. *Jede endlich-dimensionale, alternative Divisionsalgebra $\mathfrak{A}$ ist quadratisch.*

Beweis. Nach 1.4 ist $\mathfrak{A}$ eine potenz-assoziative Algebra mit Einselement e. Daher ist für jedes $x \in \mathfrak{A}$ der Substitutionshomomorphismus $\mathbb{R}[x] \to \mathfrak{A}, f \mapsto f(x)$, definiert; sein Kern ist ein *Ideal* und also ein *Hauptideal*, da $\mathbb{R}[X]$ *Hauptidealring* ist. Wegen $\dim \mathfrak{A} < \infty$ ist dieses Hauptideal $\neq 0$; da $\mathfrak{A}$ nullteilerfrei ist, handelt es sich um ein *Primideal*. Es gibt also ein normiertes Primpolynom $p \in \mathbb{R}[X]$ mit $p(x) = 0$. Da jedes solche Polynom die Form $X - \gamma$ oder $X^2 - \beta X - \alpha$ hat (siehe 4.3.4), folgt die Behauptung.

3. Quaternionen-Lemma. *Jede alternative, endlich-dimensionale reelle Divisionsalgebra $\mathfrak{A}$ enthält im Falle $\dim \mathfrak{A} \geqslant 3$ Hamiltonsche Tripel und also Unteralgebren $\mathfrak{B}$ mit $e \in \mathfrak{B}$, die zur Quaternionenalgebra $\mathbb{H}$ isomorph sind.*

Beweis. Nach Satz 2 ist $\mathfrak{A}$ quadratisch; nach dem Lemma von FROBENIUS ist daher $\operatorname{Im}\mathfrak{A}$ ein Untervektorraum von $\mathfrak{A}$, und es gilt $\dim \operatorname{Im}\mathfrak{A} \geqslant 2$ wegen $\dim \mathfrak{A} \geqslant 3$. Nach dem Existenzsatz 1.3 gibt es also HAMILTONsche Tripel in $\mathfrak{A}$ und folglich nach Satz 1.2 auch zu $\mathbb{H}$ isomorphe Unteralgebren $\mathfrak{B}$ in $\mathfrak{A}$ mit $e \in \mathfrak{B}$. □

Im assoziativen Fall erhalten wir als Verschärfung den berühmten

4. Satz von FROBENIUS (1877). *Es sei $\mathfrak{A}$ eine assoziative, endlich-dimensionale reelle Divisionsalgebra. Dann sind drei Fälle möglich:*

1) *$\mathfrak{A}$ ist isomorph zum Körper $\mathbb{R}$ der reellen Zahlen.*
2) *$\mathfrak{A}$ ist isomorph zum Körper $\mathbb{C}$ der komplexen Zahlen.*
3) *$\mathfrak{A}$ ist isomorph zur Algebra $\mathbb{H}$ der Quaternionen.*

Beweis. Wegen $\mathscr{A} \neq 0$ gilt $\dim \mathscr{A} \geqslant 1$, und nach 1.4 hat $\mathscr{A}$ ein Einselement e. Falls $\dim \mathscr{A} = 1$, so liegt der Fall 1) vor (vgl. R.3). Falls $\dim \mathscr{A} = 2$, so gibt es ein $u \in \mathscr{A}$ mit $u^2 = -e$. Dann ist $f\colon \mathbb{C} \to \mathscr{A}$, $x + yi \mapsto xe + yu$, ein Algebra-Homomorphismus. Da e, u linear unabhängig sind (!), so ist f injektiv und wegen $\dim \mathbb{C} = \dim \mathscr{A}$ bijektiv, das heißt, es liegt der Fall 2) vor.

Sei $\dim \mathscr{A} \geqslant 3$. Dann gibt es nach dem Quaternionen-Lemma ein HAMILTONsches Tripel $u, v, w \in \operatorname{Im} \mathscr{A}$ mit $w = uv$. Wäre $\dim \mathscr{A} > 4$, so gäbe es ein $x \in \operatorname{Im} \mathscr{A}$, so daß u, v, w, x linear unabhängig sind. Anwendung von Lemma 1.2 auf $U := \mathbb{R}u + \mathbb{R}x \subset \operatorname{Im} \mathscr{A}$ gibt ein $y \in U \setminus \mathbb{R}u$ mit $uy + yu = 0$. Da $y = \rho u + \tau x$ mit $\tau \neq 0$, so sind auch u, v, w, y linear unabhängig! Wegen $v, w, y \in \operatorname{Im} \mathscr{A}$ gilt (vgl. 1.1):

$$vy + yv = \alpha e \qquad \text{und} \qquad wy + yw = \beta e \qquad \text{mit} \qquad \alpha, \beta \in \mathbb{R}.$$

Da $\mathscr{A}$ assoziativ ist, so folgt jetzt:

$$\begin{aligned} yw = y(uv) = (yu)v = -(uy)v = -u(yv) = -u(\alpha e - vy) \\ = -\alpha u + wy = -\alpha u + \beta e - yw, \end{aligned}$$

also $2yw = -\alpha u + \beta e$. Rechtsmultiplikation mit w liefert $-2y = \alpha v + \beta w$ im Widerspruch zur linearen Unabhängigkeit von u, v, w, y. Es gilt also $\dim \mathscr{A} = 4$ und mithin $\mathscr{A} = \mathbb{R}e + \mathbb{R}u + \mathbb{R}v + \mathbb{R}w \cong \mathbb{H}$. □

Einen weiteren elementaren Beweis des Satzes von FROBENIUS findet man bei R. S. PALAIS: The Classification of Real Division Algebras, Amer. Math. Monthly 75, 366–368 (1968).

Der Satz von FROBENIUS hat dem Problem der Klassifizierung aller endlich-dimensionalen assoziativen Algebren einen entscheidenden Impuls gegeben. Über die Entwicklung dieser Theorie berichtet D. HAPPEL: „Klassifikationstheorie endlich-dimensionaler Algebren in der Zeit von 1880 bis 1920", L'Enseignement Math. 26, 2. Ser., 91–102 (1980).

§ 3. Satz von HOPF

> Auch wenn man die Gültigkeit des assoziativen Gesetzes der Multiplikation nicht ausdrücklich postuliert, ist der Körper der komplexen Zahlen der einzige kommutative Erweiterungskörper endlichen Grades über dem Körper der reellen Zahlen (H. HOPF 1940).

Im vorangehenden Paragraphen fanden wir mit dem Satz von FROBENIUS alle endlich-dimensionalen, reellen, *assoziativen* Divisionsalgebren. Wir interessieren uns im folgenden für endlich-dimensionale, reelle, *kommutative* Divisionsalgebren, die nicht mehr notwendig assoziativ sind. Wir beweisen den Satz von HOPF, daß alle Algebren dieser Art höchstens 2-dimensional sind; ist ein Einselement vorhanden,

so ist $\mathbb{C}$ bis auf Isomorphie die einzige solche Algebra $\neq \mathbb{R}$; das assoziative Gesetz der Multiplikation ist dann also eine Folge des kommutativen Gesetzes. Bereits bei den Vorbereitungen in den Abschnitten 1 und 2 zum Beweis des Hopfschen Satzes benötigen wir nichttriviale Ergebnisse aus der reellen Analysis. Entscheidend für den eigentlichen Beweis ist jedoch ein Satz aus der algebraischen Topologie, der besagt, daß alle Räume $\mathbb{R}^m \setminus \{0\}$ für $m \geqslant 3$ *einfach-zusammenhängend* sind.

V bezeichnet in diesem Paragraphen stets einen reellen Vektorraum.

1. Topologische Redeweisen für reelle Algebren. Eine Abbildung $V \to \mathbb{R}$, $x \mapsto |x|$, heißt eine *Norm(funktion)*, wenn für alle $x, y \in V$, $\alpha \in \mathbb{R}$ gilt:

$$|x| > 0 \text{ für } x \neq 0,\ |\alpha x| = |\alpha|\,|x|,\ |x + y| \leqslant |x| + |y| \text{ (Dreiecksungleichung).}$$

Jeder euklidische Vektorraum V mit Skalarprodukt $\langle x, y\rangle$ besitzt die Norm

$$V \to \mathbb{R}, \qquad x \mapsto |x| := +\sqrt{\langle x, x\rangle} \qquad \text{(euklidische Länge).}$$

In Vektorräumen mit Norm stehen alle topologischen Redeweisen, wie „konvergente Folge, offene bzw. abgeschlossene bzw. kompakte Menge, Stetigkeit usf." zur Verfügung. Die Abbildung $V \to \mathbb{R}$, $x \mapsto |x|$, ist stetig; ebenso sind Vektoraddition und Skalarenmultiplikation, das heißt, die Abbildungen

$$V \times V \to V, \quad (x, y) \mapsto x + y \qquad \text{und} \qquad \mathbb{R} \times V \to V, \quad (\alpha, x) \mapsto \alpha x$$

stetig (Beweis!). In endlich-dimensionalen Räumen darf man mit jeder Norm arbeiten, sie führen alle zur gleichen Topologie. In jedem *normierten* Raum V definiert man die *Einheitssphäre*

$$S := \{x \in V : |x| = 1\};$$

für endlich-dimensionale Räume V ist S immer *kompakt* (Heine-Borel).

Es sei nun $\mathscr{A} = (V, \cdot)$ eine *endlich-dimensionale*, reelle Algebra und $x \mapsto |x|$ eine Norm des unterliegenden Vektorraumes. Dann gilt:

(1) *Die Multiplikation* $\mathscr{A} \times \mathscr{A} \to \mathscr{A}$, $(x, y) \mapsto xy$ *ist stetig.*
(2) *Es gibt ein* $\sigma \geqslant 0$, *so daß* $|xy| \leqslant \sigma |x|\,|y|$ *für alle* $x, y \in V$.
(3) *Ist* $\mathscr{A}$ *eine Divisionsalgebra, so gibt es ein* $\rho > 0$, *so daß*

$$|xy| \geqslant \rho |x|\,|y| \qquad \text{für alle} \qquad x, y \in \mathscr{A}.$$

Beweis. ad (1): Ist $v_1, \ldots, v_n$ eine Basis von V, so folgt für $x := \xi_1 v_1 + \cdots + \xi_n v_n$ und $y = \eta_1 v_1 + \cdots + \eta_n v_n$, daß xy als Summe der Terme $\xi_i \eta_j v_i v_j$ stetig ist.

ad (2) bzw. (3): Wegen $|\alpha x| = |\alpha|\,|x|$ genügt es, die Existenz von Zahlen $\sigma \geqslant \rho > 0$ zu zeigen, so daß für alle $x, y \in \mathscr{A}$ mit $|x| = |y| = 1$ gilt: $\rho \leqslant |xy| \leqslant \sigma$. Das aber ist klar, denn die Abbildung $S \times S \to \mathbb{R}$, $(x, y) \mapsto |xy|$ ist nach (1) stetig und nimmt also, da $S \times S$ kompakt ist, ein *Maximum* $\sigma \geqslant 0$ und ein *Minimum* ρ an; dabei gilt im Falle einer Divisionsalgebra notwendig $\rho > 0$, da Punkte $x, y \in S \subset \mathscr{A} \setminus \{0\}$ mit $xy = 0$ nicht existieren. □

Als Anwendung von (1) und (3) zeigen wir nun

Satz. *Es sei* $\mathscr{A}$ *eine endlich-dimensionale, reelle Divisionsalgebra. Dann ist die Menge* $\{x^2 : x \in \mathscr{A} \setminus \{0\}\}$ *aller Quadrate* $\neq 0$ *abgeschlossen in* $\mathscr{A} \setminus \{0\}$.

Beweis. Es genügt zu zeigen: ist (x_n) eine Folge aus $\mathscr{A}$ und gilt $\lim x_n^2 = a$, so ist a ein Quadrat in $\mathscr{A}$. Die konvergente Folge (x_n^2) ist beschränkt, sei etwa $|x_n^2| \leqslant M$ für alle n mit einer Schranke $M > 0$. Nach (3) gibt es ein $\rho > 0$ mit $|x_n^2| \geqslant \rho |x_n|^2$ für alle n; damit ist auch die Folge (x_n) beschränkt; nach dem Satz von Weierstrass-Bolzano hat sie eine konvergente Teilfolge. Ist b deren Limes, so folgt $a = b^2$ wegen (1). □

Im nächsten Abschnitt werden wir unter anderem sehen, daß unter der zusätzlichen Annahme der Kommutativität die Menge aller Quadrate $\neq 0$ auch *offen in* $\mathscr{A} \setminus \{0\}$ ist; zum Beweis sind Hilfsmittel aus der Differentialrechnung erforderlich.

2. Die Quadratabbildung $\mathscr{A} \to \mathscr{A}$, $x \mapsto x^2$. Eine Abbildung $f\colon V \to V$ heißt *differenzierbar im Punkt* $v \in V$, wenn es eine *lineare* Abbildung $f'(v)\colon V \to V$ gibt, so daß gilt:

$$\lim_{h \to 0, h \neq 0} \frac{|f(v+h) - f(v) - f'(v)(h)|}{|h|} = 0;$$

alsdann ist die Abbildung $f'(v)$ *eindeutig* bestimmt, sie heißt das *Differential* (oder auch die *Ableitung*) von f in v.

Wir betrachten wieder eine endlich-dimensionale reelle Algebra $\mathscr{A} = (V, \cdot)$ und fixieren auf V eine Norm. Jedes Element $a \in \mathscr{A}$ bestimmt vermöge *Links-* bzw. *Rechtsmultiplikation* zwei lineare Abbildungen

$$L_a\colon V \to V, \quad x \mapsto ax; \qquad R_a\colon V \to V, \quad x \mapsto xa.$$

Lemma. *Die Quadratabbildung* $q\colon \mathscr{A} \to \mathscr{A}$, $x \mapsto x^2$ *ist in jedem Punkt* $a \in \mathscr{A}$ *differenzierbar; es gilt:* $q'(a) = L_a + R_a$.

Beweis. Es gilt $|h^2| \leqslant \sigma |h|^2$ nach 1.(2). Wegen

$$q(a+h) - q(a) = (a+h)^2 - a^2 = ah + ha + h^2 = (L_a + R_a)h + h^2$$

folgt daher

$$\frac{|q(a+h) - q(a) - (L_a + R_a)h|}{|h|} = \frac{|h^2|}{|h|} \leqslant \sigma |h|,$$

so daß sich im Limes null ergibt. □

Eine überall differenzierbare Abbildung $f\colon V \to V$ induziert die Abbildung $f'\colon V \to \mathrm{Hom}(V, V)$, $v \mapsto f'(v)$. Man nennt f *stetig differenzierbar*, wenn f' stetig ist (beachte, daß mit V auch $\mathrm{Hom}(V, V)$ endlich-dimensional ist, so daß man hier von Stetigkeit reden kann). In der Differentialrechnung beweist man als Spezialfall des Theorems über implizite Funktionen

Lokaler Umkehrsatz. *Es sei* $f\colon V \to V$ *stetig differenzierbar, und es sei* $v \in V$ *ein Punkt, so daß die Ableitung* $f'(v)\colon V \to V$ *bijektiv ist. Dann ist* f *lokal-topologisch in* v, *das heißt, es gibt eine offene Umgebung* U *von* v, *so daß die Bildmenge* $f(U)$ *offen in* V *und die induzierte Abbildung* $f \mid U\colon U \to f(U)$ *topologisch ist.*

Es folgt nun schnell:

Satz. *Ist $\mathscr{A}$ kommutativ und nullteilerfrei, so ist $q\colon \mathscr{A} \to \mathscr{A}$, $x \mapsto x^2$, in jedem Punkt $a \neq 0$ lokal-topologisch; speziell ist $q(\mathscr{A} \setminus \{0\})$ offen in $\mathscr{A}$.*

Beweis. Da stets $ah = ha$, so gilt $L_a = R_a$, das heißt, $q'(a)h = 2L_a h = 2ah$, $h \in V$, aufgrund des Lemmas. Da $\mathscr{A}$ nullteilerfrei ist, so ist $q'(a)$ also für alle $a \neq 0$ bijektiv. Da q ersichtlich stetig differenzierbar ist, so ist q wegen des lokalen Umkehrsatzes in jedem Punkt $a \neq 0$ lokal-topologisch. Die Offenheit von $q(\mathscr{A} \setminus \{0\})$ in $\mathscr{A}$ ist dann trivial.

3. Satz von HOPF. Aus den Sätzen 1 und 2 ergibt sich zunächst

Ist $\mathscr{A}$ eine endlich-dimensionale, reelle, kommutative Divisionsalgebra mit $\dim \mathscr{A} > 1$, *so ist die Abbildung $q\colon \mathscr{A} \setminus \{0\} \to \mathscr{A} \setminus \{0\}$, $x \mapsto x^2$, surjektiv und überall lokal-topologisch.*

Jeder Punkt aus $\mathscr{A} \setminus \{0\}$ hat genau zwei q-Urbilder.

Beweis. Aufgrund der Sätze 1 und 2 ist die Menge $q(\mathscr{A} \setminus \{0\})$ sowohl abgeschlossen als auch offen in $\mathscr{A} \setminus \{0\}$. *Da $\mathscr{A} \setminus \{0\}$ wegen* $\dim \mathscr{A} > 1$ *zusammenhängend ist,* folgt $q(\mathscr{A} \setminus \{0\}) = \mathscr{A} \setminus \{0\}$. Nach Satz 2 ist q überall in $\mathscr{A} \setminus \{0\}$ lokal-topologisch.

Sind $u, v \in \mathscr{A} \setminus \{0\}$ Urbilder desselben Punktes aus $\mathscr{A} \setminus \{0\}$, so gilt wegen der Kommutativität: $0 = v^2 - u^2 = (v - u)(v + u)$. Da $\mathscr{A}$ nullteilerfrei ist, folgt $v = \pm u$, so daß jeder Punkt $\neq 0$ genau zwei Urbilder hat. □

Die soeben bewiesene Aussage beinhaltet speziell, daß *jedes* Element von $\mathscr{A}$ ein Quadrat ist.

Die bisherigen Überlegungen dieses Paragraphen benutzen Hilfsmittel aus der Analysis, die heute im Prinzip jedem Studierenden mittleren Semesters vertraut sind. Um nun den eigentlichen Satz von HOPF beweisen zu können, benötigen wir zusätzlich noch folgenden Satz aus der algebraischen Topologie:

Jeder Raum $\mathbb{R}^m \setminus \{0\}$, $m \geqslant 3$, ist einfach-zusammenhängend, das heißt, jede unverzweigte und unbegrenzte Überlagerung von $\mathbb{R}^m \setminus \{0\}$ ist einblättrig.

Mit dieser zusätzlichen Information folgt nun trivial der berühmte

Satz von HOPF (1940). *Jede endlich-dimensionale, reelle, kommutative Divisionsalgebra $\mathscr{A}$ ist höchstens zweidimensional.*

Beweis. Sei $n := \dim \mathscr{A} > 1$. Dann ist die Abbildung $q\colon \mathscr{A} \setminus \{0\} \to \mathscr{A} \setminus \{0\}$, $x \mapsto x^2$, nach dem eingangs Gezeigten surjektiv und überall lokal-topologisch mit stets zweipunktigen Fasern $q^{-1}(a)$, $a \in \mathscr{A} \setminus \{0\}$. Durch q wird daher der Raum $\mathscr{A} \setminus \{0\} \simeq \mathbb{R}^n \setminus \{0\}$ sich selbst unbegrenzt und unverzweigt *zweiblättrig* überlagert. Dies ist aufgrund des zitierten Satzes aus der algebraischen Topologie nur für $n \leqslant 2$ möglich. □

Hopf hat seinen Satz im Entdeckungsjahr 1940 sofort noch wesentlich verallgemeinert. Unter Verzicht auf die Forderung der Kommutativität konnte er zeigen („Ein topologischer Beitrag zur reellen Algebra", Comment. Math. Helv. 13, 219–239 (1940), insb. S. 229):

Die Dimension einer endlich-dimensionalen, reellen Divisionsalgebra ist notwendigerweise eine Potenz von 2.

Näheres und Vertiefendes hierzu findet der Leser im Kapitel 10.

4. Der ursprüngliche HOPFsche Beweis. Unser Beweis des Hopfschen Satzes ist eine Adaption des ursprünglichen Hopfschen Beweises. Hopf selbst betrachtet 1940 die stetige Abbildung

$$g\colon \mathscr{A} \setminus \{0\} \to \mathscr{A}, \qquad x \mapsto \frac{x^2}{|x^2|}.$$

Jeder Bildvektor hat die Länge 1; daher wird $\mathscr{A} \setminus \{0\}$ in die $(n-1)$-dimensionale Sphäre

$$S^{n-1} := \{v \in V\colon |v| = 1\}, \qquad n := \dim \mathscr{A},$$

abgebildet. Für alle $x \in \mathscr{A} \setminus \{0\}$, $\alpha \in \mathbb{R} \setminus \{0\}$ gilt offensichtlich $g(\alpha x) = g(x)$; die Abbildung g bildet also jede Gerade durch 0 auf denselben Punkt ab. Nun ist der reell-projektive Raum $\mathbb{P}^{n-1}$ nichts anderes als der Raum aller Geraden in V durch 0. Damit ist die berühmte „Hopfabbildung"

$$h\colon \mathbb{P}^{n-1} \to S^{n-1}$$

konstruiert. Dies geht alles noch für beliebige Divisionsalgebren $\mathscr{A}$; zusätzlich gilt aber:

Ist $\mathscr{A}$ kommutativ, so ist $h\colon \mathbb{P}^{n-1} \to S^{n-1}$ injektiv.

Beweis. Seien $\hat{x}, \hat{y} \in \mathbb{P}^{n-1}$ Punkte mit $h(\hat{x}) = h(\hat{y})$. Wir repräsentieren $\hat{x}, \hat{y}$ durch Punkte $x, y \in V \setminus \{0\}$. Die Gleichung $h(\hat{x}) = h(\hat{y})$ bedeutet dann, wenn wir abkürzend $\xi := \sqrt{|x^2|}$, $\eta := \sqrt{|y^2|}$, $\alpha := \xi^{-1}\eta \in \mathbb{R}$ setzen:

$$(\xi^{-1}x)^2 = (\eta^{-1}y)^2, \qquad \text{das heißt,} \qquad y^2 = \alpha^2 x^2.$$

Da $\mathscr{A}$ kommutativ und nullteilerfrei ist, folgt: $0 = y^2 - \alpha^2 x^2 = (y - \alpha x)(y + \alpha x)$, also $y = \pm \alpha x$, das heißt, $\hat{y} = \hat{x}$. □

So hat Hopf jeder n-dimensionalen, reellen, kommutativen Divisionsalgebra eine *topologische* Abbildung des projektiven Raumes $\mathbb{P}^{n-1}$ *in* die Sphäre S^{n-1} zugeordnet. Und nun argumentiert er (wörtlich) wie folgt (vgl. Selecta, S. 112): „..., da $\mathbb{P}^{n-1}$ und S^{n-1} geschlossene Mannigfaltigkeiten der gleichen Dimension $n-1$ sind, muß S^{n-1} mit dem Bild von $\mathbb{P}^{n-1}$ identisch, die Mannigfaltigkeiten S^{n-1} und $\mathbb{P}^{n-1}$ müssen also homöomorph sein*). Für $n - 1 = 1$ ist dies in der Tat der Fall: sowohl der Kreis S^1 als auch die projektive

*) Hopf schreibt r statt n und P_{r-1} bzw. S_{r-1} statt $\mathbb{P}^{n-1}$ bzw. S^{n-1}; eine geschlossene Mannigfaltigkeit ist eine kompakte Mannigfaltigkeit ohne Rand. Hopf benutzt hier den tief liegenden Satz, daß eine injektive stetige Abbildung $f\colon X \to Y$ zwischen zusammenhängenden, geschlossenen und gleichdimensionalen Mannigfaltigkeiten stets eine topologische Abbildung ist.

Gerade ist eine einfach geschlossene Linie. Ist aber $n-1>1$, so ist die Sphäre S^{n-1} einfach-zusammenhängend – im Gegensatz zu dem Fall $n-1=1$ –, während der projektive Raum $\mathbb{P}^{n-1}$ niemals einfach-zusammenhängend ist, da sich in ihm die projektive Gerade nicht in einen Punkt deformieren läßt; die fragliche Homöomorphie liegt also für $n-1>1$ nicht vor." Damit hat Hopf $n-1=1$, also $n=2$, gezeigt. □

Man kennt bis heute keinen „elementaren" Beweis für den Satz von Hopf. 1954 hat der holländische Mathematiker T. A. Springer in seiner Arbeit „An algebraic proof of a theorem of H. Hopf" (Indagationes Mathematicae 16, 33–35) einen Beweis mitgeteilt, der statt des einfachen Zusammenhanges Hilfsmittel aus der algebraischen Geometrie benutzt, u. a. den Satz von Bézout.

5. Beschreibung aller 2-dimensionalen Algebren mit Einselement. Jede 2-dimensionale reelle Algebra $\mathfrak{A}$ mit Einselement e hat eine Basis e, w mit $w^2=\omega e$, wobei $\omega=0$ oder $\omega=1$ oder $\omega=-1$ (Beweis!). Hieraus folgt

Lemma. *Jede 2-dimensionale reelle Algebra $\mathfrak{A}$ mit Einselement ist kommutativ und assoziativ. Es sind drei (sich gegenseitig ausschließende) Fälle möglich:*

1) *$\mathfrak{A}$ ist isomorph zur „Algebra $(\mathbb{R}^2,\cdot)$ der dualen Zahlen", das heißt, $(1,0)\in\mathbb{R}^2$ ist Einselement, und für $\varepsilon:=(0,1)\in\mathbb{R}^2$ gilt $\varepsilon^2=0$.*

2) *$\mathfrak{A}$ ist isomorph zur „direkten Summe $\mathbb{R}\oplus\mathbb{R}$", das heißt, für $a:=(1,0)$, $b:=(0,1)\in\mathbb{R}^2$ gilt $a^2=a$, $b^2=b$, $ab=0$* (vgl. R.2, 6)).

3) *$\mathfrak{A}$ ist isomorph zur Algebra $\mathbb{C}$.*

Beweis. Die Fälle $\omega=0$, $\omega=1$, $\omega=-1$ führen zu den Fällen 1), 2), 3). Im Fall $\omega=1$ ist $u:=\frac{1}{2}(e+w)$, $v:=\frac{1}{2}(e-w)$ eine Basis von $\mathfrak{A}$ mit $u^2=u$, $v^2=v$, $uv=0$, daher ist $\mathfrak{A}\to\mathbb{R}\oplus\mathbb{R}$, $\alpha u+\beta v\mapsto\alpha a+\beta b$ ein Isomorphismus. □

Aus dem Satz von Hopf und dem Lemma folgt direkt (da $\mathfrak{A}$ in den Fällen 1) und 2) Nullteiler hat):

Korollar zum Satz von HOPF. *Jede endlich-dimensionale, reelle, kommutative Divisionsalgebra $\mathfrak{A}$ mit Einselement e ist zu $\mathbb{R}$ oder $\mathbb{C}$ isomorph.*

Aufgabe. Finden Sie den Fehler im folgenden „direkten Beweis" zum Korollar des Satzes von Hopf: „Falls $n:=\dim\mathfrak{A}>1$, so gibt es ein $j\in\mathfrak{A}$ mit $j^2=-e$. Dann ist $\mathfrak{B}:=\mathbb{R}e\oplus\mathbb{R}j$ eine zu $\mathbb{C}$ isomorphe Unteralgebra von $\mathfrak{A}$. Für jedes $a\in\mathfrak{A}$ hat das charakteristische Polynom $\det(L_a-X\cdot\mathrm{id})$ der Linksmultiplikation $L_a\colon\mathfrak{A}\to\mathfrak{A}$, $x\mapsto ax$, nach dem Fundamentalsatz der Algebra eine Nullstelle $b\in\mathfrak{B}$, das heißt, es gibt ein $c\neq0$ in $\mathfrak{A}$ mit $(a-be)c=0$. Es folgt $a=be=b\in\mathfrak{B}$, das heißt $\mathfrak{A}=\mathfrak{B}$."

In den hier angestellten Überlegungen ist die Voraussetzung, daß $\mathfrak{A}$ ein Einselement hat, wesentlich. Es gibt unendlich viele nicht-isomorphe kommutative, 2-dimensionale Divisionsalgebren, z. B. entsteht aus $\mathbb{C}$ eine solche Algebra, wenn man die Multiplikation von $w,z\in\mathbb{C}$ durch $w\circ z:=\overline{wz}$ erklärt. Die Familie aller dieser (nicht isomorphen) Algebren ist zweidimensional und nicht zusammenhängend.

Aufgabe. Zeigen Sie, daß jede 2-dimensionale, alternative und kommutative Divisionsalgebra zu $\mathbb{C}$ isomorph ist.

Kapitel 8. CAYLEY-Zahlen oder alternative Divisionsalgebren

M. Koecher, R. Remmert

> It is possible to form an analogous theory with seven imaginary roots of (-1) (A. CAYLEY 1845).

Die Erfindung der Quaternionen ist der Anfang einer neuen Epoche in der Algebra. Mit der HAMILTONschen Schöpfung eines „Systems hyperkomplexer Zahlen", das nicht mehr kommutativ ist, setzt ein Prozeß des Umdenkens ein: Mathematiker beginnen zu begreifen, daß man bei Verzicht auf das vage Permanenzprinzip auf mannigfache Weise neue Zahlensysteme „aus dem Nichts" schaffen kann, die noch weiter als die Quaternionen von den reellen und komplexen Zahlen entfernt sind. So erfand J. T. GRAVES bereits im Dezember 1843, zwei Monate nach HAMILTONS Erfindung, die *acht-dimensionale Divisionsalgebra der Oktaven* (Oktonionen), die – wie HAMILTON 1844 bemerkte – *nicht mehr assoziativ* ist. (GRAVES teilte HAMILTON seine Untersuchungen über die Oktaven in einem Brief vom 4. Jan. 1844 mit; sie wurden aber erst 1848 veröffentlicht (*Note by Professor Sir* W. R. HAMILTON, *respecting the Researches of* John T. GRAVES, *Esq.*, Trans. Roy. Irish Acad. (1848), Science 338–341)). Die Oktaven wurden 1845 von Arthur CAYLEY wiedergefunden und als Postscript in einer Arbeit über elliptische Funktionen veröffentlicht (Math. Papers 1, 127), sie heißen seither CAYLEYsche Zahlen.

Da in der CAYLEY-Algebra das Assoziativgesetz verletzt ist, kann man grundsätzlich nicht mehr (wie bei Quaternionen) den Matrizenkalkül zur Erleichterung von Rechnungen heranziehen. So ist es unvermeidbar, daß die Herleitung der wesentlichen Formeln, die uns für die Algebren $\mathbb{C}$ und $\mathbb{H}$ wohlvertraut sind, mühsamer wird. Im einleitenden Paragraphen 1 haben wir systematisch die wesentlichen Identitäten, die allgemein für *alternative* (*nullteilerfreie*) *quadratische* Algebren gelten, zusammengestellt. Im Paragraphen 2 wird die Algebra $\mathbb{O}$ der Oktaven durch Verdopplung der Algebra $\mathbb{H}$ der Quaternionen explizit konstruiert.

In Analogie zum Einzigkeitssatz von FROBENIUS für Quaternionen gibt es auch einen Einzigkeitssatz für Oktaven. Diesen Satz hat 1933 Max ZORN (bekannt durch das ZORNsche Lemma) entdeckt und in seiner Arbeit „Alternativkörper und quadratische Systeme", Abh. Math. Sem. Hamburg 9, 395–402, veröffentlicht. Wir leiten den Satz von ZORN im Paragraphen 3 her.

§ 1. Alternative quadratische Algebren

Für *jede* endlich-dimensionale, reelle quadratische Algebra $\mathfrak{A}$ gilt $\mathfrak{A} = \mathbb{R}e \oplus \operatorname{Im}\mathfrak{A}$, wobei der Imaginärraum $\operatorname{Im}\mathfrak{A}$ ein Untervektorraum von $\mathfrak{A}$ ist

(Lemma von FROBENIUS). Es gibt daher *genau eine* Linearform

$$(1) \qquad \lambda: \mathscr{A} \to \mathbb{R} \qquad \text{mit} \qquad \lambda(e) = 1 \qquad \text{und} \qquad \operatorname{Kern} \lambda = \operatorname{Im} \mathscr{A};$$

wir nennen λ *die Linearform der quadratischen Algebra*. In den Fällen $\mathscr{A} = \mathbb{C}, \mathbb{H}$ ist λ die in 3.2.2 bzw. 6.2.1 eingeführte und wesentlich benutzte *Realteil*-Linearform Re. Es ist nicht üblich, die Schreibweise Re im Allgemeinfall zu verwenden.

In den Algebren $\mathbb{C}$ und $\mathbb{H}$ leistet die Konjugierungsabbildung $x \mapsto \bar{x}$ gute Dienste. Diese Abbildung läßt sich in *jeder* quadratischen $\mathbb{R}$-Algebra $\mathscr{A}$ basisinvariant durch

$$(2) \qquad \bar{}: \mathscr{A} \to \mathscr{A}, \qquad x \mapsto \bar{x} := 2\lambda(x)e - x,$$

definieren, sie ist $\mathbb{R}$-*linear*, und es gilt:

$$\bar{x} = \lambda(x)e - u \qquad \text{für} \qquad x = \lambda(x)e + u, \qquad u \in \operatorname{Im} \mathscr{A};$$

mithin ist $x \mapsto \bar{x}$ wie bei $\mathbb{C}, \mathbb{H}$ eine *Spiegelung* an der Geraden $\mathbb{R}e$ von $\mathscr{A}$; speziell gilt stets:

$$\bar{\bar{x}} = x \quad (\textit{Involution}), \qquad \lambda(\bar{x}) = \lambda(x);$$

die *Fixpunktmenge* $\{x \in \mathscr{A} : \bar{x} = x\}$ ist die $\mathbb{R}$-Unteralgebra $\mathbb{R}e$. □

Man wird nun, wenn man sich an den Beispielen $\mathbb{C}$ und $\mathbb{H}$ orientiert, erwarten, daß durch $\langle x, y\rangle := \lambda(x\bar{y})$ ein „natürliches" Skalarprodukt für $\mathscr{A}$ gegeben wird. Wir werden in diesem Paragraphen sehen, daß dies in der Tat für *nullteilerfreie* Algebren richtig ist; allerdings werden wir $\langle x, y\rangle$ im Abschnitt 1 zunächst anders definieren und obige „Wunschgleichung" im Abschnitt 3 nur unter der zusätzlichen Annahme, daß $\mathscr{A}$ *alternativ* ist, herleiten. Für alternative Algebren gilt dann auch wie für $\mathbb{C}$ und $\mathbb{H}$ die wichtige Produktregel $|xy| = |x|\,|y|$.

Die in diesem Paragraphen gewonnenen allgemeinen Erkenntnisse werden in den Paragraphen 2 und 3 zu nicht-trivialen Einsichten in die Struktur der Algebra $\mathbb{O}$ der Oktaven führen.

1. Die Bilinearform. In den Algebren $\mathbb{C}$ und $\mathbb{H}$ besteht zwischen natürlichem Skalarprodukt und Linearform λ die (für $\mathbb{C}$ triviale) Identität $\langle x, y\rangle = 2\lambda(x)\lambda(y) - \lambda(xy)$. Für allgemeine Algebren machen wir diese Gleichung in leicht modifizierter Form nun zur Definition der Bilinearform; wir zeigen:

Lemma. *Es sei $\mathscr{A}$ eine quadratische Algebra mit Linearform λ. Dann ist*

$$\mathscr{A} \times \mathscr{A} \to \mathbb{R}, \qquad (x, y) \mapsto \langle x, y\rangle := 2\lambda(x)\lambda(y) - \tfrac{1}{2}\lambda(xy + yx)$$

eine symmetrische Bilinearform. Für alle $x, y \in \mathscr{A}$ gilt:

$$(1) \qquad \langle x, x\rangle = 2\lambda(x)^2 - \lambda(x^2),$$

$$(2) \qquad \langle x, e\rangle = \lambda(x), \qquad \langle e, e\rangle = 1,$$

$$(3) \qquad x^2 = 2\lambda(x)x - \langle x, x\rangle e,$$

$$(4) \qquad xy + yx = 2\lambda(x)y + 2\lambda(y)x - 2\langle x, y\rangle e.$$

Ist $\mathscr{A}$ zusätzlich nullteilerfrei, so gilt $\langle x, x\rangle > 0$ für alle $x \neq 0$.

Beweis. Per definitionem ist $\langle x, y\rangle$ eine symmetrische Bilinearform, für die (1) und (2) gelten. Zum Nachweis von (3) gilt zunächst $x - \lambda(x)e \in \operatorname{Kern}\lambda = \operatorname{Im}\mathscr{A}$. Laut Definition von $\operatorname{Im}\mathscr{A}$ folgt (vgl. 7.1.1): $(x - \lambda(x)e)^2 = -\omega(x)e$ mit $\omega(x) \in \mathbb{R}$, $x \in \mathscr{A}$. Schreiben wir dies in der Form $x^2 = 2\lambda(x)x - [\lambda(x)^2 + \omega(x)]e$ und wenden wir hierauf λ an, so ergibt sich $\langle x, x\rangle = \lambda(x)^2 + \omega(x)$ wegen (1). Damit folgt (3). Im nullteilerfreien Fall gilt stets $\omega(x) \geqq 0$ (vgl. 7.1.1); dann entnimmt man der letzten Gleichung $\langle x, x\rangle > 0$ für alle $x \neq 0$, denn $\lambda(x) = \omega(x) = 0$ ist nur für $x = 0$ möglich.

Gleichung (4) folgt durch Linearisieren von (3) (man setze $x+y$ statt x). □

Die im Lemma eingeführte Bilinearform $\langle x, y\rangle$ wird im folgenden eine zentrale Rolle spielen, wir nennen sie *die Bilinearform der quadratischen Algebra*. Der Leser beachte, daß die Gleichung (3) eine *universelle* quadratische Gleichung für jedes $x \in \mathscr{A}$ gibt; in der ursprünglichen Definition quadratischer Algebren (vgl. 7.2.E) wurde nur gefordert, daß zu jedem x „irgendwie" Elemente $\alpha, \beta \in \mathbb{R}$ mit $x^2 = \beta x + \alpha e$ existieren; jetzt hat sich gezeigt, daß α, β in natürlicher Weise als $\alpha = -\langle x, x\rangle$, $\beta = 2\lambda(x)$ wählbar sind.

Mit der Identität (4) kann man xy durch yx ausdrücken und damit oft Formeln vereinfachen.

2. Satz über die Bilinearform. *Für alle Elemente x, y einer quadratischen alternativen Algebra $\mathscr{A}$ gilt:*

(1) $\lambda(xy) = \lambda(yx)$ *und also* $\langle x, y\rangle = 2\lambda(x)\lambda(y) - \lambda(xy)$,

(2) $\langle xy, xy\rangle = \langle x, x\rangle\langle y, y\rangle$ (*Produktregel*).

Ist die Bilinearform von $\mathscr{A}$ positiv definit, dann ist $\mathscr{A}$ nullteilerfrei.

Beweis. Die Behauptungen (1) und (2) sind trivial, wenn e und y linear abhängig sind. Wir setzen daher e und y als linear unabhängig voraus. Da $\mathscr{A}$ alternativ ist, gilt stets $y(xy) = (yx)y$ (vgl. 7.1.3). Zweimalige Anwendung von 1.(4) gibt daher

$$\begin{aligned} 0 &= y(xy) - (yx)y = y(xy) + (xy)y - (xy + yx)y \\ &= 2[\lambda(y)xy + \lambda(xy)y - \langle y, xy\rangle e] - 2[\lambda(x)y + \lambda(y)x - \langle x, y\rangle e]y \\ &= 2\lambda(xy)y - 2\langle y, xy\rangle e - 2\lambda(x)y^2 + 2\langle x, y\rangle y. \end{aligned}$$

Wegen 1.(3) erhält man hieraus

$$0 = [\lambda(x)\langle y, y\rangle - \langle xy, y\rangle]e + [\lambda(xy) - 2\lambda(x)\lambda(y) + \langle x, y\rangle]y.$$

Da e und y linear unabhängig sind, so folgt

$$\langle xy, y\rangle = \lambda(x)\langle y, y\rangle \quad \text{und} \quad \lambda(xy) = 2\lambda(x)\lambda(y) - \langle x, y\rangle.$$

Die letzte Gleichung gibt $\lambda(xy) = \frac{1}{2}\lambda(xy + yx)$ laut Definition von $\langle x, y\rangle$ und damit (1). Linearisierung von (1) gibt (mit $y + z$ anstelle von y):

$$\langle xy, z\rangle + \langle xz, y\rangle = 2\lambda(x)\langle y, z\rangle.$$

Setzt man hier $z := xy$, so folgt wegen $x^2 = 2\lambda(x)x - \langle x, x\rangle e$ die Gleichung (2):

$$\begin{aligned}\langle xy, xy\rangle &= 2\lambda(x)\langle y, xy\rangle - \langle x^2 y, y\rangle \\ &= 2\lambda(x)\langle y, xy\rangle - 2\lambda(x)\langle xy, y\rangle + \langle x, x\rangle\langle y, y\rangle = \langle x, x\rangle\langle y, y\rangle.\end{aligned}$$

Aus $xy = 0$ folgt $\langle x, x\rangle\langle y, y\rangle = 0$ wegen (2) und also $\langle x, x\rangle = 0$ oder $\langle y, y\rangle = 0$. Im positiv definiten Fall resultiert $x = 0$ oder $y = 0$.

3. Satz über die Konjugierungsabbildung. *Für alle Elemente x, y einer quadratischen, alternativen Algebra $\mathscr{A}$ gilt:*

(1) $\qquad \overline{xy} = \bar{y}\bar{x},$

(2) $\qquad x(\bar{x}y) = \bar{x}(xy) = \langle x, x\rangle y, \qquad$ *speziell* $\qquad x\bar{x} = \bar{x}x = \langle x, x\rangle e,$

(3) $\qquad \langle x, y\rangle = \lambda(x\bar{y}) = \lambda(\bar{x}y).$

Beweis. Gleichung (1) folgt aus

$$\begin{aligned}\overline{xy} - \bar{y}\bar{x} &= 2\lambda(xy)e - xy - [2\lambda(y)e - y][2\lambda(x)e - x] \\ &= 2[\lambda(xy) - 2\lambda(x)\lambda(y)]e + 2\lambda(x)y + 2\lambda(y)x - (xy + yx),\end{aligned}$$

da wegen 1.(4) und 2.(1) rechts null steht. Gleichung (2) folgt mittels 1.(3):

$$x(\bar{x}y) = x[(2\lambda(x)e - x)y] = x(2\lambda(x)y - xy) = 2\lambda(x)xy - x^2 y = \langle x, x\rangle y;$$

analog zeigt man $\bar{x}(xy) = \langle x, x\rangle y$.

Um (3) zu verifizieren, beachten wir, daß wegen $x = 2\lambda(x)e - \bar{x}$ gilt: $\lambda(xy) = 2\lambda(x)\lambda(y) - \lambda(\bar{x}y)$. Damit folgt (3) mittels 2.(1). □

Mit (2) gelten auch die Gleichungen

(2′) $\qquad (x\bar{y})y = (xy)\bar{y} = \langle y, y\rangle x,$

das beweist man entweder durch Konjugierung von (2) oder analog wie (2). □

In jeder alternativen Algebra $\mathscr{A}$ ist das *dreifache* Produkt

$$axa := (ax)a = a(xa), \qquad x, a \in \mathscr{A},$$

von der Klammerung unabhängig (vgl. 7.1.3). Wie bei Quaternionen (vgl. 6.2.2 läßt es sich als Linearkombination in a und $\bar{x}$ ausdrücken.

Ist $\mathscr{A}$ quadratisch und alternativ, so gilt

(4) $\qquad axa = 2\langle \bar{x}, a\rangle a - \langle a, a\rangle \bar{x}, \qquad a, x \in \mathscr{A} \qquad$ *(Dreier-Identität).*

Beweis. Aus $xa + \overline{xa} = 2\lambda(xa)e$ folgt durch Linksmultiplikation mit a, wenn man $\overline{xa} = \bar{a}\bar{x}$ beachtet: $axa + a(\bar{a}\bar{x}) = 2\lambda(xa)a$. Da $a(\bar{a}\bar{x}) = \langle a, a\rangle \bar{x}$ nach (2) und $\lambda(xa) = \langle \bar{x}, a\rangle$ nach (3), so folgt (4).

4. Der euklidische Vektorraum $\mathscr{A}$ und die orthogonale Gruppe $O(\mathscr{A})$. In den Resultaten der Abschnitte 1.–3. ist enthalten:

Satz. *Ist $\mathscr{A}$ eine quadratische, alternative, nullteilerfreie Algebra, so ist $\mathscr{A}$ bezüglich der Bilinearform $\langle\ ,\ \rangle$ ein euklidischer Vektorraum. Es gilt die*

Produktregel: $|xy| = |x|\,|y|$ *für* $x, y \in \mathscr{A}$;

speziell sind alle Abbildungen $p_a: \mathscr{A} \to \mathscr{A}$, $x \mapsto axa$, $a \in \mathscr{A}$, $|a| = 1$, Isometrien von $\mathscr{A}$.

Als Verallgemeinerung des Erzeugungssatzes 6.3.2 für $O(\mathbb{H})$ gilt

Erzeugungssatz für $O(\mathscr{A})$. *Es sei $\mathscr{A}$ eine endlich-dimensionale, alternative Divisionsalgebra (also speziell eine quadratische Algebra nach 7.2.2). Dann ist jede eigentliche Isometrie $f \in O^+(\mathscr{A})$ ein Produkt aus höchstens $n := \dim \mathscr{A}$ Abbildungen p_a.*

Die volle Gruppe $O(\mathscr{A})$ wird erzeugt von den Abbildungen

$$x \mapsto axa, \qquad |a| = 1 \qquad \text{und} \qquad x \mapsto \bar{x}.$$

Beweis (analog wie der Beweis von Satz 6.3.2). 1) Jedes $f \in O^+(\mathscr{A})$ ist Produkt einer geraden Anzahl $k \leqslant n$ von Spiegelungen s_a. Für je zwei Spiegelungen s_a, s_b gilt wieder $s_a \circ s_b = p_a \circ p_{-\bar{b}}$, dies folgt analog wie früher, da auch jetzt $s_e(x) = -\bar{x}$ und (wegen 3.(4))

$$s_a \circ s_e(x) = -\bar{x} + 2\langle a, \bar{x}\rangle a = axa = p_a(x) \qquad \text{für} \qquad |a| = 1.$$

2) Für $f \in O^-(\mathscr{A})$ gilt $f \circ s_e \in O^+(\mathscr{A})$, daher wird $O(\mathscr{A})$ von den Abbildungen p_a zusammen mit der Konjugierung erzeugt. □

Warnung. Da das Assoziativgesetz nicht mehr zur Verfügung steht, läßt sich *nicht mehr ein Analogon zum Satz von* CAYLEY aus 6.3.2 aussprechen: jede Abbildung $x \mapsto a(xb)$ bzw. $x \mapsto (ax)b$, $|a| = |b| = 1$, ist zwar nach der Produktregel orthogonal, doch hat $\mathscr{A}$ im nichtassoziativen Fall weitere solche Abbildungen, so lassen sich z. B. die orthogonalen Abbildungen $x \mapsto [a(xb)]c$ mit $|a| = |b| = |c| = 1$ im allgemeinen nicht in der Form $x \mapsto u(xv)$ oder $x \mapsto (ux)v$ schreiben.

§ 2. Existenz und Eigenschaften der CAYLEY-Algebra $\mathbb{O}$

Die Algebra $\mathbb{C}$ entsteht nach HAMILTON aus der Algebra $\mathbb{R}$, wenn man im kartesischen Produkt $\mathbb{R} \times \mathbb{R}$ der reellen Zahlenpaare vermöge

$$(a_1, a_2)(b_1, b_2) = (a_1 b_1 - b_2 a_2, a_2 b_1 + b_2 a_1), \qquad a_1, a_2, b_1, b_2 \in \mathbb{R},$$

ein Produkt einführt. Durch einen analogen *Verdopplungsprozeß* läßt sich, bis auf Isomorphie, die Quaternionenalgebra $\mathbb{H}$ aus der Algebra $\mathbb{C}$ gewinnen: man führt im Produktraum $\mathbb{C} \times \mathbb{C}$ der komplexen Zahlen durch

$$(a_1, a_2)(b_1, b_2) = (a_1 b_1 - \bar{b}_2 a_2, a_2 \bar{b}_1 + b_2 a_1), \qquad a_1, a_2, b_1, b_2 \in \mathbb{C},$$

eine Multiplikation ein (dieses Vorgehen ist ganz kanonisch: mittels der Bijektion $\mathbb{C} \times \mathbb{C} \to \mathscr{H}$, $(a_1, a_2) \mapsto \begin{pmatrix} a_1 & a_2 \\ -\bar{a}_2 & \bar{a}_1 \end{pmatrix}$, wird die Matrizenmultiplikation von $\mathscr{H}$ nach $\mathbb{C} \times \mathbb{C}$ übertragen, vgl. 6.1.2). Wir werden im folgenden sehen, daß sich dieser *Verdopplungsprozeß* auch wiederum mit $\mathbb{H}$ ausführen läßt: dadurch entsteht als neue Algebra die CAYLEY-Algebra $\mathbb{O}$ der Oktaven.

1. Konstruktion der quadratischen Algebra $\mathbb{O}$ der Oktaven. Für Elemente $x = (x_1, x_2)$, $y = (y_1, y_2)$ aus $\mathbb{H} \times \mathbb{H}$ definieren wir – motiviert durch die Bemerkungen der Einleitung – ein Produkt vermöge

$$xy = (x_1, x_2)(y_1, y_2) := (x_1 y_1 - \bar{y}_2 x_2, x_2 \bar{y}_1 + y_2 x_1);$$

man verifiziert direkt, daß beide Distributivgesetze gelten. Somit ist $\mathbb{H} \times \mathbb{H}$ eine 8-*dimensionalen* $\mathbb{R}$-Algebra, wir nennen sie *die* CAYLEY-*Algebra der Oktaven* und bezeichnen sie mit $\mathbb{O}$. Es muß betont werden, daß in der Definition des Oktavenproduktes die Reihenfolge der Faktoren in der Klammer rechts für die weiteren Überlegungen ganz entscheidend ist; würde man z. B. in der zweiten Komponente $x_1 y_2$ statt $y_2 x_1$ schreiben, so erhielte man eine uninteressante Algebra.

Bezeichnet man mit e' das Einselement von $\mathbb{H}$, dann ist $e := (e', 0)$ das *Einselement von* $\mathbb{O}$. Es ergibt sich weiter direkt:

$\mathbb{O}$ ist eine quadratische Algebra: für jedes $x = (x_1, x_2) \in \mathbb{O}$ gilt:

$$x^2 = 2\,\mathrm{Re}(x_1)x - (\langle x_1, x_1\rangle + \langle x_2, x_2\rangle)e. \tag{1}$$

Beweis. Per definitionem gilt $x^2 = (x_1^2 - \bar{x}_2 x_2, x_2 \bar{x}_1 + x_2 x_1)$. Für Quaternionen wissen wir (vgl. 6.2.1 und 6.2.2): $x_1^2 = 2\,\mathrm{Re}(x_1)x_1 - \langle x_1, x_1\rangle e'$, $\bar{x}_2 x_2 = \langle x_2, x_2\rangle e'$, $\bar{x}_1 + x_1 = 2\,\mathrm{Re}(x_1)e'$. Damit folgt

$$x^2 = (2\,\mathrm{Re}(x_1)x_1 - (\langle x_1, x_1\rangle + \langle x_2, x_2\rangle)e', 2\,\mathrm{Re}(x_1)x_2).$$

2. Imaginärraum, Linearform, Bilinearform und Konjugierung von $\mathbb{O}$. Da $\mathbb{O}$ eine quadratische Algebra ist, so sind für $\mathbb{O}$ der Imaginärraum $\mathrm{Im}\,\mathbb{O}$, die Linearform λ, die Bilinearform $\langle x, y\rangle$ und die Abbildung $x \mapsto \bar{x}$ invariant definiert. Die Zusammenhänge mit den entsprechenden Bildungen in $\mathbb{H}$ sind, wenn Oktaven x, y als Quaternionenpaare (x_1, x_2), (y_1, y_2) geschrieben werden, einfach. Man verifiziert direkt:

$$\lambda(x) = \mathrm{Re}(x_1), \tag{1}$$

$$\langle x, y\rangle = \langle x_1, y_1\rangle + \langle x_2, y_2\rangle, \tag{2}$$

$$\mathrm{Im}\,\mathbb{O} = \mathrm{Im}\,\mathbb{H} \times \mathbb{H}, \tag{3}$$

$$\bar{x} = (\bar{x}_1, -x_2). \tag{4}$$

Da die Bilinearform von $\mathbb{H}$ positiv-definit ist, so hat (2) zur Folge:

Die Bilinearform von $\mathbb{O}$ ist positiv-definit, das heißt, $\mathbb{O}$ ist ein euklidischer Vektorraum.

3. $\mathbb{O}$ als alternative Divisionsalgebra. Wie für Quaternionen gelten die Identitäten

$$\overline{xy} = \bar{y}\bar{x}, \qquad x\bar{x} = \bar{x}x = \langle x, x\rangle e, \qquad x, y \in \mathbb{O}, \tag{1}$$

$$x(\bar{x}y) = \langle x, x\rangle y = (x\bar{x})y, \qquad x, y \in \mathbb{O}. \tag{2}$$

Beweis. ad (1): Die Behauptungen folgen wegen $\bar{x} = (\bar{x}_1, -x_2)$ direkt aus der Definition der Oktavenmultiplikation.

ad (2): Mit $x = (x_1, x_2)$, $y = (y_1, y_2)$ gilt

$$\bar{x}y = (\bar{x}_1, -x_2)(y_1, y_2) = (\bar{x}_1 y_1 + \bar{y}_2 x_2, -x_2\bar{y}_1 + y_2\bar{x}_1)$$

und folglich, da $\mathbb{H}$ assoziativ ist:

$$\begin{aligned} x(\bar{x}y) &= (x_1[\bar{x}_1 y_1 + \bar{y}_2 x_2] - [-y_1\bar{x}_2 + x_1\bar{y}_2]x_2,\ x_2[\bar{y}_1 x_1 + \bar{x}_2 y_2] \\ &\quad + [-x_2\bar{y}_1 + y_2\bar{x}_1]x_1) \\ &= (x_1\bar{x}_1 y_1 + y_1\bar{x}_2 x_2, x_2\bar{x}_2 y_2 + y_2\bar{x}_1 x_1) = (\langle x_1, x_1\rangle + \langle x_2, x_2\rangle)y. \end{aligned}$$

Dies ist wegen 2.(2) die Behauptung □

Satz. *Die Algebra $\mathbb{O}$ ist eine alternative Divisionsalgebra.*

Beweis. Mit $\bar{x} = 2\lambda(x)e - x$ schreibt sich (2) als $x(2\lambda(x)y - xy) = (2\lambda(x)x - x^2)y$, woraus sofort $x(xy) = x^2y$ für alle $x, y \in \mathbb{O}$ folgt. Durch Konjugation entsteht $(\bar{y}\bar{x})\bar{x} = \bar{y}\bar{x}^2$. Da mit x, y auch $\bar{x}, \bar{y}$ *alle* Elemente von $\mathbb{O}$ durchlaufen, so folgt $(yx)x = yx^2$ für alle $x, y \in \mathbb{O}$. Mithin ist $\mathbb{O}$ alternativ.

Da die Bilinearform von $\mathbb{O}$ positiv-definit ist, so ist $\mathbb{O}$ nullteilerfrei (vgl. Satz 1.2). Als endlich-dimensionale Algebra ist $\mathbb{O}$ mithin eine Divisionsalgebra (R.5). □

Für die Algebra $\mathbb{O}$ gilt nun aufgrund von Satz 1.3 stets

$$\langle x, y\rangle = \lambda(x\bar{y}) = \lambda(\bar{x}y), \qquad x, y \in \mathbb{O};$$

das läßt sich natürlich auch direkt verifizieren.

Die Algebra $\mathbb{O}$ ist nach dem Satz von FROBENIUS nicht assoziativ, so gilt z. B., wenn e, i, j, k die Standardbasis von $\mathbb{H}$ bezeichnet:

$$\begin{aligned} (0,e)[(0,i)(0,j)] &= -(0,e)(k,0) = (0,k), \\ [(0,e)(0,i)](0,j) &= (i,0)(0,j) = -(0,k); \end{aligned}$$

vgl. hierzu auch Abschnitt 6.

4. „Acht-Quadrate-Satz". Für die alternative, quadratische Algebra $\mathbb{O}$ gilt nach 1.2(3) die

Produktregel. $\quad |xy| = |x|\,|y| \quad$ *für* $\quad x, y \in \mathbb{O}.$

Dies läßt sich natürlich auch direkt, aber etwas mühsam, aus den Definitionen von $\mathbb{O}$ folgern: aufgrund der Produktregel für Quaternionen ist, wenn man $x = (x_1, x_2)$, $y = (y_1, y_2)$ schreibt und $xy = (x_1y_1 - \bar{y}_2x_2, x_2\bar{y}_1 + y_2x_1)$ sowie 2.(2) bedenkt, zu zeigen:

(*) $\quad |x_1y_1 - \bar{y}_2x_2|^2 + |x_2\bar{y}_1 + y_2x_1|^2 = (|x_1|^2 + |x_2|^2)(|y_1|^2 + |y_2|^2), \qquad x_1, x_2, y_1, y_2 \in \mathbb{H}.$

Ausrechnen führt auf $\lambda(x_2\bar{y}_1\bar{x}_1\bar{y}_2) = \lambda(x_1y_1\bar{x}_2y_2)$; was wegen 6.2.1(8) zutrifft! Man beachte die Analogie von (*) mit der Identität von GAUSS aus 6.2.3.

Aus der Produktregel für Oktaven folgt ein

„Acht-Quadrate-Satz". *Für alle $p, q, r, s, t, u, v, x \in \mathbb{R}$ und alle $P, Q, R, S, T, U, V, X \in \mathbb{R}$ gilt:*

$$
\begin{aligned}
&(P^2+Q^2+\cdots+V^2+X^2)(p^2+q^2+\cdots+v^2+x^2)\\
&\quad=(Pp+Qq+Rr+Ss+Tt+Uu+Vv+Xx)^2\\
&\qquad+(Pq-Qp+Rs-Sr+Tu-Ut+Vx-Xv)^2\\
&\qquad+(Pr-Qs-Rp+Sq\mp Tv\pm Ux\pm Vt\mp Xu)^2\\
&\qquad+(Ps+Qr-Rq-Sp\pm Tx\pm Uv\mp Vu\mp Xt)^2\\
&\qquad+(Pt-Qu\pm Rv\mp Sx-Tp+Uq\mp Vr\pm Xs)^2\\
&\qquad+(Pu+Qt\mp Rx\mp Sv-Tq-Up\pm Vs\pm Xr)^2\\
&\qquad+(Pv-Qx\mp Rt\pm Su\pm Tr\mp Us-Vp+Xq)^2\\
&\qquad+(Px+Qv\pm Ru\pm St\mp Ts\mp Ur-Vq-Xp)^2.
\end{aligned}
$$

Beweis. Man wendet die Produktregel auf die beiden Oktaven

$$(Pe+Qi+Rj+Sk, Te+Ui+Vj+Xk), (pe+qi+rj+sk, te+ui+vj+xk)$$

an, dann ergibt sich eine der behaupteten Identitäten. □

Der „Acht-Quadrate-Satz" wurde 1844 von GRAVES und 1845 von CAYLEY mit Hilfe ihrer Oktaven gefunden. Der Satz war aber schon 1818 von C. P. DEGEN entdeckt worden (Adumbratio Demonstrationis Theorematis Arithmeticae maxime generalis). DEGEN meinte irrtümlich, das Resultat auf 2^n Quadrate ausdehnen zu können; auch GRAVES glaubte zunächst an eine solche Verallgemeinerung. Weitere historische Angaben findet man bei L. E. DICKSON: On Quaternions and their Generalization and the History of the Eight Square Theorem, Ann. Math. 20, 155–171 (1919); in dieser Arbeit findet sich auch die oben angegebene „Acht-Quadrate-Formel".

5. Die Gleichung $\mathbb{O}=\mathbb{H}\oplus\mathbb{H}p$. Für die Algebren $\mathbb{C}$ bzw. $\mathbb{H}$ hat man die Darstellungen

$$\mathbb{C}=\mathbb{R}\oplus\mathbb{R}i \qquad \text{bzw.} \qquad \mathbb{H}=\mathbb{C}\oplus\mathbb{C}j$$

als direkte Summe von reellen Vektorräumen, wenn $\mathbb{R}$ in $\mathbb{C}$ mit den Paaren $(\alpha,0)$, $\alpha\in\mathbb{R}$, und $\mathbb{C}$ in $\mathbb{H}$ mit den Quaternionen $(\alpha,\beta,0,0)$, $\alpha,\beta\in\mathbb{R}$, identifiziert wird. Dabei sind $\mathbb{R}$ bzw. $\mathbb{C}$ Unteralgebren von $\mathbb{C}$ bzw. $\mathbb{H}$, die jeweils das Einselement von $\mathbb{C}$ bzw. $\mathbb{H}$ enthalten; die Summendarstellungen sind orthogonal bezüglich des natürlichen Skalarproduktes in $\mathbb{C}$ bzw. $\mathbb{H}$.

Eine analoge Situation besteht für die Oktavenalgebra $\mathbb{O}$. Zunächst ist klar:

Die Menge $\{(u,0): u\in\mathbb{H}\}$ *ist eine zur Quaternionenalgebra* $\mathbb{H}$ *isomorphe Unteralgebra von* $\mathbb{O}$, *die das Einselement* e *von* $\mathbb{O}$ *enthält.*

Wir identifizieren im folgenden diese Unteralgebra mit $\mathbb{H}$; dann gelten für alle $u\in\mathbb{H}$ und alle $(a_1,a_2)\in\mathbb{O}$ die Multiplikationsregeln

$$u(a_1,a_2)=(ua_1,a_2u), \qquad (a_1,a_2)u=(a_1u,a_2\bar{u}).$$

Für $p := (0, e')$ verifiziert man direkt:

$$p^2 = -e, \qquad (a_1, a_2) = a_1 e + a_2 p \qquad \text{für alle} \qquad (a_1, a_2) \in \mathbb{O}.$$

Jetzt folgt leicht:

Satz. *Es gilt* $\mathbb{O} = \mathbb{H}e \oplus \mathbb{H}p$ *als Vektorraum, diese Summe ist orthogonal bezüglich des euklidischen Skalarproduktes von* $\mathbb{O}$. *Für alle* $u, v \in \mathbb{H}$ *gilt*:

(1) $$u(vp) = (vu)p,$$

(2) $$(up)v = (u\bar{v})p, \qquad \textit{speziell} \qquad pv = \bar{v}p,$$

(3) $$(up)(vp) = -\bar{v}u.$$

Beweis. Wegen $(a_1, a_2) = a_1 e + a_2 p$ gilt $\mathbb{O} = \mathbb{H}e + \mathbb{H}p$. Da stets $\langle a_1 e, a_2 p\rangle = \langle (a_1, 0), (0, a_2)\rangle = \langle a_1, 0\rangle + \langle 0, a_2\rangle = 0$, so ist diese Summendarstellung orthogonal und also direkt. Die Regeln (1)–(3) bestätigt man durch Nachrechnen, z. B.

ad (1): $$u(vp) = u(0, v) = (0, vu) = (vu)p,$$

ad (2): $$(up)v = (0, u)v = (0, u\bar{v}) = (u\bar{v})p,$$

ad (3): $$(up)(vp) = (0, u)(0, v) = (-\bar{v}u, 0) = -\bar{v}u.$$ □

Gleichungen von Typ (1)–(3) spielen im nächsten Paragraphen eine wichtige Rolle.

6. Multiplikationstafel für $\mathbb{O}$. Wir wissen (R.6), daß jede Multiplikation in einem Vektorraum V mit Basis $e_1, \ldots, e_n$ durch die n^2 Einzelprodukte $e_\mu e_\nu$, $1 \leqslant \mu, \nu \leqslant n$, festgelegt ist. Im Fall $V := \mathbb{R}^8$ mit der natürlichen Basis $e_1 := (1, 0, \ldots, 0), \ldots, e_8 := (0, \ldots, 0, 1)$ wird, wenn e_1 Einselement ist, durch die Tabelle

	e_2	e_3	e_4	e_5	e_6	e_7	e_8
e_2	$-e_1$	e_4	$-e_3$	e_6	$-e_5$	$-e_8$	e_7
e_3	$-e_4$	$-e_1$	e_2	e_7	e_8	$-e_5$	$-e_6$
e_4	e_3	$-e_2$	$-e_1$	e_8	$-e_7$	e_6	$-e_5$
e_5	$-e_6$	$-e_7$	$-e_8$	$-e_1$	e_2	e_3	e_4
e_6	e_5	$-e_8$	e_7	$-e_2$	$-e_1$	$-e_4$	e_3
e_7	e_8	e_5	$-e_6$	$-e_3$	e_4	$-e_1$	$-e_2$
e_8	$-e_7$	e_6	e_5	$-e_4$	$-e_3$	e_2	$-e_1$

die Oktavenmultiplikation bestimmt. Hieraus liest man sofort die Nicht-Assoziativität von $\mathbb{O}$ ab, z. B. $e_5(e_6 e_7) = e_8$, aber $(e_5 e_6)e_7 = -e_8$, indessen wäre es sehr mühsam, mittels dieser Tabelle festzustellen, daß $\mathbb{O}$ alternativ ist.

§ 3. Einzigkeit der CAYLEY-Algebra

Ein größeres System kann nicht mehr alternativ sein (M. ZORN 1933).

In diesem Paragraphen sei $\mathscr{A}$ eine reelle, endlich-dimensionale alternative Divisionsalgebra. Das Ziel ist ein Beweis des Satzes von ZORN, wonach $\mathscr{A}$ assoziativ oder isomorph zur CAYLEY-Algebra ist.

Nach Satz 7.2.2 ist $\mathscr{A}$ zunächst quadratisch, wie in 8.1.1 wird die Bilinearform von $\mathscr{A}$ mit $(x,y)\mapsto\langle x,y\rangle$ bezeichnet, sie ist nach Lemma 8.1.1 positiv definit.

Eine zentrale Rolle spielt der

1. Verdopplungssatz. *Es sei $\mathscr{B}$ eine Unteralgebra von $\mathscr{A}$, die e enthält. Dann ist für jedes q aus $\mathscr{A}$ mit*

$$q^2 = -e \qquad \textit{und} \qquad q\in\mathscr{B}^\perp := \{v\in\mathscr{A} : \langle v,u\rangle = 0 \textit{ für alle } u\in\mathscr{B}\}$$

die Menge

$$\mathscr{B} + \mathscr{B}q = \{u + vq : u,v\in\mathscr{B}\}$$

eine Unteralgebra von $\mathscr{A}$, die e enthält. Für alle $u,v\in\mathscr{B}$ gilt

(1) $$(uq)v = (u\bar{v})q, \qquad \textit{speziell} \qquad qv = \bar{v}q,$$

(2) $$u(vq) = (vu)q,$$

(3) $$(uq)(vq) = -\bar{v}u.$$

Die Summe $\mathscr{B} + \mathscr{B}q$ ist direkt, und es gilt $\dim(\mathscr{B} + \mathscr{B}q) = 2\dim\mathscr{B}$.

Der *Beweis* wird wie folgt organisiert:

a) *$\mathscr{B}$ ist konjugationsstabil*: $\bar{\mathscr{B}} = \mathscr{B}$. Denn mit x und e gehört auch $\bar{x} = 2\lambda(x)e - x$ zu $\mathscr{B}$.

b) *Es gilt $\bar{q} = -q\notin\mathscr{B}$ und $\langle xq, xq\rangle = \langle x,x\rangle$ für alle $x\in\mathscr{A}$.* Aus $q^2 = -e$ folgt $q\in\operatorname{Im}\mathscr{A}$, also $\bar{q} = -q \neq 0$, also $q\notin\mathscr{B}$ (da sonst $\langle q,q\rangle = 0$). Somit sind e und q linear unabhängig, und $q^2 = 2\lambda(q)q - \langle q,q\rangle e$ ergibt $\lambda(q) = 0$ und $\langle q,q\rangle = 1$. Mit 1.2(2) ist alles bewiesen.

c) *Die Summe $\mathscr{B} + \mathscr{B}q$ der Vektorräume ist direkt.* Denn für $u\in\mathscr{B}\cap\mathscr{B}q$ gilt $u = vq$ mit $v\in\mathscr{B}$. Es folgt

$$vu = v(vq) = v^2q = 2\lambda(v)vq - \langle v,v\rangle q = 2\lambda(v)u - \langle v,v\rangle q, \text{ das heißt } \langle v,v\rangle q\in\mathscr{B}.$$

Nach b) erhält man $\langle v,v\rangle = 0$, also $v = 0$ und $u = 0$.

d) $\dim(\mathscr{B} + \mathscr{B}q) = 2\dim\mathscr{B}$. Wegen c) ist zu zeigen: $\dim\mathscr{B} = \dim\mathscr{B}q$. Das ist klar, da $\mathscr{B}\to\mathscr{B}q$, $u\mapsto uq$, wegen der Nullteilerfreiheit injektiv und also bijektiv ist.

e) *Beweis von* (1): Nach 1.3.(2′) gilt $(x\bar{y})y = \langle y,y\rangle x$. Linearisieren (mit $y + z$ statt y) gibt: $(x\bar{y})z + (x\bar{z})y = 2\langle y,z\rangle x$. Wählt man hier $y := q$ und $z := v\in\mathscr{B}$, so gilt $\langle q,v\rangle = 0$ und also $-(x\bar{q})v = (x\bar{v})q$. Da $q\in\operatorname{Im}\mathscr{A}$, so gilt $\bar{q} = -q$, womit (1) verifiziert ist.

f) *Beweis von* (2): Konjugieren von (1) gibt $\bar{v}(q\bar{u}) = q(v\bar{u})$. Wegen $\bar{\mathscr{B}} = \mathscr{B}$ gilt auch $q\bar{u} = uq$ und $q(v\bar{u}) = (u\bar{v})q$. Es folgt $\bar{v}(uq) = (u\bar{v})q$. Dies ist (2), wenn man u,v statt $\bar{v},u$ schreibt.

g) *Beweis von* (3): Es gilt $(uq)[uq\cdot vq + \bar{v}u] = (uq)^2\cdot vq + uq\cdot\bar{v}u$. Da $\lambda(uq) = \langle\bar{u},q\rangle$ nach Satz 1.3, und da stets $\langle\bar{u},q\rangle = 0$ wegen $q\in\mathscr{B}^\perp$, so gilt wegen b): $(uq)^2 = 2\lambda(uq)uq - \langle u,u\rangle e = -\langle u,u\rangle e$. Damit sehen wir:

(○) $$(uq)[uq\cdot vq + \bar{v}u] = -\langle u,u\rangle vq + uq\cdot\bar{v}u.$$

Nun folgt aus (1), wenn man dort $\bar{v}u$ statt v einträgt: $uq\cdot\bar{v}u = (u(\bar{u}v))q$. Wegen

$u(\bar{u}v) = \langle u, u\rangle v$ heißt dies $uq \cdot \bar{v}u = \langle u, u\rangle vq$. Die rechte Seite von (○) verschwindet also, woraus wegen der Nullteilerfreiheit von $\mathscr{A}$ die Identität (3) folgt.

h) *$\mathscr{B} + \mathscr{B}q$ ist Unteralgebra.* Das folgt direkt aus a), e), f) und g). Mit a)–h) ist der Satz bewiesen.

2. Anwendung des Verdopplungssatzes. Um den Verdopplungssatz anwenden zu können, benötigt man bei vorgegebener Unteralgebra $\mathscr{B} \neq \mathscr{A}$ ein Element $q \in \mathscr{B}^{\perp}$ mit $q^2 = -e$.

Lemma. *Es sei $\mathscr{B} \neq \mathscr{A}$ eine Unteralgebra von $\mathscr{A}$ mit $e \in \mathscr{B}$. Dann gibt es stets ein Element $q \in \mathscr{B}^{\perp}$ mit $q^2 = -e$.*

Beweis. Nach dem Lemma von FROBENIUS gilt $\mathscr{A} = \mathbb{R}e \oplus \operatorname{Im}\mathscr{A}$ und $\mathscr{B} = \mathbb{R}e \oplus \operatorname{Im}\mathscr{B}$. Ersichtlich besteht die Inklusion $\operatorname{Im}\mathscr{B} \subset \operatorname{Im}\mathscr{A}$; wegen $\mathscr{B} \neq \mathscr{A}$ gilt also $\operatorname{Im}\mathscr{B} \subsetneq \operatorname{Im}\mathscr{A}$. Das orthogonale Komplement von $\operatorname{Im}\mathscr{B}$ in $\operatorname{Im}\mathscr{A}$ enthält somit wenigstens einen Vektor $q \neq 0$; wegen 7.1.1 darf man $q^2 = -e$ annehmen. Jedes $u \in \mathscr{B}$ hat die Form $\alpha e + w$, $\alpha \in \mathbb{R}$, $w \in \operatorname{Im}\mathscr{B}$. Aus $\langle q, w\rangle = 0$ für alle $w \in \operatorname{Im}\mathscr{B}$ folgt nun $\langle q, u\rangle = \langle q, \alpha e\rangle + \langle q, w\rangle = \alpha\lambda(q) = 0$ wegen $q \in \operatorname{Im}\mathscr{A} = \operatorname{Kern}\lambda$, also $q \in \mathscr{B}^{\perp}$. □

Verdopplungssatz und Lemma weisen den tieferen Grund dafür auf, warum bei den klassischen Algebren $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, $\mathbb{O}$ die letzten drei jeweils durch „Verdopplung" der vorangehenden entstehen. Als Anwendung des Verdopplungssatzes und des Lemmas folgt

Satz. *Jede echte Teilalgebra $\mathscr{B}$ von $\mathscr{A}$, die e enthält, ist assoziativ.*

Beweis. Man wählt ein $q \in \mathscr{B}^{\perp}$ mit $q^2 = -e$ und betrachtet die Verdopplungsalgebra $\mathscr{B} + \mathscr{B}q$. Man linearisiert die Identität $x^2z = x \cdot xz$ und erhält

$$(*) \qquad (xy + yx)z = x(yz) + y(xz).$$

Man setzt hier $x := u$, $y := vq$, $z := \bar{w}$, wobei $u, v, w \in \mathscr{B}$. Mit 1.(1)–1.(3) folgt:

$$(xy + yx)z = (u \cdot vq + vq \cdot u)\bar{w} = (vu \cdot q + v\bar{u} \cdot q)w = [v(u + \bar{u}) \cdot q]\bar{w}$$
$$= 2\lambda(u)vq \cdot \bar{w} = 2\lambda(u)vw \cdot q,$$
$$x(yz) = u(vq \cdot \bar{w}) = u(vw \cdot q) = (vw \cdot u)q, \qquad y(xz) = vq \cdot u\bar{w} = (v \cdot w\bar{u})q.$$

Wegen (*) ist damit: $2\lambda(u)vw \cdot q = (vw \cdot u + v \cdot w\bar{u})q$. Da $\mathscr{A}$ nullteilerfrei ist, erhält man $vw \cdot u = 2\lambda(u)vw - v \cdot w\bar{u} = v \cdot wu$, also die Assoziativität von $\mathscr{B}$.

3. Einzigkeit der CAYLEY-Algebra (ZORN 1933). *Jede reelle, endlich-dimensionale, alternative, nicht assoziative Divisionsalgebra $\mathscr{A}$ ist zur* CAYLEY*-Algebra $\mathbb{O}$ isomorph.*

Beweis. Nach dem Quaternionenlemma 7.2.3 gibt es eine Unteralgebra $\mathscr{B}$ von $\mathscr{A}$ und einen Algebramonomorphismus $f: \mathbb{H} \to \mathscr{A}$ mit $f(\mathbb{H}) = \mathscr{B}$. Da $\mathscr{A}$ nicht assoziativ ist, gilt $\mathscr{B} \neq \mathscr{A}$. Nach Lemma 2 gibt es ein Element $q \in \mathscr{B}^{\perp}$ mit $q^2 = -e$.

Nach dem Verdopplungssatz ist dann $\mathscr{B} \oplus \mathscr{B}q$ eine Unteralgebra von $\mathscr{A}$, die e enthält. Nach Satz 2.5 gilt $\mathbb{O} = \mathbb{H}e \oplus \mathbb{H}p$. Die Abbildung

$$h: \mathbb{O} = \mathbb{H}e \oplus \mathbb{H}p \to \mathscr{B} \oplus \mathscr{B}q, \qquad ue + vp \mapsto f(u) + f(v)q,$$

ist ein Algebra-Isomorphismus, denn sie ist bijektiv, $\mathbb{R}$-linear, und – da die Regeln (1)–(3) der Sätze 2.5 und 3.1 dieselbe Gestalt haben – auch Multiplikationstreu.

Im Falle $\mathscr{B} \oplus \mathscr{B}q \neq \mathscr{A}$ wäre $\mathscr{B} \oplus \mathscr{B}q$ nach Satz 2 assoziativ, was nicht geht, da $\mathbb{O}$ nicht assoziativ ist. Es folgt $\mathscr{A} = \mathscr{B} + \mathscr{B}q \cong \mathbb{O}$. □

Wir können nun folgende Verallgemeinerung des Satzes von FROBENIUS aussprechen:

Struktursatz. *Jede reelle, alternative, endlich-dimensionale Divisionsalgebra ist isomorph zu* $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ *oder* $\mathbb{O}$.

4. Beschreibung von $\mathbb{O}$ durch ZORNsche Vektormatrizen. Wir haben in 2.1 Oktaven als Paare (x_1, x_2) von Quaternionen eingeführt. Max ZORN hat 1933 in seiner klassischen Arbeit eine Beschreibung von alternativen Algebren gegeben, die dem Wunsch, explizit zu rechnen, noch mehr nachkommt. Um die ZORNsche Definition zu motivieren, gehen wir aus vom Oktavenprodukt (vgl. 2.1)

$$(1) \qquad xy = (x_1, x_2)(y_1, y_2) = (x_1 y_1 - \bar{y}_2 x_2, x_2 \bar{y}_1 + y_2 x_1).$$

Mit $x_k = \alpha_k e + u_k$, $y_k = \beta_k e + v_k$, wobei $\alpha_k, \beta_k \in \mathbb{R}$ und $u_k, v_k \in \operatorname{Im}\mathbb{H}$, hat dieses Produkt xy, wenn man noch $uv = -\langle u, v\rangle e + u \times v$ beachtet, die Form

$$\begin{aligned}(2) \qquad &([\alpha_1\beta_1 - \alpha_2\beta_2 - \langle u_1, v_1\rangle + \langle u_2, v_2\rangle]e \\ &\quad + \alpha_1 v_1 + \beta_1 u_1 + \alpha_2 v_2 - \beta_2 u_2 + u_1 \times v_1 - u_2 \times v_2, \\ &[\alpha_2\beta_1 + \alpha_1\beta_2 + \langle u_2, v_1\rangle - \langle u_1, v_2\rangle]e \\ &\quad - \alpha_2 v_1 + \beta_2 u_1 + \alpha_1 v_2 + \beta_1 u_2 - u_2 \times v_1 - u_1 \times v_2).\end{aligned}$$

Wir fassen nun u_k, v_k als Vektoren des $\mathbb{R}^3$ auf und „komplexifizieren“:

$$(3) \quad \alpha := \alpha_1 + i\alpha_2, \quad \beta := \beta_1 + i\beta_2 \in \mathbb{C}; \quad u := u_1 + iu_2, \quad v := v_1 + iv_2 \in \mathbb{C}^3.$$

Von nun an wird der Querstrich nur noch zur Konjugierung im Komplexen benutzt; für Vektoren $w = (w_1, w_2, w_3)$, $z = (z_1, z_2, z_3) \in \mathbb{C}^3$ schreiben wir

$$(4) \quad \langle w, z\rangle := \sum_1^3 w_\nu z_\nu, \qquad w \times z =: (w_2 z_3 - w_3 z_2, w_3 z_1 - w_1 z_3, w_1 z_2 - w_2 z_1).$$

Dann nimmt (2) die Gestalt an

$$(5) \quad (\operatorname{Re}([\alpha\beta - \langle \bar{u}, v\rangle]e + \bar{\alpha}v + \beta u + \bar{u} \times \bar{v}),\ \operatorname{Im}([\alpha\beta - \langle \bar{u}, v\rangle]e + \bar{\alpha}v + \beta u + \bar{u} \times \bar{v})),$$

wenn man Identitäten wie $\bar{u} \times \bar{v} = u_1 \times v_1 - u_2 \times v_2 + i(-u_2 \times v_1 - u_1 \times v_2)$ verwendet. Die Formel (5) läßt sich besonders bequem ausdrücken, wenn man den acht-dimensionalen reellen Vektorraum $\mathscr{Z} := \mathbb{C}e \oplus \mathbb{C}^3$ mit Elementen $\alpha e + u$, $\alpha \in \mathbb{C}$, $u \in \mathbb{C}^3$, her nimmt und die Abbildung

$$(6) \quad F: \mathbb{O} \to \mathscr{Z}, \qquad x = (x_1, x_1) = (\alpha_1 e + u_1, \alpha_2 e + u_2) \mapsto \alpha e + u = x_1 + ix_2,$$

einführt. Dann können wir zusammenfassend sagen:

Die Abbildung F ist ein $\mathbb{R}$-Vektorraum-Isomorphismus; für zwei Oktaven x, y mit $F(x) = \alpha e + \boldsymbol{u}$, $F(y) = \beta e + v$ gilt:

$$F(xy) = [\alpha\beta - \langle \bar{\boldsymbol{u}}, v\rangle]e + [\bar{\alpha}v + \beta\boldsymbol{u} + \bar{\boldsymbol{u}} \times \bar{v}]. \tag{7}$$

Jetzt ist klar, wie man in $\mathfrak{X}$ zu multiplizieren hat: man setzt

$$(\alpha e + \boldsymbol{u})(\beta e + v) := [\alpha\beta - \langle \bar{\boldsymbol{u}}, v\rangle]e + [\bar{\alpha}v + \beta\boldsymbol{u} + \bar{\boldsymbol{u}} \times \bar{v}] \tag{8}$$

und weiß aufgrund des Vorangehenden:

$\mathfrak{X}$ mit der „Zornschen" Multiplikation (8) ist eine $\mathbb{R}$-Algebra; die Abbildung $F: \mathbb{O} \to \mathfrak{X}$ ist ein $\mathbb{R}$-Algebra-Isomorphismus.

Nach ZORNschem Vorbild schreibt man die Elemente $\alpha e + \boldsymbol{u}$ von $\mathfrak{X}$ auch als Vektormatrizen $\begin{pmatrix} \alpha & \boldsymbol{u} \\ -\bar{\boldsymbol{u}} & \bar{\alpha} \end{pmatrix}$; ihr Produkt ist dann das „Matrizenprodukt"

$$\begin{pmatrix} \alpha & \boldsymbol{u} \\ -\bar{\boldsymbol{u}} & \bar{\alpha} \end{pmatrix}\begin{pmatrix} \beta & v \\ -\bar{v} & \bar{\beta} \end{pmatrix} = \begin{pmatrix} \alpha\beta - \langle \bar{\boldsymbol{u}}, v\rangle & \bar{\alpha}v + \beta\boldsymbol{u} + \bar{\boldsymbol{u}} \times \bar{v} \\ -\alpha\bar{v} - \bar{\beta}\bar{\boldsymbol{u}} - u \times v & \bar{\alpha}\bar{\beta} - \langle \boldsymbol{u}, \bar{v}\rangle \end{pmatrix}.$$

Bemerkung. In der Literatur findet man manchmal das Oktavenprodukt anders als hier oder in 2.1 definiert. Oft wird ein Isomorphismus durch Änderung des Produktes $(x, y) \mapsto xy$ in das „entgegengesetzte" Produkt $(x, y) \mapsto yx$ gegeben. Nach dem Einzigkeitssatz 3 sind natürlich alle diese Darstellungen isomorph.

Kapitel 9. Kompositionsalgebren. Satz von HURWITZ

M. Koecher, R. Remmert

Durch diesen Nachweis wird die alte Streitfrage, ob sich die bekannten Produktformeln für Summen von 2, 4 und 8 Quadraten auf Summen von mehr als 8 Quadraten ausdehnen lassen, endgültig, und zwar in verneinendem Sinne entschieden (A. HURWITZ 1898).

1. Für die Multiplikation in den Algebren $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ und $\mathbb{O}$ gilt $|xy|^2 = |x|^2|y|^2$, wobei $|\ |$ die euklidische Länge bezeichnet. Schreibt man die Vektoren $x, y, z := xy$ bezüglich einer Orthonormalbasis in Koordinaten (ξ_ν), (η_ν), (ζ_ν), so erhält man wegen der Bilinearität des Produktes xy den

Quadratesatz. *Falls $n = 1, 2, 4, 8$, so gibt es n reelle (sogar ganz-rationale) Bilinearformen*

$$\zeta_\nu = \sum_{\lambda,\mu=1}^{n} \alpha_{\lambda\mu}^{(\nu)} \xi_\lambda \eta_\mu, \qquad \alpha_{\lambda\mu}^{(\nu)} \in \mathbb{Z}, \qquad \nu = 1, \ldots, n,$$

so daß für alle Zahlen $\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_n \in \mathbb{R}$ gilt:

$$(*) \qquad \zeta_1^2 + \zeta_2^2 + \cdots + \zeta_n^2 = (\xi_1^2 + \xi_2^2 + \cdots + \xi_n^2)(\eta_1^2 + \eta_2^2 + \cdots + \eta_n^2).$$

Wir haben diesen Satz in 3.3.4, 6.2.3 und 8.2.4 eingehend diskutiert.

Der französische Mathematiker A. M. LEGENDRE (1752–1833) war der erste, der einen Unmöglichkeitsbeweis für den Quadratesatz im Fall $n = 3$ führte. In seinem 1830 in Paris erschienenen großen Werk „Théorie des Nombres" bemerkt er auf Seite 198, daß zwar $3 = 1^2 + 1^2 + 1^2$ und $21 = 4^2 + 2^2 + 1^2$, daß aber das Produkt $3 \cdot 21 = 63$ nicht Summe von drei Quadraten natürlicher Zahlen ist; daraus folgt, daß der Quadratesatz für $n = 3$ mit *rationalen* Bilinearformen $\zeta_1, \zeta_2, \zeta_3$ nicht gelten kann. „Hätte HAMILTON diese Bemerkung von LEGENDRE gekannt, so hätte er vielleicht den Versuch, Tripel zu multiplizieren, gleich aufgegeben. Zum Glück hat er LEGENDRE nicht gelesen: er war ein Autodidakt." (B. L. VAN DER WAERDEN in „Hamiltons Entdeckung der Quaternionen", S. 14).

2. Die sich aufdrängende Frage, für welche Werte von $n \geqslant 1$ die Gleichung $(*)$

$$\zeta_1^2 + \zeta_2^2 + \cdots + \zeta_n^2 = (\xi_1^2 + \xi_2^2 + \cdots + \xi_n^2)(\eta_1^2 + \eta_2^2 + \cdots + \eta_n^2)$$

durch geeignete reelle Bilinearformen $\zeta_1, \ldots, \zeta_n$ in den $\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_n$ lösbar ist, wurde endgültig erst 1898 von Adolf HURWITZ (geb. 1859 in Hildesheim; 1877 Studium bei KLEIN, WEIERSTRASS, KRONECKER; 1881 Promotion in Leipzig; 1882 Habilitation in Göttingen, da sich in Leipzig Realgymnasialabiturienten nicht habilitieren durften; 1884 mit 25 Jahren Extraordinarius in Königsberg, dort Freundschaft mit HILBERT und MINKOWSKI; 1892 Ablehnung der SCHWARZ-Nachfolge in Göttingen und Annahme des Rufes als FROBENIUSnachfolger an das Eidgenössische Polytechnikum Zürich, gest. 1919 in Zürich. Arbeiten zur Funktio-

nentheorie, zur Theorie der Modulfunktionen, zur Algebra und zur algebraischen Zahlentheorie) entschieden. In seiner in den Nachrichten der k. Gesellschaft der Wissenschaften zu Göttingen veröffentlichten Arbeit „Über die Komposition der quadratischen Formen von beliebig vielen Variablen" (1898), 309–316 (Math. Werke 2, 565–571) bewies er mit Hilfe des Matrizenkalküls, daß die im Quadratesatz zugelassenen Fälle $n = 1, 2, 4, 8$ die einzig möglichen sind. Wir werden das HURWITZsche Resultat in diesem Kapitel aus einem Struktursatz für Kompositionsalgebren ableiten, der seinerseits auf dem Hauptsatz über alternative Divisionsalgebren beruht.

§ 1. Kompositionsalgebren

Eine reelle Algebra $\mathscr{A} = (V, \cdot) \neq 0$ heißt eine *Kompositionsalgebra*, wenn V euklidisch und die Multiplikation in $\mathscr{A}$ längentreu ist:

$$|xy| = |x|\,|y| \qquad \text{für alle} \qquad x, y \in V \qquad \text{(Produktregel)}.$$

Jede Kompositionsalgebra ist nullteilerfrei. Die Algebren $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ und $\mathbb{O}$ sind Kompositionsalgebren mit Einselement; das Ziel dieses Paragraphen ist zu zeigen, daß dies bis auf Isomorphie bereits alle solchen Algebren endlicher Dimension sind.

1. Historisches zur Kompositionstheorie. Um die Wahl des Wortes „Kompositionsalgebra" zu verstehen, skizzieren wir kurz die historischen Ursprünge. In den berühmten *Disquisitiones Arithmeticae*, dem Hauptwerk des jungen GAUSS aus dem Jahre 1801 (Werke 1), beginnt mit Artikel 153 ein systematisches Studium der Arithmetik *binärer* quadratischer Formen, das heißt, der Polynome $f(\xi_1, \xi_2) = a\xi_1^2 + 2b\xi_1\xi_2 + c\xi_2^2$ mit ganzzahligen Koeffizienten. Seit FERMAT hat man sich für die Frage der Darstellbarkeit ganzer Zahlen durch solche Formen interessiert, das heißt, für die Frage, ob die Gleichung $f(\xi_1, \xi_2) = n$ bei vorgegebener Zahl $n \in \mathbb{Z}$ ganzzahlige Lösungen ξ_1, ξ_2 besitzt. Dieses zahlentheoretische Problem geht wesentlich über die Zahlentheorie der Griechen (EUKLID, Buch 9) hinaus, die ersten allgemeinen Resultate erzielte LAGRANGE.

Im Zusammenhang mit dem Darstellbarkeitsproblem natürlicher Zahlen durch quadratische Form führte GAUSS den Begriff der *Komposition von quadratischen Formen* ein (Disq. Arith., Art. 235 ff.): Sind drei quadratische Formen f, g, h mit *ganzzahligen* Koeffizienten a, b, c bzw. a', b', c' bzw. A, B, C gegeben, so sagt er, daß h durch *Komposition* aus f und g entsteht, wenn für alle $\xi_1, \xi_2, \eta_1, \eta_2$ gilt

$$(*) \qquad A\zeta_1^2 + 2B\zeta_1\zeta_2 + C\zeta_2^2 = (a\xi_1^2 + 2b\xi_1\xi_2 + c\xi_2^2)(a'\eta_1^2 + 2b'\eta_1\eta_2 + c'\eta_2^2),$$

wobei ζ_1 und ζ_2 geeignete Bilinearformen in $\xi_1, \xi_2, \eta_1, \eta_2$ mit *ganzzahligen* Koeffizienten sind. Die GAUSSsche Kompositionstheorie bildet einen Schwerpunkt seiner *Disquisitiones*. Heute weiß man, daß diese Theorie im wesentlichen äquivalent ist zur Idealtheorie quadratischer Zahlkörper (wegen Einzelheiten vgl. W. SCHARLAU und H. OPOLKA: Von FERMAT bis MINKOWSKI, S. 97 ff.).

Ein Hauptresultat der GAUSSschen Theorie ist (bei starker Simplifizierung), daß die „Äquivalenzklassen" ganzzahliger quadratischer Formen zu vorgegebener

Diskriminante $d := b^2 - ac$ eine endliche abelsche Gruppe bilden (sogenannte Klassengruppe); GAUSS weist die Gruppeneigenschaften nach, ohne den Gruppenbegriff zu kennen. Die Theorie ist echt arithmetisch; läßt man reelle Koeffizienten zu und beschränkt man sich auf positiv definite Formen (das heißt, a, c, $-d$ positiv), so wird (*) bei Übergang zu geeigneten neuen Variablen zur Zwei-Quadrate-Formel $(\zeta_1'^2 + \zeta_2'^2) = (u^2 + v^2)(x^2 + y^2)$, die durch $\zeta_1' = ux - vy$, $\zeta_2' = uy + vx$ gelöst wird (Zwei-Quadrate-Satz 3.3.4).

Im Anschluß an GAUSS wurden allgemein Kompositionen von quadratischen Formen in n Variablen betrachtet. Es ergeben sich auch bei Verzicht auf die einschränkende arithmetische Forderung der Ganzzahligkeit der Koeffizienten interessante Probleme. HURWITZ beginnt seine Arbeit zum Quadratesatz, der er nicht von ungefähr den Titel „Über die Komposition der quadratischen Formen ..." gab, wie folgt: „Im Gebiete der quadratischen Formen von n Variablen wird eine Kompositionstheorie stattfinden, wenn für irgend drei quadratische Formen φ, ψ, χ von nicht verschwindender Determinante die Gleichung

$$\varphi(x_1, x_2, \ldots, x_n)\psi(y_1, y_2, \ldots, y_n) = \chi(z_1, z_2, \ldots, z_n) \tag{1}$$

dadurch befriedigt werden kann, daß man die Variablen $z_1, z_2, \ldots, z_n$ durch geeignet gewählte bilineare Funktionen der Variablen $x_1, x_2, \ldots, x_n$ und $y_1, \ldots, y_n$ ersetzt. Da eine quadratische Form durch lineare Transformation der Variablen in eine Summe von Quadraten überführt werden kann*), so darf man, ohne die Allgemeinheit zu beeinträchtigen, an Stelle der Gleichung (1) die folgende:

$$(x_1^2 + x_2^2 + \cdots + x_n^2)(y_1^2 + y_2^2 + \cdots + y_n^2) = z_1^2 + z_2^2 + \cdots + z_n^2 \tag{2}$$

betrachten. Hiernach ist die Frage, ob für quadratische Formen mit n Variablen eine Kompositionstheorie existiert, im wesentlichen identisch mit der andern, ob man der Gleichung (2) durch geeignete bilineare Funktionen $z_1, \ldots, z_n$ der $2n$ unabhängigen Variablen $x_1, \ldots, x_n, y_1, \ldots, y_n$ genügen kann".

Wir notieren ein einfaches Existenzkriterium für Kompositionstheorien.

Für n Bilinearformen $\phi_\nu(x, y)$, $1 \leqslant \nu \leqslant n$, gilt die Identität

$$(*) \qquad (\xi_1^2 + \cdots + \xi_n^2)(\eta_1^2 + \cdots + \eta_n^2) = \phi_1(\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_n)^2 + \cdots + \phi_n(\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_n)^2$$

genau dann, wenn $(\mathbb{R}^n, \cdot)$ *mit* $x \cdot y := (\phi_1(x, y), \ldots, \phi_n(x, y))$ *eine Kompositionsalgebra ist.*

Beweis. Je n Bilinearformen $\phi_1, \ldots, \phi_n$ machen $\mathbb{R}^n$ wie angegeben zu einer Algebra. Die Produktregel $|xy| = |x|\,|y|$ gilt genau dann, wenn (*) erfüllt ist. □

Da jeder n-dimensionale euklidische Vektorraum V zum Zahlenraum $\mathbb{R}^n$ der n-tupel $x = (\xi_1, \ldots, \xi_n)$, $y = (\eta_1, \ldots, \eta_n)$ mit seinem kanonischen Skalarprodukt $\langle x, y \rangle = \sum_1^n \xi_\nu \eta_\nu$ längentreu isomorph ist, folgt weiter:

Es gibt genau dann eine Kompositionstheorie für reelle quadratische Formen in n Variablen, wenn es eine n-dimensionale Kompositionsalgebra gibt.

---

*) Für HURWITZ sind quadratische Formen hier stets positiv definit.

2. Beispiele. Die Algebren $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ und $\mathbb{O}$ geben die klassischen Kompositionstheorien für $n = 1, 2, 4, 8$. Für $\mathbb{C}$ ist dies die Identität

$$(1) \qquad (x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 - x_2y_2)^2 + (x_1y_2 + x_2y_1)^2;$$

für $\mathbb{H}$ und $\mathbb{O}$ schreiben wir die entsprechenden Gleichungen nicht mehr hin (vgl. 6.2.3 und 8.2.4).

Jede eindimensionale Kompositionsalgebra ist zu $\mathbb{R}$ isomorph und hat insbesondere ein Einselement (vgl. R.4). Es ist leicht, Kompositionsalgebren der Dimension 2, 4 oder 8 anzugeben, die kein Einselement haben. Sei zunächst $n = 2$. Wir definieren auf $\mathbb{R}^2$ drei Multiplikationen unter Verwendung des gewöhnlichen komplexen Produktes wz in $\mathbb{R}^2 = \mathbb{C}$. Wir setzen:

$$w \underset{1}{\square} z := \bar{w}z, \qquad w \underset{2}{\square} z := w\bar{z}, \qquad w \underset{3}{\square} z = \overline{wz}.$$

Dann verifiziert man mühelos:

$\mathscr{A}_\nu := (\mathbb{R}^2, \underset{\nu}{\square})$, $1 \leqslant \nu \leqslant 3$, ist eine 2-dimensionale, nicht alternative Kompositionsalgebra ohne Einselement; lediglich die Algebra $\mathscr{A}_3$ ist kommutativ. Die zugehörigen Kompositionsformeln sind:

$$(2) \qquad (x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 + x_2y_2)^2 + (x_1y_2 - x_2y_1)^2,$$

$$(3) \qquad (x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 + x_2y_2)^2 + (-x_1y_2 + x_2y_1)^2,$$

$$(4) \qquad (x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_1 - x_2y_2)^2 + (-x_1y_2 - x_2y_1)^2.$$

Die Gleichungen (1), (4) und (2), (3) unterscheiden sich nur unwesentlich durch Vorzeichen, für die Algebren läßt sich leicht zeigen:

Von den Algebren $\mathbb{C}$, $\mathscr{A}_1$, $\mathscr{A}_2$, $\mathscr{A}_3$ sind keine zwei zueinander isomorph. Jede weitere 2-dimensionale Kompositionsalgebra $\mathscr{A}$ ist zu einer dieser vier Algebren längentreu isomorph.

Sei nun $n = 4$. Für je zwei Quaternionen $a, b \in \mathbb{H}$ der Länge 1 erhält man (wenn rechts die gewöhnliche Quaternionenmultiplikation steht) vermöge

$$x \square y := axyb \text{ bzw. } x \square y := a\bar{x}yb \text{ bzw. } x \square y := ax\bar{y}b \text{ bzw. } x \square y := a\overline{xy}b$$

unendlich viele nicht-isomorphe Kompositionsalgebren $(\mathbb{R}^4, \square)$ ohne Einselement. Auch für $n = 8$ lassen sich nach diesem Muster unendlich viele nicht-isomorphe Kompositionsalgebren $(\mathbb{R}^8, \square)$ ohne Einselement konstruieren.

3. Kompositionsalgebren mit Einselement. In diesem Abschnitt bezeichnet $\mathscr{A} = (V, \cdot)$ eine reelle nicht notwendig endlich-dimensionale Kompositionsalgebra. Wir benutzen die Längenidentität in der (quadrierten) Form

$$(*) \qquad \langle xy, xy\rangle = \langle x, x\rangle\langle y, y\rangle, \qquad x, y \in V.$$

Um die einschneidenden Konsequenzen dieser Bedingung zu erkennen, wenden wir zweimal das *Verfahren der Linearisierung* an. Wir schreiben $x + x'$ statt x in $(*)$ und erhalten durch Ausrechnen:

$$\langle xy, xy\rangle + 2\langle xy, x'y\rangle + \langle x'y, x'y\rangle = (\langle x, x\rangle + 2\langle x, x'\rangle + \langle x', x'\rangle)\langle y, y\rangle.$$

Hieraus folgt wegen $(*)$:

$$(0) \qquad \langle xy, x'y\rangle = \langle x, x'\rangle\langle y, y\rangle.$$

Schreibt man nun $y + y'$ statt y, so ergibt sich weiter:

$$\langle x'y, xy\rangle + \langle x'y', xy\rangle + \langle x'y, xy'\rangle + \langle x'y', xy'\rangle$$
$$= \langle x, x'\rangle(\langle y, y\rangle + 2\langle y, y'\rangle + \langle y', y'\rangle).$$

Nach (0) sind rechts und links jeweils erste und letzte Terme gleich, damit folgt:

(1) $\quad \langle xy, x'y'\rangle + \langle xy', x'y\rangle = 2\langle x, x'\rangle\langle y, y'\rangle \quad$ für alle $\quad x, x', y, y' \in V.$

Setzt man hier $x' := z$, $y' := y$ bzw. $x' := x$, $y' := z$, so erhält man noch

(2) $\quad \langle xy, zy\rangle = \langle x, z\rangle\langle y, y\rangle$

(3) $\quad \langle xy, xz\rangle = \langle x, x\rangle\langle y, z\rangle$ $\qquad$ *für alle* $\qquad x, y, z \in V.$

Nach diesen Vorbereitungen zeigen wir den fundamentalen

Satz. *Jede reelle Kompositionsalgebra $\mathscr{A}$ mit Einselement e ist quadratisch und alternativ.*

Beweis. Aus (1) folgt (mit $x' := z$, $y' := e$ bzw. $x' := e$, $y' := z$)

$$\langle xy, z\rangle + \langle x, zy\rangle = 2\langle y, e\rangle\langle x, z\rangle \quad \text{bzw.} \quad \langle xy, z\rangle + \langle xz, y\rangle = 2\langle x, e\rangle\langle y, z\rangle.$$

Schreibt man hier in der ersten Gleichung xy statt x und in der zweiten Gleichung xz statt z, so erhält man wegen (2) bzw. (3):

$$\langle xy \cdot y, z\rangle + \langle y, y\rangle\langle x, z\rangle = 2\langle y, e\rangle\langle xy, z\rangle,$$
$$\langle x \cdot xz, y\rangle + \langle x, x\rangle\langle y, z\rangle = 2\langle x, e\rangle\langle y, xz\rangle.$$

Aus diesen Formeln für reelle Zahlen gewinnt man, da das Skalarprodukt $\langle x, y\rangle$ *nicht ausgeartet**) ist, Identitäten für alle Elemente $x, y \in \mathscr{A}$:

(4) $$xy \cdot y = 2\langle y, e\rangle xy - \langle y, y\rangle x,$$

(5) $$x \cdot xy = 2\langle x, e\rangle xy - \langle x, x\rangle y.$$

Setzt man in der letzten Gleichung $y := e$, so folgt

(6) $$x^2 = 2\langle x, e\rangle x - \langle x, x\rangle e \qquad \text{für alle} \qquad x \in \mathscr{A},$$

mithin ist $\mathscr{A}$ quadratisch. Rechtsmultiplikation von (6) mit y liefert $x^2 \cdot y = x \cdot xy$ wegen (5); Linksmultiplikation von $y^2 = 2\langle y, e\rangle y - \langle y, y\rangle e$ mit x führt wegen (4) zu $x \cdot y^2 = xy \cdot y$. Daher ist $\mathscr{A}$ auch alternativ. □

Beweisanalyse: Die Herleitung der Gleichungen (1)–(3) benutzt *nur* die Symmetrie der Bilinearform $\langle x, y\rangle$ und die Tatsache, daß $\mathbb{R}$ ein Körper der Charakteristik $\neq 2$ ist. Die Herleitung der Gleichungen (4), (5) benutzt, daß $\langle x, y\rangle$ nicht ausgeartet ist. Wir haben also allgemeiner bewiesen:

---

*) Eine Bilinearform $\langle x, y\rangle$ heißt *nicht ausgeartet*, wenn aus $\langle w, v\rangle = 0$ für alle $v \in V$ stets $w = 0$ folgt; positiv definite Bilinearformen sind nicht ausgeartet. Man gewinnt (4) aus der für alle $z \in V$ geltenden Identität $\langle xy \cdot y - 2\langle y, e\rangle xy + \langle y, y\rangle x, z\rangle = 0$; analog folgt (5), wenn man zuletzt y statt z schreibt.

Es sei K ein kommutativer Körper der Charakteristik $\neq 2$, und es sei $\mathscr{A} = (V, \cdot) \neq 0$ eine K-Algebra mit Einselement. Es sei $\langle x, y\rangle$ eine nicht ausgeartete K-Bilinearform auf V, so daß für die Algebramultiplikation gilt:

$$\langle xy, xy\rangle = \langle x, x\rangle\langle y, y\rangle \qquad \text{für alle} \qquad x, y \in V.$$

Dann ist $\mathscr{A}$ quadratisch und alternativ.

4. Struktursatz für endlich-dimensionale Kompositionsalgebren mit Einselement. Da Kompositionsalgebren nullteilerfrei sind, so ist jede endlich-dimensionale Kompositionsalgebra eine Divisionsalgebra (Kriterium R.5). Aufgrund von Satz 3 und des Struktursatzes 8.3.3 folgt daher:

Struktursatz. *Es sei $\mathscr{A}$ eine endlich-dimensionale Kompositionsalgebra mit Einselement. Dann ist $\mathscr{A}$ längentreu isomorph zu einer der vier Algebren* $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, $\mathbb{O}$.

Es läßt sich sogar zeigen, daß in diesem Satz die Voraussetzung, daß $\mathscr{A}$ endlich-dimensional ist, überflüssig ist. Weiter läßt sich der Satz verallgemeinern auf beliebige Grundkörper (auch mit Charakteristik 2). Wir verweisen den Leser auf den Artikel von I. KAPLANSKY „Infinite-dimensional quadratic forms admitting composition" in Proceed. Amer. Math. Soc. 4, 956–960 (1953). □

Für die Gültigkeit des Struktursatzes ist die Existenz eines Einselementes in $\mathscr{A}$ ganz wesentlich, wie die Beispiele der Algebren $\mathscr{A}_1$, $\mathscr{A}_2$, $\mathscr{A}_3$ des Abschnittes 2 zeigen.

§ 2. Mutation von Kompositionsalgebren

Aufgrund von 1.1 ist die Existenz einer Kompositionstheorie für Formen in n Variablen gleichbedeutend mit der Existenz einer n-dimensionalen Kompositionsalgebra. Solche Algebren brauchen kein Einselement zu haben, und es gibt sie, wie wir im Abschnitt 2 sahen, in verwirrender Vielfalt. Der Struktursatz 1.4 scheint daher zur Lösung des HURWITZschen Formenproblems keine echte Hilfe zu sein. Und dennoch ist das Problem damit schon im wesentlichen gelöst! Es gibt nämlich ein einfaches Verfahren, *von einer beliebigen Kompositionsalgebra $(V, \cdot)$ zu einer Kompositionsalgebra $(V, \square)$ mit Einselement überzugehen.* Wir beschreiben zunächst eine allgemeine Methode, in einer beliebigen Algebra die Multiplikation abzuändern.

1. Mutationen von Algebren. *Es sei $(V, \cdot)$ eine K-Algebra, es seien $f\colon V \to V$, $g\colon V \to V$ zwei K-lineare Abbildungen. Man setze*

$$x \square y := f(x)g(y) \qquad \text{für alle} \qquad x, y \in V.$$

Dann ist $(V, \square)$ eine K-Algebra.

Beweis. Die Distributivgesetze für $\square$ folgen aus den Distributivgesetzen für $\cdot$ und der Linearität von f, g. □

Jedes Element a einer Algebra $\mathscr{A} = (V, \cdot)$ bestimmt vermöge Links- bzw. Rechtsmultiplikation zwei K-lineare Abbildungen

$$L_a\colon V \to V, \quad x \mapsto ax; \qquad R_a\colon V \to V, \quad x \mapsto xa.$$

Sind *beide Abbildungen* L_a, R_a *bijektiv*, so existiert nach dem Vorangehenden die Algebra $\mathscr{A}(a) := (V, \square)$ mit dem Produkt

$$x \square y := R_a^{-1}(x) \cdot L_a^{-1}(y);$$

wir nennen $\mathscr{A}(a)$ *die Mutation von* $\mathscr{A}$ *bezüglich* a. Die Abbildungen R_a^{-1}, L_a^{-1} dienen als Ersatz für das im allgemeinen nicht vorhandene Inverse a^{-1} von a in $\mathscr{A}$; falls a^{-1} existiert, so gilt $x \square y = (xa^{-1})(a^{-1}y)$. Es gilt stets

$$xa \square ay = xy. \tag{1}$$

Hat $\mathscr{A}$ ein Einselement e, so existiert $\mathscr{A}(e)$, und es gilt $\mathscr{A}(e) = \mathscr{A}$. Im allgemeinen unterscheiden sich Mutationen aber ganz wesentlich von der Ausgangsalgebra (vgl. hierzu Abschnitt 2).

Existenzkriterium für Mutationen. *Ist* $\mathscr{A}$ *endlich-dimensional, so existiert für jeden Nichtnullteiler* $a \in \mathscr{A}$ *die Mutation* $\mathscr{A}(a)$.

Speziell existiert in einer endlich-dimensionalen Kompositionsalgebra für jedes $a \in \mathscr{A} \setminus \{0\}$ *die Mutation* $\mathscr{A}(a)$.

Beweis. Für Nichtnullteiler a sind beide Abbildungen L_a, R_a injektiv und also, wenn $\mathscr{A}$ endlich-dimensional ist, bijektiv. □

Im nächsten Abschnitt benötigen wir folgende Aussagen über Mutationen.

1) *Jede Mutation* $\mathscr{A}(a)$ *hat* a^2 *als Einselement.*

2) *Eine Mutation* $\mathscr{A}(a)$ *einer Kompositionsalgebra* $\mathscr{A}$ *ist im Falle* $|a| = 1$ *wieder eine Kompositionsalgebra.*

Beweis. ad 1). Wegen (1) gilt $a^2 \square ax = ax$ und $xa \square a^2 = xa$. Da aber die Abbildungen $x \mapsto ax$ und $x \mapsto xa$ bijektiv sind, folgt $a^2 \square x = x = x \square a^2$, $x \in \mathscr{A}$.

ad 2). Mit (1) folgt $|xa|\,|ay| = |xy| = |xa \square ay|$ und daher $|u|\,|v| = |u \square v|$ für alle $u, v \in \mathscr{A}$.

2. Mutationssatz für endlich-dimensionale Kompositionsalgebren. *Jede endlich-dimensionale Kompositionsalgebra* $\mathscr{A}$ *besitzt eine Mutation* $\mathscr{A}(a)$ *mit* $|a| = 1$, *so daß* $\mathscr{A}(a)$ *isometrisch isomorph ist zu einer der vier Algebren* $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$, $\mathbb{O}$; *speziell gilt* $\dim \mathscr{A} = 1, 2, 4$ *oder* 8.

Beweis. Da $\mathscr{A} \neq 0$, so gibt es Elemente $a \in \mathscr{A}$ mit $|a| = 1$. Nach 1.1) und 1.2) ist $\mathscr{A}(a)$ eine Kompositionsalgebra mit Einselement. Die Behauptung folgt nun aus dem Struktursatz 1.4. □

Alle in 1.2 angegebenen Kompositionsalgebren, so z. B. die 2-dimensionalen Algebren $\mathscr{A}_\nu = (\mathbb{R}^2, \underset{\nu}{\square})$, $1 \leqslant \nu \leqslant 3$, fallen unter den Mutationssatz. Eine direkte Verifikation gibt sofort:

Jede Mutation $\mathscr{A}_1(1)$, $\mathscr{A}_2(1)$, $\mathscr{A}_3(1)$, wo $1 := (1,0)$ die „komplexe Eins" bezeichnet, ist isometrisch zur Algebra $\mathbb{C}$ isomorph.

Diese Aussage zeigt besonders instruktiv, wie sich durch Mutation die Multiplikation verändern läßt: *Algebren, die weder kommutativ noch alternativ sind, werden durch Mutation kommutativ und assoziativ.*

Als Kontrast hierzu sei bemerkt:

Ist $\mathscr{A} = (V, \cdot)$ eine endlich-dimensionale, assoziative Divisionsalgebra, so gilt:

$$\mathscr{A}(a) = (V, \square) \quad \text{mit} \quad x \square y = xa^{-2}y \quad \text{für jedes} \quad a \in \mathscr{A} \setminus \{0\};$$

die Abbildung $f\colon \mathscr{A}(a) \to \mathscr{A}$, $x \mapsto a^{-2}x$, ist ein Algebraisomorphismus.

Beweis. Nach Lemma R.5 existiert $a^{-1} \in \mathscr{A}$ für alle $a \neq 0$, daher gilt $x \square y = (xa^{-1})(a^{-1}y)$ nach 1.(2). Da $\mathscr{A}$ assoziativ ist, folgt: $x \square y = xa^{-2}y$. Die Abbildung f ist linear und bijektiv, weiter gilt: $f(x \square y) = a^{-2}(xa^{-2}y) = (a^{-2}x)(a^{-2}y) = f(x)f(y)$.

3. Satz von HURWITZ (1898). *Es sei $n \geqslant 1$ eine natürliche Zahl; es seien $\zeta_1, \ldots, \zeta_n$ reelle Bilinearformen in reellen Variablen $\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_n$, so daß gilt:*

$$\zeta_1^2 + \zeta_2^2 + \cdots + \zeta_n^2 = (\xi_1^2 + \xi_2^2 + \cdots + \xi_n^2)(\eta_1^2 + \eta_2^2 + \cdots + \eta_n^2).$$

Dann folgt $n = 1, 2, 4$ oder 8.

Beweis. Man benutze den Mutationssatz 2 und das Kriterium 1.1. □

Am Schluß seiner klassischen Arbeit hat HURWITZ eine Verallgemeinerung des Kompositionsproblems formuliert:

Es seien $m \geqslant 1, n \geqslant 1$ vorgegebene natürliche Zahlen. Man bestimme die größte natürliche Zahl p, so daß die Gleichung

$$\zeta_1^2 + \cdots + \zeta_m^2 = (\xi_1^2 + \xi_2^2 + \cdots + \xi_p^2)(\eta_1^2 + \eta_2^2 + \cdots + \eta_n^2)$$

durch Bilinearformen $\zeta_1, \ldots, \zeta_m$ in $\xi_1, \ldots, \xi_p, \eta_1, \ldots, \eta_n$ lösbar ist.

Für den Fall $m = n$ hat er 1923 diese Frage vollständig gelöst in „Über die Komposition quadratischer Formen", Math. Ann. 88, 1–25 (veröffentlicht nach seinem Tode, auch Math. Werke 2, 641–666). Mit einer anderen Methode löste J. RADON 1923 die Aufgabe in „Lineare Scharen orthogonaler Matrizen", Abh. Math. Sem. Hamburg 1, 1–14; in seiner Formulierung lautet der

Satz von HURWITZ-RADON (1923). *Es sei $n = u2^{4\alpha+\beta}$, $1 \leqslant u$ ungerade, $0 \leqslant \alpha$, $0 \leqslant \beta \leqslant 3$. Dann sind folgende Aussagen äquivalent:*

i) *Es gibt n reelle Bilinearformen $\zeta_1, \ldots, \zeta_n$ in $\xi_1, \ldots, \xi_p, \eta_1, \ldots, \eta_n$ mit*

$$\zeta_1^2 + \zeta_2^2 + \cdots + \zeta_n^2 = (\xi_1^2 + \xi_2^2 + \cdots + \xi_p^2)(\eta_1^2 + \eta_2^2 + \cdots + \eta_n^2).$$

ii) *Es gilt $p \leqslant 8\alpha + 2^\beta$.*

Es ist trivial, daß stets $p \leqslant n$. Eine elementare Überlegung lehrt, daß der Fall $p = n$ genau für $n = 1$, 2, 4 und 8 gilt, wie es nach dem ursprünglichen

HURWITZschen Satz sein muß. Der Fall $n := 4k$, $p := 4$ bzw. $n := 8k$, $p := 8$ wird durch die $\mathbb{R}$-Vektorräume $\mathbb{H}^k$ bzw. $\mathbb{O}^k$ mit der Norm

$$|x|^2 := |x_1|^2 + \cdots + |x_k|^2, \qquad \text{falls} \qquad x := (x_1, \ldots, x_k),$$

realisiert: man setzt $qx := (qx_1, \ldots, qx_n)$ für $q \in \mathbb{H}$, $x \in \mathbb{H}^k$ bzw. $q \in \mathbb{O}$, $x \in \mathbb{O}^k$, und verifiziert: $|qx|^2 = |q|^2|x|^2$. Da $\dim_{\mathbb{R}} \mathbb{H}^k = 4k$, $\dim_{\mathbb{R}} \mathbb{O}^k = 8k$, so ist die Aussage i) des Hurwitz-Radonschen Satzes in beiden Fällen klar.

Die Beweismethoden von HURWITZ und RADON sind ad hoc ersonnen. 1943 veröffentlichte B. ECKMANN eine Arbeit „Gruppentheoretischer Beweis des Satzes von HURWITZ-RADON über die Komposition quadratischer Formen", Comm. Math. Helv. 15, 358–366, wo der Satz in Gedankengänge der Darstellungstheorie eingeordnet wird. Heute weiß man (Satz von ADAMS), daß – bei Übersetzung in die Sprache der Vektorfelder auf Sphären – die HURWITZ-RADON-Zahl $p - 1$ eine obere Schranke für die Zahl der unabhängigen Vektorfelder auf der $(n-1)$-Sphäre ist, vgl. hierzu das folgende Kapitel.

Kapitel 10. Divisionsalgebren und Topologie

F. Hirzebruch

In den vorangehenden Kapiteln wurden die Divisionsalgebren der reellen Zahlen, der komplexen Zahlen, der Quaternionen und der Oktaven studiert. Sie haben die Dimensionen 1, 2, 4, 8. Bis heute können die Algebraiker nicht zeigen, daß jede Divisionsalgebra die Dimension 1, 2, 4 oder 8 haben muß. Mit topologischen Methoden läßt sich diese erstaunliche Tatsache beweisen. H. HOPF konnte 1940 zeigen [7], daß die Dimension einer Divisionsalgebra eine Potenz von 2 sein muß. Sein Beweis, der die Homologiegruppen der projektiven Räume benutzt, soll in § 1 angedeutet werden. Im Jahre 1958 bewiesen M. KERVAIRE und J. MILNOR unabhängig voneinander, daß die Potenz von 2 gleich 1, 2, 4 oder 8 sein muß [9]. Sie benutzten dabei den Periodizitätssatz von R. BOTT über die Homotopiegruppen der unitären und orthogonalen Gruppen. Der Periodizitätssatz hat zur Entwicklung der K-Theorie geführt ([4], [3]), einer neuen Kohomologietheorie, mit deren Hilfe viele klassische Probleme der Topologie, bei denen die gewöhnliche Homologie- und Kohomologietheorie versagte, gelöst werden konnten. Wir werden in § 2 einen Beweis für den (1, 2, 4, 8)-Satz schildern, der im Rahmen der K-Theorie verläuft [5].

§ 1. Die Dimension einer Divisionsalgebra ist eine Potenz von 2

H. HOPF folgend werden wir einen Satz über stetige ungerade Abbildungen von Sphären beweisen, der den angestrebten Satz über Divisionsalgebren als Korollar hat. Da die Homologie der projektiven Räume benutzt wird, bringt Abschnitt 2 eine kurze Einführung in die Homologietheorie (vgl. A. DOLD: Lectures on Algebraic Topology. Springer, Berlin-Heidelberg-New York 1980, 2. Aufl.).

Die projektiven Räume sind Beispiele für Mannigfaltigkeiten. Das sind topologische Räume, die in der Umgebung eines jeden Punktes n reelle Koordinaten zulassen, n = Dimension der Mannigfaltigkeit. Diese Koordinaten kann man auch als Homöomorphismus von der Umgebung auf eine offene Teilmenge des euklidischen Raumes $\mathbb{R}^n$ auffassen. (Genau genommen sollte man die Bedingungen „hausdorffsch" und „abzählbare Basis der Topologie" hinzufügen. Ferner behandeln wir nur differenzierbare Mannigfaltigkeiten, das heißt solche, bei denen die verschiedenen Koordinatensysteme durch beliebig häufig differenzierbaren Koordinatenwechsel miteinander zusammenhängen.) Hier sei bereits festgehalten, daß wir nur zusammenhängende und kompakte Mannigfaltigkeiten betrachten. Diese Voraussetzung ist in Abschnitt 2 wichtig.

1. Ungerade Abbildungen und der Satz von HOPF. Wenn A eine Divisionsalgebra der Dimension n ist, dann kann man einen Vektorraum-Isomorphismus von A auf den $\mathbb{R}^n$ wählen und die in A definierte Multiplikation auf den $\mathbb{R}^n$ übertragen:

$$\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n, \qquad (x, y) \to z = x \cdot y.$$

Der Vektor $z = (\zeta_1, \ldots, \zeta_n)$ hängt von $x = (\xi_1, \ldots, \xi_n)$ und $y = (\eta_1, \ldots, \eta_n)$ ab, und zwar ist ζ_i eine Bilinearform in den ξ_r und η_s. Im $\mathbb{R}^n$ ist die übliche Euklidische Länge $\|x\| = \xi_1^2 + \xi_2^2 + \cdots + \xi_n^2$ definiert, und die $(n-1)$-dimensionale Sphäre S^{n-1} der Vektoren der Länge 1 kann eingeführt werden. Die obige Multiplikationsabbildung $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ kann auf $S^{n-1} \times S^{n-1}$ eingeschränkt werden. Die Beschränkung heiße f. Da die Algebra nullteilerfrei ist, nimmt f den Wert $0 \in \mathbb{R}^n$ nicht an, und die Abbildung $g = f/\|f\|$ ist wohldefiniert:

$$g\colon S^{n-1} \times S^{n-1} \to S^{n-1}.$$

HOPF benötigt für seinen Beweis nur, daß g eine stetige *ungerade* Abbildung ist, das heißt,

$$g(-x, y) = g(x, -y) = -g(x, y) \qquad \text{für} \qquad x, y \in S^{n-1}.$$

(Die Abbildung $z \mapsto -z$ ordnet jedem Punkt der Sphäre seine Antipode zu.)

Satz. *Wenn es eine stetige ungerade Abbildung von $S^{n-1} \times S^{n-1}$ in S^{n-1} gibt, dann ist n eine Potenz von 2.*

Korollar. *Die Dimension einer Divisionsalgebra über $\mathbb{R}$ ist eine Potenz von 2.*

Der Satz ist wesentlich allgemeiner als sein Korollar. Die ungerade Abbildung könnte zum Beispiel durch reelle algebraische Formen gegeben sein, welche homogen in den ξ_r von ungeradem Grad sowie homogen in den η_s von ungeradem Grad sind. HOPF wendet sich auch einer weiteren Verallgemeinerung zu. Er betrachtet nämlich stetige ungerade Abbildungen $S^{p-1} \times S^{n-1} \to S^{m-1}$. Für welche p, n, m solche Abbildungen existieren, scheint bis heute nicht vollständig bekannt zu sein. Wenn das in 9.2.3 erwähnte Kompositionsproblem von HURWITZ für p, n, m lösbar ist, dann existiert eine ungerade Abbildung $S^{p-1} \times S^{n-1} \to S^{m-1}$. Wir wollen uns hier auf den Fall $p = n = m$ beschränken.

Bis zum Beweis des obigen Satzes von HOPF ist noch ein recht langer Weg, da man die Homologie der reellen projektiven Räume verwenden muß und im nächsten Paragraphen zunächst eine kleine Einführung in die Homologietheorie folgen soll.

Der reelle projektive Raum $\mathbb{P}^{n-1}$ ist die $(n-1)$-dimensionale Mannigfaltigkeit, welche entsteht, wenn man auf der Sphäre S^{n-1} jeden Punkt mit seinem Antipodenpunkt identifiziert. Mit $\alpha\colon S^{n-1} \to \mathbb{P}^{n-1}$ soll die Abbildung der Antipodenidentifikation bezeichnet werden. Jeder k-dimensionale lineare Unterraum des $\mathbb{R}^n$ definiert eine Einbettung der Sphäre S^{k-1} in S^{n-1} und deshalb (als Bild unter α) einen $(k-1)$-dimensionalen projektiven Unterraum von $\mathbb{P}^{n-1}$, der häufig einfach mit $\mathbb{P}^{k-1}$ bezeichnet werden soll. Für $k = 2$ erhält man die Großkreise auf S^{n-1}, deren Bilder unter α die projektiven Geraden in $\mathbb{P}^{n-1}$ sind. Eine ungerade

Abbildung $g\colon S^{n-1} \times S^{n-1} \to S^{n-1}$ induziert eine Abbildung

$$G\colon \mathbb{P}^{n-1} \times \mathbb{P}^{n-1} \to \mathbb{P}^{n-1},$$

auf welche HOPF die Homologietheorie anwendet. (Die Homologie der Sphären ist zu trivial, um Ergebnisse zu erhalten!) Aber nun zunächst zur Homologietheorie überhaupt.

2. Homologie und Kohomologie mit Koeffizienten in F_2. Es sei X ein topologischer Raum. Zwei Punkte P und Q in X heißen homolog, wenn sich P und Q durch einen Weg in X verbinden lassen. Die Menge S_0X der Homologieklassen von Punkten ist also gleich der Menge der (wegweisen) Zusammenhangskomponenten von X. Die nullte oder nulldimensionale Homologiegruppe $H_0(X)$ mit Koeffizienten in F_2, dem Körper mit 2 Elementen, ist der F_2-Vektorraum aller formalen Linearkombinationen der Elemente von S_0X mit Koeffizienten in F_2. Wenn X wegzusammenhängend ist, dann ist $H_0(X) \cong F_2$.

Um die q-dimensionale Homologiegruppe $H_q(X)$ für $q > 0$ zu definieren, muß man anstatt der Punkte q-dimensionale Gebilde (Zyklen) in X betrachten, mit ihnen formale Linearkombinationen mit Koeffizienten in F_2 bilden und modulo einer „Homologie" genannten Äquivalenzrelation rechnen. Einzelheiten können hier nicht näher ausgeführt werden. Folgendes muß aber erwähnt werden.

a) Jeder geschlossene Weg w in X repräsentiert eine Homologieklasse $|w| \in H_1(X)$.

b) Jede q-dimensionale Untermannigfaltigkeit M einer n-dimensionalen Mannigfaltigkeit X repräsentiert eine Homologieklasse $|M| \in H_q(X)$. Wenn $q = 0$, dann kommt man auf die anfangs erwähnten Homologieklassen von Punkten zurück. Es ist $H_n(X) \cong F_2$, und $|X|$ ist das von 0 verschiedene Element von $H_n(X)$.

Im allgemeinen werden durch a) und b) nicht alle Homologieklassen erfaßt, aber bei den Sphären S^n ($n > 0$) und den projektiven Räumen $\mathbb{P}^n$ ist es der Fall (beachte, daß im Vergleich zu Abschnitt 1 der Dimensionsindex $n - 1$ durch n ersetzt wurde):

Die Homologiegruppen $H_0(S^n)$ und $H_n(S^n)$ haben beide den Rang 1 als F_2-Vektorräume. Sonst ist $H_q(S^n) = 0$. Die von 0 verschiedenen Elemente sind $|P| \in H_0(S^n)$, wo P ein beliebiger Punkt ist, und $|S^n| \in H_n(S^n)$. Für die projektiven Räume $\mathbb{P}^n$ gilt:

Die Homologiegruppen $H_q(\mathbb{P}^n)$ haben den Rang 1 für $0 \leqslant q \leqslant n$ (das heißt, $H_q(\mathbb{P}^n) \cong F_2$). Sonst ist $H_q(\mathbb{P}^n) = 0$. Alle q-dimensionalen projektiven Unterräume $\mathbb{P}^q \subset \mathbb{P}^n$ ($0 \leqslant q \leqslant n$) sind zueinander homolog, und zwar ist $|\mathbb{P}^q|$ das von 0 verschiedene Element von $H_q(\mathbb{P}^n)$.

Wir fahren in der allgemeinen Beschreibung der Homologie fort: Jede stetige Abbildung $f\colon X \to Y$ zwischen den topologischen Räumen X und Y *induziert einen Homomorphismus* $f_*\colon H_q(X) \to H_q(Y)$. Die Definition von f_* lautet für Homologieklassen von Punkten bzw. von geschlossenen Wegen folgendermaßen: Für einen Punkt $P \in X$ ist $f_*|P| = |f(P)|$. Für einen geschlossenen Weg w in X ist $f \cdot w$ ein geschlossener Weg in Y und $f_*|w| = |f \cdot w|$.

Der Übergang von f nach f_* ist mit dem Hintereinanderschalten verträglich: Für $f\colon X \to Y$ und $g\colon Y \to Z$ gilt $(g \cdot f)_* = g_* \cdot f_*$. Durch (H_q, f_*) wird ein „kovarianter Funktor" gegeben. Der Hopfsche Beweis erfordert neben der

Homologie die Kohomologie: Die q-te *Kohomologiegruppe* von X mit Koeffizienten in F_2 ist der zu $H_q(X)$ duale Vektorraum

$$H^q(X) = \mathrm{Hom}(H_q(X), F_2),$$

dessen Elemente die linearen Abbildungen $u\colon H_q(X) \to F_2$ sind. Wenn $x \in H_q(X)$ ist, dann wird der Wert von u auf x mit $\langle u, x\rangle \in F_2$ bezeichnet. Einer stetigen Abbildung $f\colon X \to Y$ ordnet man den zu $f_*\colon H_q(X) \to H_q(Y)$ dualen Homomorphismus

$$f^*\colon H^q(Y) \to H^q(X), \qquad f^*(u) = u \cdot f_*, \qquad \text{also} \qquad \langle f^*(u), x\rangle = \langle u, f_* x\rangle,$$

zu. Beim Übergang von f nach f^* wird die Richtung umgekehrt, und beim Hintereinanderschalten gilt $(g \cdot f)^* = f^* \cdot g^*$.

Soweit bringt die Kohomologie nicht wesentlich Neues. Aber man kann jetzt ein *Produkt*, das heißt, eine bilineare Abbildung

$$H^p(X) \times H^q(X) \to H^{p+q}(X), \qquad (u, v) \to u \cdot v,$$

definieren, wodurch die direkte Summe $H^*(X) = \bigoplus_{p \geq 0} H^p(X)$ zu einem assoziativen und kommutativen Ring (graduierten F_2-Algebra) wird. Wie das gemacht wird, kann hier nicht ausgeführt werden. Der Homomorphismus f^* ist mit dem Produkt verträglich, $f^*(u \cdot v) = f^*(u) \cdot f^*(v)$, ist also ein Ringhomomorphismus

$$f^*\colon H^*(Y) \to H^*(X).$$

Der Rang von $H_p(X)$ als F_2-Vektorraum wird mit $b_p(X)$ bezeichnet (p-te Bettische Zahl). Wenn die Bettischen Zahlen endlich sind, was bei uns immer der Fall sein wird, ist $b_p(X)$ auch gleich dem Rang von $H^p(X)$. Für eine n-dimensionale Mannigfaltigkeit X ist $b_p(X) = b_{n-p}(X)$. Das ist der *Poincarésche Dualitätssatz* (1895), der heute so formuliert werden kann: Es gibt eine kanonische Isomorphie

$$\pi\colon H_{n-p}(X) \xrightarrow{\cong} H^p(X),$$

die eine *geometrische Deutung* des Kohomologieproduktes in Mannigfaltigkeiten ermöglicht: Wenn M und N Untermannigfaltigkeiten von X der Kodimension p und q sind, die transversal zueinanderliegen, so daß ihr Durchschnitt eine Untermannigfaltigkeit der Kodimension $p + q$ ist, dann gilt

$$\pi(|M|) \cdot \pi(|N|) = \pi(|M \cap N|).$$

Dem Produkt in $H^*(X)$ entspricht also die Durchschnittsbildung. Die direkte Summe $H_*(X) = \bigoplus_{p \geq 0} H_p(X)$ ist aufgrund der Poincaréschen Isomorphie für Mannigfaltigkeiten ebenfalls ein Ring. Die Multiplikation

$$H_{n-p}(X) \times H_{n-q}(X) \to H_{n-(p+q)}(X)$$

heißt Schnittprodukt. Wenn X zusammenhängend ist ($H_0(X) \cong F_2$), dann ist für $x \in H_{n-p}(X)$, $y \in H_p(X)$ das Schnittprodukt $x \cdot y \in H_0(X)$ als Element von F_2 anzusehen (Schnittzahl). Es gilt

$$\langle \pi(x), y\rangle = x \cdot y.$$

Der Schnittring für Mannigfaltigkeiten ist viel länger bekannt als der Kohomologiering eines beliebigen topologischen Raumes (vgl. Abschnitt 4). Der Kohomologiering des projektiven Raumes $\mathbb{P}^n$ läßt sich nun leicht bestimmen: Der Durch-

schnitt von q projektiven Unterräumen der Dimension $n-1$, die sich in allgemeiner Lage befinden, ist ein projektiver Unterraum $\mathbb{P}^{n-q}$ der Dimension $n-q$ $(0 \leqslant q \leqslant n)$. Bezeichnet man $\pi(|\mathbb{P}^{n-1}|)$ mit u, dann ist u das von 0 verschiedene Element von $H^1(\mathbb{P}^n)$ und $u^q = \pi(|\mathbb{P}^{n-q}|)$ das von 0 verschiedene Element von $H^q(\mathbb{P}^n)$. Also ist $H^*(\mathbb{P}^n)$ der Polynomring über F_2 in u mit der Relation $u^{n+1} = 0$, die sich aus $H^q(\mathbb{P}^n) = 0$ für $q > n$ ergibt.

Der Kohomologiering des cartesischen Produktes $\mathbb{P}^n \times \mathbb{P}^n$ läßt sich ebenfalls leicht angeben: Die Homologieklassen $|\mathbb{P}^r \times \mathbb{P}^s|$ mit $r+s=q$ und $0 \leqslant r \leqslant n$, $0 \leqslant s \leqslant n$ bilden eine Basis für $H_q(\mathbb{P}^n \times \mathbb{P}^n)$. Aufgrund des Schnittproduktes $|\mathbb{P}^r \times \mathbb{P}^s| \cdot |\mathbb{P}^k \times \mathbb{P}^l| = |\mathbb{P}^r \cap \mathbb{P}^k \times \mathbb{P}^s \cap \mathbb{P}^l|$ folgt: Der Kohomologiering ist der Polynomring über F_2 mit Unbestimmten u und v modulo der Relationen $u^{n+1} = 0$, $v^{n+1} = 0$, wobei u und v mittels des Poincaréschen Isomorphismus von $\mathbb{P}^n \times \mathbb{P}^n$ aus den Homologieklassen $|\mathbb{P}^{n-1} \times \mathbb{P}^n|$ bzw. $|\mathbb{P}^n \times \mathbb{P}^{n-1}|$, die eine Basis von $H_{2n-1}(\mathbb{P}^n \times \mathbb{P}^n)$ bilden, hervorgehen. Durch Schnittzahlenbildung sieht man, daß

$$\langle u, |\mathbb{P}^1 \times \text{Punkt}| \rangle = 1, \qquad \langle u, |\text{Punkt} \times \mathbb{P}^1| \rangle = 0$$

und entsprechendes für v gilt. Dies wird in Abschnitt 3 wichtig sein.

3. Beweis des Satzes von HOPF. In Abschnitt 1 wurde die Abbildung

$$G\colon \mathbb{P}^{n-1} \times \mathbb{P}^{n-1} \to \mathbb{P}^{n-1}$$

betrachtet. Die erste Homologie von $\mathbb{P}^{n-1} \times \mathbb{P}^{n-1}$ hat $|\mathbb{P}^1 \times \text{Punkt}|$ und $|\text{Punkt} \times \mathbb{P}^1|$ als Basis, wo $\mathbb{P}^1$ irgendein eindimensionaler projektiver Teilraum von $\mathbb{P}^{n-1}$ ist (er entsteht aus einem Großkreis durch Antipodenidentifikation). Wir behaupten, daß

$$G_*(|\mathbb{P}^1 \times \text{Punkt}|) = G_*(|\text{Punkt} \times \mathbb{P}^1|) = |\mathbb{P}^1|.$$

Warum ist $G_*(|\mathbb{P}^1 \times \text{Punkt}|)$ *das von* 0 *verschiedene Element von* $H_1(\mathbb{P}^{n-1})$? Man benutzt folgendes Kriterium für die Homologieklasse $|w| \in H_1(\mathbb{P}^{n-1})$ eines geschlossenen Weges in $\mathbb{P}^{n-1}$: Es gibt einen Weg $\tilde{w}$ in S^{n-1}, welcher bei der Antipodenidentifikation $\alpha\colon S^{n-1} \to \mathbb{P}^{n-1}$ in den (*einmal* durchlaufenen) Weg w übergeht. Der Weg $\tilde{w}$ ist entweder geschlossen, oder er verbindet zwei Antipodenpunkte. Im ersten Fall ist $|w| = 0$, im zweiten $|w| \neq 0$.

Nun ist $\mathbb{P}^1 \times \text{Punkt}$ ein geschlossener Weg $w \times \text{Punkt}$ in $\mathbb{P}^{n-1} \times \mathbb{P}^{n-1}$, der Weg $\tilde{w}$ ist ein „halber Großkreis", verbindet also zwei Antipodenpunkte. Da für die Abbildung $g\colon S^{n-1} \times S^{n-1} \to S^{n-1}$, die G induziert, die Gleichung $g(-x, y) = -g(x, y)$ gilt, verbindet auch der Bildweg $g \cdot (\tilde{w} \times \text{Punkt})$ in S^{n-1} zwei Antipodenpunkte. Bei der Antipodenidentifikation geht dieser Weg in $G \cdot (w \times \text{Punkt})$ über, welcher die Homologieklasse $G_*(|\mathbb{P}^1 \times \text{Punkt}|)$ repräsentiert, die also wie behauptet verschieden von 0 ist.

Der Kohomologiering von $\mathbb{P}^{n-1}$ werde als Polynomring über F_2 in der Unbestimmten t mit der Relation $t^n = 0$ geschrieben, während der Kohomologiering von $\mathbb{P}^{n-1} \times \mathbb{P}^{n-1}$ der Polynomring in u, v mit den Relationen $u^n = 0, v^n = 0$ ist (vgl. Abschnitt 2). Wir behaupten

$$G^*(t) = u + v.$$

Zum Beweis beachten wir, daß

$$\langle G^*(t), |\text{Punkt} \times \mathbb{P}^1| \rangle = \langle t, G_* |\text{Punkt} \times \mathbb{P}^1| \rangle = \langle t, |\mathbb{P}^1| \rangle = 1 \in F_2$$

und entsprechend bei $|\mathbb{P}^1 \times \text{Punkt}|$. Andererseits ist auch

$$\langle u + v, |\text{Punkt} \times \mathbb{P}^1| \rangle = \langle u + v, |\mathbb{P}^1 \times \text{Punkt}| \rangle = 1,$$

vgl. die letzte Formel in Abschnitt 2. Eine Kohomologieklasse ist aber durch ihren Wert auf den Homologieklassen bestimmt.

Nun kommt der eigentliche Beweis des Satzes von HOPF [7], der, wenn die Homologietheorie bekannt ist, beeindruckend kurz ist: Aus $t^n = 0$ folgt

$$G^*(t^n) = (G^*(t))^n = (u + v)^n = 0.$$

Es ist

$$0 = (u + v)^n = \sum_{k=1}^{n-1} \binom{n}{k} u^k v^{n-k}$$

(wegen der Relationen $u^n = 0$, $v^n = 0$). Also sind alle Binomialkoeffizienten $\binom{n}{k}$ gerade ($1 \leqslant k < n$), also ist n eine Potenz von 2.

4. Historische Bemerkungen zur Homologie- und Kohomologietheorie. Die Entwicklung der algebraischen Topologie begann mit H. POINCARÉ (1854–1912). Man findet bei ihm jedoch noch keine Homologiegruppen $H_q(X)$ sondern nur die Bettischen Zahlen $b_q(X)$. (POINCARÉ benutzt die ganzen Zahlen als „Koeffizientenbereich". Die von uns verwendeten Koeffizienten F_2 haben den Vorteil, daß Vorzeichen- und Orientierungsfragen entfallen.) Anstatt von algebraischer Topologie sprach man von kombinatorischer analysis situs. Die gruppentheoretische Formulierung der Homologie verdankt man Emmy NOETHER (1882–1935). Das Schnittprodukt in Mannigfaltigkeiten war Jahrzehnte vor dem Kohomologiering bekannt. POINCARÉ benutzte bereits Schnittzahlen für seinen Dualitätssatz. Der Ringhomomorphismus f^*: $H^*(Y) \to H^*(X)$ für beliebige topologische Räume X, Y und stetige Abbildungen f: $X \to Y$ hat einen Vorläufer in dem von HOPF (1928–1930) für Mannigfaltigkeiten M und N und eine stetige Abbildung f: $M \to N$ definierten Umkehrhomomorphismus ϕ: $H_*(N) \to H_*(M)$. H. SAMELSON [12] schreibt darüber:

„Eine stetige Abbildung f: $M \to N$ induziert eine Abbildung der Schnittringe, die linear, aber unglücklicherweise im allgemeinen *nicht* multiplikativ ist. Für Hopf erhob sich die Aufgabe herauszufinden, ob man f nicht *etwas* Multiplikatives zuordnen kann. Nun weiß man, daß in der Mengentheorie eine Abbildung f zwar die Summe aber nicht den Durchschnitt erhält (in Analogie zu dem eben erwähnten Tatbestand), während die Operation f^{-1} (= totales Urbild) additiv und multiplikativ ist. Wohl davon angeregt, definierte Hopf für jede Homologieklassse von N ein „Umkehrbild" als das richtig aufgefaßte totale Urbild in M. (Etwas ungenau gesagt, schneidet man, für einen Zyklus z in N, den Zyklus $M \times z$ (in $M \times N$) mit dem Graphen von f und projiziert das Resultat in M hinein.) Die Konstruktion ergibt einen Ring(Algebra-)-Homomorphismus ϕ des Ringes von N in den von M, der mit f durch die Formel $f_*(\phi(z) \cdot x) = z \cdot f_*(x)$ (analog zu der Formel

$f(f^{-1}(A) \cap B) = A \cap f(B)$ in der Mengentheorie) verbunden ist. (Die Konstruktion erfordert nicht, daß M und N gleiche Dimension haben; ϕ erhöht die Dimension durch $\dim M - \dim N$.)...

Die Konstruktion des Umkehrungshomomorphismus, die Hopf mit sicherem Spürsinn von der elementaren mengentheoretischen Analogie ablas, erwies sich später als „das erste Erscheinen der Kohomologie". Man interpretiert ϕ heute als die Komposition von (a) Poincaré-Dualität in N (von Homologie nach Kohomologie), (b) der Kohomologieabbildung f^*, die von f induziert wird, und (c) Poincaré-Dualität in M (von Kohomologie nach Homologie). Die Kohomologieabbildung ist für beliebige Räume multiplikativ, und Poincaré-Dualität bildet den Schnitt (in Homologie) auf das Cup-Produkt (in Kohomologie) ab (heute *definiert* man oft den Schnitt auf diese Weise). Es dauerte aber noch einige Jahre, bis Kohomologie entdeckt wurde (1935, Alexander, Kolmogoroff, Whitney)."

Über „das erste Erscheinen der Kohomologie" sagt Hopf 1966 in seinem Vortrag „Einige persönliche Erinnerungen aus der Vorgeschichte der heutigen Topologie" [8] folgendes:

„Das Jahr 1935 war für die Entwicklung der Topologie aus mehreren Gründen besonders bedeutungsvoll. Im September fand in Moskau die „Erste Internationale Konferenz über Topologie" statt. Die Vorträge, die auf dieser Konferenz völlig unabhängig voneinander von J. W. Alexander, I. Gordon und A. N. Kolmogoroff gehalten wurden, darf man als den Beginn der Cohomologie-Theorie ansehen – (für welche allerdings Lefschetz von 1930 mit seinen „Pseudo-Zyklen" die Rolle eines Vorläufers gespielt hatte).

Was mich – und wahrscheinlich manche andere Topologen – damals vollständig überraschte, waren nicht die Cohomologie-Gruppen – diese sind ja nichts anderes als die Charakterengruppen der Homologiegruppen – als vielmehr die Tatsache, daß man zwischen ihnen, in beliebigen Komplexen und allgemeineren Räumen, eine Multiplikation erklären kann, also den Cohomologie-Ring, der den Schnittring der Mannigfaltigkeiten verallgemeinert. Wir hatten geglaubt, so etwas sei nur, dank der lokalen Euklidizität, in Mannigfaltigkeiten möglich."

5. Charakteristische Homologieklassen nach STIEFEL. Ergänzend zu den vorangegangenen Ausführungen soll die Frage nach den Dimensionen der Divisionsalgebren zu anderen topologischen Problemen in Zusammenhang gebracht werden.

Es sei M eine Mannigfaltigkeit der Dimension n. Der Tangentialraum T_xM ist für jeden Punkt $x \in M$ wohldefiniert. Er ist ein n-dimensionaler Vektorraum. Ein Vektorfeld v in M ist eine Funktion, die jedem $x \in M$ einen Vektor $v(x) \in T_xM$ zuordnet. Natürlich soll v stetig sein. Wann existiert ein Vektorfeld v, das nirgendwo verschwindet? Die Antwort ist in dem berühmten Satz von Hopf aus dem Jahre 1926 enthalten (Vorläufer bei Poincaré, Brouwer and Hadamard). Vergleiche das Buch von Milnor [10].

Ein Vektorfeld v ohne Nullstellen existiert genau dann, wenn die Euler-Poincarésche Charakteristik von M verschwindet.

Die Euler-Poincarésche Charakteristik $\chi(M)$ ist die Wechselsumme der Bettischen Zahlen (Abschnitt 2), also $\chi(M) = \sum_{i=0}^{n} (-1)^i b_i(M)$. Für die Sphäre S^n

($n \geqslant 1$) ist $\chi(S^n) = 2$ für n gerade und $\chi(S^n) = 0$ für n ungerade. Für die projektiven Räume gilt $\chi(\mathbb{P}^n) = 1$ für n gerade und $\chi(\mathbb{P}^n) = 0$ für n ungerade. Für n ungerade existieren also nirgendwo verschwindende Vektorfelder auf S^n und $\mathbb{P}^n$, für n gerade dagegen nicht.

Unter einem k-Feld wollen wir ein k-tupel $v_1, \ldots, v_k$ von Vektorfeldern auf M verstehen, so daß die Vektoren $v_1(x), \ldots, v_k(x)$ in jedem Punkt $x \in M$ linear-unabhängig sind. Ein 1-Feld ist also ein Vektorfeld ohne Nullstellen. Das maximale k, für das ein k-Feld existiert, soll Span(M) heißen. Natürlich ist $0 \leqslant \text{Span}(M) \leqslant n = \dim M$.

Wenn $\text{Span}(M) = n$, dann heißt die Mannigfaltigkeit parallelisierbar (man sagt auch, es existiert ein Fernparallelismus). Man kann nämlich dann zwei Vektoren in verschiedenen Punkten x und y von M parallel nennen, wenn sie bezüglich der Basen $v_1(x), \ldots, v_n(x)$ von $T_x(M)$ und $v_1(y), \ldots, v_n(y)$ von T_yM gleiche Koeffizienten haben. Der Raum aller Tangentialvektoren, das heißt, $\bigcup_{x \in M} T_xM$, kann also dann bijektiv auf $M \times \mathbb{R}^n$ abgebildet werden.

Es ist ein schwieriges Problem, für eine vorgegebene Mannigfaltigkeit Span(M) zu bestimmen, z. B. für die Sphären und die projektiven Räume. Offensichtlich ist $\text{Span}(S^n) \geqslant \text{Span}(\mathbb{P}^n)$, denn ein Vektorfeld auf $\mathbb{P}^n$ ist nichts anderes als ein Vektorfeld auf S^n, das bei der Antipodenabbildung in sich übergeht.

Heute weiß man, daß $\text{Span}(S^n) = \text{Span}(\mathbb{P}^n)$, und auch den genauen Wert kennt man. Darüber einige Bemerkungen in § 3.

Es besteht ein enger Zusammenhang zwischen Divisionsalgebren und verwandten algebraischen Strukturen und der Existenz von Vektorfeldern auf Sphären und projektiven Räumen (siehe § 3). Einfachstes Beispiel ist der folgende **Satz**:

Wenn eine Divisionsalgebra der Dimension n über $\mathbb{R}$ *existiert, dann sind der projektive Raum* $\mathbb{P}^{n-1}$ *und die Sphäre* S^{n-1} *parallelisierbar.*

Beweis. Wie in Abschnitt 1 betrachten wir die Multiplikation $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ ($(x, y) \to z = x \cdot y$). Es seien $e_1, \ldots, e_n$ die Standard-Basisvektoren des $\mathbb{R}^n$ und $y \in S^{n-1}$. Dann sind die Vektoren $e_1 \cdot y, \ldots, e_n \cdot y$ linear-unabhängig. Orthonormalisiert man sie, dann erhält man n Vektoren $w_1(y), \ldots, w_n(y)$ mit $w_1(y) = e_1 \cdot y / \|e_1 \cdot y\|$. Die Vektoren $w_2(y), \ldots, w_n(y)$ sind Tangentialvektoren an S^{n-1} in Punkte $w_1(y)$. Da $y \to w_1(y)$ eine bijektive Abbildung von S^{n-1} auf sich ist, hat man ein $(n-1)$-Feld auf S^{n-1} gefunden, das offensichtlich antipodentreu ist.

H. Hopf (seit 1931 Professor an der ETH in Zürich) schlug einem seiner ersten Schüler, nämlich E. Stiefel, das folgende Problem vor:

In welchen Mannigfaltigkeiten M der Dimension n gibt es ein m-Feld? In anderen Worten: *Wann ist* $\text{Span}(M) \geqslant m$?

In seiner Dissertation [14] entwickelte Stiefel die Theorie der charakteristischen Homologieklassen: Er ließ m-Felder mit Singularitäten zu. Ein singulärer Punkt eines m-Feldes $v_1, \ldots, v_m$ ist ein Punkt $x \in M$, wo $v_1(x), \ldots, v_m(x)$ linear abhängig sind. Stiefel zeigte, daß es stets ein m-Feld gibt, dessen Singularitätenmenge $(m-1)$-dimensional ist und als $(m-1)$-dimensionaler Zyklus für die Homologie mit Koeffizienten in F_2 aufgefaßt werden kann.

Hauptergebnis. *Die Homologieklasse* $s_{m-1} \in H_{m-1}(M)$ $(m = 1, \ldots, n)$ *der Singularitätenmenge ist unabhängig von der Wahl des m-Feldes.*

(Wir haben STIEFELS Theorie vereinfacht, für gewisse m benutzt STIEFEL Homologie mit ganzzahligen Koeffizienten, die aber mod 2 reduziert werden kann, so daß sich dann stets unser s_{m-1} ergibt.)

Die s_{m-1} sind die charakteristischen Homologieklassen von STIEFEL. In der Homologie von M werden also bestimmte Elemente durch die Eigenschaften des Tangentialbündels $\bigcup_{x \in M} T_x(M)$ ausgezeichnet. Es gilt

$$\mathrm{Span}(M) \geqslant m \;\rightarrow\; s_0 = 0, s_1 = 0, \ldots, s_{m-1} = 0.$$

Nach dem Satz von HOPF über Vektorfelder ist $s_0 = 0$ genau dann, wenn $\chi(M)$ gerade ist.

Die Berechnung der Stiefelschen Klassen führt also zu Aussagen über $\mathrm{Span}(M)$, was z. B. für $M = \mathbb{P}^n$ Aussichten auf Erfolg hat, weil die Homologie von $\mathbb{P}^n$ nicht trivial ist, für $M = S^n$ jedoch zum Scheitern verurteilt ist.

In einer späteren Arbeit [15] berechnet STIEFEL die charakteristischen Homologieklassen des $\mathbb{P}^n$ durch Konstruktion spezieller m-Felder mit Singularitäten. Diese Arbeit ist am gleichen Tage bei den Commentarii eingegangen wie die Arbeit von HOPF [7]. STIEFELS Ergebnis kann durch ein fast wörtliches Zitat wiedergegeben werden:

Die charakteristische Homologieklasse s_{m-1} *des* $\mathbb{P}^n$ *ist die Nullklasse oder die Klasse, welche* $\mathbb{P}^{m-1}$ *enthält, je nachdem* $\binom{n+1}{m}$ *gerade oder ungerade ist.*

(Beachte, daß für $m = 1$ in der Tat $\binom{n+1}{m} = n + 1 = \chi(\mathbb{P}^n) \bmod 2$.)

Was folgt jetzt aus $\mathrm{Span}(\mathbb{P}^{n-1}) \geqslant m - 1$?
Antwort:

$$\mathrm{Span}(\mathbb{P}^{n-1}) \geqslant m - 1 \;\rightarrow\; \binom{n}{k} \text{ gerade für } \quad 0 < k < m.$$

Insbesondere impliziert die Parallelisierbarkeit von $\mathbb{P}^{n-1}$, daß $\binom{n}{k}$ gerade ist für $0 < k < n$ und also n eine Potenz von 2 sein muß. Damit haben wir aufgrund des vorstehenden Satzes erneut bewiesen, daß eine Divisionsalgebra nur existieren kann, wenn n eine Potenz von 2 ist.

§ 2. Die Dimension einer Divisionsalgebra ist gleich 1, 2, 4 oder 8

Die folgenden 8 Abschnitte enthalten einen Beweis dafür, daß es nur in den Dimensionen 1, 2, 4 und 8 Divisionsalgebren geben kann. Der Beweis benutzt Methoden der algebraischen Topologie, die in 10.1.2 besprochene Kohomologietheorie, so wie die Theorien der Vektorraumbündel und charakteristischen Klassen, die im 3. und 4. Abschnitt eingeführt werden sollen (für eine ausführliche Darstellung siehe [11]). Das entscheidende Mittel zum Beweis ist der BOTTsche

Periodizitätssatz. Er wird ohne jede Beweisandeutung im 6. Abschnitt angegeben. Alle Beweise des (1, 2, 4, 8)-Satzes benutzen die Bottsche Periodizität. Der hier zu schildernde Beweis stammt von Atiyah und Hirzebruch [5]. Die ersten Beweise wurden, wie schon zu Beginn dieses Kapitels erwähnt, kurz nach dem Erscheinen des Periodizitätssatzes unabhängig voneinander von Kervaire und Milnor 1958 gefunden.

1. Die mod 2-Invariante $\alpha(f)$. Bei einer stetigen Abbildung $\phi\colon S^{n-1} \to S^{n-1}$ nennt man y einen regulären Wert, wenn es zu jedem x mit $\phi(x) = y$ eine Umgebung gibt, welche durch ϕ homöomorph auf eine Umgebung von y abgebildet wird. Dann ist die Anzahl

$$\#\phi^{-1}(y) < \infty.$$

Die Parität

$$\#\phi^{-1}(y) \text{ modulo } 2$$

hängt nicht von der Wahl von y ab. Man nennt sie den mod 2-Abbildungsgrad von ϕ. Bei einem Homöomorphismus ist der Grad $\neq 0$. Der Satz von Sard besagt, daß jede C^∞-Abbildung (das heißt, beliebig häufig differenzierbare Abbildung) ϕ stets reguläre Werte und somit einen Abbildungsgrad besitzt. Wenn ϕ lediglich stetig ist, kann man ϕ durch C^∞-Abbildungen approximieren. Alle hinreichend guten Approximationen haben denselben Grad, so daß man auch bei jeder stetigen Abbildung $\phi\colon S^{n-1} \to S^{n-1}$ von ihrem mod 2-Abbildungsgrad sprechen kann. Siehe J. Milnor [10]. Man kann ϕ mit der gleichen Methode eine ganze Zahl als Abbildungsgrad zuordnen, die bei Reduktion modulo 2 den mod 2-Abbildungsgrad ergibt. (Man muß in $\phi^{-1}(y)$ die Punkte mit Multiplizität -1 oder $+1$ zählen, je nachdem Orientierungswechsel eintritt oder nicht.)

Es sei $GL(n)$ die topologische Gruppe der $(n \times n)$-reihigen invertierbaren Matrizen. Man kann, wenn man die n Spalten betrachtet, $GL(n)$ auch als Menge der Basen des $\mathbb{R}^n$ ansehen. Es werden nun stetige Abbildungen

$$f\colon S^{n-1} \to GL(n)$$

betrachtet. Nach Wahl eines festen Vektors $v \in S^{n-1}$ definiert man die stetige Abbildung

$$\phi\colon S^{n-1} \to S^{n-1}, \qquad \phi(x) = \frac{f(x) \cdot v}{\|f(x) \cdot v\|}.$$

Ihr mod 2-Abbildungsgrad hängt nicht von der Wahl von v ab. Man nennt ihn die mod 2-Invariante $\alpha(f)$ von f. Man hat folgendes tiefliegendes Ergebnis

Satz. *Wenn die* mod 2*-Invariante einer stetigen Abbildung* $f\colon S^{n-1} \to GL(n)$ *von* 0 *verschieden ist, dann ist* $n = 1, 2, 4$ *oder* 8.

Im folgenden Abschnitt wird dieses Ergebnis auf Divisionsalgebren angewandt. Vom übernächsten Abschnitt an wird geschildert, mit welchen Methoden man diesen Satz beweist.

2. Parallelisierbarkeit der Sphären und Divisionsalgebren. Die Sphäre S^{n-1} sei parallelisierbar, für jeden Vektor $x \in S^{n-1}$ gibt es dann $n-1$ linear unabhängige Vektoren $w_2(x), \ldots, w_n(x)$, die senkrecht auf x stehen und stetig von x abhängen. Die n „Spalten" $x, w_2(x), \ldots, w_n(x)$ bilden ein Element $f(x) \in GL(n)$. Wenn v der Vektor $(1, 0, \ldots, 0)$ des $\mathbb{R}^n$ ist, dann gilt $f(x)v = x$ und deshalb $\alpha(f) = 1$. Setzt man den Satz des 1. Abschnitts voraus, dann folgt:

Die Sphäre S^{n-1} ist nur für $n = 1, 2, 4, 8$ parallelisierbar. Eine Divisionsalgebra gibt es höchstens in den Dimensionen $1, 2, 4, 8$ (vgl. den Satz in 10.1.5).

(Eigentlich sollte man den Fall $n = 1$ ausschließen, weil er eventuell zu trivialen Zusatzbetrachtungen führt.)

Geht man direkt von der Divisionsalgebra aus (Multiplikation $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, $(x, y) \to x \cdot y$), dann definiert man

$$f: S^{n-1} \to GL(n)$$

durch $f(x) \cdot v = x \cdot v$ für $x \in S^{n-1}$ und $v \in \mathbb{R}^n$. Es ist $\alpha(f) = 1$.

3. Vektorraumbündel. Der Satz im 1. Abschnitt wird am Ende des 4. Abschnitts zu einer Aussage über charakteristische Klassen von Vektorraumbündeln umformuliert. Unter einem n-dimensionalen Vektorraumbündel über dem topologischen Raum X versteht man einen weiteren topologischen Raum E zusammen mit einer stetigen Abbildung (Bündelprojektion) $p: E \to X$, so daß jede Faser $E_x = p^{-1}(x)$ $(x \in X)$ ein n-dimensionaler reeller Vektorraum ist. Ferner wird verlangt, daß E im folgenden Sinne lokal trivial ist: Zu jedem Punkt in X gibt es eine Umgebung U und n Schnitte $v_1, \ldots, v_n: U \to E$ (das heißt, stetige Abbildungen mit $p \cdot v_i = \mathrm{id}$), so daß für jedes $x \in U$ die n Vektoren $v_1(x), \ldots, v_n(x)$ eine Basis von E_x bilden. Vektorraumbündel gehören zu den grundlegenden Begriffen der Differentialtopologie und -geometrie. Das vielleicht wichtigste Beispiel ist das Tangentialbündel $E = TM$ einer n-dimensionalen Mannigfaltigkeit M, welches durch Zusammenfassen aller Tangentialräume $TM = \bigcup_{x \in M} T_x M$ gebildet wird (siehe 10.1.5).

Im vorliegenden Fall interessieren jedoch andere Bündel, nämlich die m-dimensionalen Bündel E_f über S^n, die mittels einer Abbildung $f: S^{n-1} \to GL(m)$ zusammengeklebt werden: Man zerlegt S^n in die obere und untere Halbkugel

$$S^n = H^+ \cup H^-, \qquad H^+ \cap H^- = S^{n-1} \qquad (\text{Äquator}),$$

bildet die trivialen Bündel $H^+ \times \mathbb{R}^m$ und $H^- \times \mathbb{R}^m$ und verklebt sie längs S^{n-1}, indem man jeden Punkt $(x, v) \in S^{n-1} \times \mathbb{R}^m$ mit $(x, f(x) \cdot v) \in S^{n-1} \times \mathbb{R}^m$ identifiziert. Der verklebte Identifikationsraum wird mit E_f bezeichnet. Die Bündelprojektion $p: E_f \to S^n$ entsteht aus den Projektionen $H^+ \times \mathbb{R}^m \to H^+$ und $H^- \times \mathbb{R}^m \to H^-$ auf den ersten Faktor. Jedes Bündel über S^n kann man auf diese Weise erhalten. Wenn $f: S^{n-1} \to GL(n)$ wie im 2. Abschnitt von einer Divisionsalgebra herkommt, nennt man das geklebte Bündel E_f das *Hopfsche Bündel* der Algebra. (Für $n = 1$ handelt es sich um das bekannte MÖBIUSband.)

4. Charakteristische Kohomologieklassen nach WHITNEY. In die Definition des n-dimensionalen Vektorraumbündels E über X wurde aufgenommen, daß man

lokal stets n linear unabhängige Schnitte hat. H. WHITNEY, der die Theorie der Bündel begründete, befaßte sich mit dem Problem, welche Hindernisse dem Versuch entgegenstehen, bei einem n-dimensionalen Bündel k globale, das heißt, auf ganz X definierte, und überall linear unabhängige Schnitte zu finden. Es gelang ihm, solche Hindernisse kohomologisch zu beschreiben: Es bezeichne $H^i(X)$, wie bereits früher eingeführt (10.1.2), die i-te Kohomologie von X mit Koeffizienten im zweielementigen Körper F_2. Dann definiert WHITNEY zum Bündel E sogenannte charakteristische Kohomologieklassen

$$w_i(E) \in H^i(X), \qquad i = 1, \ldots, n.$$

Wie diese Definition aussieht, wird hier nicht ausgeführt. Als einschlägiges Lehrbuch sei [11] empfohlen.

Es sei jedoch erwähnt, daß $w_i(E) = 0$ für $i > n - k$, wenn k überall linearunabhängige globale Schnitte existieren. Wenn $w_{n-k+1}(E) \neq 0$, dann existiert ein solcher k-Schnitt nicht.

Wir haben in 10.1.5 die Stiefelschen Klassen $s_0, s_1, \ldots, s_{n-1}$ des Tangentialbündels TM einer n-dimensionalen Mannigfaltigkeit M betrachtet. Es ist $s_{k-1} \in H_{k-1}(M)$. Unter der Poincaréschen Dualität (10.1.2) geht s_{k-1} in die Whitneysche Klasse w_{n-k+1} des Tangentialbündels über (s_{k-1} bzw. w_{n-k+1} sind das erste Hindernis gegen einen k-Schnitt).

Zurück zu den n-dimensionalen Bündeln E_f über S^n, die durch Verkleben mittels $f\colon S^{n-1} \to GL(n)$ entstanden. Da $H^i(S^n) = 0$ für $i \neq 0$, $i \neq n$ ist, interessiert nur die Klasse $w_n(E_f) \in H^n(S^n) \cong F_2$. Es ist $w_n(E_f) = \alpha(f)$, die im 1. Abschnitt eingeführte mod 2-Invariante. Der Satz am Ende des ersten Abschnitts lautet also umformuliert:

Satz. *Wenn es über S^n ein n-dimensionales Vektorraumbündel E mit $w_n(E) \neq 0$ gibt, dann ist $n = 1, 2, 4$, oder 8.*

Daß es in diesen vier Dimensionen tatsächlich solche Bündel E gibt, belegen die Hopfschen Bündel zu den Divisionsalgebren der reellen und komplexen Zahlen, der Quaternionen und der Oktaven.

Um den Satz zu beweisen, benutzt man den Überblick über alle möglichen Vektorraumbündel über den Sphären S^n, den R. BOTT in seinem Periodizitätssatz ausspricht. Wir formulieren ihn mit Hilfe des Ringes $KO(X)$, der im nächsten Abschnitt eingeführt wird.

5. Der Ring der Vektorraumbündel. Aus den zwei Vektorräumen E und F kann man bekanntlich neue Vektorräume herstellen, indem man die direkte Summe $E \oplus F$ oder das Tensorprodukt $E \otimes F$ bildet. Das geht auch für Vektorraumbündel E und F über X: Man kann das direkte Summenbündel $E \oplus F$ über X bilden, so daß für die Fasern über $x \in X$ gilt $(E \oplus F)_x = E_x \oplus F_x$. Entsprechend kann man das Tensorprodukt $E \otimes F$ mit $(E \otimes F)_x = E_x \otimes F_x$ einführen.

Man betrachtet nun die Menge $N(X)$ der Isomorphieklassen von Vektorraumbündeln über X. In dieser Menge sind die Verknüpfungen $\oplus$ und $\otimes$ erklärt. Sie erfüllen wie die übliche Addition und Multiplikation in $\mathbb{N}$ die Assoziativ-, Kommutativ- und Distributivgesetze, und man hat neutrale Elemente für $\oplus$ und

$\otimes$. Es liegt also nahe, wie von $\mathbb{N}$ zu $\mathbb{Z}$ auch von $N(X)$ zu einem Ring überzugehen: Man bildet $N(X) \times N(X)$ und definiert folgende Äquivalenzrelation

$$(a, b) \sim (c, d).$$

Es gibt ein f mit

$$a \oplus d \oplus f = c \oplus b \oplus f.$$

(Man muß f benutzen, weil in $N(X)$ anders als in $\mathbb{N}$ die Kürzungsregel nicht gilt.) Die Menge der Äquivalenzklassen wird mit $KO(X)$ bezeichnet. Man überträgt, wie man es beim Übergang von $\mathbb{N}$ auf $\mathbb{Z}$ tut, die Verknüpfungen $\oplus$ und $\otimes$ auf $N(X)$ zu $+$ und $\cdot$ auf $KO(X)$ und macht dadurch $KO(X)$ zu einem kommutativen Ring mit Einselement. Die natürliche Abbildung $N(X) \to KO(X)$ ist nicht injektiv, da die Kürzungsregel in $N(X)$ nicht gilt.

Die Zuordnung $X \to KO(X)$ verhält sich wie eine Kohomologietheorie: Wenn $f: Y \to X$ eine stetige Abbildung ist, hebt man jedes n-dimensionale Vektorraumbündel E über X (Bündelprojektion $p: E \to X$) zu folgendem ebenfalls n-dimensionalen Vektorraumbündel f^*E über Y an:

$$f^*E = \{(y, v) \in Y \times E: f(y) = p(v)\}.$$

Dadurch wird eine Abbildung $f^!: N(X) \to N(Y)$ induziert, die mit $\oplus$ und $\otimes$ verträglich ist, und folglich erhält man einen Ringhomomorphismus

$$f^!: KO(X) \to KO(Y)$$

mit $(f \cdot g)^! = g^! \cdot f^!$ für $Z \xrightarrow{g} Y \xrightarrow{f} X$.

6. Die Bottsche Periodizität. Wenn man jedem Vektorraumbündel seine Faserdimension zuordnet, erhält man einen Epimorphismus

$$\varepsilon: KO(X) \to \mathbb{Z}.$$

(Hierzu wird X als zusammenhängend vorausgesetzt.) Der Kern von ε wird mit $\widetilde{KO}(X)$ bezeichnet.

Bottscher Periodizitätssatz. *Es ist*

$$\widetilde{KO}(S^1) = \widetilde{KO}(S^2) = \mathbb{Z}/2, \qquad \widetilde{KO}(S^3) = 0, \qquad \widetilde{KO}(S^4) = \mathbb{Z},$$
$$\widetilde{KO}(S^5) = \widetilde{KO}(S^6) = \widetilde{KO}(S^7) = 0, \qquad \widetilde{KO}(S^8) = \mathbb{Z}.$$

In den Dimensionen $n = 1, 2, 4$ und 8 werden die erzeugenden Elemente durch die Hopfschen Bündel zu den Divisionsalgebren der reellen und komplexen Zahlen, der Quaternionen und der Oktaven (vermindert um das n-dim. triviale Bündel) repräsentiert.

Es ist $\widetilde{KO}(S^n) \cong \widetilde{KO}(S^{n+8})$ für alle n.

(Alle Isomorphismen sind nur additiv und nicht als Ringisomorphismen zu verstehen.)

Über den Beweis dieses Satzes, den R. Bott 1957 in einer anderen Formulierung (vgl. Abs. 9) veröffentlichte, kann hier nichts gesagt werden. Was wir aber benötigen, ist eine Beschreibung des Isomorphismus $\widetilde{KO}(S^n) \cong \widetilde{KO}(S^{n+8})$:

Zu zwei Sphären S^n und S^m bildet man das kartesische Produkt $S^n \times S^m$. Außerdem wählt man in S^n und S^m je einen Basispunkt x_0 bzw. y_0. Dann liegt in $S^n \times S^m$ das „Achsenkreuz" $S^n \vee S^m = \{x_0\} \times S^m \cup S^n \times \{y_0\}$. Man zieht es zu einem Punkt zusammen. Dann wird $S^n \times S^m$ zu S^{n+m}, und man erhält die Abbildungen $S^n \vee S^m \xrightarrow{i} S^n \times S^m \xrightarrow{p} S^{n+m}$. Das liefert, wie hier nicht bewiesen wird, eine exakte Sequenz

$$0 \to \widetilde{KO}(S^{n+m}) \xrightarrow{p^!} \widetilde{KO}(S^n \times S^m) \xrightarrow{i^!} \widetilde{KO}(S^n \vee S^m) \to 0,$$

das heißt, $p^!$ ist injektiv, $i^!$ ist surjektiv und $\operatorname{Kern} i^! = \operatorname{Bild} p^!$. Es seien π_1, π_2 die Projektionen von $S^n \times S^m$ auf die beiden Faktoren. Zu $a \in \widetilde{KO}(S^n)$ und $b \in \widetilde{KO}(S^m)$ bildet man

$$a \cdot b = \pi_1^! a \cdot \pi_2^! b \in KO(S^n \times S^m).$$

Da $i^!(a \cdot b) = 0$ ist, ist $a \cdot b$ das $p^!$-Bild genau eines Elementes in $\widetilde{KO}(S^{n+m})$, welches auch mit $a \cdot b$ bezeichnet wird.

Der Bottsche Isomorphismus $\widetilde{KO}(S^n) \cong \widetilde{KO}(S^{n+8})$ *wird durch* $a \mapsto a \cdot (\rho_8 - 8)$ *beschrieben, wobei* ρ_8 *das Hopfsche Bündel zu den Oktaven und* 8 *das* 8*-dimensionale triviale Bündel über* S^8 *bedeuten.*

Der für den Beweis des Satzes im 4. Abschnitt angekündigte Überblick über Vektorraumbündel über den Sphären ist nun erreicht. Zum Beweis fehlen nur noch Methoden zur Berechnung der charakteristischen Klassen.

Bemerkung. Man findet unsere Formulierung des Bottschen Periodizitätssatzes im wesentlichen bei

Bott, R.: Lectures on $K(X)$. W. A. Benjamin, New York 1969 auf Seite 73, allerdings ohne Beweise.

Einen ausführlichen Beweis im Rahmen der K-Theorie enthält das Lehrbuch

Karoubi, M.: K-Theory. An Introduction. Springer, Berlin-Heidelberg-New York 1978. Allerdings wird es dem Leser einige Mühe machen, aus Karoubis Formulierungen die hier benutzte herauszufinden.

Viel einfacher ist die K-Theorie für komplexe Vektorraumbündel zugänglich, siehe

Atiyah, M.: K-Theory. W. A. Benjamin New York 1967. Im Anhang dieses Buches (On K-Theory and Reality) wird knapp dargelegt, wie man durch geschickte Modifikationen KO erreichen kann.

7. Charakteristische Klassen von direkten Summen und Tensorprodukten. Nach Whitney wurden im 4. Abschnitt einem n-dimensionalen Bündel E über X die Klassen $w_i(E) \in H^i(X)$ für $i = 1, \ldots, n$ zugeordnet. Es ist zweckmäßig, dies durch $w_0(E) = \text{Einselement} \in H^0(X)$ und $w_i(E) = 0$ für $i > n$ zu ergänzen und alle Klassen zur totalen Stiefel-Whitneyschen Klasse

$$w(E) = 1 + w_1(E) + \cdots + w_n(E) = \sum_{i=0}^{\infty} w_i(E) \in H^*(X)$$

zusammenzufassen. Für die direkte Summe gilt dann nach WHITNEY (1941):

$$w(E \oplus F) = w(E) \cdot w(F),$$

ausgeschrieben

$$w_i(E \oplus F) = \sum_{r+s=i} w_r(E) \cdot w_s(F).$$

Wir beschränken uns jetzt auf Räume X mit $H^i(X) = 0$ für fast alle i. Sämtliche Elemente $a \in H^*(X)$, deren 0-dimensionale Komponente $= 1$ ist, bilden dann eine multiplikative Gruppe $G(X)$. Die Whitneysche Summenformel bedeutet:

Die totale Stiefel-Whitneysche Klasse bestimmt einen Homomorphismus von der additiven Gruppe $KO(X)$ in die multiplikative Gruppe $G(X)$,

$$w\colon KO(X) \to G(X).$$

Die Stiefel-Whitneyschen Klassen sind außerdem mit dem Anheben von Bündeln verträglich, $w_i(f^*E) = f^*w_i(E)$. Man hat daher ein kommutatives Diagramm (für eine stetige Abbildung $f\colon Y \to X$)

$$\begin{array}{ccc} KO(X) & \xrightarrow{f^!} & KO(Y) \\ {\scriptstyle w}\downarrow & & \downarrow{\scriptstyle w} \\ G(X) & \xrightarrow{f^*} & G(Y). \end{array}$$

Die Vektorraumbündel verhalten sich bei stetigen Abbildungen kontravariant. Wie gut, daß WHITNEY seine Klassen als Kohomologieklassen und damit kontravariant einführte! Für beliebige Vektorraumbündel hatte er jedoch gar keine andere Möglichkeit. Definieren in der Mathematik ist eben kein willkürliches Spiel.

Nun zu den charakteristischen Klassen von Tensorprodukten. Das ist bei eindimensionalen Bündeln E, F einfach:

$$w_1(E \otimes F) = w_1(E) + w_1(F).$$

Wenn $E = E_1 \oplus \cdots \oplus E_m$ und $F = F_1 \oplus \cdots \oplus F_n$ direkte Summen von eindimensionalen Bündeln sind, ist $E \otimes F = \bigoplus (E_i \otimes F_j)$, wobei über alle $1 \leqslant i \leqslant m$ und $1 \leqslant j \leqslant n$ summiert wird. Nach der Whitneyschen Summenformel ist dann $w(E \otimes F) = \prod_{i,j} (1 + w_1(E_i) + w_1(F_j))$.

Für beliebige Vektorraumbündel E und F der Dimensionen m und n bleibt dieses Ergebnis in folgendem Sinne richtig:

Man betrachtet das Polynom $\prod_{i,j}(1 + x_i + y_i)$ mit Koeffizienten in F_2. Da es in den x_i und in den y_j symmetrisch ist, läßt es sich als Polynom in den elementarsymmetrischen Funktionen $\sigma_1, \ldots, \sigma_m$ der x_i und $\tau_1, \ldots, \tau_n$ der y_j schreiben:

$$\prod_{i,j}(1 + x_i + y_i) = P(\sigma_1, \ldots, \sigma_m, \tau_1, \ldots, \tau_n).$$

Es ist

$$w(E \otimes F) = P(w_1(E), \ldots, w_m(E), w_1(F), \ldots, w_n(F)).$$

8. Schluß des Beweises. Wir erinnern den Leser daran, daß das Hauptergebnis des 2. Abschnitts auf dem Satz des 1. Abschnitts beruhte und daß dieser Satz im

4. Abschnitt umformuliert wurde. Mit dem nun folgenden Beweis des Satzes im 4. Abschnitt ist das Ziel also erreicht.

Wenn man ein n-dimensionales Vektorraumbündel E über X durch $E - n \in \widetilde{KO}(X)$ ersetzt, ändert sich wegen der Whitneyschen Summenformel die Stiefel-Whitneysche Klasse nicht. Zum Beweis des Satzes genügt es also, $w(c) = 1$ (das heißt, $w_n(c) = 0$) für alle $c \in \widetilde{KO}(S^n)$ und $n \neq 1, 2, 4$ oder 8 zu zeigen. Wegen der im Bottschen Satz angegebenen Periode ist dies für $n = 3, 5, 6$ und 7 richtig. Wenn nun $n > 9$ ist, schreibt man $n = m + 8$ und kann wegen der Periodizität $c = a \cdot (\rho_8 - 8)$ schreiben, wobei $a \in \widetilde{KO}(S^m)$ ist. Aufgrund der Definition von $\widetilde{KO}(S^m)$ läßt sich a als $a = E - F$ darstellen, wobei E und F gleichdimensionale Bündel über S^m sind, also gilt in $KO(S^m \times S^8)$ die Gleichung

$$c = (E - F) \cdot (\rho_8 - 8) = E \cdot \rho_8 - F \cdot \rho_8 - 8 \cdot E + 8 \cdot F,$$

und wegen der Whitneyschen Summenformel somit

$$w(c) = w(E \cdot \rho_8) \cdot w(F \cdot \rho_8)^{-1} \cdot w(8 \cdot E)^{-1} \cdot w(8 \cdot F).$$

Es wird behauptet, daß jeder der vier Faktoren $= 1$ ist: Betrachten wir etwa $w(E \cdot \rho_8)$ wobei $E \cdot \rho_8 \in KO(S^m \times S^8)$ das durch $\pi_1^* E \otimes \pi_2^* \rho_8$ bestimmte Element ist. Man benutzt nun das folgende

Lemma. *Es seien ξ und η geradedimensionale Vektorraumbündel über X mit $w(\xi) = 1 + w_r(\xi)$ und $w(\eta) = 1 + w_s(\eta)$, s gerade, und mit $w_r(\xi)^2 = w_s(\eta)^2 = 0$. Dann ist $w(\xi \otimes \eta) = 1$.*

Das Lemma folgt aus der Beschreibung der Stiefel-Whitneyschen Klassen eines Tensorproduktes mittels symmetrischer Polynome mit Koeffizienten in F_2, siehe das Ende des 7. Abschnitts.

Man wendet dieses Lemma auf $X = S^m \times S^8$, $\xi = \pi_1^* E$ und $\eta = \pi_2^* \rho_8$ an. Man kann ohne weiteres annehmen, daß E und F geradedimensional sind. Sonst addiert man bei E und F das triviale 1-dimensionale Bündel, ohne daß sich $a = E - F$ ändert. Man erhält $w(\pi_1^* E \otimes \pi_2^* \rho_8) = 1$.

9. Historische Anmerkungen. Das erste Lehrbuch über Faserbündel verdankt man N. STEENROD [13], der in seinem Vorwort schreibt:

„The recognition of the domain of mathematics called fibre bundles took place in the period 1935–1940. The first general definitions were given by H. Whitney. His work and that of H. Hopf and E. Stiefel demonstrated the importance of the subject for the applications of topology to differential geometry. Since then, some seventy odd papers dealing with bundles have appeared. The subject has attracted general interest, for it contains some of the finest applications of topology to other fields, and gives promise of many more. It also marks a return of algebraic topology to its origin; and after many years of introspective development, a revitalization of the subject from its roots in the study of classical manifolds."

H. HOPF berichtet [8], daß er auf der Moskauer Konferenz 1935 über STIEFELS Theorie vorgetragen habe, und schreibt dann anschließend „Nachdem ich in

Moskau alles dies vorgetragen hatte, machte H. Whitney in der Diskussion darauf aufmerksam, daß ein großer Teil davon in seiner soeben erschienenen Note „Sphere Spaces" (Proc. Nat. Acad. Sci. 21 (1935)) enthalten sei; er hatte recht, aber Stiefel und ich kannten diese Note nicht; jedenfalls ist es ganz berechtigt, daß man die „charakteristischen" Klassen heute meistens die „Stiefel-Whitney-Klassen" nennt. Ich finde, daß bei Whitney alles etwas allgemeiner ist als bei Stiefel, während Stiefels Interesse mehr auf spezielle Probleme gerichtet ist, die bei Whitney nicht vorkommen."

In der Tat, WHITNEYS Theorie ist allgemeiner. Er führt die charakteristischen Klassen für ein beliebiges Vektorraumbündel über einem Basisraum X ein und nicht nur für das Tangentialbündel einer Mannigfaltigkeit. Er mußte Kohomologie verwenden. Nur für Mannigfaltigkeiten kommt man mit der Homologie zurecht.

Es hat lange gedauert, bis man mit den Stiefel-Whitneyschen Klassen wirklich rechnen konnte. Wir verzichten auf Literaturangaben zur historischen Entwicklung. Es sei aber auf das Lehrbuch von MILNOR und STASHEFF [11] verwiesen.

HOPF [8] würde den § 1 dieses Kapitels zur Vorgeschichte der Topologie rechnen, den § 2 mit Kohomologie, Vektorraumbündeln, detaillierter Theorie der charakteristischen Klassen, Bottscher Periodizität, K-Theorie aber wohl zur Neuzeit. BOTT hat seinen Satz ursprünglich in der Sprache der Homotopiegruppen formuliert und differentialgeometrisch bewiesen (erste Ankündigung in Proc. Nat. Acad. Sci. USA 43, 933–935 (1957), ausführlicher dargestellt in: The Stable Homotopy of the Classical Groups, Ann. of Math. 70, 313–337 (1959). Siehe auch J. MILNOR: Morse Theory, Princeton Univ. Press 1963.) Wesentliches Hilfsmittel war die Theorie von M. MORSE.

A. GROTHENDIECK hat 1958 im Rahmen der algebraischen Geometrie mit Hilfe algebraischer Vektorraumbündel einen Ring $K(X)$ für eine algebraische Varietät X eingeführt und für seine verallgemeinerte Fassung des Satzes von RIEMANN-ROCH-HIRZEBRUCH benutzt. Sein Ring der Vektorraumbündel verhielt sich kontravariant wie der Kohomologiering $H^*(X)$ eines topologischen Raumes. (GROTHENDIECK suchte sich einen Buchstaben in der Nähe von H und kam auf K.) Motiviert durch GROTHENDIECK wurde für topologische Räume X der Ring $K(X)$ mit Hilfe topologischer Vektorraumbündel eingeführt, deren Fasern *komplexe* Vektorräume sind ([4], [3]). Nimmt man reelle Vektorräume, kommt man zu $KO(X)$, das O erinnert dabei an die Rolle der orthogonalen Gruppe bei reellen Vektorräumen. Um K und KO zu einer kompletten Kohomologietheorie zu machen, benötigt man die Bottsche Periodizität, die für K einfacher ist als für KO. Es ist nämlich $\tilde{K}(S^n) = \mathbb{Z}$ für n gerade und $\tilde{K}(S^n) = 0$ für n ungerade.

§ 3. Ergänzungen

Hauptziel dieses Kapitels war es natürlich anzudeuten, wie der (1, 2, 4, 8)-Satz für Divisionsalgebren mit topologischen Methoden bewiesen werden kann. Gleichzeitig haben wir einen kleinen Spaziergang von der „Vorgeschichte" (30er–40er

Jahre) bis zum Anfang der 60er Jahre gemacht. Dieser Bericht wäre jedoch sehr unvollständig, wenn wir die Hopfsche Invariante nicht erwähnen würden (HOPF [6]). Wir gehen auch kurz auf Vektorfelder auf Sphären ein (vgl. 10.2.5).

1. Definition der Hopfschen Invarianten (vgl. [12]). Es sei $F: S^{2n-1} \to S^n$ $(n \geqslant 2)$ eine stetige Abbildung. Nach Deformation kann man annehmen, daß F beliebig häufig differenzierbar ist. Das Urbild $F^{-1}(x)$ eines Punktes $x \in S^n$ ist im allgemeinen eine $(n-1)$-dimensionale Untermannigfaltigkeit von S^{2n-1}, in die man eine n-dimensionale berandete Mannigfaltigkeit M einspannen kann, die dann durch F mit einem Abbildungsgrad γ_F auf S^n abgebildet wird (vgl. 10.2.1). Die *ganze* Zahl γ_F heißt Hopfsche Invariante von F. Nach dieser Definition ist γ_F auch die Schnittzahl von $F^{-1}(y)$ $(y \neq x$, y in allgemeiner Lage) mit M, oder auch die Verschlingungszahl von $F^{-1}(x)$ und $F^{-1}(y)$. Eine Orientierungsbetrachtung zeigt, daß γ_F verschwindet, wenn n ungerade ist. Die Zahl γ_F hängt nur von der Homotopieklasse von F ab und ist ein Homomorphismus der Homotopiegruppe $\pi_{2n-1}(S^n)$ in die ganzen Zahlen.

2. Die Hopfsche Konstruktion (vgl. [12]). HOPF hat folgendes Problem gestellt. *Für festes gerades n bestimme man die additive Untergruppe der ganzen Zahlen, die als Invariante γ für eine stetige Abbildung $F: S^{2n-1} \to S^n$ auftreten?*

Mit Hilfe der Hopfschen Konstruktion wird aus einer vorgegebenen Abbildung $g: S^{n-1} \times S^{n-1} \to S^{n-1}$ eine Abbildung $F: S^{2n-1} \to S^n$ konstruiert:

Die Sphäre S^{2n-1} läßt sich topologisch als Rand von $E^n \times E^n$ beschrieben, wo E^n der n-dimensionale Ball ist. Deshalb ist

$$S^{2n-1} = \partial(E^n \times E^n) = S^{n-1} \times E^n \cup E^n \times S^{n-1}.$$

Diese Sphäre ist also in die beiden Produkte $S^{n-1} \times E^n$ und $E^n \times S^{n-1}$ zerlegt mit $S^{n-1} \times S^{n-1}$ als gemeinsamen Rand.

Die Sphäre S^n wird durch S^{n-1} in zwei Halbkugeln H^+ und H^- zerlegt. Man erweitert g in naheliegender Weise zu einer Abbildung von $E^n \times S^{n-1}$ in H^+ und $S^{n-1} \times E^n$ in H^-. Die so gewonnene Abbildung $F: S^{2n-1} \to S^n$ ist die Hopfsche Konstruktion zu g. Nach HOPF ist $\gamma_F = c_1 \cdot c_2$, falls g vom Doppelgrad (c_1, c_2) ist (hier ist c_1 der Grad, mit dem $S^{n-1} \times$ Punkt auf S^{n-1} abgebildet wird, entsprechend für c_2).

Wenn $g: S^{n-1} \times S^{n-1} \to S^{n-1}$ ungerade ist (10.1.1), *dann ist γ_F ungerade.*

Jede Funktion $f: S^{n-1} \to GL(n)$ (siehe 10.2.1) definiert eine Abbildung $g: S^{n-1} \times S^{n-1} \to S^{n-1}$ durch

$$g(x, v) = \frac{f(x)v}{\|f(x)v\|}$$

vom Doppelgrad $(c, 1)$ und mittels Hopfscher Konstruktion eine Abbildung $F: S^{2n-1} \to S^n$ mit $\gamma_F = c$. Hier ist c gerade oder ungerade, je nachdem, ob $w_n(E_f) = 0$ oder $w_n(E_f) \neq 0$ (siehe 10.2.4). Nimmt man (für n gerade) die Klebefunktion f des Tangentialbündels von S^n, dann ist $\gamma_F = 2$. Nimmt man für f

die Funktionen $S^{n-1} \to GL(n)$, die von den Divisionsalgebren herrühren (10.2.1), dann ist $\gamma_F = 1$.

Zum Hopfschen Problem hat sich also ergeben:

Alle ganzen Zahlen treten für $n = 2, 4, 8$ als Hopfsche Invariante auf. Für die anderen geraden n treten jedenfalls alle geraden ganzen Zahlen als Hopfsche Invariante auf.

3. Der Satz von ADAMS über die HOPFsche Invariante. F. ADAMS [1] zeigte, daß *Abbildungen f: $S^{2n-1} \to S^n$ mit ungeradem γ_f nur für $n = 2, 4, 8$ existieren.* Er benutzte sogenannte sekundäre Kohomologieoperationen. Inzwischen gibt es einen K-theoretischen Beweis (ADAMS und ATIYAH 1966, vgl. [3]), der sehr einfach ist, sobald man die K-Theorie voll entwickelt hat.

4. Zusammenfassung. *Es sei $n \geqslant 2$.* Die Ergebnisse von § 2 und § 3 haben gezeigt: Die folgenden „mathematischen Objekte" existieren nur für $n = 2, 4, 8$.

Divisionsalgebren der Dimension n,
Ungerade Abbildungen $S^{n-1} \times S^{n-1} \to S^{n-1}$,
Parallelisierung von $\mathbb{P}^{n-1}$,
Parallelisierung von S^{n-1},
Vektorraumbündel E über S^n mit Stiefel-Whitneyscher Klasse $w_n(E) \neq 0$,
Abbildungen f: $S^{2n-1} \to S^n$ mit ungerader Hopfscher Invariante.

Aus einer Divisionsalgebra der Dimension n kann man, wie wir gesehen haben, die anderen angegebenen mathematischen Objekte recht elementar konstruieren. Aus jedem angegebenen Objekt kann man ziemlich einfach (mit der HOPFschen Konstruktion) eine Abbildung mit ungerader HOPFscher Invariante erhalten. In diesem Sinne ist der Satz von ADAMS über die Nichtexistenz von Abbildungen ungerader Hopfscher Invariante das allgemeinste Resultat, das die Nichtexistenz der anderen Objekte impliziert.

5. Der Satz von ADAMS über Vektorfelder auf Sphären. Zunächst sei an den Satz von HURWITZ-RADON erinnert (9.2.3). Wenn die Aussage i) erfüllt ist, dann kann man leicht $p - 1$ linear-unabhängige tangentielle Vektorfelder auf $\mathbb{P}^{n-1}$ konstruieren. Dies ist völlig analog zu dem Satz in 10.1.5. Also ist $\mathrm{Span}(S^{n-1}) \geqslant \mathrm{Span}(\mathbb{P}^{n-1}) \geqslant 8\alpha + 2^\beta - 1$, falls $n = u \cdot 2^{4\alpha+\beta}$, $1 \leqslant u$ ungerade, $0 \leqslant \alpha$, $0 \leqslant \beta \leqslant 3$. ADAMS [2] zeigt, daß $\mathrm{Span}(S^{n-1}) = \mathrm{Span}(\mathbb{P}^{n-1}) = 8\alpha + 2^\beta - 1$. Hierbei gehen viele weitentwickelte Methoden der K-Theorie ein. ADAMS' Satz ist eine phantastische Verallgemeinerung des Satzes, daß nur die Sphären S^1, S^3, S^7 parallelisierbar sind (z. B. ist $\mathrm{Span}(S^{15}) = 8$), und damit auch eine Verallgemeinerung unseres $(1, 2, 4, 8)$-Satzes.

Der Satz von ADAMS hatte viele Vorläufer. Die Ungleichung

$$\mathrm{Span}(\mathbb{P}^{n-1}) \leqslant \max\left\{k > 0 \,\middle|\, \binom{n}{k} \text{ gerade}\right\}$$

ist ein solcher Vorläufer (siehe 10.1.5). Zusammen mit der Formel $\mathrm{Span}(\mathbb{P}^{n-1}) \geqslant 8\alpha + 2^\beta - 1$ liefert dies zum Beispiel unmittelbar $\mathrm{Span}(\mathbb{P}^{4m+1}) = 1$ und $\mathrm{Span}(\mathbb{P}^{8m+3}) = 3$.

B. ECKMANN und G. B. WHITEHEAD bewiesen bereits in den 40er Jahren, daß sogar $\mathrm{Span}(S^{4m+1}) = 1$ und $\mathrm{Span}(S^{8m+3}) = 3$.

Literatur

[1] ADAMS, J. F.: On the non-existence of elements of Hopf invariant one. Ann. of Math. 72, 20–104 (1960)

[2] ADAMS, J. F.: Vector fields on spheres. Ann. of Math. 75, 603–632 (1962)

[3] ATIYAH, M. F.: *K*-Theory. W. A. Benjamin, Inc., New York, Amsterdam 1967

[4] ATIYAH, M. F., HIRZEBRUCH, F.: Vector bundles and homogeneous spaces. Proc. of Symposia in Pure Mathematics, Vol. 3, p. 7–38. Am. Math. Soc. 1961

[5] ATIYAH, M. F., HIRZEBRUCH, F.: Bott periodicity and the parallelisability of the spheres. Proc. Cambridge Phil. Soc. 57, 223–226 (1961)

[6] HOPF, H.: Über die Abbildungen von Sphären auf Sphären niedrigerer Dimension. Fundamenta Math. 25, 427–440 (1935)

[7] HOPF, H.: Ein topologischer Beitrag zur reellen Algebra. Comm. Math. Helvetici 13, 219–239 (1940/41)

[8] HOPF, H.: Einige persönliche Erinnerungen aus der Vorgeschichte der heutigen Topologie. Colloque de Topologie, Centre Belge de Recherches Mathématiques 1966, p. 9–20

[9] MILNOR, J.: Some consequences of a theorem of Bott. Ann. of Math. 68, 444–449 (1958)

[10] MILNOR, J.: Topology from the differentiable viewpoint. The University Press of Virginia 1965

[11] MILNOR, J., STASHEFF, J.: Characteristic classes. Ann. of Math. Studies 76, Princeton University Press 1974

[12] SAMELSON, H.: Heinz Hopf zum Gedenken. II. Zum wissenschaftlichen Werk von Heinz Hopf. Jber. Deutsch. Math.-Verein 78, 126–146 (1976)

[13] STEENROD, N.: The topology of fibre bundles. Princeton University Press 1951

[14] STIEFEL, E.: Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten. Comm. Math. Helvetici 8, 305–353 (1936)

[15] STIEFEL, E.: Über Richtungsfelder in den projektiven Räumen und einen Satz aus der reellen Algebra. Comm. Math. Helvetici 13, 201–218 (1940/41)

Teil C

Ausblicke

Kapitel 11. Non-Standard Analysis

A. Prestel

§ 1. Einführung

In diesem Kapitel wollen wir den Körper $\mathbb{R}$ der reellen Zahlen zu einem Körper $^*\mathbb{R}$ erweitern, in dem es unendlich kleine und unendlich große „Zahlen" gibt. Insbesondere lassen sich in $^*\mathbb{R}$ die Leibnizschen Differentiale dx, dy exakt definieren und ein Zusammenhang des Differentialquotienten dy/dx mit der Ableitung $f'(x)$ einer Funktion $y = f(x)$ an der Stelle x herstellen.

Das Rechnen mit unendlich kleinen Größen wie dx war in früheren Jahrhunderten in der Mathematik und Physik selbstverständlich, jedoch nicht immer unumstritten. (Einen Einblick in den Gebrauch infinitesimaler Größen und die Kritik an diesem Gebrauch findet man etwa in dem Buch von Edwards [1].) Erst durch die „Epsilontik", dem Aufbau der Analysis auf dem Limesbegriff durch Weierstraß, wurden diese Größen aus der Mathematik verbannt. Genauer gesagt, wurden sie aus den exakten Beweisen verbannt; in der mathematischen Heuristik und in der Physik behaupten sie nach wie vor ihren Platz.

In der „Epsilontik" wird etwa der Differentialquotient dy/dx einer Funktion $y = f(x)$ durch den Limes des Differenzenquotienten

$$\frac{f(x+h) - f(x)}{h}$$

für $h \to 0$ erklärt, falls dieser existiert. Sein Wert wird dann mit $f'(x)$ bezeichnet; die Bezeichnung durch dy/dx ist zwar ebenfalls üblich, es wird jedoch immer betont, daß die Größen dy und dx für sich genommen keinen Sinn haben. Dies ist natürlich richtig, wenn man nur die reellen Zahlen im Auge hat: es gibt keine reelle Zahl, die etwa zwischen 0 und allen positiven reellen Zahlen liegt.

Will man also mit unendlich kleinen Größen wie etwa dx rechnen, so muß man diese offenbar aus einem größeren Bereich als $\mathbb{R}$ nehmen. Mathematiker früherer Jahrhunderte rechneten mit solchen Größen ganz selbstverständlich so wie mit reellen Zahlen, waren sich aber durchaus über den Unterschied im Klaren. Die Tatsache, daß man sich damals nicht um die Konstruktion eines entsprechenden Bereiches bemühte, sollte nicht weiter verwundern; Konsistenzfragen dieser Art waren damals noch nicht üblich. Es genügte den meisten Mathematikern, daß solche Größen in der mathematischen Intuition existierten und ihre Benutzung zu richtigen Ergebnissen führte. Diejenigen unter den Mathematikern und Philosophen, die den Umgang damit ablehnten, taten dies wohl hauptsächlich deshalb, weil sie einen Widerspruch darin erblickten, diese Größen wie reelle Zahlen zu behandeln, ihnen aber keine „Endlichkeit" zuzuerkennen.

Im Jahre 1960 konstruierte Abraham Robinson einen Erweiterungskörper $^*\mathbb{R}$ des Körpers $\mathbb{R}$ der reellen Zahlen, in dem es zwar unendlich kleine Größen gibt, dessen Eigenschaften sich aber von denen von $\mathbb{R}$ innerhalb eines gewissen, jedoch sehr weiten Rahmens nicht unterscheiden. Robinson benutzte dazu eine Konstruktion, die 1933 zum ersten Mal von Th. Skolem angewandt worden war, um damit eine Erweiterung der natürlichen Zahlen zu erhalten, die immer noch ein Modell der Peanoschen Axiome war [6]. Dabei waren die Peanoschen Axiome in der sogenannten 1. Stufe der Prädikatenlogik formuliert, was gegenüber der üblichen mengentheoretischen Formulierung (vgl. Kap. 1, § 2) eine gewisse Einschränkung darstellt. Solche Modelle wurden als „Non-Standard"-Modelle des Peanoschen Axiomensystems bezeichnet. Die benützte Konstruktionsmethode, die im wesentlichen die heutige Ultrapotenz-Methode darstellt, läßt sich auf eine beliebige Struktur anwenden und führt immer zu einer Erweiterung, die innerhalb eines gewissen Rahmens – der Logik 1. Stufe – die gleichen Eigenschaften wie die Ausgangsstruktur besitzt. A. Robinson wandte diese Methode auf den Körper $\mathbb{R}$ an und erhielt einen Oberkörper $^*\mathbb{R}$, dessen Elemente er Non-Standard Zahlen nannte. Das Arbeiten in und mit $^*\mathbb{R}$ wurde von ihm als *Non-Standard Analysis* bezeichnet. Eine ausführliche Darstellung dieser Methode und ihrer Anwendungen findet man in Robinsons Buch [5].

Wir wollen nun kurz einige Eigenschaften von $^*\mathbb{R}$ erwähnen, um erklären zu können, wie in $^*\mathbb{R}$ das Differential dy einer Funktion $y = f(x)$ definiert werden kann und wie der Quotient dy/dx mit dem Limes $f'(x)$ des Differenzenquotientens zusammenhängt.

Der Körper $^*\mathbb{R}$ ist ein angeordneter Oberkörper von $\mathbb{R}$, in dem es Elemente a mit $r < a$ für alle $r \in \mathbb{R}$ gibt, das heißt, a ist unendlich groß. Selbstverständlich ist dann $\frac{1}{a}$ unendlich klein: es liegt zwischen 0 und allen positiven $\varepsilon \in \mathbb{R}$. Die Elemente der Menge

$$\mathfrak{O} = \{x \in {}^*\mathbb{R} \colon |x| \leqslant r \text{ für ein } r \in \mathbb{R}\}$$

heißen *endlich*; die Elemente von

$$\mathfrak{M} = \{x \in {}^*\mathbb{R} \colon |x| \leqslant \varepsilon \text{ für alle } \varepsilon \in \mathbb{R}^+\}$$

heißen *unendlich klein*. Dabei ist $\mathbb{R}^+$ die Menge der positiven reellen Zahlen. Man nennt Elemente $x, y \in {}^*\mathbb{R}$ *benachbart*, falls $x - y \in \mathfrak{M}$ ist, und schreibt $x \approx y$. Jede endliche Zahl x ist zu genau einer Zahl $r \in \mathbb{R}$ benachbart. Man schreibt $r = \mathrm{st}(x)$ und nennt r den *Standardteil* von x. Um eine Konfusion mit dem Realteil komplexer Zahlen zu vermeiden, sieht man davon ab, für $\mathrm{st}(x)$ die sehr viel suggestivere Bezeichnung „Realteil" zu verwenden. Eine wesentliche Eigenschaft von $^*\mathbb{R}$ ist nun, daß sich jede reelle Funktion $y = f(x)$ „kanonisch" zu einer Funktion $^*f$ in $^*\mathbb{R}$ fortsetzen läßt; dies soll heißen, daß ihre (in der 1. Stufe formulierbaren) Eigenschaften erhalten bleiben. Unter Benutzung dieser Fortsetzung läßt sich für jedes von 0 verschiedene Element dx aus $\mathfrak{M}$ der Wert $^*f(x + dx)$ in $^*\mathbb{R}$ bilden, falls x aus dem Definitionsbereich von f ist. Die Differenz

$$df = {}^*f(x + dx) - f(x)$$

nennt man dann das Differential der Funktion f. Da sowohl df als auch dx Elemente von $^*\mathbb{R}$ sind, kann der Quotient df/dx (für $dx \neq 0$) selbstverständlich gebildet

werden. Dieser Differentialquotient ist also ein Element aus $^*\mathbb{R}$. Ist f an der reellen Stelle x differenzierbar, das heißt, $\lim_{h\to 0}(f(x+h)-f(x))/h = f'(x)$ existiert, dann läßt sich zeigen, daß

$$\frac{df}{dx} \approx f'(x)$$

gilt, das heißt, $f'(x)$ ist der Standardteil von df/dx. Man beachte, daß im Gegensatz zum Limes des Differenzenquotienten der Differentialquotient df/dx immer existiert. Er muß jedoch nicht für jedes dx endlich sein und, falls er für jedes dx endlich ist, muß sein Standardteil nicht notwendig unabhängig von der Wahl von dx sein. Genau dann, wenn dies jedoch der Fall ist, existiert der Limes des Differenzenquotienten und ist gleich dem Standardteil von df/dx.

Bevor wir in den nächsten Paragraphen alles dies und mehr zeigen werden, wollen wir jetzt, ausgehend von den Aussagen, die etwa Leibniz und l'Hospital über das Unendlich Kleine und den Umgang damit gemacht haben, einen Weg aufzeigen, der basierend auf einer einzigen, sehr natürlichen Forderung fast zwangsläufig zu dem von Robinson benützten $^*\mathbb{R}$ führen wird.

Wie wir schon erwähnten, war man sich früher durchaus darüber im Klaren, daß Größen wie etwa dx oder $f(x+dx)$ nicht einfach reelle Zahlen sein können. Zu Beginn seines Lehrbuches „G.F.A. de Analyse des infiniments petits" (Paris 1696) gibt der Marquis de l'Hospital folgende Definitionen:

„*Definition I.* Variable Größen sind solche, die stetig ab- oder zunehmen. Und konstante Größen sind solche, die beständig gleich bleiben, während andere sich ändern ..."

„*Definition II.* Der unendliche kleine Teil, um den variable Größen ständig zu- oder abnehmen, heißt das Differential dieser Größe."

Eine Größe wie etwa dx ist also etwas Variables, etwas, was „mit der Zeit" variiert werden kann. Leibniz schreibt im Jahre 1702 in einem Brief an den Pariser Professor Pierre Varignon u. a.:

„Hierbei ist jedoch zu berücksichtigen, daß die unvergleichlich kleinen Größen, selbst in ihrem populären Sinn genommen, keineswegs konstant und bestimmt sind, daß sie vielmehr, da man sie so klein annehmen kann wie man will, in geometrischen Erwägungen dieselbe Rolle wie die Unendlichkleinen im strengen Sinne spielen."

Mit den unvergleichlich kleinen Größen meint Leibniz hier die von ihm eingeführten Differentiale.

Diese Zitate legen unserer Ansicht nach nahe, die Größe dx als eine (variable) Funktion anzusehen (sagen wir – als eine Funktion der Zeit –), die auf die Dauer immer kleinere reelle Werte annimmt. Dagegen kann etwa die Zahl 2 als eine Funktion der Zeit angesehen werden, die konstant den Wert zwei hat. So gesehen wären unsere „Größen" also alle Abbildungen von einer Zeitachse T in $\mathbb{R}$. Dabei ist es unerheblich (wovon man sich später überzeugen kann), ob die Zeit kontinuierlich abläuft oder nicht. Wir könnten zwar $\mathbb{R}^+$ für T nehmen, entscheiden uns aber aus schreibtechnischen Gründen für $T = \mathbb{N}' = \{1, 2, \ldots\}$. Unsere Größen sind also

Folgen reeller Zahlen. Die Zuordnung $t \mapsto \frac{1}{t}$ liefert z. B. die Folge

$$\left(1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{t}, \frac{1}{t+1}, \ldots\right).$$

Diese Folge nimmt im Laufe der Zeit immer kleinere Werte an; wir möchten sie deshalb gern als unendlich klein gegenüber den konstanten Folgen

$$(\varepsilon, \varepsilon, \ldots, \varepsilon, \ldots)$$

ansehen, die die positiven reellen Zahlen ε repräsentieren. Die Quotientenfolge

$$\left(\frac{f(x + (1/t)) - f(x)}{1/t}\right)_{t \in \mathbb{N}'}$$

würden wir gerne für $x \in \mathbb{R}$ als Differentialquotienten bezeichnen. Wohl gemerkt, die Quotientenfolge selbst und nicht ihren eventuell existierenden Limes für $t \to \infty$! Hier taucht jedoch eine Schwierigkeit auf. Während man für Folgen

$$(a_1, a_2, \ldots) \qquad \text{und} \qquad (b_1, b_2, \ldots)$$

kanonisch Addition, Subtraktion und Multiplikation durch

$$(a_1 \pm b_1, a_2 \pm b_2, \ldots) \qquad \text{und} \qquad (a_1 \cdot b_1, a_2 \cdot b_2, \ldots)$$

erklären kann, hapert es mit der Division. Mit anderen Worten: die Menge R aller Folgen von $\mathbb{N}'$ in $\mathbb{R}$ ist bezüglich der angegebenen Operationen ein Ring, jedoch kein Körper. Um zu einem Körper zu kommen – dies ist die *einzige* Forderung, die wir stellen wollen – müssen wir den Begriff der „Größe" etwas weiter fassen, das heißt, wir werden gewisse Folgen als die gleiche Größe repräsentierend ansehen müssen. Unterscheiden sich zwei Folgen von einer bestimmten Stelle ab nicht mehr, so wollen wir sie als gleich ansehen, da sie sich in diesem Falle offenbar nur unwesentlich unterscheiden. Würden wir zwei Folgen nur in diesem Falle als gleich ansehen, so bedeutete dies eine Restklassenbildung des Ringes R nach dem Ideal D der Folgen $(a^{(t)})_{t \in \mathbb{N}'}$, bei denen $a^{(t)}$ fast immer gleich 0 ist. Diese Restklassenbildung führt jedoch noch immer nicht zu einem Körper. Wir erreichen dies schließlich dadurch, daß wir ein maximales Ideal M über D (das heißt $M \supset D$) wählen und den Restklassenring R/M als unseren neuen Zahlbereich $^*\mathbb{R}$ nehmen. Bekannterweise (vgl. 2.3.4) ist R/M ein Körper, falls M ein maximales Ideal ist. Unsere Größen sind also die Kongruenzklassen von Folgen $(a^{(t)})_{t \in \mathbb{N}'}$ nach dem Ideal M, das heißt, zwei Folgen werden als gleich angesehen, falls ihre Differenz in M liegt, sie sich also nur um ein Element aus M unterscheiden.

Wir werden in den nächsten Paragraphen zeigen, daß der so definierte Körper $^*\mathbb{R} = R/M$ nicht nur unendlich kleine und große Elemente besitzt, sondern auch, daß sich alle Funktionen von $\mathbb{R}$ nach $^*\mathbb{R}$ kanonisch fortsetzen lassen und daß $\mathbb{R}$ und $^*\mathbb{R}$ alle in einem bestimmten Rahmen formulierbaren Eigenschaften gemeinsam haben. Wir betonen noch einmal, daß dies alles allein aus der Forderung folgt, daß R/M ein Körper sein solle. Es ist dabei unerheblich, welches maximale Ideal M wir wählen, solange es nur D umfaßt. Auf diese Zusammenhänge werden wir im Epilog noch etwas genauer eingehen.

§ 2. Der Non-Standard Zahlbereich *ℝ

1. Konstruktion von *ℝ. Wie wir schon in der Einführung festlegten, ist R der Ring der Folgen $a = (a^{(n)})_{n\in\mathbb{N}'}$ reeller Zahlen, wobei Addition, Subtraktion und Multiplikation komponentenweise definiert sind. Weiter ist D das Ideal in R, das genau aus den Folgen $(a^{(n)})_{n\in\mathbb{N}'}$ besteht, für die fast immer $a^{(n)} = 0$ gilt. Schließlich ist M ein maximales Ideal in R, das D umfaßt. Die Existenz eines solchen Ideales sichert das Zornsche Lemma (vgl. 13.3.2).

Der Ring R enthält ein kanonisches, isomorphes Abbild des Körpers $\mathbb{R}$ der reellen Zahlen. Diese kanonische Einbettung wird geliefert durch

$$r \mapsto (r, r, r, \ldots)$$

für $r \in \mathbb{R}$. Wir wollen $\mathbb{R}$ mit seinem Bild identifizieren, das heißt, wir sehen die konstanten Folgen $(r, r, r, \ldots)$ als reelle Zahlen an. Für Folgen $a, b \in R$ definieren wir

$$a \equiv b \bmod M :\Leftrightarrow a - b \in M.$$

Dies ist eine Äquivalenzrelation auf R. Die Menge der Äquivalenzklassen bezeichnen wir mit *ℝ. Also ist

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

Bekannterweise übertragen sich die Operationen $+, -, \cdot$ auf den Quotienten R/M, der dadurch wieder zu einem Ring wird. Aus der Maximalität von M ergibt sich sogar, daß R/M wieder ein Körper ist. Da jede konstante Folge $\neq 0$ in R invertierbar ist, wird der Unterkörper $\mathbb{R}$ von R bei der Restklassenbildung nicht beeinträchtigt, das heißt, wir finden $\mathbb{R}$ als isomorphes Bild in R/M wieder. Auch hier wollen wir identifizieren. Also ist schließlich $*\mathbb{R} = R/M$ ein Oberkörper von $\mathbb{R}$ (wie wir später sehen werden, natürlich ein echter Oberkörper).

Als nächstes wollen wir zeigen, daß sich jede Funktion $f\colon \mathbb{R}^m \to \mathbb{R}$ zu einer Funktion $*f\colon *\mathbb{R}^m \to *\mathbb{R}$ so fortsetzen läßt, daß sie alle Eigenschaften, die im Rahmen der Logik 1. Stufe ausdrückbar sind (dies werden wir in § 3 präzisieren), behält. Zu einer gegebenen Funktion $f\colon \mathbb{R}^m \to \mathbb{R}$ definieren wir zuerst eine Fortsetzung $\bar{f}$ auf R komponentenweise durch

$$\bar{f}(a_1, \ldots, a_m) = (f(a_1^{(1)}, \ldots, a_m^{(1)}), f(a_1^{(2)}, \ldots, a_m^{(2)}), \ldots),$$

wobei $a_i = (a_i^{(n)})_{n\in\mathbb{N}'}$ für $1 \leqslant i \leqslant m$ Folgen aus R sind. Wir setzen dann weiter

$$*f(a_1, \ldots, a_m) \equiv \bar{f}(a_1, \ldots, a_m) \bmod M,$$

wobei hier die Folgen $a_1, \ldots, a_m$ Vertreter gewisser Restklassen nach M sind. Es bleibt zu zeigen, daß diese Definition unabhängig von der Wahl dieser Vertreter ist. Seien also

$$a_1 \equiv b_1, \ldots, a_m \equiv b_m \bmod M.$$

Es ist dann zu zeigen, daß

$$\bar{f}(a_1, \ldots, a_m) \equiv \bar{f}(b_1, \ldots, b_m) \bmod M$$

ist. Der Beweis dieser ziemlich allgemeinen Aussage liegt nicht auf der Hand, wie

etwa für den Fall der Addition: falls $a_1 - b_1$, $a_2 - b_2 \in M$, so gilt natürlich auch $(a_1 + b_1) - (a_2 + b_2) \in M$. Der Beweis gelingt für beliebige Ideale M, die Maximalität von M ist dabei unerheblich. Wir führen erst einmal den Beweis für das Ideal D aus, dies wird uns dann den Weg weisen. Offensichtlich bedeutet $a \equiv b \bmod D$ gerade, daß für fast alle n (das heißt, für alle bis auf endlich viele) $a^{(n)} = b^{(n)}$ gilt. Damit gilt aber auch in unserem Falle

$$a_1^{(n)} = b_1^{(n)} \qquad \text{und} \qquad \ldots \qquad \text{und} \qquad a_m^{(n)} = b_m^{(n)}$$

für fast alle $n \in \mathbb{N}'$. Also gilt für fast alle n

$$f(a_1^{(n)}, \ldots, a_m^{(n)}) = f(b_1^{(n)}, \ldots, b_m^{(n)}),$$

was natürlich $\bar{f}(a_1, \ldots, a_m) \equiv \bar{f}(b_1, \ldots, b_m) \bmod D$ zur Folge hat.

Für ein beliebiges Ideal M gehen wir jetzt so vor: zu $a \in R$ definieren wir

$$Z(a) = \{n \in \mathbb{N}' : a^{(n)} = 0\}.$$

Weiter bilden wir die Menge

$$U = U_M = \{Z(a) : a \in M\}.$$

Für U gelten die folgenden Eigenschaften:

(0) $$\emptyset \notin U,$$

(i) $$\mathbb{N} \in U,$$

(ii) $$Z_1, Z_2 \in U \Rightarrow Z_1 \cap Z_2 \in U,$$

(iii) $$Z \in U, \qquad Z \subset A \subset \mathbb{N}' \Rightarrow A \in U,$$

das heißt, U ist ein *Filter auf* $\mathbb{N}'$. Ist M maximales Ideal, so ist U_M ein *Ultrafilter*, das heißt, es gilt zusätzlich

(iv) $$A \subset \mathbb{N}' \Rightarrow A \in U \text{ oder } \mathbb{N}' \setminus A \in U.$$

Umfaßt M außerdem noch D, so ist U ein *nicht-trivialer* Ultrafilter, das heißt, es gilt zusätzlich

(v) $$A \subset \mathbb{N}', \ |\mathbb{N}' \setminus A| < \infty \Rightarrow A \in U.$$

Der Nachweis dieser Eigenschaften ist sehr leicht, wir führen ihn z. B. für (iv) durch: Wir wählen eine Folge a aus Nullen und Einsen, so daß $Z(a) = A$ gilt, und nehmen $A \notin U$ an. Also ist insbesondere $a \notin M$. Da M maximal ist, gibt es Elemente $b \in M$ und $c \in R$ mit $1 = b + ac$. Es folgt sofort $Z(b) = Z(1 - ac) \subset \mathbb{N}' \setminus A$. Wegen $b \in M$ erhalten wir $Z(b) \in U$ und dann mit (iii) auch $\mathbb{N}' \setminus A \in U$.

Für $a \in R$ gilt nun generell

$$a \in M \Leftrightarrow Z(a) \in U.$$

Wegen der Definition von U bleibt nur eine Richtung zu zeigen: Sei $Z(a) \in U$. Wir müssen zeigen, daß $a \in M$ ist. Wegen $Z(a) \in U$ gibt es ein $b \in M$ mit $Z(a) = Z(b)$, das heißt, a und b haben die gleichen 0-Komponenten. Wir definieren eine Folge $c = (c^{(n)})_{n \in \mathbb{N}'}$ durch

$$c^{(n)} = \begin{cases} a^{(n)}/b^{(n)} & \text{für} \quad n \notin Z(b), \\ 0 & \text{sonst.} \end{cases}$$

Dann gilt offenbar $a = bc \in M$.

Beachtet man nun, daß $Z(a-b) = \{n: a^{(n)} = b^{(n)}\}$ ist, so erhält man für $a, b \in R$

(vi) $$a \equiv b \bmod M \Leftrightarrow \{n: a^{(n)} = b^{(n)}\} \in U_M.$$

Diese wichtige Beziehung erlaubt uns, nun den noch anstehenden Unabhängigkeitsbeweis zu führen; gleichzeitig wird sie wegweisend für das Folgende sein.

Zurück also zur Definition von $*f$. Nach Voraussetzung gilt $a_i \equiv b_i \bmod M$ für $1 \leqslant i \leqslant m$, das heißt, wir haben $\{n: a_i^{(n)} = b_i^{(n)}\} \in U$ für $1 \leqslant i \leqslant m$. Damit gilt auch

$$\bigcap_{i=1}^{m} \{n: a_i^{(n)} = b_i^{(n)}\} = \{n: a_1^{(n)} = b_1^{(n)}, \ldots, a_m^{(n)} = b_m^{(n)}\} \in U.$$

Wegen

$$\{n: a_1^{(n)} = b_1^{(n)}, \ldots, a_m^{(n)} = b_m^{(n)}\} \subset \{n: f(a_1^{(n)}, \ldots, a_m^{(n)}) = f(b_1^{(n)}, \ldots, b_m^{(n)})\}$$

folgt mit (iii) daraus

$$\{n: f(a_1^{(n)}, \ldots, a_m^{(n)}) = f(b_1^{(n)}, \ldots, b_m^{(n)})\} \in U.$$

Mit (vi) ist dies jedoch zu der zu beweisenden Behauptung äquivalent. Damit wissen wir, wie wir reelle Funktionen auf R/M für jedes Ideal M fortsetzen können. Welche ihrer Eigenschaften sich dabei von $\mathbb{R}$ auf $*\mathbb{R}$ übertragen, werden wir in § 3 genau analysieren. Wie man sich leicht überlegt, gilt für die Fortsetzung zweier Funktionen f und g

$$*(f \circ g) = *f \circ *g.$$

Von dieser Eigenschaft werden wir im folgenden Gebrauch machen, ohne dies besonders hervorzuheben.

Die Tatsache, daß wir hier nur Funktionen betrachten, die auf ganz $\mathbb{R}^m$ definiert sind, sollte nicht weiter stören, da wir ja jede auf einer Teilmenge des $\mathbb{R}^m$ definierte Funktion trivial auf ganz $\mathbb{R}^m$ fortsetzen können.

2. Eigenschaften von $*\mathbb{R}$. Wir wollen ab jetzt wieder annehmen, daß M ein maximales Ideal über D ist. Damit ist zum einen sichergestellt, daß $*\mathbb{R} = R/M$ ein Körper ist. Zum anderen werden wir jetzt zeigen, daß sich die Anordnung $\leqslant$ der reellen Zahlen kanonisch zu einer Anordnung von $*\mathbb{R}$ fortsetzen läßt. Bei der Definition der Fortsetzung von $\leqslant$, die wir wieder mit $\leqslant$ bezeichnen wollen, lassen wir uns von (vi) leiten. Für $a, b \in R$ setzen wir

$$a \leqslant b \bmod M :\Leftrightarrow \{n: a^{(n)} \leqslant b^{(n)}\} \in U_M.$$

Wie (vi) drückt dies aus, daß die Beziehung $a \leqslant b$ modulo M gelten soll, falls die entsprechende Eigenschaft für „sehr viele" Komponenten gilt. Es bleibt noch zu zeigen, daß diese Definition ebenfalls unabhängig von den gewählten Vertretern ist. Sei also $a \equiv a_1$ und $b \equiv b_1 \bmod M$. Dann gilt

$$\{n: a^{(n)} \leqslant b^{(n)}\} \cap \{n: a^{(n)} = a_1^{(n)}\} \cap \{n: b^{(n)} = b_1^{(n)}\} \subset \{n: a_1^{(n)} \leqslant b_1^{(n)}\}.$$

Nach Voraussetzung und mit (ii) ist die linke Seite in U; also mit (iii) auch die rechte, was zu zeigen war. Als nächstes folgt mit den Eigenschaften (iv) und (iii) von U sofort für beliebige $a, b \in R$:

$$\{n: a^{(n)} \leqslant b^{(n)}\} \in U \qquad \text{oder} \qquad \{n: b^{(n)} \leqslant a^{(n)}\} \in U,$$

das heißt, es gilt $a \leqslant b \bmod M$ oder $b \leqslant a \bmod M$. Die weiteren Eigenschaften einer Anordnung

$$\begin{aligned} &a \leqslant a, \\ &a \leqslant b, \quad b \leqslant a \Rightarrow a = b, \\ &a \leqslant b, \quad b \leqslant c \Rightarrow a \leqslant c, \\ &a \leqslant b \qquad\quad \Rightarrow a + c \leqslant b + c, \\ &0 \leqslant a, \quad 0 \leqslant b \Rightarrow 0 \leqslant a \cdot b, \end{aligned}$$

weist man sofort unter Benutzung der Filter-Eigenschaften von U_M nach.

Bis jetzt wurde die Voraussetzung $D \subset M$ noch nicht benützt. Dies wird jetzt zum ersten Mal geschehen bei dem Nachweis, daß $^*\mathbb{R}$ ein Element besitzt, das größer ist als alle reellen Zahlen. Es gilt nämlich für alle $r \in \mathbb{R}$

$$r \leqslant \omega \bmod M,$$

falls wir setzen

$$\omega = (1, 2, 3, \ldots, n, n+1, \ldots).$$

Dies ist klar; wegen (v) gilt nämlich $\{n\colon r^{(n)} \leqslant \omega^{(n)}\} = \{n\colon r \leqslant n\} \in U_M$.

Wie schon in der Einführung angedeutet, wollen wir jetzt den Ring der endlichen Elemente von $^*\mathbb{R}$ definieren. Dabei benützen wir die Fortsetzung $^*|\ |$ des Absolutbetrages der reellen Zahlen. Wie in $\mathbb{R}$, so gilt auch in $^*\mathbb{R}$

$$^*|a| = \begin{cases} a, & \text{falls} \quad 0 \leqslant a, \\ -a, & \text{falls} \quad a \leqslant 0. \end{cases}$$

Ist dabei $a \in R$ ein Repräsentant, so müßten wir eigentlich sowohl die Gleichung als auch die Ungleichung modulo M verstehen. Da wir aber M ein für alle Mal fest gewählt haben, wollen wir hier und im folgenden den Zusatz „mod M" wegfallen lassen; jedenfalls dann, wenn dies zu keiner Konfusion führt. Wir wollen auch im folgenden sehr oft den Stern an den Fortsetzungen von Funktionen weglassen, insbesondere dann, wenn diese Funktionen einen bestimmten Namen haben, z. B. $|\ |$.

Die eben angegebene Eigenschaft der Fortsetzung von $|\ |$ folgt nach dem allgemeinen Übertragungsprinzip von § 3 aus der entsprechenden des reellen Absolutbetrages. Man sieht dies allerdings auch unmittelbar so ein: gilt für eine Folge $a = (a^{(n)})_{n \in \mathbb{N}}$, z. B. $0 \leqslant a$, so heißt dies $\{n\colon 0 \leqslant a^{(n)}\} \in U$. Dann ist natürlich auch die Obermenge $\{n\colon |a^{(n)}| = a^{(n)}\}$ Element von U. Also folgt mit (vi), $|a| \equiv a \bmod M$. Analog schließt man bei $a \leqslant 0$.

Definieren wir nun

$$\mathfrak{O} = \{a \in {}^*\mathbb{R}\colon |a| \leqslant r \text{ für ein } r \in \mathbb{R}\},$$

so sieht man sofort, daß $\mathfrak{O}$ ein echter konvexer Teilring von $^*\mathbb{R}$ ist. Dabei meinen wir mit Konvexität von $\mathfrak{O}$ die Eigenschaft

$$0 \leqslant b \leqslant a \in \mathfrak{O} \Rightarrow b \in \mathfrak{O}.$$

Weiter definieren wir

$$\mathfrak{M} = \{a \in {}^*\mathbb{R} : |a| \leqslant \varepsilon \text{ für alle } \varepsilon \in \mathbb{R}^+\}.$$

Man rechnet sofort nach, daß $\mathfrak{M}$ ein konvexes Ideal in $\mathfrak{O}$ ist, das heißt, es gilt

$$a, b \in \mathfrak{M} \Rightarrow a + b \in \mathfrak{M},$$

$$a \in \mathfrak{M},\ b \in \mathfrak{O} \Rightarrow a \cdot b \in \mathfrak{M},$$

$$0 \leqslant b \leqslant a \in \mathfrak{M} \Rightarrow b \in \mathfrak{M}.$$

$\mathfrak{M}$ besteht nicht nur aus der Null, denn wegen $n \leqslant \omega$ folgt natürlich $0 < 1/\omega \leqslant 1/n$ für alle $n \in \mathbb{N}'$. Also ist $1/\omega \in \mathfrak{M}$. Die Elemente von $\mathfrak{M}$ bezeichnen wir als *unendlich kleine* oder *infinitesimale* Größen. Die Elemente von $\mathfrak{O}$ heißen *endliche* Größen. Alle anderen Elemente von *ℝ werden als *unendliche* oder *infinite* Größen bezeichnet. Für $a, b \in {}^*\mathbb{R}$ setzen wir

$$a \approx b \Leftrightarrow a - b \in \mathfrak{M},$$

das heißt, a und b unterscheiden sich nur um eine infinitesimale Größe. Wir nennen deshalb a und b *benachbart*. Offensichtlich ist $\approx$ eine Äquivalenzrelation auf *ℝ. Wir zeigen jetzt den wichtigen

Satz. *Jede endliche Größe $a \in {}^*\mathbb{R}$ ist zu genau einer reellen Zahl r benachbart. r wird dann als der Standardteil* $\mathrm{st}(a)$ *von a bezeichnet.*

Beweis. Zum Beweis der *Existenz* betrachten wir die Mengen $X_a = \{r \in \mathbb{R} : r \leqslant a\}$ und $Y_a = \{s \in \mathbb{R} : a \leqslant s\}$. Offensichtlich definieren X_a, Y_a einen Schnitt in $\mathbb{R}$. Wegen der Schnittvollständigkeit von $\mathbb{R}$ (siehe Kap. 2, § 2.2) gibt es ein $t \in \mathbb{R}$ mit $r \leqslant t \leqslant s$ für $r \in X_a$, $s \in Y_a$. Hieraus erhält man sofort $|t - a| \leqslant \varepsilon$ für alle $\varepsilon \in \mathbb{R}^+$. Also gilt $a \approx t \in \mathbb{R}$.

Um die *Eindeutigkeit* nachzuweisen, nehmen wir an, es sei $t_1 \approx a \approx t_2$ für $t_1, t_2 \in \mathbb{R}$. Dann folgt $t_1 \approx t_2$, das heißt, $|t_1 - t_2| < \varepsilon$ für alle $\varepsilon \in \mathbb{R}^+$. Dies ist jedoch nur für $t_1 - t_2 = 0$ möglich. □

Aus diesem Satz ersieht man insbesondere, daß $\mathrm{st} : \mathfrak{O} \to \mathbb{R}$ ein ordnungserhaltender Ringhomomorphismus mit Kern $\mathfrak{M}$ und $\mathrm{st} \mid \mathbb{R} = \mathrm{id}$ ist. Damit wissen wir insbesondere, daß wir mit endlichen Größen alle Körperoperationen durchführen können, vorausgesetzt, wir schreiben immer $\approx$ statt $=$ und dividieren nur dann durch ein $a \in \mathfrak{O}$, falls $a \not\approx 0$ gilt.

Es sei noch bemerkt, daß man analog zur Konstruktion von *ℝ aus $\mathbb{R}$ auch *K für jeden Teilkörper K von $\mathbb{R}$ bilden kann. Dann läßt sich der obige Satz verschärfen zu: „In *K besitzt jede endliche Größe einen Standardteil genau dann, wenn $K = \mathbb{R}$ gilt". Der obige Satz drückt also die Schnittvollständigkeit von $\mathbb{R}$ aus; das Rechnen mit Standardteilen ersetzt somit die explizite Anwendung der Schnittvollständigkeit der klassischen Analysis.

Wir wenden uns jetzt der *Stetigkeit* einer Funktion $f : \mathbb{R} \to \mathbb{R}$ im Punkte $x \in \mathbb{R}$ zu. In der uns heute geläufigen Definition heißt dies, daß es zu jedem $\varepsilon \in \mathbb{R}^+$ ein $\delta \in \mathbb{R}^+$ gibt, so daß für alle $h \in \mathbb{R}$ gilt

$$|h| \leqslant \delta \Rightarrow |f(x + h) - f(x)| \leqslant \varepsilon.$$

Wie wir vorher gesehen haben, läßt sich die Funktion f zu einer Funktion ${}^*f\colon {}^*\mathbb{R} \to {}^*\mathbb{R}$ so fortsetzen, daß für $a = (a^{(n)})_{n\in\mathbb{N}'}$ gilt

$$ {}^*f(a) \equiv (f(a^{(n)}))_{n\in\mathbb{N}'} \bmod M. $$

Wir wollen jetzt ein $h \in {}^*\mathbb{R}$ mit $|h| \leqslant \delta$ betrachten. Es gilt also $\{n\colon |h^{(n)}| \leqslant \delta\} \in U$. Wegen $\{n\colon |h^{(n)}| \leqslant \delta\} \subset \{n\colon |f(x + h^{(n)}) - f(x)| \leqslant \varepsilon\}$ ist auch die letztere Menge in U, das heißt, es gilt $|{}^*f(x + h) - {}^*f(x)| \leqslant \varepsilon$. Ist insbesondere $h \in \mathfrak{M}$, so ist $|h| \leqslant \delta$ für alle $\delta \in \mathbb{R}^+$ richtig. Also ist auch die Folgerung $|{}^*f(x + h) - {}^*f(x)| \leqslant \varepsilon$ für alle $\varepsilon \in \mathbb{R}^+$ richtig, das heißt, es gilt

$$ {}^*f(x + h) \approx {}^*f(x). $$

Damit haben wir eine Richtung der folgenden Äquivalenz bewiesen.

Satz. *Die Funktion $f\colon \mathbb{R} \to \mathbb{R}$ ist genau dann in dem Punkt $x \in \mathbb{R}$ stetig, falls ${}^*f(x + h) \approx f(x)$ für alle $h \approx 0$ gilt.*

Beweis der Richtung „⇐“. Wir nehmen an, f sei in x nicht stetig. Dann gibt es ein $\varepsilon \in \mathbb{R}^+$, so daß es zu jedem $n \in \mathbb{N}'$ ein $h^{(n)} \in \mathbb{R}$ gibt mit $|h^{(n)}| \leqslant \frac{1}{n}$ und $|f(x+h^{(n)})-f(x)| \geqslant \varepsilon$. Setzen wir nun $h=(h^{(n)})_{n\in\mathbb{N}'}$, so gilt offenbar $|h| \leqslant 1/\omega$. Wegen $1/\omega \in \mathfrak{M}$ gilt auch $h \in \mathfrak{M}$, das heißt, $h \approx 0$. Wegen der Wahl von $h^{(n)}$ gilt andererseits $\{n\colon |f(x + h^{(n)}) - f(x)| \geqslant \varepsilon\} \in U$, das heißt, $|{}^*f(x + h) - f(x)| \geqslant \varepsilon$. Dies widerspricht aber offensichtlich der Voraussetzung ${}^*f(x + h) \approx f(x)$ für $h \approx 0$. □

§ 3. Gemeinsamkeiten von $\mathbb{R}$ und ${}^*\mathbb{R}$

Der Beweis des letzten Satzes und (noch deutlicher) der Nachweis der charakteristischen Eigenschaft des Absolutbetrages zeigen, daß sich gewisse Eigenschaften von Funktionen auf $\mathbb{R}$ über die Komponenten auf die Fortsetzungen dieser Funktionen nach ${}^*\mathbb{R}$ vererben. In diesem Paragraphen wollen wir versuchen, ein allgemeines Übertragungsprinzip von Eigenschaften von $\mathbb{R}$ nach ${}^*\mathbb{R}$ zu formulieren. Es ist dabei von vornherein klar, daß man nicht jede Eigenschaft übertragen können wird; schließlich sind $\mathbb{R}$ und ${}^*\mathbb{R}$ verschieden. Man wird jedoch versuchen, einen möglichst großen Bereich von übertragbaren Eigenschaften zu finden. Wir werden einen solchen Bereich induktiv aufbauen: wir werden, von sehr einfachen Eigenschaften ausgehend, mit bestimmten Verfahren zu immer komplexeren fortschreiten.

Um dieses Programm durchführen zu können, müssen wir uns erst überlegen, wie wir generell Eigenschaften beschreiben können. Eine Möglichkeit ist die folgende: Wir führen eine Kunstsprache – eine formale Sprache – ein, in der wir die uns interessierenden Eigenschaften beschreiben. Der eben angedeutete induktive Prozeß läuft dann durch eine Induktion über diese „Beschreibungen“, z. B. über die Länge der Beschreibung. Die angesprochene formale Sprache wird natürlich der mathematischen Umgangssprache sehr ähnlich sein. Dies empfiehlt sich schon deshalb, damit sie möglichst leicht gelesen werden kann und dabei die richtige Interpretation unmittelbar suggeriert wird.

Betrachten wir ein Beispiel. Eine Eigenschaft, die $\mathbb{R}$ und $^*\mathbb{R}$ gemeinsam haben, ist die Kommutativität der Addition. In der von uns noch einzuführenden formalen Sprache werden wir dies beschreiben durch den formalen Ausdruck

$$\forall x \forall y\, x + y = y + x.$$

Dies ist eine Zeichenreihe, bestehend aus 11 Einzelzeichen. Von der Schreibweise her sind die Zeichen so gewählt, daß man augenblicklich an die richtige Interpretation dieser Zeichenreihe denkt. Es bleibt eigentlich nur eine Interpretation offen, nämlich die von $\forall x$. Dies hängt nun davon ab, ob wir die obige Formel in $\mathbb{R}$ oder in $^*\mathbb{R}$ interpretieren wollen. Im ersten Fall soll $\forall x$ gerade durch „für alle $a \in \mathbb{R}$" interpretiert werden, im zweiten Fall durch „für alle $a \in {}^*\mathbb{R}$". Wir werden also auch eine Beziehung $\models$ zwischen $\mathbb{R}$ und Formeln bzw. zwischen $^*\mathbb{R}$ und Formeln definieren müssen, die die Gültigkeit einer Formel in $\mathbb{R}$ bzw. $^*\mathbb{R}$ zum Ausdruck bringt.

Für die zu definierende formale Sprache verwenden wir als *Grundzeichen* (Einzelzeichen) die bekannten Zeichen

$$\neg \quad \wedge \quad \exists \quad = \quad \leqslant \quad) \quad , \quad ($$

sowie Zeichen für Variablen

$$v_0 \quad v_1 \quad v_2 \quad \cdots .$$

Weiter benützen wir für jede Funktion $f\colon \mathbb{R}^m \to \mathbb{R}$ ein Zeichen, um damit f bzw. $^*f$ benennen zu können. Bequemlichkeitshalber verwenden wir f selbst als Zeichen. Man beachte, daß also die Interpretation von f über $\mathbb{R}$ gerade f ist, während über $^*\mathbb{R}$ die Interpretation von f eben $^*f$ ist. Schließlich führen wir für jedes $a \in {}^*\mathbb{R}$ ein Zeichen $\underline{a}$ (den Namen von a) ein. Die Interpretation von $\underline{a}$ ist natürlich a.

Aus diesen Grundzeichen bauen wir nun zuerst *Terme* induktiv auf.

(1) Variablen und $\underline{a}$ für $a \in {}^*\mathbb{R}$ sind Terme;

(2) sind $t_1, \ldots, t_m$ Terme und ist f eine m-stellige Funktion, so ist auch $f(t_1, \ldots, t_m)$ ein Term.

Enthält ein Term t keine Variablen, so heißt er *konstant*. Die Interpretation eines konstanten Termes in $\mathbb{R}$ bzw. in $^*\mathbb{R}$ liegt auf der Hand und ist in beiden Fällen dieselbe, falls nur Konstanten $\underline{a}$ mit $a \in \mathbb{R}$ vorkommen.

Als nächstes definieren wir *Formeln*.

(1) Sind t_1 und t_2 Terme, so sind $t_1 = t_2$ und $t_1 \leqslant t_2$ Formeln (sogenannte *Primformeln*);

(2) sind φ_1 und φ_2 Formeln und ist v eine Variable, so sind auch $\neg\varphi_1$, $(\varphi_1 \wedge \varphi_2)$, $\exists v \varphi_1$ Formeln.

Eine Formel, in der keine Variablen mehr frei vorkommen, nennen wir eine *Aussage*. Dabei kommt eine Variable v, die in einer Formel φ noch frei vorkommt, in der Formel $\exists v \varphi$ nicht mehr frei vor. Dies läßt sich etwa so präzisieren: Für einen Term t sei $F(t)$ die (endliche) Menge aller in t vorkommenden Variablen. Wir

definieren dann rekursiv

$$F(t_1 = t_2) = F(t_1 \leqslant t_2) = F(t_1) \cup F(t_2),$$
$$F(\neg \varphi) = F(\varphi),$$
$$F(\varphi_1 \wedge \varphi_2) = F(\varphi_1) \cup F(\varphi_2),$$
$$F(\exists v \varphi) = F(\varphi) \setminus \{v\}.$$

φ heißt demnach eine Aussage, falls $F(\varphi) = \emptyset$ ist.

Für Aussagen α definieren wir nun ihre *Gültigkeit* in $\mathbb{R}$ bzw. $^*\mathbb{R}$, das heißt, wir definieren die Beziehung $\mathbb{R} \models \alpha$ („in $\mathbb{R}$ gilt α") bzw. $^*\mathbb{R} \models \alpha$ („in $^*\mathbb{R}$ gilt α"). Der erste Fall macht offenbar nur dann einen Sinn, wenn in α keine Konstanten $\underline{a}$ mit $a \in {}^*\mathbb{R} \setminus \mathbb{R}$ vorkommen. In diesem Fall nennen wir α eine $\mathbb{R}$-*Aussage*. Für Primformeln α liegt ihre Gültigkeit in $\mathbb{R}$ bzw. in $^*\mathbb{R}$ auf der Hand. Sei nun die Gültigkeit von α_1 und α_2 in $^*\mathbb{R}$ schon definiert. Dann setzen wir

$$^*\mathbb{R} \models \neg \alpha_1 \quad \Leftrightarrow {}^*\mathbb{R} \not\models \alpha_1,$$
$$^*\mathbb{R} \models (\alpha_1 \wedge \alpha_2) \Leftrightarrow [{}^*\mathbb{R} \models \alpha_1 \text{ und } {}^*\mathbb{R} \models \alpha_2],$$
$$^*\mathbb{R} \models \exists v \varphi \quad \Leftrightarrow \text{es gibt } a \in {}^*\mathbb{R} \text{ mit } {}^*\mathbb{R} \models \varphi(\underline{a}).$$

Dabei ist $\varphi(\underline{a})$ das Ergebnis der Einsetzung von $\underline{a}$ für v in die Formel φ. Eine möglicherweise in φ auftretende andere Quantifikation $\exists v \psi$ darf dabei natürlich nicht durch $\exists \underline{a}\, \psi(\underline{a})$ ersetzt werden, da dies nach obigem Aufbau überhaupt keine Formel mehr ist. In dem letzten Fall der Definition sichert die Voraussetzung, daß $\exists v \varphi$ eine Aussage ist, natürlich die gleiche Eigenschaft für $\varphi(\underline{a})$. Der induktive Aufbau der Definition für die Gültigkeit von $\mathbb{R}$-Aussagen in $\mathbb{R}$ verläuft analog zur obigen Definition; es heißt dann z. B.

$$\mathbb{R} \models \exists v \varphi \Leftrightarrow \text{es gibt } a \in \mathbb{R} \text{ mit } \mathbb{R} \models \varphi(\underline{a}).$$

Wir sind nun endlich in der Lage, das allgemeine Übertragungsprinzip zu formulieren und zu beweisen.

Allgemeines Übertragungsprinzip. *Es sei α eine Aussage, in der höchstens die Konstanten $\underline{a}_1, \ldots, \underline{a}_m$ vorkommen, also $\alpha = \alpha(\underline{a}_1, \ldots, \underline{a}_m)$. Dann gilt*

$$^*\mathbb{R} \models \alpha(\underline{a}_1, \ldots, \underline{a}_m) \Leftrightarrow \{n \colon \mathbb{R} \models \alpha(\underline{a}_1^{(n)}, \ldots, \underline{a}_m^{(n)})\} \in U_M.$$

Beweis. Sei α zuerst eine Primaussage, also $t_1 = t_2$ oder $t_1 \leqslant t_2$. Dann liefern (vi) und die Definition von $\leqslant$ in $^*\mathbb{R}$ gerade die behauptete Äquivalenz.

Wir schließen jetzt weiter durch Induktion über den Aufbau der Aussage α. Ist α von der Gestalt $\neg \alpha_1$ und setzen wir obige Äquivalenz für α_1 voraus, so können wir folgendermaßen weiterschließen:

$$^*\mathbb{R} \models \neg \alpha_1(\underline{a}_1, \ldots) \Leftrightarrow {}^*\mathbb{R} \not\models \alpha_1(\underline{a}_1, \ldots)$$
$$\Leftrightarrow \{n \colon \mathbb{R} \models \alpha_1(\underline{a}_1^{(n)}, \ldots)\} \notin U$$
$$\Leftrightarrow \{n \colon \mathbb{R} \not\models \alpha_1(\underline{a}_1^{(n)}, \ldots)\} \in U$$
$$\Leftrightarrow \{n \colon \mathbb{R} \models \neg \alpha_1(\underline{a}_1^{(n)}, \ldots)\} \in U.$$

Dies ist klar, da für einen Ultrafilter U auf $\mathbb{N}'$ immer

$$A \notin U \Leftrightarrow \mathbb{N}' \setminus A \in U$$

gilt. Ist α von der Gestalt $(\alpha_1 \wedge \alpha_2)$, so schließen wir so:

$$\begin{aligned} {}^*\mathbb{R} \models (\alpha_1(\underline{a_1}, \ldots) \wedge \alpha_2(\underline{a_1}, \ldots)) &\Leftrightarrow [{}^*\mathbb{R} \models \alpha_1(\underline{a_1}, \ldots) \text{ und } {}^*\mathbb{R} \models \alpha_2(\underline{a_1}, \ldots)] \\ &\Leftrightarrow [\{n\colon \mathbb{R} \models \alpha_1(\underline{a_1^{(n)}}, \ldots)\} \in U \text{ und } \ldots] \\ &\Leftrightarrow \{n\colon \mathbb{R} \models \alpha_1((\underline{a_1^{(n)}}, \ldots) \wedge \alpha_2(\underline{a_1^{(n)}}, \ldots))\} \in U_M. \end{aligned}$$

Dies ist wiederum richtig wegen

$$A \cap B \in U \Leftrightarrow A \in U \text{ und } B \in U.$$

Ist schließlich α von der Gestalt $\exists v \varphi$, so gilt offensichtlich

$$\begin{aligned} {}^*\mathbb{R} \models \exists v \varphi(\underline{a_1}, \ldots) &\Leftrightarrow \text{es gibt } a \in {}^*\mathbb{R} \text{ mit } {}^*\mathbb{R} \models \varphi(\underline{a}, \underline{a_1}, \ldots) \\ &\Leftrightarrow \text{es gibt } a \in {}^*\mathbb{R} \text{ mit } \{n\colon \mathbb{R} \models \varphi(\underline{a^{(n)}}, \underline{a_1^{(n)}}, \ldots)\} \in U \\ &\Rightarrow \{n\colon \mathbb{R} \models \exists v \varphi(\underline{a^{(n)}}, \ldots)\} \in U. \end{aligned}$$

Dies folgt aus der Inklusion

$$\{n\colon \mathbb{R} \models \varphi(\underline{a_1^{(n)}}, \ldots)\} \subset \{n\colon \mathbb{R} \models \exists v \varphi(\ldots)\}.$$

Es bleibt die Umkehrung der letzten Implikation zu zeigen. Sei also $\{n\colon \exists v \varphi(\underline{a_1^{(n)}}, \ldots)\} \in U$. Wir definieren eine Folge $a = (a^{(n)})_{n \in \mathbb{N}'}$ folgendermaßen. Es sei $a^{(n)}$ ein $r \in \mathbb{R}$, für das $\mathbb{R} \models \varphi(\underline{r}, \underline{a_1^{(n)}}, \ldots)$ gilt, falls es ein solches r überhaupt gibt. Andernfalls setzen wir $a^{(n)} = 0$. Offenbar gilt mit dieser Folge a

$$\{n\colon \mathbb{R} \models \exists v \varphi(\underline{a^{(n)}}, \ldots)\} \subset \{n\colon \mathbb{R} \models \varphi(\underline{a^{(n)}}, \underline{a_1^{(n)}}, \ldots)\}.$$

Ist die erste Menge in U_M, so auch die zweite. Dies beweist die Umkehrung der letzten Inklusion. □

Als Korollar erhalten wir das

Übertragungsprinzip. *Sei α eine $\mathbb{R}$-Aussage. Dann gilt α in $\mathbb{R}$ genau dann, wenn α in $^*\mathbb{R}$ gilt.*

Beweis. Da für Zahlen $r \in \mathbb{R}$ alle Komponenten $r^{(n)} = r$ sind, ist für $a_i \in \mathbb{R}$

$$\{n\colon \mathbb{R} \models \alpha(\underline{a_1^{(n)}}, \ldots, \underline{a_m^{(n)}})\} = \mathbb{N}' \text{ oder } \emptyset,$$

je nachdem $\alpha(\underline{a_1^{(n)}}, \ldots, \underline{a_m^{(n)}})$ in $\mathbb{R}$ gilt oder nicht. Wegen $\mathbb{N}' \in U$ und $\emptyset \notin U$ folgt also

$$\begin{aligned} \mathbb{R} \models \alpha &\Leftrightarrow \{n \mid \mathbb{R} \models \alpha(\underline{a_1^{(n)}}, \ldots)\} \in U \\ &\Leftrightarrow {}^*\mathbb{R} \models \alpha. \end{aligned}$$

□

Im folgenden werden wir nur noch dieses Übertragungsprinzip verwenden. Wir können also die Entstehungsgeschichte von $^*\mathbb{R}$ vollständig vergessen. Dies erweist

sich tatsächlich meistens als nützlich, da das Hantieren mit Indizes sehr unübersichtlich werden kann. Bei der alleinigen Anwendung des Übertragungsprinzipes können die Elemente von $\mathbb{R}$ und $^*\mathbb{R}$ gleichermaßen als „Urelemente" behandelt werden. Arbeitet man dagegen mit der speziellen Konstruktion von $^*\mathbb{R}$, so hat man auf der einen Seite reelle Zahlen, während man auf der anderen Seite Äquivalenzklassen von Folgen reeller Zahlen hat.

Die einzige, nicht zu unterschätzende Schwierigkeit bei der Anwendung des Übertragungsprinzipes ist die Formalisierung der zu übertragenden Eigenschaft. Es bedarf einiger Übung, bis man dafür ein „Gefühl" entwickelt hat. Dies ist der Preis, den man zahlen muß für die Annehmlichkeiten des Rechnens mit infinitesimalen Größen.

Bei der Formalisierung der zu übertragenden Eigenschaften wird man natürlich mehr und mehr die benützte formale Sprache durch Abkürzungen besser lesbar machen, z. B. verwendet man

$$\begin{array}{lll} (\varphi \vee \psi) & \text{für} & \neg(\neg\varphi \wedge \neg\psi), \\ (\varphi \to \psi) & \text{für} & \neg(\varphi \wedge \neg\psi), \\ (\varphi \leftrightarrow \psi) & \text{für} & (\varphi \to \psi) \wedge (\psi \to \varphi), \\ \forall v\varphi & \text{für} & \neg\exists v \neg\varphi. \end{array}$$

Zum Abschluß dieses Paragraphen wollen wir noch einmal einen Beweis für den Satz über die Stetigkeit aus § 2 geben; diesmal unter alleiniger Benutzung des Übertragungsprinzipes. Für eine spätere Verwendung nehmen wir gleich eine kleine Verallgemeinerung vor. Wir zeigen den

Limessatz. *Für* $g: \mathbb{R} \to \mathbb{R}$ *und* $x_0, b \in \mathbb{R}$ *ist die Aussage* $\lim_{0 \neq h \to 0} g(x_0 + h) = b$ *äquivalent zu* $^*g(x_0 + h) \approx b$ *für alle* $h \approx 0$ *mit* $h \neq 0$.

Beweis. Gilt die Limesaussage für g, so gibt es zu jedem $\varepsilon \in \mathbb{R}^+$ ein $\delta \in \mathbb{R}^+$ mit

$$\mathbb{R} \models \forall h(0 < |h| < \underline{\delta} \to |g(\underline{x_0} + h) - \underline{b}| < \underline{\varepsilon}).$$

Mit dem Übertragungsprinzip erhalten wir daraus

$$^*\mathbb{R} \models \forall h(0 < |h| < \underline{\delta} \to |g(\underline{x_0} + h) - \underline{b}| < \underline{\varepsilon}).$$

Für $h \approx 0$ und $h \neq 0$ gilt wegen $|h| < \delta$ also $|^*g(x_0 + h) - b| < \varepsilon$. Da dies sogar für alle $\varepsilon \in \mathbb{R}^+$ gilt, erhalten wir $^*g(x_0 + h) \approx b$. Umgekehrt setzen wir die Gültigkeit von $^*g(x_0 + h) \approx b$ für $h \approx 0$ mit $h \neq 0$ voraus. Sei $h_0 \approx 0$ mit $0 < h_0$ fest gewählt. Für $0 < |h| < h_0$ gilt dann also $^*g(x_0 + h) \approx b$, das heißt, wir haben insbesondere

$$^*\mathbb{R} \models \exists\delta(0 < \delta \wedge \forall h(0 < |h| < \delta \to |g(\underline{x_0} + h) - \underline{b}| < \underline{\varepsilon}),$$

wenn man an $\delta = h_0$ denkt. Dabei ist $\varepsilon \in \mathbb{R}^+$ beliebig. Das Übertragungsprinzip ergibt dann

$$\mathbb{R} \models \exists\delta(0 < \delta \wedge \forall h(0 < |h| < \delta \to |g(\underline{x_0} + h) - \underline{b}| < \underline{\varepsilon}).$$

Dies ist aber gerade die behauptete Limesaussage. □

Dieser Beweis zeigt deutlich die Überlegenheit der Anwendung des Übertragungsprinzipes über die ad hoc Konstruktionen mit Folgen. Dies ist nicht verwunderlich, da sich die Konstruktionen jetzt im Beweis des Übertragungsprinzipes befinden.

§ 4. Differential- und Integralrechnung

1. Differentiation. Für eine Funktion $f: \mathbb{R} \to \mathbb{R}$ führen wir jetzt ihr *Differential* $df(x)$ an einer Stelle $x \in \mathbb{R}$ ein. Dazu fixieren wir ein $h \approx 0$ mit $h \neq 0$ und setzen

$$df(x) = {}^*f(x + h) - f(x).$$

Für die Identitätsfunktion $f(x) = x$ erhalten wir damit speziell $dx = (x + h) - x = h$. Im folgenden wollen wir deshalb stets dx an Stelle von $h \approx 0$ mit $h \neq 0$ verwenden. Damit erhalten wir für das Differential von f an der Stelle x

$$df(x) = {}^*f(x + dx) - f(x).$$

Man beachte jedoch, daß dieses Differential von der Wahl der Größe $dx \in \mathfrak{M} \setminus \{0\}$ abhängt.

Der *Differentialquotient* $df(x)/dx$ kann für *jede* Funktion $f: \mathbb{R} \to \mathbb{R}$ gebildet werden; er ist ein bestimmtes Element aus ${}^*\mathbb{R}$. Den Zusammenhang mit der Ableitung der Funktion f an der Stelle x beschreibt folgender

Satz. *Existiert für $f: \mathbb{R} \to \mathbb{R}$ der Limes des Differenzenquotienten an der Stelle $x \in \mathbb{R}$ mit Wert $f'(x)$, so gilt $df(x)/dx \approx f'(x)$ für alle $dx \in \mathfrak{M} \setminus \{0\}$ und umgekehrt.*

Beweis. Setzt man im Limessatz (§ 3) $g(h) = \dfrac{f(x + h) - f(x)}{h}$ und $x_0 = 0$, so erhält man die Gleichwertigkeit der behaupteten Limesaussage mit

$$\frac{df(x)}{dx} = \frac{{}^*f(x + dx) - f(x)}{dx} = {}^*g(dx) \approx f'(x)$$

für alle $dx \in \mathfrak{M} \setminus \{0\}$. □

Für die Ableitungen von Funktionen ergeben sich nun leicht die üblichen Regeln:

(1) *Ist f in x differenzierbar, so ist f in x stetig.*

Wegen $df(x)/dx \approx f'(x)$ folgt ${}^*f(x + dx) - f(x) = df(x) \approx f'(x)dx \approx 0$. Also ist ${}^*f(x + dx) \approx f(x)$ für alle dx. Dies ist die Stetigkeit von f in x.

(2) *Sind f und g in x differenzierbar, so auch $(f + g)$ und $(f \cdot g)$, und es gilt $(f + g)'(x) = f'(x) + g'(x)$, $(f \cdot g)'(x) = (f' \cdot g)(x) + (f \cdot g')(x)$.*

Wir führen gleich den Fall der Multiplikation durch:

$$\begin{aligned} d(f \cdot g)(x) &= {}^*(f \cdot g)(x + dx) - (f \cdot g)(x) \\ &= {}^*f(x + dx) \cdot {}^*g(x + dx) - f(x)g(x) \\ &= (df(x) + f(x)) \cdot (dg(x) + g(x)) - f(x)g(x) \\ &= df(x)\,dg(x) + f(x)\,dg(x) + g(x)\,df(x). \end{aligned}$$

Division durch dx ergibt dann

$$\begin{aligned} \frac{d(f \cdot g)(x)}{dx} &= f(x) \cdot \frac{dg(x)}{dx} + g(x) \cdot \frac{df(x)}{dx} + df(x) \cdot \frac{dg(x)}{dx} \\ &\approx f(x)g'(x) + g(x)f'(x). \end{aligned}$$

Dies ergibt sich aus der Voraussetzung der Differenzierbarkeit von f und g an der Stelle x, die insbesondere (nach (1)) $df(x) \approx 0$ impliziert. Insgesamt erhalten wir damit

$$(f \cdot g)'(x) \approx f(x) \cdot g'(x) + g(x) \cdot f'(x).$$

Da aber beide Seiten Elemente von $\mathbb{R}$ sind, muß die Gleichheit gelten.

(3) *Ist f in x differenzierbar und $f(x) \neq 0$, so ist $1/f$ in x differenzierbar und es gilt $(1/f)'(x) = -f'(x)/f(x)^2$.*

Wegen $f(x) \approx {}^*f(x + dx)$ ist natürlich auch ${}^*f(x + dx) \neq 0$. Wir erhalten also

$$d\frac{1}{f}(x) = \frac{1}{{}^*f(x + dx)} - \frac{1}{f(x)} = \frac{f(x) - {}^*f(x + dx)}{f(x) \cdot {}^*f(x + dx)}.$$

Daraus folgt

$$\frac{d\frac{1}{f}(x)}{dx} = \frac{\frac{-df(x)}{dx}}{f(x)\,{}^*f(x + dx)} \approx \frac{-f'(x)}{f(x)\,{}^*f(x + dx)} \approx \frac{-f'(x)}{f(x)^2}.$$

Hieraus folgt wie in (2) die Behauptung.

(4) *Sind f in x und g in $f(x)$ differenzierbar, so ist auch $g \circ f$ in x differenzierbar, und es gilt $(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$.*

Im Falle $df(x) \neq 0$ erhalten wir mit

$$\begin{aligned} d(g \circ f)(x) &= {}^*g({}^*f(x + dx)) - g(f(x)) \\ &= {}^*g(f(x) + df(x)) - g(f(x)) \\ &= dg(f(x)) \end{aligned}$$

ein Differential von g an der Stelle $f(x)$, gebildet mit $h = df(x)$. Division durch dx ergibt dann

$$\frac{d(g \circ f)(x)}{dx} = \frac{dg(f(x))}{df(x)} \cdot \frac{df(x)}{dx}.$$

Der Übergang zum Standardteil ergibt

$$\mathrm{st}\left(\frac{d(g\circ f)(x)}{dx}\right) = g'(f(x))\cdot f'(x).$$

Diese Gleichung gilt jedoch auch im Falle $df(x)=0$. Hieraus folgt nämlich einerseits $f'(x)=\mathrm{st}(df(x)/dx)=0$ und andererseits $d(g\circ f)(x)={}^*g(f(x))-g(f(x))=0$. Da also obige Gleichung für alle $dx\in\mathfrak{M}\setminus\{0\}$ gilt, ist nach dem letzten Satz $(g\circ f)$ in x differenzierbar mit der behaupteten Ableitung. □

Dem aufmerksamen Leser wird sicherlich nicht entgangen sein, daß wir in den eben geführten Beweisen an einigen Stellen, z. B. bei

$${}^*(f\cdot g)(y) = {}^*f(y)\cdot {}^*g(y),$$

implizit das Übertragungsprinzip benützt haben. Da nämlich die Aussage

$$\forall v(f\cdot g)(v) = f(v)\cdot g(v)$$

in $\mathbb{R}$ gilt, muß sie auch in ${}^*\mathbb{R}$, das heißt, für die Fortsetzungen der drei Funktionen f, g und $(f\cdot g)$ gelten.

2. Integration. In diesem letzten Abschnitt wollen wir skizzieren, wie sich das Integral einer in dem abgeschlossenen Intervall $[a,b]$ stetigen Funktion f als eine Summe von Rechtecksinhalten mit unendlich kleiner Breite beschreiben läßt. Verständlicherweise kann es sich dabei nicht um eine endliche Summe handeln, das heißt, die Summation kann nicht einfach von 0 bis zu einem $n\in\mathbb{N}$ laufen. Dies ist klar, da eine endliche Summe von Elementen aus $\mathfrak{M}$ wieder in $\mathfrak{M}$ liegt. Wir werden statt dessen als obere Summationsschranke eine unendlich große „natürliche Zahl" nehmen. Es bleibt dann allerdings noch zu klären, was in diesem Falle „Summe" bedeutet.

Zuerst zu den unendlich großen „natürlichen" Zahlen. Die charakteristische Funktion χ von $\mathbb{N}$, definiert für $x\in\mathbb{R}$ durch

$$\chi(x) = \begin{cases} 1, & \text{falls } x\in\mathbb{N} \\ 0, & \text{sonst,} \end{cases}$$

besitzt wie jede Funktion eine Fortsetzung ${}^*\chi$ auf ${}^*\mathbb{R}$. Diese Fortsetzung behält (nach dem Übertragungsprinzip) selbstverständlich die Eigenschaft, eine 0, 1-Funktion zu sein. Wir definieren nun

$${}^*\mathbb{N} = \{a\in{}^*\mathbb{R}\colon {}^*\chi(a)=1\}.$$

Wegen $\{n\colon \chi(\omega^{(n)})=1\}=\mathbb{N}'\in U$ folgt nach dem Allgemeinen Übertragungsprinzip sofort $\omega\in{}^*\mathbb{N}$. Es gibt also unendlich große natürliche Zahlen in ${}^*\mathbb{R}$. Man überzeugt sich leicht davon, daß $\mathbb{N}$ Teilmenge von ${}^*\mathbb{N}$ ist und daß die neuen Elemente von ${}^*\mathbb{N}$ größer sind als alle Elemente von $\mathbb{N}$.

Bevor wir zur Integration kommen, wollen wir unter Benutzung unendlich feiner Intervallteilungen das folgende, wohlbekannte Lemma beweisen.

Lemma. *Ist eine Funktion f im abgeschlossenen Intervall $[a,b]$ stetig, so nimmt sie dort ihr Maximum (und ihr Minimum) an.*

Beweis. Ist $n \in \mathbb{N}$ und setzt man $a_i = a + ((b-a)/n)(i-1)$ für $1 \leqslant i \leqslant n+1$, so gibt es natürlich unter den endlich vielen Werten $f(a_i)$ mit $1 \leqslant i \leqslant n+1$ einen maximalen. Daß dies für alle $n \in \mathbb{N}$ gilt, läßt sich in $\mathbb{R}$ durch eine Aussage mit den Parametern $\underline{a}$ und $\underline{b}$ ausdrücken (Übungsaufgabe!), gilt also auch für alle Elemente von $^*\mathbb{N}$, z. B. für ω. Sei etwa $^*f(a_j)$ maximal unter den Werten $^*f(a_i)$ für $1 \leqslant i \leqslant \omega + 1$. Sei weiter $x = \mathrm{st}(a_j) \in [a, b]$. Zu jeder reellen Zahl $y \in [a, b]$ gibt es ein $i \leqslant \omega$ mit $a_i \leqslant y \leqslant a_{i+1}$. Dies folgt wieder mit dem Übertragungsprinzip, da es für jedes $n \in \mathbb{N}$ an Stelle von ω richtig ist. Wegen $a_{i+1} - a_i = (b-a)/\omega \approx 0$ erhalten wir insbesondere $y = \mathrm{st}(a_i)$. Mit der Stetigkeit von f folgt dann $f(y) \approx {}^*f(a_i) \leqslant {}^*f(a_j) \approx f(x)$. Dies impliziert sofort $f(y) \leqslant f(x)$. □

Zurück zum Integral. Es seien $f\colon \mathbb{R} \to \mathbb{R}$ und $a, b \in \mathbb{R}$ mit $a < b$. Zu jedem $h \in \mathbb{R}^+$ mit $h \leqslant b - a$ gibt es ein $n \in \mathbb{N}$ mit $nh \leqslant b - a < (n+1)h$. Die Funktion

$$S_f(a, b, h) = \sum_{i=1}^{n} f(a_i) \cdot h + f(a_{n+1})(b - a_{n+1}),$$

wobei wir hier $a_i = a + (i-1)h$ setzen, ist die Summe der Rechtecksinhalte der Breite h bzw. $(b - a_{n+1})$ und Höhe $f(a_i)$. Für unendlich kleines h sollte eine solche Summe zur Beschreibung des Flächeninhaltes unter der „Kurve" $y = f(x)$ von a bis b geeignet sein. Dies ist in der Tat richtig.

Wie jede reelle Funktion, so läßt sich auch S_f auf $^*\mathbb{R}$ fortsetzen (und zwar unter Beibehaltung ihrer nach § 3 ausdrückbaren Eigenschaften). Ist für jedes positive $h \in \mathfrak{M}$ der Wert $^*S_f(a, b, h)$ endlich und besitzt immer den gleichen Standardteil c, so nennen wir c das *Integral* der Funktion f von a bis b. Es läßt sich ohne große Mühe nachweisen, daß c gleich dem Riemann-Integral von f über $[a, b]$ ist, falls dieses im üblichen Sinne existiert.

Wir wollen jetzt den folgenden Satz (wenigstens teilweise) nachweisen.

Satz. *Ist die Funktion f in dem reellen Intervall $[a, b]$ stetig, so existiert ihr Integral von a bis b.*

Beweis. Wir zeigen zuerst, daß $^*S_f(a, b, h)$ endlich ist für $h \in \mathfrak{M} \setminus \{0\}$. Da f auf $[a, b]$ in $\mathbb{R}$ stetig ist, ist nach dem obigen Lemma $|f|$ beschränkt. Für $h \in \mathbb{R}^+$ mit $h \leqslant b - a$ gilt also

$$|S_f(a, b, h)| \leqslant \sum_{i=1}^{n} |f(a_i)| \cdot |h| + |f(a_{n+1})|(b - a_{n+1}) \leqslant (b-a)c,$$

wobei $c \in \mathbb{R}^+$ eine obere Schranke für $|f|$ auf $[a, b]$ ist. Also haben wir

$$\mathbb{R} = \forall h\, (0 < h \leqslant (\underline{b} - \underline{a}) \to |S_f(\underline{a}, \underline{b}, h)| \leqslant (\underline{b} - \underline{a})\underline{c}).$$

Die gleiche Aussage gilt in $^*\mathbb{R}$. Also ist $^*S_f(a, b, h)$ endlich für jedes positive $h \in \mathfrak{M}$.

Der Nachweis der Unabhängigkeit von h ist etwas schwieriger; wir überlassen ihn dem interessierten Leser. □

Sei wieder f in $[a, b]$ stetig. Dann ist auch f in $[a, x]$ stetig für jedes $x \in [a, b]$. Wir setzen

$$I(a, x) = \mathrm{st}({}^*S_f(a, x, h)),$$

wobei h ein beliebiges positives Element aus $\mathfrak{M}$ ist. Wir wollen jetzt die Additivität des Integrals zeigen, das heißt, für $\varepsilon \in \mathbb{R}^+$ mit $x + \varepsilon \in [a, b]$ zeigen wir

$$I(a, x) + I(x, x + \varepsilon) = I(a, x + \varepsilon).$$

Dies folgt sofort aus der Beziehung

$${}^*S_f(a, x, h) + {}^*S_f(x, x + \varepsilon, h) = {}^*S_f(a, x + \varepsilon, h),$$

wobei wir wegen der Unabhängigkeit des Standardteils $h \in \mathfrak{M}$ passend positiv wählen können. In der Tat gilt diese Beziehung für $h = (x - a)/\omega$, da die Beziehung

$$S_f(a, x, h) + S_f(x, x + \varepsilon, h) = S_f(a, x + \varepsilon, h)$$

in $\mathbb{R}$ für $h = (x - a)/n$ mit jedem hinreichend großen $n \in \mathbb{N}$ gilt.

Es ist nun leicht, den Hauptsatz der Differential- und Integralrechnung zu beweisen.

Hauptsatz. *Ist f eine im Intervall $[a, b]$ stetige Funktion, so ist $F(x) = I(a, x)$ eine Stammfunktion von f, das heißt, für $x \in (a, b)$ gilt $F'(x) = f(x)$.*

Beweis. Für $dx \in \mathfrak{M} \setminus \{0\}$ ist

$$\frac{{}^*F(x + dx) - F(x)}{dx} \approx f(x)$$

zu zeigen. Wegen der Additivität von I, die sich natürlich auf ${}^*\mathbb{R}$ überträgt, bedeutet dies

$$\frac{{}^*I(x, x + dx)}{dx} \approx f(x),$$

wobei wir für $dx < 0$ unter ${}^*I(x, x + dx)$ natürlich $-{}^*I(x + dx, x)$ verstehen wollen. Dies zeigen wir nun so: Nach dem Übertragungsprinzip besitzt ${}^*f$ auf dem Intervall $[x, x + dx]$ in ${}^*\mathbb{R}$ ein Maximum c_1 und ein Minimum c_2, und es gilt (für positives dx)

$$c_2\, dx \leqslant {}^*I(x, x + dx) \leqslant c_1\, dx.$$

Also folgt

$$f(x_2) = c_2 \leqslant \frac{{}^*I(x, x + dx)}{dx} \leqslant c_1 = f(x_1),$$

wobei x_1, x_2 passende Elemente des Intervalls $[x, x + dx]$ in ${}^*\mathbb{R}$ sind. Wegen der Stetigkeit von f gilt aber $f(x_1) \approx f(x) \approx f(x_2)$. Dies ergibt die Behauptung. □

Epilog

Wir wollen hier noch kurz auf drei Punkte in Zusammenhang mit der dargestellten Einführung des Non-Standard Zahlbereiches $^*\mathbb{R}$ eingehen: die Einzigkeit von $^*\mathbb{R}$, Erweiterungen des Rahmens und andere Zugänge.

Einzigkeit von $^*\mathbb{R}$. Setzt man die Kontinuumshypothese $2^{\aleph_0} = \aleph_1$ voraus, so folgt aus allgemeinen Sätzen der Modelltheorie (siehe etwa [4], Kapitel 5, Korollar 23.6), daß der angeordnete Körper $^*\mathbb{R}$ bis auf Isomorphie eindeutig festgelegt ist. Dies meint, daß $^*\mathbb{R}$ nicht von der Wahl des maximalen Ideals M in R abhängt, solange es über dem Ideal D liegt. In diesem Falle ist der resultierende Ultrafilter U_M nämlich nicht-trivial (in [4] als „frei" bezeichnet). Wählt man dagegen ein maximales Ideal M mit $D \not\subset M$, so wird U_M zu einem sogenannten Hauptultrafilter, was $R/M \simeq \mathbb{R}$ zur Folge hat.

Die Unabhängigkeit (bis auf Isomorphie) von der Wahl des maximalen Ideals $M \supset D$ gilt jedoch nicht mehr für die Fortsetzungen aller reellen Funktionen auf $^*\mathbb{R}$.

Erweiterung des Rahmens. Das in § 3 bewiesene Prinzip erlaubt die Übertragung gewisser Eigenschaften von $\mathbb{R}$ nach $^*\mathbb{R}$. Der Rahmen dieser Übertragung ist durch die dort betrachtete formale Sprache festgelegt. Dieser Rahmen ist in gewisser Weise willkürlich gewählt. Insbesondere haben wir ihn hier so einfach wir nur möglich gewählt. Er läßt sich wesentlich erweitern. Bei einer Erweiterung kann und wird in der Regel eine bisher unbemerkt gebliebene Schwierigkeit neu auftreten. Wir wollen dies kurz erläutern.

Zu jeder Teilmenge A von $\mathbb{R}$ gibt es (analog zu $^*\mathbb{N}$) eine Erweiterung $^*A$ in $^*\mathbb{R}$. Nicht jede Teilmenge von $^*\mathbb{R}$ ist jedoch von dieser Gestalt. Mehr noch: erweitert man die formale Sprache so, daß eine Quantifikation über alle Teilmengen von $\mathbb{R}$ möglich wird, so läuft die Quantifikation, interpretiert in $^*\mathbb{R}$, nicht mehr über *alle* Teilmengen, sondern nur noch über die sogenannten „internen" Teilmengen von $^*\mathbb{R}$. So ist z. B. zwar $^*\mathbb{N}$ eine interne Teilmenge, jedoch nicht $\mathbb{N}$. Dies kann man leicht einsehen, wenn man etwa die folgende Aussage in dem erweiterten Rahmen formalisiert:

> „jede Teilmenge, die 0 und mit x auch $x + 1$ enthält, überschreitet jedes Element."

In $\mathbb{R}$ ist diese Aussage selbstverständlich richtig. In $^*\mathbb{R}$ interpretiert kann sie sich nicht auf alle Teilmengen beziehen, da $\mathbb{N}$ zwar die Voraussetzungen erfüllt, jedoch nicht jedes Element von $^*\mathbb{R}$ überschreitet.

Andere Zugänge. In dem hier ausgeführten Zugang zur Non-Standard Analysis haben wir den Bereich $^*\mathbb{R}$ aus dem schon vorhandenen Bereich $\mathbb{R}$ konstruiert. Dieser Zugang entspricht der Konstruktion der reellen Zahlen aus den rationalen mit Hilfe von Folgen. Eine andere Möglichkeit ist – analog zur axiomatischen Einführung der reellen Zahlen (die dann die rationalen als Teilmenge enthalten) –, auch $^*\mathbb{R}$ axiomatisch einzuführen. $\mathbb{R}$ ist dann eine ausgezeichnete Teilmenge. Diesen Zugang findet man etwa in dem Buch „Elementary Calculus" von

H. J. Keisler ausgeführt [2]. Während bei Keisler die Axiomatik speziell auf $^*\mathbb{R}$ abgestimmt ist, wird in [3] von E. Nelson eine sehr viel allgemeinere, mengentheoretische Axiomatik gewählt.

Literatur

[1] Edwards, C. H., Jr.: The historical development of the calculus. Springer-Verlag, New York-Heidelberg-Berlin 1979

[2] Keisler, H. J.: Elementary calculus. Prindle, Weber & Schmidt, Incorporated, Boston 1976

[3] Nelson, E.: Internal set theory: a new approach to non-standard analysis. Bull. of the Amer. Math. Soc. 83, 1165–1198 (1977)

[4] Potthoff, K.: Einführung in die Modelltheorie und ihre Anwendungen. Wiss. Buchges., Darmstadt 1981

[5] Robinson, A.: Non-standard analysis. North-Holland Publ. Comp., Amsterdam, London 1966

[6] Skolem, Th.: Über die Nichtcharakterisierbarkeit der Zahlreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschließlich Zahlvariablen. Fund. Math. 23, 150–161 (1934)

Kapitel 12. Zahlen und Spiele

H. Hermes

§ 1. Einleitung

In diesem vorletzten Kapitel des ersten Bandes soll eine neue Methode vorgestellt werden, mit der reelle Zahlen eingeführt werden können. Sie wurde in den siebziger Jahren von dem englischen Mathematiker J. H. CONWAY publiziert. Anders als in den vorangehenden Kapiteln wird im folgenden keine systematische Darstellung gegeben. Es sollen vielmehr vor allem die Ideen dargelegt werden, auf welchen die Conwaysche Theorie beruht. Einzelheiten der technischen Durchführung findet man in Conways Buch [1] sowie – in einer „popularisierten" Version – in [4].

1. Der traditionelle Aufbau der reellen Zahlen. Hier werden nur einige charakteristische Züge hervorgehoben (eine ausführliche Darstellung wird in den Kapiteln 1 und 2 gegeben). Grundlage ist die Mengenlehre. Die reellen Zahlen werden schrittweise aufgebaut. Es gibt dabei mehrere Varianten, die sich jedoch nicht grundsätzlich voneinander unterscheiden. Eine dieser Varianten führt in drei Schritten zum Ziel:

Im ersten Schritt werden die natürlichen Zahlen eingeführt. Bei der von Neumannschen Konstruktion wird die Zahl 0 mit der leeren Menge $\emptyset$ identifiziert und die Zahl $n + 1$ mit der Menge $n \cup \{n\}$ (vgl. 13.1.3 und 13.2.1).

Im zweiten Schritt (in Kap. 1 in zwei Teilschritte zerlegt) werden die rationalen Zahlen als Klassen von Tripeln von natürlichen Zahlen aufgefaßt; zur Klasse $-\frac{2}{3}$ gehört z. B. das Tripel (13, 17, 6), da $-\frac{2}{3} = (13\text{–}17)/6$. Die rationalen Zahlen bilden einen geordneten Körper.

Der dritte Schritt führt von den rationalen zu den reellen Zahlen. Eine reelle Zahl ist ein *Dedekindscher Schnitt* [2] (vgl. Kap. 2.2), das heißt, ein (geordnetes) Paar (x_1, x_2), wobei x_1 und x_2 Mengen von rationalen Zahlen sind. Es ist üblich, x_1 die Unterklasse und x_2 die Oberklasse des Schnittes (x_1, x_2) zu nennen. Wir folgen hier Conway und sprechen von der *linken Klasse* x_1 und der *rechten Klasse* x_2 des Schnittes (x_1, x_2).

An einen Schnitt (x_1, x_2) stellt DEDEKIND vier Forderungen:

(D1) *Jede rationale Zahl liegt in genau einer der Klassen* x_1, x_2.
(D2) x_1 *und* x_2 *sind nicht leer.*
(D3) *Jedes Element von* x_1 *ist kleiner als jedes Element von* x_2.
(D4) x_1 *hat kein größtes Element.*

Bei dem dreistufigen Aufbau der reellen Zahlen werden die Rechenoperationen dreimal erklärt, und es muß jede Stufe in die folgende isomorph eingebettet werden.

2. Die Conwaysche Methode. Auch hier geht man von der Mengenlehre aus. Die reellen Zahlen werden in einem einzigen Schritt gewonnen. Dabei verwendet Conway zwei Ideen. Die erste Idee besteht darin, die Dedekindschen Schnitte in geeigneter Weise zu verallgemeinern. Dabei ist es zunächst nicht klar, wie für die verallgemeinerten Schnitte die Ordnung zu definieren ist. Hier hilft die zweite Idee Conways. Er sieht nämlich, daß die von ihm verallgemeinerten Schnitte als Spiele zwischen zwei Personen aufgefaßt werden können, und daß die Theorie solcher Spiele einen Schlüssel zur Definition der Ordnung liefert.

Zu den Vorzügen von Conways Methode gehört, daß sie den schrittweisen Aufbau der reellen Zahlen und die damit zusammenhängenden ermüdenden Wiederholungen vermeidet. Ein weiterer Vorzug kann darin gesehen werden, daß der Zahlbegriff mit dem Spielbegriff in Verbindung gebracht wird. So schlägt Conway eine Brücke zu ältesten Erfahrungen der Menschheit und zu frühesten Erfahrungen jedes einzelnen Individuums (vgl. [3]). Jede solche Verbindung ist wertvoll für eine Wissenschaft wie die Mathematik mit der Tendenz zu immer größerer Abstraktion.

Es wird hier keineswegs behauptet oder auch nur vermutet, daß Conways Methode den traditionellen Aufbau der reellen Zahlen verdrängen wird. Man kann nämlich nicht leugnen, daß diese Methode neben den angedeuteten Vorzügen auch Schattenseiten hat. Dazu gehört der oft langwierige Nachweis für die Gültigkeit der Rechenregeln. Ferner liefert das Conwaysche Verfahren primär nicht nur die reellen Zahlen, sondern einen geordneten Zahlkörper, der den Körper der reellen Zahlen echt umfaßt. Die dabei auftretenden „Nicht-Standard-Zahlen" sind entweder unendlich groß oder unendlich klein oder unendlich benachbart zu einer reellen Zahl (vgl. Kap. 11). Will man zu den reellen Zahlen kommen, so muß man diese aus dem mit der Conwayschen Methode gewonnenen Oberkörper aussondern (vgl. § 8.3).

3. Übersicht. In § 2 werden die Dedekindschen Postulate (D1)–(D4) im Hinblick auf die beabsichtigte Verallgemeinerung von Conway diskutiert, und es wird der Begriff des Conwayspiels eingeführt. Conwayspiele lassen sich als Spiele auffassen; der hier relevante Spielbegriff wird in § 3 definiert. Einige grundlegende Sätze über solche Spiele werden auf der Basis des Begriffs der Gewinnstrategie in § 4 bewiesen. In § 5 wird gezeigt, daß die Spiele (modulo einer Äquivalenzrelation $=$) eine halbgeordnete Gruppe bilden. Wir schließen die Spieltheorie ab in § 6 mit der Erkenntnis, daß man die Conwayspiele als „Normalformen" der (hier betrachteten) Spiele ansehen kann.

Die beiden grundlegenden Postulate (C1), (C2) von Conway werden in § 7 formuliert. Sie sind motiviert durch die Überlegungen in § 2 und verwenden die für Spiele eingeführte Halbordnung. § 8 endlich enthält die Definition der Rechenoperationen für den geordneten Körper der Conwayzahlen und schließt mit einem kurzen Überblick über Conways Ergebnisse.

Eine *Warnung*: Obwohl die Grundideen von Conways Theorie recht einfach und einleuchtend sind, erweist sich ihre genaue Durchführung – auf welche hier meist verzichtet wird – des öfteren als ziemlich mühsam (vgl. etwa 8.2) oder als nichttrivial (wie etwa der Nachweis für die in 8.3 behauptete reelle Abgeschlossenheit).

§ 2. Conwayspiele

Poesis doctrinae tamquam somnium.
(Poesie ist wie der Traum einer Wissenschaft) Francis BACON

Die erste Idee von Conway besteht – wie bereits in 1.2 bemerkt – darin, die Dedekindschen Schnitte zu verallgemeinern. Wir wollen hier Dedekinds Postulate (D1)–(D4) näher betrachten, um das für die Verallgemeinerung Wichtige herauszuschälen. Wir kommen schließlich zu der Definition der Conwayspiele. Diese Definition kann als Vorstufe für die (D1)–(D4) entsprechenden Conwayschen Postulate (C1), (C2) (siehe § 7) angesehen werden.

1. Diskussion der Dedekindschen Postulate. (D4) soll verhindern, daß eine rationale Zahl r durch die beiden verschiedenen Mengenpaare

(Menge der rationalen Zahlen $\leqslant r$, Menge der rationalen Zahlen $> r$), und
(Menge der rationalen Zahlen $< r$, Menge der rationalen Zahlen $\geqslant r$)

dargestellt werden. Läßt man zu, daß eine reelle Zahl durch verschiedene Mengenpaare gegeben werden kann, so wird (D4) überflüssig.

(D2) verbietet z. B. das Mengenpaar

(Menge aller rationalen Zahlen, leere Menge).

Eine durch dieses Paar gegebene „Zahl" wäre – anschaulich gesprochen – eine positive unendliche Zahl. Sie wäre sogar die einzige solche Zahl, womit man offenbar mit den Axiomen eines geordneten Körpers in Konflikt käme. Ein solcher Konflikt könnte möglicherweise vermieden werden, wenn man durch eine Verallgemeinerung der Dedekindschen Konstruktion unendlich viele positive unendlich große Zahlen erzeugen könnte. Solche Zahlen wären vor nicht allzu langer Zeit in der Mathematik nicht zugelassen worden, als eine Periode grundlagenkritischer Untersuchungen einen „horror infiniti" erzeugt hatte. Heute hat man jedoch den Schrecken vor „unendlichen" Objekten verloren (vgl. auch Kapitel 11).

(D1) impliziert, daß z. B. die rechte Klasse eines Schnittes durch die linke eindeutig bestimmt ist. Logisch einfacher wäre es daher, stets nur mit der linken Klasse zu operieren und auf Dedekinds „poetische" Auffassung einer reellen Zahl als Mengen*paar* zu verzichten. Conway nimmt jedoch Dedekinds „Dichtung" ernst: Bei ihm bestimmen sich die linke und die rechte Klasse eines Mengenpaares, welches eine Zahl erzeugt, nicht gegenseitig. Conway muß also (D1) verwerfen. Damit wird es möglich, z. B. die reelle Zahl 0 auch durch das Mengenpaar

(Menge der rationalen Zahlen $-1/2^n$, Menge der rationalen Zahlen $1/2^n$)

darzustellen. Ferner definiert bei ihm aber auch z. B. das Mengenpaar $(\{0\}, \{1\})$ eine Zahl (vgl. dazu 8.2).

(D3) bleibt als letztes Postulat. Diese Forderung garantiert bei Dedekind, daß die reellen Zahlen einen total geordneten Bereich bilden. Entsprechendes fordert auch Conway. Formuliert man (D3) um, indem man $\leqslant$ an Stelle von $<$ verwendet, so erhält man die folgende Version, an die wir später anknüpfen

werden:

(D3′) Kein Element der rechten Klasse ist kleiner oder gleich einem Element der linken Klasse.

2. Conways Modifikation der Dedekindschen Postulate. Conway faßt wie Dedekind seine Zahlen als Mengenpaare (x, y) auf. Während jedoch Dedekind als Elemente von x und y nur rationale Zahlen zuläßt – die bereits früher konstruiert worden sind –, gestattet Conway bei der Bildung einer Zahl (x, y), daß als Elemente von x und y irgendwelche bereits „früher" nach seiner Methode konstruierte Zahlen auftreten können.

Die Paarbildung wird jedoch (wie bei Dedekind) eingeschränkt durch die Forderung (D3′). Hier entsteht ein Problem, auf welches bereits in 1.2 hingewiesen wurde: Bei Dedekind ist (D3) (bzw. (D3′)) sinnvoll, weil für die rationalen Zahlen bereits eine Ordnung definiert ist. Bei der beabsichtigten Conwayschen Verallgemeinerung muß man voraussetzen, daß für die Elemente von x und y bereits eine $\leqslant$-Beziehung definiert worden ist.

Solange man sich noch nicht vorstellen kann, wie eine solche Definition aussehen soll, liegt es nahe, zunächst auf die durch (D3′) ausgesprochene Einschränkung der Paarmengenbildung zu verzichten und zu untersuchen, welche Mengen man ohne sie erzeugen kann. Es ist zu erwarten, daß dann neben den Zahlen noch weitere Objekte erzeugt werden können. In § 3 werden wir sehen, daß man alle so erzeugbaren Objekte zwanglos als *Spiele* auffassen kann. Unter Vorwegnahme der dort gegebenen Erklärungen wollen wir daher die durch die Conwaysche Paarmengenbildung erzeugbaren Objekte *Conwayspiele* nennen.

Die Spieltheorie liefert in naheliegender Weise eine Halbordnung $\leqslant$ zwischen Spielen (§ 6). Diese Halbordnung wird schließlich bei der Formulierung der Einschränkung (D3′) verwendet (§ 7).

3. Conwayspiele. Gemäß den Erläuterungen in Abschnitt 2 wollen wir die Conwayspiele einführen durch das Postulat

(CS) *Wenn x und y Mengen von Conwayspielen sind, so ist das (geordnete) Paar (x, y) ein Conwayspiel.*

(CS) ist eine induktive Definition. Mit Hilfe bekannter Techniken der Mengenlehre könnte man (CS) in eine explizite Definition verwandeln. Es ist jedoch bequemer, mit der induktiven Definition zu arbeiten.

Bekannt sind induktive Definitionen z. B. aus der elementaren Zahlentheorie, wo z. B. die Addition sich durch die beiden Forderungen $x + 0 = x$ und $x + S(y) = S(x + y)$ induktiv definieren läßt (vgl. 1.2.3).

Einige *Beispiele* sollen erläutern, wie (CS) angewandt werden kann.

Wenn man bedenkt, daß zur Erzeugung eines Conwayspiels (x, y) die Elemente von x und y als Conwayspiele erzeugt worden sein müssen, so könnte man meinen, daß es überhaupt unmöglich wäre, mit (CS) Conwayspiele zu erzeugen. Dies wäre jedoch ein Fehlschluß: Wenn nämlich x und y leer sind, so sind trivialerweise x

und y Mengen von Conwayspielen (das heißt, es ist jedes Element von x und y ein Conwayspiel). Also ist gemäß (CS) das Mengenpaar $(\emptyset, \emptyset)$ ein Conwayspiel. Es wird sich später zeigen, daß man dieses Spiel mit der Zahl 0 identifizieren kann:

$$0 = (\emptyset, \emptyset). \tag{1}$$

Da 0 ein Conwayspiel ist, so ist $\{0\}$ eine Menge von Conwayspielen. Damit erhält man mit (CS) die Conwayspiele $(\{0\}, \emptyset)$, $(\emptyset, \{0\})$ und $(\{0\}, \{0\})$. Insbesondere sieht man, daß die folgenden Mengen Conwayspiele sind:

$$\begin{aligned} 1 &= (\{0\}, \emptyset), & 2 &= (\{0,1\}, \emptyset), \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ n+1 &= (\{0,\ldots,n\}, \emptyset), & \omega &= (\{0,1,2,\ldots\}, \emptyset). \end{aligned} \tag{2}$$

Man erkennt, daß die Methode, mit der von Neumann die Ordinalzahlen erzeugt hat (vgl. 13.1.3), auch Conwayspiele liefert, und hat daher

(3) *Alle Ordinalzahlen sind Conwayspiele.*

Um zu zeigen, daß eine Menge z ein Conwayspiel ist, hat man als einziges Postulat (CS) zur Verfügung. Daher muß z ein Mengenpaar (x, y) sein, wobei x und y Mengen von Conwayspielen sind. Die Elemente von x wollen wir *linke Elemente von* z und die Elemente von y *rechte Elemente von* z nennen. Wir haben also

(4) *Jedes Conwayspiel ist ein Mengenpaar. Die linken und die rechten Elemente eines Conwayspiels sind selbst Conwayspiele.*

§ 3. Spiele

Wir interessieren uns für eine spezielle Klasse von Spielen zwischen zwei Personen. Dazu gehören viele bekannte Spiele und, was hier besonders interessiert, alle Conwayspiele. Später (§ 6) werden wir sogar zeigen können, daß man jedem Spiel der hier betrachteten Klasse ein „gleiches" Conwayspiel zuordnen kann.

Wenn wir im folgenden von „Spielen" sprechen, meinen wir im allgemeinen Spiele der hier betrachteten Klasse.

1. Der Spielbegriff. Wir betrachten *Spiele* zwischen zwei Personen, dem *linken Spieler* L und dem *rechten Spieler* R. Vor einer *Partie* wird verabredet, welcher Spieler beginnt. Danach wird abwechselnd gezogen. Ein *Zug* führt von einer *Stellung* zu einer anderen. Es gibt ein *Menge* S *von Stellungen*, von denen eine als *Ausgangsstellung* s_0 ausgezeichnet ist. Zwischen den Stellungen bestehen zwei zweistellige *Spielrelationen* $\to_L$ und $\to_R$. Hat eine Partie zu einer Stellung s geführt, bei der z. B. der Spieler L am Zug ist, so besteht ein *Zug* von L darin, zu einer Stellung s' überzugehen, für welche $s \to_L s'$. Wenn es kein solches s' gibt, so hat L keinen Zug und damit verabredungsgemäß die Partie *verloren* (und R die Partie *gewonnen*). – Entsprechendes gilt für R.

Wir definieren

(1) $s \to s'$ *genau dann, wenn* $s \to_L s'$ *oder* $s \to_R s'$,

und stellen an den Spielbegriff die

Endlichkeitsforderung. *Es gibt keine unendliche Folge* $s_0, s_1, s_2, \ldots$ *von Stellungen derart, daß* $s_0 \to s_1 \to s_2 \to \cdots$.

Es folgt, daß jede Partie nach endlich vielen Zügen abbricht und daß dann einer der beiden Spieler die Partie gewonnen hat. Es gibt also keine unentschiedenen Partien. Es kann auch keine Partie geben, bei der nach endlich vielen Zügen wieder die Ausgangsstellung erreicht wird.

Ein Spiel ist durch die Menge S der Stellungen, die Anfangsstellung s_0 und die beiden Spielrelationen $\to_L$ und $\to_R$ gegeben, so daß wir es mit dem Tupel $(S, s_0, \to_L, \to_R)$ identifizieren können.

2. Beispiele für Spiele. Man überzeuge sich davon, daß es sich in den folgenden Beispielen wirklich um Spiele handelt, welche unter die in Abschnitt 1 gegebene Definition fallen.

(a) NIM-*Spiele*. Z. B. die folgende Version: Die Ausgangsstellung s_0 sei ein vorgegebenes m-Tupel $(N_1, \ldots, N_m)$ natürlicher Zahlen. Stellungen seien die m-Tupel $(n_1, \ldots, n_m)$ mit $n_i \leqslant N_i$ $(i = 1, \ldots, m)$. Zwischen zwei Stellungen $(n_1, \ldots, n_m)$ und $(n'_1, \ldots, n'_m)$ bestehe die Relation $\to_L$ *und* die Relation $\to_R$, falls $n_i = n'_i$ für alle i, abgesehen von *einem* Index i_0, für welchen $n'_{i_0} < n_{i_0}$. (Ein Zug besteht also darin, daß der ziehende Spieler von einem „Haufen" etwas wegnimmt.)

(b) DOMINO-*Spiele* der folgenden Art: Die Ausgangsstellung s_0 sei eine endliche Menge von Quadraten einer karierten Ebene. Die Stellungen seien die Teilmengen von s_0. Es gelte $s \to_L s'$ bzw. $s \to_R s'$, falls s' aus s entsteht durch Wegnahme zweier senkrecht bzw. waagrecht benachbarter Quadrate. (Dies läßt sich in praxi vollziehen durch Überdecken mit einem Dominostein.)

(c) *Conwayspiele*. Jedes Conwayspiel x läßt sich als ein Spiel auffassen. Die Ausgangsstellung s_0 wird mit x identifiziert. Stellungen seien neben der Ausgangsstellung x die linken und rechten Elemente von x, ferner deren linke und rechte Elemente, usf. Alle Stellungen sind also nach (2.4) selbst Conwayspiele. Es sei $s \to_L s'$ bzw. $s \to_R s'$ genau dann, wenn s' ein linkes bzw. rechtes Element von s ist. Es gilt die Endlichkeitsforderung: Gäbe es nämlich eine unendliche Folge $s_0 \to s_1 \to s_2 \to \cdots$ von Stellungen, so wäre dies eine unendliche Folge $s_0, s_1, s_2, \ldots$ von Mengen, bei der jede Menge linkes oder rechtes Element der vorangehenden wäre. Eine solche Folge kann es aber nach dem Fundierungsaxiom der Mengenlehre nicht geben (vgl. 13.2.2).

3. Ein Induktionsprinzip für Spiele. Gegeben sei ein Spiel $x = (S, s, \to_L, \to_R)$. Jedem s'_0 mit $s_0 \to s'_0$ ordnen wir ein Spiel $x' = (S', s'_0, \to'_L, \to'_R)$ zu wie folgt: Es sei $s \in S'$ genau dann, wenn es eine Kette $s'_0 \to \cdots \to s$ gibt ($s = s'_0$ ist dabei eingeschlossen). Für $s, s' \in S'$ sei $s \to'_L s'$ genau dann, wenn $s \to_L s'$. Entsprechend sei $s \to'_R s'$ definiert.

Jedes so erzeugte Spiel x' soll ein *Vorgängerspiel von* x heißen. Genauer sprechen wir von einem *linken* bzw. *rechten Vorgängerspiel*, wenn $s_0 \to_L s'$ bzw. $s_0 \to_R s'_0$.

Das Induktionsprinzip für Spiele befaßt sich mit einer Eigenschaft P, welche für Spiele erklärt ist. Px soll heißen, daß das Spiel x die Eigenschaft P hat.

Induktionsprinzip für Spiele. *Wenn aus der Induktionsvoraussetzung, daß P für jedes Vorgängerspiel x' von x gilt, die Induktionsbehauptung Px folgt, so hat jedes Spiel x die Eigenschaft P.*

Beweis. Wir nehmen an, daß aus der Induktionsvoraussetzung die Induktionsbehauptung folgt, daß es aber ein Spiel x_0 gibt, auf welches P nicht zutrifft. Dann gibt es ein Vorgängerspiel x_0' von x_0, auf das P nicht zutrifft, ein Vorgängerspiel x_0'' von x_0', auf das P nicht zutrifft, usf. Für die Ausgangsstellungen $s_0, s_0', s_0'', \ldots$ der Spiele $x_0, x_0', x_0'', \ldots$ gilt dann

$$s_0 \to s_0' \to s_0'' \to \cdots,$$

entgegen der Endlichkeitsforderung für Spiele. Dieser Widerspruch widerlegt die Annahme. □

§ 4. Zur Theorie der Spiele

> It signifies nothing to play well if you loose (Englisches Sprichwort).

Wir zeigen, daß in jedem Spiel entweder der Spieler L oder der Spieler R oder der beginnende („erste") Spieler oder der nicht beginnende („zweite") Spieler den Gewinn erzwingen kann. Hier spielt der Begriff der Gewinnstrategie eine entscheidende Rolle. Dieser Begriff kann insbesondere dazu verwendet werden, für Spiele die Eigenschaften „positiv" und „negativ" zu definieren. Dabei sind die natürlichen Zahlen, welche wir in § 2 bereits als Conwayspiele kennengelernt haben, in diesem Sinne positiv.

1. Gewinnstrategien. Fundamental für die Spieltheorie ist der Begriff der *Strategie*. Sei bei einer Partie im Spiel x der Spieler A (also $A = L$ oder $A = R$) am Zuge. Wenn A überhaupt einen Zug hat – sonst ist die Partie beendet und für A verloren –, so hat er im allgemeinen mehrere mögliche Züge. Eine *Strategie σ für A in x* schreibt in einem solchen Fall einen Zug eindeutig vor.

Der von einer Strategie vorgeschriebene Zug kann von dem bisherigen Partieverlauf abhängen. (Es wäre möglich, den Strategiebegriff so einzuengen, daß der vorgeschriebene Zug nur von der erreichten Stellung abhängt. Wir wollen jedoch von diesem einfacheren Strategiebegriff nicht Gebrauch machen, da wir in diesem Fall mehr beweisen müßten.)

Wir sagen, daß *der Spieler A im Spiel x eine Partie mit der Strategie σ* spielt, wenn σ eine Strategie für A in x ist und wenn A jeden seiner Züge gemäß der Vorschrift σ ausführt.

Bei dem Begriff der *Gewinnstrategie* unterscheiden wir, um welchen Spieler es sich handelt und welcher Spieler die Partie beginnt.

σ heiße eine Gewinnstrategie für L im Spiel x, falls R beginnt, genau dann, wenn σ eine Strategie für L in x ist und wenn L jede Partie gewinnt, bei der R beginnt, falls L mit der Strategie σ spielt.

LxR soll heißen, daß L eine Gewinnstrategie im Spiel x hat, falls R beginnt. Analog definiert man LxL, RxL und RxR.

Anschließend benötigen wir die beiden folgenden Hilfssätze:

(1) *x' sei ein rechtes Vorgängerspiel von x. Es gelte $Rx'L$. Dann gilt RxR.*

(2) *Für jedes rechte Vorgängerspiel x' von x sei $Lx'L$. Dann gilt LxR.*

Zu (1): σ' sei eine Gewinnstrategie für R in x', falls L beginnt. Eine Gewinnstrategie σ für R in x, falls R beginnt, erhält man so: Im ersten Zug gehe R über zu der Anfangsstellung von x'. Danach spiele R nach der Strategie σ'. Damit garantiert ihm die Voraussetzung über σ' den Gewinn.

Zu (2): Eine Gewinnstrategie σ für L in R, falls R beginnt, besteht darin, daß zu Beginn einer Partie der Spieler L zunächst R einen Zug machen läßt (falls R keinen Zug hat, gewinnt L sofort). Dieser Zug führt zu einem rechten Vorgängerspiel x' von x, in welchem L am Zug ist. Nun kann L eine Gewinnstrategie σ' anwenden, die wegen der Voraussetzung $Lx'L$ existiert. □

Es sei bemerkt, daß aus Symmetriegründen neben den Aussagen (1), (2) (und späteren mit demselben Vokabular) auch die *dualen Aussagen* gelten, die man erhält, wenn man „L" mit „R" vertauscht und ebenso „linkes Vorgängerspiel" mit „rechtes Vorgängerspiel".

Wenn in einem Spiel etwa der Spieler R beginnt, so können nicht beide Spieler L und R eine Gewinnstrategie haben. Die folgende Aussage zeigt, daß immer wenigstens einer der beiden Spieler eine Gewinnstrategie besitzt. Es gilt nämlich für jedes Spiel x:

(3) *(LxR oder RxR) und (LxL oder RxL).*

Beweis mit dem Induktionsprinzip für Spiele. Nachweis der ersten Klammer: Wenn es ein rechtes Vorgängerspiel x' von x gibt mit $Rx'L$, so gilt RxR nach (1), also die Behauptung. *Sonst* gilt für jedes rechte Vorgängerspiel x' von x die Aussage *nicht* $Rx'L$, also nach der Induktionsvoraussetzung die Aussage $Lx'L$, und damit LxR nach (2), wie behauptet. – Die zweite Klammer ergibt sich analog mit den zu (1), (2) dualen Aussagen. □

2. Positive und negative Spiele. Wenn zu Beginn eines Spieles x der Spieler R keinen Zug hat, so ist trivialerweise LxR. Dies trifft auf alle in 2.3(2) genannten Conwayspiele zu. Alle diese Zahlen sind positiv (im Sinne von $\geqslant 0$). Diese Beispiele motivieren die Einführung einer Eigenschaft „$0 \leqslant$" durch die

Definition. $0 \leqslant x$ *genau dann, wenn* LxR.

Dual zu „$\geqslant 0$" führen wir eine Eigenschaft „negativ", kurz „$\leqslant 0$" ein durch die

Definition. $x \leqslant 0$ *genau dann, wenn* RxL.

Mit Hilfe dieser Definition lassen sich unter Berücksichtigung von (3) die Aussagen (1) und (2) umformulieren. Wir folgen dabei Conway und verwenden x^L bzw. x^R *als Variablen für die linken bzw. rechten Vorgängerspiele von* x. Wir erhalten so:

(1′) *Wenn ein* $x^R \leqslant 0$, so nicht $0 \leqslant x$.

(2′) *Wenn für alle* x^R nicht $x^R \leqslant 0$, so $0 \leqslant x$.

Durch Zusammenfassung dieser beiden Aussagen erhält man gemeinsame induktive Charakterisierungen von „$0 \leqslant$“ und „$\leqslant 0$“, nämlich

(4) $0 \leqslant x$ genau dann, wenn für alle x^R *nicht* $x^R \leqslant 0$,

sowie dual dazu:

(5) $x \leqslant 0$ genau dann, wenn für alle x^L *nicht* $0 \leqslant x^L$.

3. Eine Einteilung der Spiele. Gleichwertigkeit von Spielen. Wendet man auf (3) das distributive Gesetz der booleschen Operation *und* an, so sieht man, daß für jedes Spiel x gilt:

(LxR und LxL) oder (LxR und RxL) oder (RxR und LxL)

oder (RxR und RxL).

Wenn die erste Klammer gilt, so hat L eine Gewinnstrategie für das Spiel x, wenn L beginnt, und auch eine Gewinnstrategie, wenn R beginnt. Wir wollen sagen, daß ein solches Spiel zur *Klasse* **L** gehört. Entsprechend rechnen wir ein Spiel zu *Klasse* **R**, wenn die letzte Klammer gilt.

Wenn die dritte Klammer zutrifft, so hat der Spieler, der *beginnt*, eine Gewinnstrategie, also der *erste* Spieler. Solche Spiele rechnen wir zur Klasse **E**.

Wenn schließlich die zweite Klammer gilt, so hat der Spieler, der *nicht beginnt*, eine Gewinnstrategie, also der *zweite* Spieler. Solche Spiele rechnen wir zur Klasse **Z**.

Man sieht leicht, daß kein Spiel in zwei verschiedenen Klassen liegen kann, so daß wir zusammenfassend haben:

(6) *Jedes Spiel liegt in genau einer der Klassen* **L, R, E, Z**.

Definition. Wir nennen Spiele *gleichwertig*, wenn sie in derselben Klasse liegen.

Beispiele. Die Dominospiele mit den Anfangsstellungen  bzw. bzw.  bzw.   liegen jeweils in den Klassen **L, R, E, Z**, wie man leicht sieht. D_n sei ein Dominospiel, dessen Ausgangsstellung aus einem Quadrat der Seitenlänge n besteht. D_0, D_1, D_5 gehören zu **Z** und D_2, D_3, D_4 zu **E**. Das in § 2(1) definierte Conwayspiel $0 = (\emptyset, \emptyset)$ gehört zu **Z**, da kein Spieler in der Ausgangsstellung einen Zug hat.

Man kann natürlich die genannten Klassen auch mit Hilfe von $\leqslant 0$ und $\geqslant 0$ kennzeichnen. Damit hat man:

(a) $x \in \mathbf{Z}$ *genau dann, wenn* $x \leqslant 0$ *und* $0 \leqslant x$,
(b) $x \in \mathbf{L}$ *genau dann, wenn* $0 \leqslant x$ *und* $x \not\leqslant 0$,
(c) $x \in \mathbf{R}$ *genau dann, wenn* $x \leqslant 0$ *und* $0 \not\leqslant x$,
(d) $x \in \mathbf{E}$ *genau dann, wenn* $x \not\leqslant 0$ *und* $0 \not\leqslant x$.

Man hat insbesondere $0 \in \mathbf{Z}$, also im Sinne der beiden Definitionen für $0 \leqslant x$ und $x \leqslant 0$ die Aussage

(7) $0 \leqslant 0$.

Definiert man $0 < x$ durch $0 \leqslant x$ *und* $x \not\leqslant 0$, so sieht man, daß in **L** genau die (echt) positiven Spiele liegen. In **R** liegen entsprechend die (echt) negativen Spiele.

§ 5. Eine halbgeordnete Gruppe äquivalenter Spiele

Im letzten Paragraphen haben wir die Spieleigenschaften „positiv" und „negativ" eingeführt. Statt x *ist positiv* haben wir auch geschrieben: x *hat die Eigenschaft* „$\leqslant 0$" oder „$0 \leqslant$"x oder noch kürzer $0 \leqslant x$; entsprechend haben wir für x *ist negativ* geschrieben: x hat die Eigenschaft „$\leqslant 0$", x„$\leqslant 0$", $x \leqslant 0$. Die Schreibweisen $0 \leqslant x$ und $x \leqslant 0$ suggerieren, daß x mit dem Spiel 0 verglichen wird, wovon aber in den Definitionen dieser Eigenschaft nicht die Rede war.

In diesem Paragraphen wollen wir eine *zweistellige* Relation $\leqslant$ zwischen Spielen einführen und zeigen, daß „$0 \leqslant$"x genau dann, wenn $0 \leqslant x$, und daß x„$\leqslant 0$" genau dann, wenn $x \leqslant 0$.

Dazu werden wir für Spiele zwei Operationen $-x$ und $x + y$ definieren. Wir erklären dann $x \leqslant y$ durch „$0 \leqslant$" $y - x$, wobei $y - x$ wie üblich eine Abkürzung für $y + (-x)$ ist.

Die Relation $\leqslant$ ist (übrigens neben $-$ und $+$) ein Beitrag der Spieltheorie zu Conways Theorie der Zahlen. Es ist die Relation, welche wir in § 2 vermißt haben.

Die Relation, welche zwischen zwei Spielen x und y besteht, wenn sowohl $x \leqslant y$ als auch $y \leqslant x$, ist eine mit $\leqslant$, $-$ und $+$ verträgliche Äquivalenzrelation. Der Übergang zu den Kongruenzklassen liefert eine halbgeordnete abelsche Gruppe mit dem Nullelement **Z**.

1. Das Negative eines Spiels. Das Negative eines Spiels

$$x = (S, s_0, \rightarrow_L, \rightarrow_R)$$

sei das Spiel

$$-x = (S, s_0, \rightarrow_R, \rightarrow_L),$$

also das Spiel, welches aus x dadurch entsteht, daß man die Spielrelationen für R und L miteinander vertauscht. Offenbar gilt:

(1) $$-(-x) = x, \qquad -0 = 0,$$

wobei man (vgl. § 2, 3(1)) das Conwayspiel 0 als das Spiel mit der einzigen Stellung $(\emptyset, \emptyset)$ aufzufassen hat, in welchem keiner der beiden Spieler einen Zug hat.

(2) *Wenn* $0 \leqslant x$, *so* $-x \leqslant 0$ (und umgekehrt).

Beweis. Es ist zu zeigen, daß $R(-x)L$, falls LxR. Dies folgt aus der Bemerkung, daß eine Gewinnstrategie für L in x, falls R beginnt, auch eine Gewinnstrategie für R in $-x$ ist, falls L beginnt. □

2. Die Summe zweier Spiele. Zunächst ein Beispiel: x_1 sei ein NIM-Spiel und x_2 ein Dominospiel. Dann soll $x_1 + x_2$ das Spiel sein, welches man dadurch erhält, daß man die Spiele x_1 und x_2 *simultan* spielt. Dazu verabredet man, daß der am Zug befindliche Spieler nach eigener Wahl einen Zug in x_1 oder einen Zug in x_2 machen kann.

Allgemein definieren wir: Ist

$$x_i = (S_i, s_{0i}, \to_{Li}, \to_{Ri}) \qquad (i = 1,2).$$

Dann sei

$$x_1 + x_2 = (S, s_0, \to_L, \to_R),$$

wobei $S = S_1 \times S_2$ die Menge der Paare der Stellungen der Spiele x_1, x_2 ist, s_0 das Paar (s_{01}, s_{02}) und

$$(s_1, s_2) \to_L (s_1', s_2')$$

genau dann, wenn

$$(s_1 \to_{L1} s_1' \text{ und } s_2 = s_2') \qquad \text{oder} \qquad (s_1 = s_1' \text{ und } s_2 \to_{L2} s_2').$$

(Entsprechend sei $\to_R$ definiert.) Offenbar gilt:

(3) $$-(x + y) = -x - y \qquad (= -x + (-y)).$$

Ferner hat man:

(4) a) $0 \leqslant x - x$ *und* $x - x \leqslant 0$.
b) *Wenn* $0 \leqslant x$ *und* $0 \leqslant y$, *so* $0 \leqslant x + y$.
c) *Wenn* $0 \leqslant x + y$ *und* $y \leqslant 0$, *so* $0 \leqslant x$.

Beweis. a) $0 \leqslant x - x$ heißt, daß $L(x - x)R$. Wenn R beginnt, so kann L im Spiel $x - x$ gewinnen, wenn L jeden Zug von R in der anderen Komponente übernimmt. Die zweite Behauptung folgt dual.

b) Sei LxR und LyR. Es ist zu zeigen, daß $L(x + y)R$. Man erhält eine Gewinnstrategie für L im Spiel $x + y$, falls R beginnt, durch die Vorschrift, daß L auf jeden Zug von R in derselben Komponente antwortet, in der R gezogen hat, und zwar so, wie es eine nach Voraussetzung für diese Komponente vorliegende Gewinnstrategie verlangt.

c) Wir zeigen, daß aus $0 \not\leqslant x$ und $y \leqslant 0$ die Aussage $0 \not\leqslant x + y$ folgt. Wegen (4.3) genügt dazu der Nachweis für

Wenn RxR *und* RyL, *so* $R(x + y)R$.

R zieht zunächst in der Komponente x, wo R eine Gewinnstrategie hat. Danach zieht R jeweils in der vom Gegner gewählten Komponente gemäß einer dort für R vorliegenden Gewinnstrategie. □

3. Isomorphe Spiele. Für Spiele kann man den Isomorphiebegriff nach dem üblichen Muster einführen.

Man sieht leicht, daß das Spiel $x + y$ isomorph zu $y + x$ ist und das Spiel $(x + y) + z$ isomorph zu $x + (y + z)$.

Ist y isomorph zu x und gilt LxR, so natürlich auch LyR, usf.

Ein *Beispiel*: Das Dominospiel mit der Ausgangsstellung  ist isomorph zur Summe der Dominospiele mit den Ausgangsstellungen  und .

4. Eine Halbordnung der Spiele.

Definition. $x \leqslant y$ *genau dann, wenn* $0 \leqslant y - x$ (wobei auf der rechten Seite natürlich die in § 4 eingeführte Eigenschaft „$0 \leqslant$“ gemeint ist).

Wir wollen zeigen, daß $0 \leqslant y$ genau dann, wenn „$0 \leqslant$“y (vgl. die einleitenden Bemerkungen). (Entsprechend beweist man, daß $x \leqslant 0$ genau dann, wenn x„$\leqslant 0$“.) Es ist zu beweisen, daß für die Eigenschaft $0 \leqslant$ gilt:

$$0 \leqslant y - 0 \qquad \text{genau dann, wenn} \qquad 0 \leqslant y.$$

Wenn $0 \leqslant y$, so folgt $0 \leqslant y - 0$ aus $0 \leqslant 0$ (4.7), $-0 = 0$ (1) und (4b). Wenn $0 \leqslant y - 0$, so folgt $0 \leqslant y$ aus $0 \leqslant 0$, $-0 = 0$ und (4c). □

$\leqslant$ *ist eine Halbordnung.* Die *Reflexivität* ergibt sich aus (4a). Es bleibt die *Transitivität*: Sei $x \leqslant y$ und $y \leqslant z$. Daher hat man

$$0 \leqslant y - x \qquad \text{und} \qquad 0 \leqslant z - y$$

$$0 \leqslant (z - y) + (y - x) \qquad \text{(4b)}$$

$$0 \leqslant (z - x) + (y - y) \qquad \text{(Isomorphie)}$$

$$0 \leqslant z - x \qquad \text{wegen (4a) und (4c)}$$

und damit $x \leqslant z$.

(5) a) *Wenn* $x \leqslant y$, *so* $-y \leqslant -x$.
b) *Wenn* $x \leqslant y$, *so* $x + z \leqslant y + z$.

Beweis. a) Sei $x \leqslant y$. Es folgt $0 \leqslant y - x$, $0 \leqslant -x - (-y)$ (Isomorphie), $-y \leqslant -x$.

b) Sei $x \leqslant y$. Es folgt $0 \leqslant y - x$, $0 \leqslant (y - x) + (z - z)$ (4b), $0 \leqslant (y + z) - (x + z)$ (Isomorphie) und damit die Behauptung. □

(6) *Nie* $x^R \leqslant x$ *und nie* $x \leqslant x^L$.

(Zu den Bezeichnungen x^R und x^L vgl. 4.2.) Wir zeigen die erste Behauptung (die zweite ist dazu dual). $x^R - x^R$ ist ein rechtes Vorgängerspiel von $x - x^R$. Nach (4a) hat man $R(x^R - x^R)L$. Aus § 4(1) folgt nun $R(x - x^R)R$, also *nicht* $L(x - x^R)R$, das heißt, *nicht* $x^R \leqslant x$. □

In § 4 haben wir die Eigenschaft „$\leqslant 0$“ induktiv gekennzeichnet. Eine entsprechende induktive Kennzeichnung gibt es für die zweistellige Relation $\leqslant$.

Satz. *$x \leqslant y$ genau dann, wenn* (a) *nie $y^R \leqslant x$ und* (b) *nie $y \leqslant x^L$.*

Beweis. Sei $x \leqslant y$. Zu (a): Wäre $x \leqslant y$ und $y^R \leqslant x$, so $y^R \leqslant y$ wegen der Transitivität, entgegen (6). Entsprechend zeigt man (b).

Sei nie $y^R \leqslant x$ und nie $y \leqslant x^L$, aber *nicht* $x \leqslant y$. Dann hätte man $R(y - x)R$. R hat also eine Gewinnstrategie für das Spiel $y - x$, falls R beginnt. Für den ersten Zug von R nach dieser Gewinnstrategie sind zwei Fälle denkbar:

(i) R zieht in der Komponente y. Dieser Zug liefert ein y^R, und es ist $R(y^R - x)L$, also $L(x - y^R)R$, das heißt, $y^R \leqslant x$ entgegen der Voraussetzung.

(ii) R zieht in der Komponente $-x$. Dieser Zug liefert ein rechtes Vorgängerspiel von $-x$, also ein linkes Vorgängerspiel x^L von x. Es ist $R(y - x^L)L$, also $L(x^L - y)^R$, das heißt, $y \leqslant x^L$ entgegen der Voraussetzung. □

5. Gleichheit von Spielen. Im Vorangehenden haben wir alle Eigenschaften von $\leqslant$ nachgewiesen, welche die zweistellige Beziehung *$x \leqslant y$ und $y \leqslant x$* als eine Äquivalenzrelation kennzeichnen, welche mit $\leqslant$, $-$ und $+$ verträglich ist. In der Bezeichnungsweise folgen wir nun Conway und nennen zwei Spiele *gleich* ($=$), wenn zwischen ihnen die genannte Beziehung besteht. *Man beachte, daß wir bisher unter der Gleichheit stets die logische Identität verstanden haben.* Falls eine Verwechslung zu befürchten ist, werden wir letztere von nun an durch das Symbol „$\equiv$“ kennzeichnen. – Wir haben also die

Definition. *$x = y$ genau dann, wenn $x \leqslant y$ und $y \leqslant x$.*

Wir ersparen uns die triviale Durchführung der *Klassenbildung modulo der Gleichheit* und die Übertragung von $\leqslant$, $-$ und $+$ auf diese Klassen und formulieren das Ergebnis in dem

Satz. *Die Klassen gleicher Spiele bilden in bezug auf $\leqslant$, $-$, $+$ eine halbgeordnete abelsche Gruppe mit dem Nullelement* **Z**.

Gleiche Spiele sind natürlich gleichwertig im Sinne von 4.3. Jede der Klassen **Z, L, R, E** zerfällt also in Klassen gleicher Spiele. Alle Spiele von **Z** sind untereinander gleich, die anderen Klassen zerfallen jedoch in mehrere (sogar unendlich viele) Klassen gleicher Spiele. So liegen z. B. die beiden Dominospiele x mit der Anfangsstellung □□ und y mit der Anfangsstellung □□ □□ offenbar beide in **R**. Sie sind aber nicht gleich. Es ist nämlich x isomorph zu einem y^R. Für ein solches y^R gilt trivialerweise $y^R \leqslant x$. Daraus kann man mit dem Satz aus Abschnitt 4 schließen, daß nicht $x \leqslant y$, also $x \neq y$.

§ 6. Spiele und Conwayspiele

In 3.2 haben wir gesehen, daß man jedes Conwayspiel c als Spiel auffassen kann. Dies bedeutet genauer, daß wir jedem Conwayspiel c ein Spiel $c_{\mathbf{S}}$ zugeordnet haben.

Wir wollen nun auch *umgekehrt* jedem Spiel x ein Conwayspiel $x_{\mathbf{C}}$ zuordnen und zeigen, daß dieses Conwayspiel $x_{\mathbf{C}}$, als Spiel aufgefaßt, also das Spiel $x_{\mathbf{CS}}$, gleich x ist, wobei die im letzten Paragraphen eingeführte Gleichheit gemeint ist.

Man könnte $x_{\mathbf{C}}$ als *die Normalform von* x bezeichnen. Conway legt seiner Theorie von vornherein die Normalformen zugrunde. Dies hat den Vorteil größerer mathematischer Einfachheit, allerdings auf Kosten der Anschaulichkeit.

Die beiden Abbildungen $c \mapsto c_{\mathbf{S}}$ und $x \mapsto x_{\mathbf{C}}$ ermöglichen es, die zunächst für Spiele erklärten Beziehungen $\leqslant$ und $=$ und Operationen $+$, $-$ kanonisch auf Conwayspiele zu übertragen.

1. Die grundlegenden Abbildungen. Zunächst wiederholen wir die im Prinzip bereits in 3.2 gegebene Definition von $c_{\mathbf{S}}$:

$$c_{\mathbf{S}} \equiv (S_c, c, \rightarrow_{c L}, \rightarrow_{c R}), \tag{1}$$

wobei die Stellungen von $c_{\mathbf{S}}$ neben der Ausgangsstellung c die linken und die rechten Elemente von c sind, ferner deren linke und rechte Elemente, usf. Es gilt $s \rightarrow_{c L} s'$ genau dann, wenn s, s' Stellungen sind und wenn s' linkes Element von s ist; entsprechend wird $\rightarrow_{c R}$ erklärt.

In 4.2 haben wir x^L, x^R als Variablen für die linken und rechten Vorgänger eines Spiels x eingeführt. Analog wollen wir c^L, c^R *als Variablen für die linken und rechten Elemente eines Conwayspiels* (vgl. 2.3) verwenden. Man verifiziert leicht:

(2) *Die* ${c_{\mathbf{S}}}^L$ *stimmen mit den* ${c^L}_{\mathbf{S}}$ *überein, ebenso die* ${c_{\mathbf{S}}}^R$ *mit den* ${c^R}_{\mathbf{S}}$.

Wir wollen nun jedem Spiel x ein Conwayspiel $x_{\mathbf{C}}$ zuordnen. Diese Zuordnung definieren wir induktiv, indem wir $x_{\mathbf{C}}$ erklären unter der Voraussetzung, daß für alle Vorgängerspiele z von x bereits $z_{\mathbf{C}}$ erklärt worden ist. (Man kann solche induktiven Definitionen „rechtfertigen" mit Hilfe des Induktionsprinzips für Spiele in 3.3.) Wir definieren also:

$$x_{\mathbf{C}} \equiv (\text{Menge der } {x^L}_{\mathbf{C}}, \text{ Menge der } {x^R}_{\mathbf{C}}). \tag{3}$$

Durch Induktion über Spiele sieht man sofort, daß $x_{\mathbf{C}}$ stets ein Conwayspiel ist. Aus (6.3) kann man unmittelbar ablesen:

(4) *Die* ${x_{\mathbf{C}}}^L$ *stimmen mit den* ${x^L}_{\mathbf{C}}$ *überein, ebenso die* ${x_{\mathbf{C}}}^R$ *mit den* ${x^R}_{\mathbf{C}}$.

(5) *Für jedes Conwayspiel* c *gilt* $c_{\mathbf{SC}} \equiv c$.

Zum *Beweis* benötigen wir ein dem Induktionsprinzip für Spiele analoges und analog beweisbares

Induktionsprinzip für Conwayspiele. Wenn aus der *Induktionsvoraussetzung*, daß Px' für jedes linke oder rechte Element x' eines beliebigen Conwayspieles x, die *Induktionsbehauptung* Px folgt, so hat jedes Conwayspiel x die Eigenschaft P.

Damit erhalten wir:

$$\begin{aligned} c_{\mathbf{SC}} &\equiv (\text{Menge der } {c_{\mathbf{S}}}^L{}_{\mathbf{C}}, \text{ Menge der } {c_{\mathbf{S}}}^R{}_{\mathbf{C}}) && (3)\\ &\equiv (\text{Menge der } {c^L}_{\mathbf{SC}}, \text{ Menge der } {c^R}_{\mathbf{SC}}) && (2)\\ &\equiv (\text{Menge der } c^L, \text{ Menge der } c^R) && (\text{Induktionsvoraussetzung})\\ &\equiv c. && \square \end{aligned}$$

(6) *Für jedes Spiel* x *gilt* $x = x_{\mathbf{CS}}$.

Wir beweisen, daß $x \leqslant x_{\mathbf{CS}}$ (der Beweis für $x_{\mathbf{CS}} \leqslant x$ ist analog). Wir verwenden dabei die in 5.4 gegebene induktive Kennzeichnung der Relation $\leqslant$, ferner (2), (4) und die Induktionsvoraussetzung.

$$\begin{aligned} x \leqslant x_{\mathbf{CS}} \quad & gdw \quad \text{kein} \quad {x_{\mathbf{CS}}}^R \leqslant x \quad \text{und} \quad x_{\mathbf{CS}} \leqslant \quad \text{kein} \quad x^L, \\ & gdw \quad \text{kein} \quad {x^R}_{\mathbf{CS}} \leqslant x \quad \text{und} \quad x_{\mathbf{CS}} \leqslant \quad \text{kein} \quad {x^L}_{\mathbf{CS}}, \\ & gdw \quad \text{kein} \quad x^R \leqslant x \quad \text{und} \quad x_{\mathbf{C}} \leqslant \quad \text{kein} \quad {x_{\mathbf{CS}}}^L, \end{aligned}$$

und die letzte Konjunktion gilt wegen § 5(6). □

2. Übertragung der für Spiele definierten Relationen und Operationen auf Conwayspiele. Zunächst definieren wir die Relation $\leqslant$ zwischen Conwayspielen c, c':

(7) $c \leqslant c'$ *genau dann, wenn* ${c_{\mathbf{S}}} \leqslant {c'_{\mathbf{S}}}$.

Analog zu den Spielen setzen wir $c = c'$, wenn $c \leqslant c'$ und $c' \leqslant c$.

Die Übertragung der für Spiele definierten Operationen $-$ und $+$ auf Conwayspiele erfolgt kanonisch durch die beiden folgenden Definitionen:

(8) $$-c \equiv (-{c_{\mathbf{S}}})_{\mathbf{C}},$$

(9) $$c_1 + c_2 \equiv ({c_{1\mathbf{S}}} + {c_{2\mathbf{S}}})_{\mathbf{C}}.$$

Man kann die eingeführte Relation $\leqslant$ und die eingeführten Operationen $-$, $+$ auch induktiv charakterisieren:

(7I) $c \leqslant c'$ *genau dann, wenn* (a) *nie* ${c'}^R \leqslant c$, *und*

(b) *nie* $c' \leqslant c^L$.

(8I) $$-c \equiv (\text{Menge der } -(c^R), \text{ Menge der } -(c^L)).$$

(9I) $$\begin{aligned} c_1 + c_2 \equiv (&\text{Menge der } (c_1^L + c_2) \cup \text{Menge der } (c_1 + c_2^L), \\ &\text{Menge der } (c_1^R + c_2) \cup \text{Menge der } (c_1 + c_2^R)). \end{aligned}$$

(7I) ergibt sich sofort aus der induktiven Charakterisierung der $\leqslant$-Beziehung zwischen Spielen mit Hilfe von (2).

Wir beweisen (8I). Nach (8) und (3) hat man

$$-c \equiv (\text{Menge der } {(-{c_{\mathbf{S}}})^L}_{\mathbf{C}}, \text{ Menge der } {(-{c_{\mathbf{S}}})^R}_{\mathbf{C}}).$$

Aus (8) erhält man mit (6) $(-c)_{\mathbf{S}} = -{c_{\mathbf{S}}}$ und damit

$$-c \equiv (\text{Menge der } {(-c)_{\mathbf{S}}^L}_{\mathbf{C}}, \text{ Menge der } {(-c)_{\mathbf{S}}^R}_{\mathbf{C}})$$

und hieraus mit (2) und (8)

$$-c \equiv (\text{Menge der } (-c)^L, \text{ Menge der } (-c)^R).$$

Damit hat man (8I), wenn man berücksichtigt, daß die Menge der $(-c)^L$ mit der Menge der $-(c^R)$ übereinstimmt, und entsprechend die Menge der $(-c)^R$ mit der Menge der $-(c^L)$. □

Aus dem Ergebnis von 5.5 gewinnt man nun leicht den

Satz. *Die Klassen gleicher Conwayspiele bilden in bezug auf* $\leqslant$, $-$, $+$ *eine halbgeordnete abelsche Gruppe.*

3. Beispiele. Für einige der am Ende von 4.2 betrachteten Dominospiele wollen wir die zugeordneten Conwayspiele bestimmen. Da D_0 und D_1 keine Vorgänger haben, ist $D_{0\mathbf{C}} \equiv D_{1\mathbf{C}} \equiv (\emptyset, \emptyset) \equiv 0$. Das Dominospiel mit der Ausgangsstellung  hat D_0 als linken und keinen rechten Vorgänger. Diesem Spiel ist also das Conwayspiel $(\{0\}, \emptyset) \equiv 1$ zugeordnet. Analog sieht man, daß den Dominospielen mit den Ausgangsstellungen  bzw.  die Conwayspiele $(\emptyset, \{0\}) \equiv -1$ bzw. $(\{-1\}, \{1\})$ entsprechen. Das Dominospiel mit der Ausgangsstellung  , das D_1 als einzigen linken und einzigen rechten Vorgänger hat, wird auf das Conwayspiel $(\{0\}, \{0\})$ abgebildet.

§ 7. Conwayzahlen

In § 2 haben wir die Dedekindschen Postulate (D1) bis (D4) diskutiert. Bei der beabsichtigten Verallgemeinerung sollte – neben der grundlegenden Auffassung einer Zahl als ein Paar von Mengen, deren Elemente bereits vorher gebildete Zahlen sind – nur das Postulat (D3) (bzw. die Version (D3′), vgl. 2.1) bleiben. Dabei entstand das Problem, wie nunmehr die $\leqslant$-Beziehung zu verstehen sei. Dieses Problem ist gelöst. Die Conwayschen Zahlen sind ihrer Bildung gemäß jedenfalls Conwayspiele, und für Conwayspiele haben wir in 6.2 eine spieltheoretisch motivierte Halbordnung eingeführt. Damit können wir jetzt die beiden Conwaypostulate (C1) und (C2) formulieren. (C1) verallgemeinert das Dedekindsche Postulat (D3′), und (C2) enthält die induktive Kennzeichnung von $\leqslant$ (siehe 6.2).

1. Die Conwayschen Postulate (C1) und (C2). Die Conwayschen Zahlen – im folgenden kurz *Zahlen* genannt – führen wir ein durch die beiden folgenden Postulate. Dabei verwenden wir wie in 6.1 z^L und z^R als Variablen für die linken bzw. rechten Elemente eines Mengenpaares.

(C1)	*Wenn* $z = (x, y)$, *wobei* x *und* y *Mengen von Zahlen sind, und wenn nie* $z^R \leqslant z^L$, *so ist* z *eine Zahl.*
(C2)	*Für Zahlen* x, y ist $x \leqslant y$ *genau dann, wenn nie* $y^R \leqslant x$ *und nie* $y \leqslant x^L$.

Conway entwickelt seine Theorie nur auf der Grundlage dieser beiden Postulate – abgesehen natürlich von den Definitionen der Rechenoperationen (vgl. 8.1). Damit müssen wir alle Eigenschaften von $\leqslant$ aus diesen Postulaten ableiten und dürfen nicht auf die spieltheoretische Definition aus § 5 zurückgreifen.

Wir befinden uns hier in einer analogen Situation wie in § 2, wo wir die Conwayspiele mit dem Postulat (CS) definiert hatten.

Aus (C1) folgt:

(1) Jede Zahl ist ein Mengenpaar. Die linken und die rechten Elemente einer Zahl sind selbst Zahlen. Jede Zahl ist ein Conwayspiel.

Wenn x eine Menge von Zahlen ist, so sind $(x, \emptyset)$ und $(\emptyset, x)$ Zahlen, da die einschränkende Bedingung in (C1) trivialerweise erfüllt ist. Damit folgt insbesondere (vgl. 2.3):

(2) Alle Ordinalzahlen sind Zahlen.

Im folgenden werden wir des öfteren induktive Beweise führen. Dazu formulieren wir ein Induktionsprinzip für Zahlen, welches dem Induktionsprinzip für Conwayspiele (6.2) und dem Induktionsprinzip für Spiele (3.4) entspricht und am einfachsten ebenso bewiesen wird. Neben dem Induktionsprinzip *für eine Eigenschaft* formulieren wir anschließend auch ein Induktionsprinzip *für eine Relation.*

Induktionsprinzip für Zahlen (für eine *Eigenschaft P*). Wenn aus der *Induktionsvoraussetzung*, daß Px' für jedes linke oder rechte Element x' einer Zahl x, für jede solche Zahl die *Induktionsbehauptung* Px folgt, so hat jede Zahl die Eigenschaft P.

Induktionsprinzip für Zahlen (für eine *Relation R*).

Induktionsbehauptung: $Rx_1, \ldots, x_n$.

Induktionsvoraussetzung: $Rx_1', \ldots, x_n'$ für alle n-Tupel $x_1', \ldots, x_n'$, wobei für jedes i $x_i' = x_i$ oder x_i' linkes oder rechtes Element von x_i ist und wobei für wenigstens ein i x_i' linkes oder rechtes Element von x_i ist.

Wenn (für alle $x_1, \ldots, x_n$) aus der Induktionsvoraussetzung die Induktionsbehauptung folgt, so $Rx_1, \ldots, x_n$ für alle Zahlen $x_1, \ldots, x_n$.

2. Elementare Eigenschaften der Ordnung. Wir zeigen zunächst mit dem Induktionsprinzip, daß $\leqslant$ *reflexiv* ist. Simultan beweisen wir zwei weitere Aussagen:

Für jede Zahl x gilt:

$$(3) \qquad \begin{cases} \text{(a) } x^R \not\leqslant x \text{ für jedes } x^R, \\ \text{(b) } x \not\leqslant x^L \text{ für jedes } x^L, \\ \text{(c) } x \leqslant x. \end{cases}$$

Beweis. Zu (a) ((b) ergibt sich analog): Wäre ein $x^R \leqslant x$, so nach (C2) insbesondere $z \leqslant x^R$ für kein rechtes Element von x. Nun ist aber $z = x^R$ ein rechtes Element von x. Es wäre also $x^R \not\leqslant x^R$, entgegen der Induktionsvoraussetzung (Teil (c)).

Zu (c): Wäre $x \not\leqslant x$, so nach § 6 (7I) ein $x^R \leqslant x$ oder $x \leqslant$ ein x^L, im Widerspruch zu (a) bzw. (b). □

Wie bei Spielen und bei Conwayspielen wollen wir eine Äquivalenzrelation $=$ für Zahlen einführen durch die

Definition. *$x = y$ genau dann, wenn $x \leqslant y$ und $y \leqslant x$.*

Damit ergibt sich aus (3):

(4) *Für jede Zahl x ist $x = x$.*

Wir wollen nun zeigen, daß $\leqslant$ *transitiv* ist:

(5) *Für alle Zahlen x, y, z gilt: Wenn $x \leqslant y$ und $y \leqslant z$, so $x \leqslant z$.*

(Wir wissen natürlich bereits aus den früheren Paragraphen, daß dies gilt, wenn wir $\leqslant$ spieltheoretisch definieren. Hier handelt es sich darum, dies aus den Conwayschen Postulaten herzuleiten.) Wir verwenden das Induktionsprinzip für die *drei*stellige Relation R, welche definiert ist durch

$Rxyz$ genau dann, wenn (wenn $x \leqslant y$ und $y \leqslant z$, so $x \leqslant z$)
und (wenn $y \leqslant z$ und $z \leqslant x$, so $y \leqslant x$)
und (wenn $z \leqslant x$ und $x \leqslant y$, so $z \leqslant y$).

Wir haben zu zeigen, daß aus der Induktionsvoraussetzung die Induktionsbehauptung $Rxyz$ folgt. Aus Symmetriegründen genügt es zu zeigen, daß aus der Induktionsvoraussetzung folgt, daß $x \leqslant z$, wenn $x \leqslant y$ und $y \leqslant z$. Sei also $x \leqslant y$ und $y \leqslant z$. Wäre $x \nleqslant z$, gäbe es nach (C2) ein z^R mit $z^R \leqslant x$ oder ein x^L mit $z \leqslant x^L$. Wir beschränken uns auf den ersten Fall (der zweite erledigt sich analog). Aus $z^R \leqslant x$ und $x \leqslant y$ ergibt sich nach der Induktionsvoraussetzung, wobei wir uns auf das *dritte* Konjunktionsglied von $Rxyz$ beziehen, daß $z^R \leqslant y$. Aus $z^R \leqslant y$ und $y \leqslant z$ ergibt sich aus der Induktionsvoraussetzung, wobei wir uns auf das *erste* Konjunktionsglied von $Rxyz$ beziehen, daß $z^R \leqslant z$. Dies widerspricht (3). □

Beim Beweis für die Reflexivität und die Transitivität haben wir nicht davon Gebrauch gemacht, daß die Paarmengenbildung durch eine Bedingung in (C1) eingeschränkt ist. Diese Einschränkung wird aber im folgenden wesentlich sein.

Wir definieren $x < y$ wie üblich durch $x \leqslant y$ *und* $y \nleqslant x$ (oder äquivalent dazu: $x \leqslant y$ *und* $x \neq y$) und behaupten:

(6) *Für jede Zahl x ist $x^L < x$ und $x < x^R$.*

(Man beachte, daß die entsprechende Aussage *für Conwayspiele falsch* ist. Aus ihr folgt nämlich, daß stets $x^L < x^R$, und es gibt natürlich Conwayspiele x und z, so daß z zugleich linkes und rechtes Element von x ist.)

Induktionsbeweis für $x^L < x$. In (3) haben wir bereits gezeigt, daß $x \nleqslant x^L$. Es genügt daher der Nachweis für $x^L \leqslant x$. Wäre $x^L \nleqslant x$, so gäbe es nach (C2) ein x^R mit $x^R \leqslant x^L$, oder ein x^{LL} mit $x \leqslant x^{LL}$.

$x^R \leqslant x^L$ widerspricht (C1).

Wäre $x \leqslant x^{LL}$, so mit der Induktionsvoraussetzung $x^{LL} < x^L$ und der Transitivität von $\leqslant$ auch $x \leqslant x^L$, entgegen (3). □

Jetzt wollen wir zeigen, daß $\leqslant$ die Zahlen *total ordnet.* (Dies gilt nicht für beliebige Spiele. Wir haben nämlich in 4.3 ein Beispiel für ein Spiel x aus der Klasse **E** angegeben. Damit ist $x \not\leqslant 0$ und $0 \not\leqslant x$).

(7) *Für beliebige Zahlen x, y ist $x \leqslant y$ oder $y \leqslant x$.*

Beweis. Wir nehmen an, daß $y \not\leqslant x$, und haben zu zeigen, daß $x \leqslant y$. Aus $y \not\leqslant x$ folgt mit (C2), daß ein $x^R \leqslant y$ oder $x \leqslant$ ein y^L.

Aus $x \leqslant x^R$ (6) und $x^R \leqslant y$ folgt $x \leqslant y$.

Aus $x \leqslant y^L$ und $y^L \leqslant y$ (6) folgt $x \leqslant y$. □

3. Beispiele. Wir haben gesehen, daß alle Ordinalzahlen Zahlen sind. Wenn man die Ordinalzahlen sukzessive erzeugt (vgl. 2.3, wo die ersten Ordinalzahlen definiert sind), so sieht man, daß jede Ordinalzahl mit keiner vorangehenden identisch ($\equiv$) ist. Darüber hinaus gilt aber auch, daß jede Ordinalzahl mit keiner vorangehenden gleich ($=$) ist. Wir begnügen uns damit, dies für die natürlichen Zahlen n zu zeigen. Dazu genügt es zu beweisen, daß stets $n < n + 1$.

(a) $n \leqslant n + 1$: Wir verwenden dazu (C2): (a_1) Nie $(n + 1)^R \leqslant n$, da es kein rechtes Element von $n + 1$ gibt. (a_2) Wäre $n + 1 \leqslant$ ein n^L, so wäre nach der Definition von $n + 1$ ein solches n^L auch ein $(n + 1)^L$; es wäre also $n + 1 \leqslant$ ein $(n + 1)^L$, entgegen (3).

(b) $n + 1 \not\leqslant n$: Wegen (C2) genügt es zu zeigen, daß $n \leqslant$ ein $(n + 1)^L$. Es ist aber n ein $(n + 1)^L$ und $n \leqslant n$ nach (3).

§ 8. Der Körper der Conwayzahlen

Im vorangehenden Paragraphen haben wir die Conwayzahlen eingeführt, einschließlich der Ordnung $\leqslant$ und der Äquivalenzrelation $=$. Wir geben nun die Definitionen für die Rechenoperationen, einige Beispiele (mehr findet man in [1] und [4]) und eine Übersicht über Eigenschaften des Körpers der Conwayzahlen.

1. Die Rechenoperationen für Zahlen. Die Rechenoperationen werden induktiv definiert. Bei $-$ und $+$ erinnern wir uns, daß wir solche Operationen bereits für Conwayspiele in 6.2 eingeführt haben. Wir übernehmen die dort genannten induktiven Definitionen (8I) und (9I) und geben die entsprechenden Definitionen für Zahlen in Gestalt zweier Postulate (C$-$) und (C$+$):

(C$-$) *Für jede Zahl x sei*

$$-x \equiv (\textit{Menge aller } -x^R,\ \textit{Menge aller } -x^L).$$

(C$+$) *Für je zwei Zahlen x, y sei*

$$x + y \equiv (\textit{Menge aller } (x^L + y) \cup \textit{Menge aller } (x + y^L),$$
$$\textit{Menge aller } (x^R + y) \cup \textit{Menge aller } (x + y^R)).$$

Man kann zeigen, daß die Operationen $-$ und $+$ nicht aus dem Bereich der Zahlen herausführen und daß die in 7.2 eingeführte Gleichheit eine Kongruenzrelation für diese Operationen ist.

Für die Multiplikation scheint es kein Vorbild im Bereich der Spiele und der Conwayspiele zu geben. Mit einiger Mühe (vgl. [1]) gelingt Conway die folgende zu (C−) und (C+) analoge Formulierung von (C*) als induktive Definition der Multiplikation:

(C*) *Für je zwei Zahlen* x, y *sei*

$$xy \equiv (\textit{Menge aller } x^L y + xy^L - x^L y^L) \cup \textit{Menge aller } (x^R y + xy^R - x^R y^R),$$
$$\textit{Menge aller } x^L y + xy^R - x^L y^R) \cup \textit{Menge aller } (x^R y + xy^L - x^R y^L)).$$

Die Multiplikation führt nicht aus dem Bereich der Zahlen heraus, und die in 7.2 eingeführte Gleichheit ist eine Kongruenzrelation für diese Operation.

Conway zeigt, daß die Gesamtheit aller Zahlen modulo der Gleichheit in bezug auf $\leqslant$, $-$, $+$, $*$, einen geordneten Körper bildet.

2. Beispiele. Die folgenden Beispiele sollen die Definitionen in Abschnitt 1 illustrieren. Wir zeigen durch Induktion, daß $x + 0 \equiv x$ und $x + y \equiv y + x$, daß $x + - x = 0$ (nicht $\equiv$), $1 + 1 = 2$ und daß $\frac{1}{2} + \frac{1}{2} = 1$, für $\frac{1}{2} \equiv (\{0\}, \{1\})$.

Wir kürzen im folgenden „Menge der (...)“ ab durch „$M(\ldots)$“.

(a) $x + 0 \equiv x$ (und ebenso $0 + x \equiv x$).

$$\begin{aligned} x + 0 &\equiv (M(x^L + 0) \cup M(x + 0^L), M(x^R + 0) + M(x + 0^R)) \qquad (C+) \\ &\equiv (M(x^L + 0), M(x^R + 0)) \\ &\equiv (M(x^L), M(x^R)) \qquad \text{(Induktionsvoraussetzung)} \\ &\equiv x. \end{aligned}$$

□

(b) $x + y \equiv y + x.$

Induktion über y:

$$\begin{aligned} x + y &\equiv (M(x^L + y) \cup M(x + y^L), M(x^R + y) \cup M(x + y^R)) \qquad (C+) \\ &\equiv (M(x + y^L) \cup M(x^L + y), M(x + y^R) \cup M(x^R + y)) \\ &\equiv M(y^L + x) \cup M(y + x^L), M(y^R + x) \cup M(y + x^R)) \quad \text{(Induktionsvoraussetzung)} \\ &\equiv y + x \qquad (C+) \end{aligned}$$

□

(c) $x + - x = 0.$

(Hier kann $=$ nicht durch $\equiv$ ersetzt werden, wie man z. B. für $x \equiv 1$ sieht.)

Wir verwenden die Definition von $=$ in 7.2 und begnügen uns damit zu zeigen, daß $x + - x \leqslant 0$. Es ist klar, daß kein $0^R \leqslant x + - x$, da es kein 0^R gibt. Gäbe es ein $z \equiv (x + - x)^L$ mit $0 \leqslant z$, so wäre nach (C+) $z \equiv x^L + - x$ oder $z \equiv x + (- x)^L$. *Im ersten Fall* wäre $0 \leqslant x^L + - x$, also nach (C2) nie $(x^L + - x)^R \leqslant 0$. Es ist aber nach (C+) und (C−) $x^L + - x^L$ ein solches $(x^L + - x)^R$ und $x^L + - x^L \leqslant 0$ nach Induktionsvoraussetzung. *Im zweiten Fall* gäbe es ein x^R mit $z \equiv x + - x^R$, und es wäre $0 \leqslant x + - x^R$. Damit wäre nach (C2) kein $(x + - x^R)^R \leqslant 0$; es ist aber nach Induktionsvoraussetzung $x^R + - x^R \leqslant 0$. □

(d) $$1 + 1 = 2.$$

In 2.2 haben wir definiert $1 \equiv (\{0\}, \emptyset), 2 = (\{0, 1\}, \emptyset)$. Es ist

$$\begin{aligned} 1 + 1 &\equiv (M(1^L + 1) \cup M(1 + 1^L), M(1^R + 1) \cup M(1 + 1^R)) \qquad (C+) \\ &\equiv (\{0 + 1\} \cup \{1 + 0\}, \emptyset) \\ &\equiv (\{1\}, \emptyset) \qquad (a). \end{aligned}$$

(d_1) $(\{1\}, \emptyset) \leqslant (\{0, 1\}, \emptyset)$. Da es kein $\emptyset^R$ gibt, genügt der Nachweis dafür, daß $2 \leqslant$ kein $(\{1\}, \emptyset)^L$, das heißt $2 \not\leqslant 1$. Dies folgt aus $1 < 2$ (vgl. 7.3).

(d_2) $(\{0, 1\}, \emptyset) \leqslant (\{1\}, \emptyset)$. Es genügt zu zeigen, daß $(\{1\}, \emptyset) \leqslant$ kein $(\{0, 1\}, \emptyset)$, das heißt, daß $(\{1\}, \emptyset) \not\leqslant 0$ und $(\{1\}, \emptyset) \not\leqslant 1$. Wäre $(\{1\}, \emptyset) \leqslant 0$, so wäre $0 \leqslant$ kein $(\{1\}, \emptyset)^L$ entgegen $0 \leqslant 1$. Wäre $(\{1\}, \emptyset) \leqslant 1$, so wäre $1 \leqslant$ kein $(\{1\}, \emptyset)^L$ entgegen $1 \leqslant 1$. □

(e) Wir definieren

$$\tfrac{1}{2} \equiv (\{0\}, \{1\}).$$

$\frac{1}{2}$ ist nach (C1) eine Zahl, da $1 \not\leqslant 0$. Die Bezeichnungsweise wird gerechtfertigt durch den Nachweis, daß $\frac{1}{2} + \frac{1}{2} = 1$.

(f) $$0 \leqslant \tfrac{1}{2}.$$

Es genügt zu zeigen, daß kein $(\frac{1}{2})^R \leqslant 0$. Dies gilt wegen $1 \not\leqslant 0$.

(g) $$1 \not\leqslant \tfrac{1}{2}.$$

Dies folgt daraus, daß 1 ein rechtes Element von $\frac{1}{2}$ ist, und $1 \leqslant 1$.

(h) $$1 + \tfrac{1}{2} \equiv (\{\tfrac{1}{2}, 1\}, \{1 + 1\}).$$

Dies ergibt sich aus (C +) mit (a) und (b).

(i) $$\tfrac{1}{2} + \tfrac{1}{2} = 1.$$

$$\begin{aligned} \tfrac{1}{2} + \tfrac{1}{2} &\equiv (\{0 + \tfrac{1}{2}\} \cup \{\tfrac{1}{2} + 0\}, \{1 + \tfrac{1}{2}\} \cup \{\tfrac{1}{2} + 1\}) \qquad (C+) \\ &\equiv (\{\tfrac{1}{2}\}, \{1 + \tfrac{1}{2}\}) \qquad (a), (b). \end{aligned}$$

Es genügt daher der Nachweis von (d_1) und (d_2).

(d_1) $1 \leqslant (\{\frac{1}{2}\}, \{1 + \frac{1}{2}\})$. Dazu zeigen wir: ($d_{11}$) $1 + \frac{1}{2} \not\leqslant 1$; dies ergibt sich aus $1 \leqslant 1$, weil 1 nach (h) ein $(1 + \frac{1}{2})^L$ ist. (d_{12}) $(\{\frac{1}{2}\}, \{1 + \frac{1}{2}\}) \not\leqslant 0$; dies ergibt sich aus $0 \leqslant \frac{1}{2}$ (f).

(d_2) $(\{\frac{1}{2}\}, \{1 + \frac{1}{2}\}) \leqslant 1$. Dazu genügt es zu zeigen, daß $1 \leqslant$ kein $(\{\frac{1}{2}\}, \{1 + \frac{1}{2}\})^L$, das heißt, daß $1 \not\leqslant \frac{1}{2}$, was wir in (g) gezeigt haben. □

3. Eigenschaften des Körpers der Zahlen. Die Gesamtheit aller Zahlen bildet eine echte Klasse, also keine Menge (vgl. Kap. 13). Dies folgt schon daraus, daß jede Ordinalzahl eine Conwayzahl ist und die Ordinalzahlen keine Menge bilden (vgl. 13.2.4).

Wir haben schon erwähnt, daß die Klasse aller Zahlen in bezug auf die in Abschnitt 1 eingeführten Operationen einen geordneten Körper K_0 bildet. K_0 ist reell-abgeschlossen. K_0 ist (bis auf Isomorphie) eindeutig bestimmt durch

die Eigenschaft, ein universell-einbettender geordneter Körper zu sein. Dies bedeutet folgendes: Zu jedem geordneten Teilkörper K_1 von K_0, der eine Menge ist, und zu jeder Erweiterung K_2 von K_1, wobei K_2 ein geordneter Körper und eine Menge ist, gibt es einen bezüglich der Körperoperationen und der Ordnung zu K_2 isomorphen Teilkörper K_2' von K_0, wobei der Isomorphismus auf K_1 die Identität ist. Daraus ergibt sich insbesondere, daß jeder geordnete Körper in K_0 einbettbar ist. Dazu gehören alle in Kap. 11 betrachteten „Nicht-Standard-Modelle".

Läßt man bei der Konstruktion von Zahlen mit Hilfe von (C1) nur *endliche* Mengen zu, so erhält man genau die *dyadischen* Zahlen, das heißt, die Zahlen der Gestalt $\pm m/2^n$, wobei m, n natürliche Zahlen sind. Für jede *reelle* Zahl x sei x_1 bzw. x_2 die Menge der dyadischen Zahlen $< x$ bzw. $> x$. Dann ist $x = (x_1, x_2)$. Eine Zahl x ist genau dann reell, wenn es eine natürliche Zahl n gibt mit $-n < x < n$ und wenn

$$x = (\text{Menge aller Zahlen } x - 1/2^k, \text{ Menge aller Zahlen } x + 1/2^k).$$

Für Ordinalzahlen liefern die mit (C +) bzw. (C*) eingeführten Operationen deren sogenannte natürliche Summe bzw. natürliches Produkt.

Es gibt unendliche Zahlen, z. B. ω. Damit gibt es auch unendliche kleine Zahlen, z. B. $1/\omega$.

Literatur

[1] Conway, J. H.: On Numbers and Games. Academic Press 1976, ^{3}1979
[2] Dedekind, R.: Stetigkeit und irrationale Zahlen. Vieweg, 1872, ^{7}1965
[3] Huizinga, J.: Homo Ludens. Rowohlt 1956
[4] Knuth, D. E.: Insel der Zahlen, Vieweg 1978

Kapitel 13. Mengenlehre und Mathematik

H.-D. Ebbinghaus

> Gesetzt, es gebe eine große nützliche mathematische Wahrheit, auf die der Erfinder durch einen offenbaren Trugschluß gekommen wäre; – wenn es dergleichen nicht gibt, so könnte es doch dergleichen geben – leugnete ich darum diese Wahrheit, entsagte ich dann, mich dieser Wahrheit zu bedienen? (LESSING, Theologische Streitschriften)

Einleitung. Am 7. Dezember 1873 entwuchs die Mengenlehre den Kinderschuhen. An diesem Tag nämlich bewies Georg CANTOR, daß die Menge der reellen Zahlen überabzählbar ist, also nicht in „abzählender" Gestalt $\{r_0, r_1, r_2, \ldots\}$ geschrieben werden kann [2, S. 115 ff.]. Er legte damit zu einem Zeitpunkt, als der Begriff des *aktual Unendlichen*, die Existenz unendlicher Mengen als *fertiger* Gesamtheiten, in der Mathematik noch kontrovers war, den Grundstein zur *Theorie der unendlichen Mächtigkeiten*. 1878 zeigte er, daß das lineare Kontinuum der reellen Zahlen bijektiv auf die höherdimensionalen Kontinua Ebene, Raum, ... abgebildet werden kann, daß demnach die Kontinua verschiedener Dimension gleichmächtig sind [2, S. 119 ff.]. Mit diesem unerwarteten Resultat gab er den Anstoß zur Entwicklung der Dimensionstheorie. In der Folgezeit führten ihn Untersuchungen über die Bildung $H(A)$ der Menge der Häufungspunkte einer reellen Zahlenmenge A, indem er den Bildungsprozeß gemäß

$$A^{(0)} := A, \quad A^{(1)} := H(A), \quad \ldots, \quad A^{(n+1)} := H(A^{(n)}), \ldots,$$

$$A^{(\infty)} := \bigcap_{n \in \mathbb{N}} A^{(n)}, \qquad A^{(\infty+1)} := H(A^{(\infty)}), \ldots$$

ins Transfinite fortsetzte, zur Schöpfung der *Theorie der transfiniten Ordinalzahlen* [2, S. 145 ff.]. Anknüpfungspunkt war dabei eine Arbeit über den Identitätssatz für trigonometrische Reihen [2, S. 92 ff.], ein Umstand, den ZERMELO [2, S. 102] zum Anlaß nimmt, „in der Theorie der trigonometrischen Reihen die Geburtsstätte der CANTORschen ‚Mengenlehre' zu erblicken".

Allerdings waren bereits vor CANTORS bahnbrechenden Arbeiten der Mengen- und der Unendlichkeitsbegriff Gegenstand scharfsinniger Untersuchungen. So führte im Hochmittelalter die Diskussion über das aktual Unendliche zu Betrachtungen über den Vergleich unendlicher Mengen mittels bijektiver Zuordnungen. ALBERT VON SACHSEN (ca. 1320–1390) beweist z. B. in seinen *Questiones subtilissime in libros de celo et mundo*, daß ein einseitig unendlich langer Holzbalken dasselbe Volumen besitzt wie der unendliche dreidimensionale Raum: In einem Gedankenexperiment zersägt er den Balken in endlich lange Stücke, die er zu sich jeweils anschließenden Kugelschalen umformt, um auf diese Weise den gesamten Raum mit Holz auszufüllen.

Große Klarheit prägt die Ausführungen des bedeutenden Prager Theologen, Philosophen und Mathematikers Bernhard BOLZANO (1781–1848). In seiner Definition einer Menge oder „Vielheit" als „Inbegriff, den wir einem Begriff unterstellen, bei dem die Anordnung seiner Teile gleichgültig ist" (1847, [1, S. 4]),

erkennen wir einen Vorläufer unserer heutigen extensionalen Auffassung, der zufolge eine Menge allein durch ihre Elemente bestimmt ist. BOLZANO verteidigt die Existenz unendlicher Mengen gegen Kritiker. Auch er zeigt an Beispielen, daß unendliche Mengen im Gegensatz zu endlichen Mengen (!) gleichmächtig zu einer echten Teilmenge sein können [1, S. 28 ff.] – eine Einsicht, die DEDEKIND 1888 zur Grundlage seiner Endlichkeitsdefinition macht.

Richard DEDEKIND (1831–1916) entwickelte unabhängig von CANTOR klare Vorstellungen über den Mengenbegriff und seine Bedeutung für die Grundlagen der Mathematik. 1871 schlägt er vor, die KUMMERschen idealen Zahlen – nach seiner Meinung lediglich „fingierte" Zahlen – durch die uns heute vertrauten Ideale zu ersetzen [3, Bd. III, S. 251]; Ideale als Mengen wirklicher Zahlen unterliegen bei ihm hinsichtlich ihrer Existenz keinem Zweifel. Noch konsequenter folgt er dieser Auffassung 1872 in seiner Schrift *Stetigkeit und irrationale Zahlen* (konzipiert 1858), in der die reellen Zahlen mengentheoretisch, nämlich durch DEDEKINDsche Schnitte, „erschaffen" werden [3, Bd. III, S. 315 ff.], und den reinsten Ausdruck verleiht er ihr 1888 in der Abhandlung *Was sind und was sollen die Zahlen*, in der auch die natürlichen Zahlen mengentheoretisch definiert werden [3, Bd. III, S. 335 ff.]. Gerade mit der letzten Schrift hat DEDEKIND einen maßgeblichen Einfluß auf die Entwicklung der Mengenlehre ausgeübt.

Trotz der beträchtlichen Leistungen anderer muß Georg CANTOR (geboren 1845 in Petersburg, gestorben 1918 in Halle) als der eigentliche Begründer der Mengenlehre gelten. Seine Ergebnisse ließen viele naive Vorstellungen zusammenbrechen und öffneten die Tür zu weitreichenden Entwicklungen. Mit seinen Untersuchungen über unendliche Mächtigkeiten und über transfinite Ordinalzahlen schuf er nach HILBERTS Worten [10, S. 167] „die bewundernswerteste Blüte mathematischen Geistes und überhaupt eine der höchsten Leistungen rein verstandesmäßiger menschlicher Tätigkeit".

CANTORS Mengenlehre ist anschaulicher Natur. Sie fußt auf Vorstellungen, denen er in verschiedener Weise Ausdruck verliehen hat. So ist eine Menge für ihn ein „Vieles, welches sich als Eines denken läßt", ein „Inbegriff bestimmter Elemente, welcher durch ein Gesetz zu einem Ganzen verbunden werden kann" (1883, [2, S. 204]), eine „Zusammenfassung von bestimmten wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens zu einem Ganzen" (1895, [2, S. 282]).

Unter Geburtswehen (vgl. 2.1) konnten in den ersten Jahrzehnten dieses Jahrhunderts die intuitiven Vorstellungen CANTORS präzisiert und taugliche *Axiomensysteme* für die Mengenlehre aufgestellt werden. Neue und anspruchsvolle Techniken, wie die Theorie der *konstruktiblen Mengen* (K. GÖDEL 1938) und die *Erzwingungs-* oder *Forcing-Methode* (P. J. COHEN 1963), eröffneten eine Epoche stürmischer Entwicklungen, die noch heute andauert. So gelang es z. B. mit diesen Techniken, die von CANTOR 1878 geäußerte *Kontinuumshypothese*, der zufolge jede überabzählbare Menge reeller Zahlen gleichmächtig mit der Menge aller reellen Zahlen ist, als *unabhängig* nachzuweisen, das heißt, als *unbeweisbar* (COHEN 1963) und *unwiderlegbar* (GÖDEL 1938) auf der Basis der heutigen mengentheoretischen Axiomensysteme.

Rund 100 Jahre nach CANTORS wegbereitenden Arbeiten hat sich die Mengenlehre zu einer ausgewachsenen mathematischen Disziplin entwickelt. Daneben

hat sie jedoch auch Bedeutung für die gesamte Mathematik gewonnen: In einer konsequenten Befolgung DEDEKINDscher Intentionen hat die Mathematik eine immer stärkere mengentheoretische Ausprägung erfahren. Dies hat einmal zu einer schärferen Fassung mancher mathematischer Begriffe geführt und zum anderen die methodischen Hilfsmittel der Mathematik beträchtlich erweitert. HILBERT spricht von einem „Paradies, das CANTOR uns geschaffen" [10, S. 170].

Zudem gestatten es die Axiomatisierungen der Mengenlehre, im axiomatischen Aufbau mathematischer Theorien die Lücken zu schließen, die dort bezüglich des mengentheoretischen Teils klaffen. So berufen sich ja z. B. die Axiomensysteme für topologische Räume auf mengentheoretische Sachverhalte, ohne diese selbst zu axiomatisieren. Und nicht zuletzt schaffen erst Axiomensysteme der Mengenlehre jenes Maß an Genauigkeit, das notwendig ist, um Unabhängigkeitsresultate, wie die Unabhängigkeit der Kontinuumshypothese, zu beweisen.

Wir wollen mit den folgenden Ausführungen die hier angedeuteten Aspekte in den Beziehungen von Mathematik und Mengenlehre näher erörtern und dazu insbesondere auch einen axiomatischen Aufbau der Mengenlehre beschreiben. Dabei müssen wir uns auf grundsätzliche Sachverhalte beschränken und auf manche Details verzichten. Der Leser sei zur weiteren Information auf die Bücher [4], [6], [8] und [15] verwiesen.

§ 1. Mengen und die Objekte der Mathematik

1. Urelemente und höhere Objekte. Die mengentheoretische Ausprägung der heutigen Mathematik beruht insbesondere auf einer mengentheoretischen Beschreibung ihrer Gegenstände. Bevor wir eine solche Beschreibung systematisch in Angriff nehmen, wollen wir uns einen Überblick über die Vielfalt mathematischer Objekte verschaffen. Dabei betrachten wir zunächst eine „konkrete" Theorie, etwa die *Analysis*. Ausgangsobjekte sind hier die reellen Zahlen. Hinzu treten n-Tupel von reellen Zahlen und „kompliziertere" Objekte, wie reelle Funktionen, Intervalle und andere Mengen von reellen Zahlen, Relationen zwischen reellen Zahlen usf.

Reelle Funktionen besitzen eine für die Analysis bedeutsame innere Struktur: sie stellen Zuordnungen zwischen reellen Zahlen her. Dagegen spielen die reellen Zahlen für den Analytiker die Rolle von „Atomen"; nicht ihre innere Struktur ist von Interesse, bedeutsam sind allein die Beziehungen *zwischen* ihnen, wie sie in den üblichen Axiomensystemen der Analysis formuliert werden. Gerade deshalb ist es möglich, Analysis zu betreiben, ohne zu wissen, was reelle Zahlen eigentlich sind. Ähnlich verhält es sich mit den natürlichen Zahlen in der Arithmetik oder den Punkten in der euklidischen Geometrie.

In der Mengenlehre nennt man die Gegenstände einer Theorie, welche solch einen „atomaren" Charakter haben, häufig *Urelemente* (ZERMELO 1930). Im Sinne einer durch diesen Namen bereits intendierten Hierarchie der Objekte bilden also die Urelemente den Ausgangspunkt. Hinzu treten dann sogenannte *Objekte höheren Typs*, wie Eigenschaften von Urelementen, Relationen zwischen Urelementen, Mengen und Funktionen von Urelementen oder auch von n-Tupeln von Urelementen. Darüber türmen sich abermals kompliziertere Objekte, wie Mengen

von Mengen von Urelementen, z. B. offene Überdeckungen in der Analysis oder Restklassenringe in der Arithmetik. Offenbar läßt sich dieser Übergang zu immer komplizierteren Objekten beliebig weit fortsetzen, und es entsteht auf diese Weise über den Urelementen ein hierarchisches Gebäude mathematischer Objekte von zunehmender Komplexität. Zum Teil können wir Schichtungen erkennen (Urelemente, Mengen von Urelementen, Mengen von Mengen von Urelementen). Doch stellen wir auch verwickeltere Beziehungen fest. So können Funktionen auftreten, die Funktionen von Urelementen auf Urelemente abbilden, in der Analysis z. B. die Bildung des bestimmten Integrals für feste Grenzen. Technisch nennt man einen solchen Turm von Objekten, der auf einem Bereich von Urelementen gründet, eine *Typenhierarchie*.

In einer *abstrakten* mathematischen Theorie, wie z. B. der Gruppentheorie, spielen die Elemente von Gruppen eine den Urelementen einer „konkreten" Theorie vergleichbare Rolle. Doch wird hier nicht die Existenz eigener Urelemente gefordert; man geht vielmehr davon aus, daß als Elemente von Gruppen alle mathematischen Objekte in Frage kommen, ohne daß man zusätzliche Forderungen erhebt oder Abgrenzungen trifft.

2. Mengentheoretische Definition höherer Objekte. Es hat sich herausgestellt, daß man Eigenschaften, Relationen und Funktionen, die in der Mathematik meistens intuitiv benutzt werden, auf den Mengenbegriff zurückführen kann. Dadurch wird es möglich, die ganze Vielfalt der Typenhierarchie über einem Bereich von Urelementen mengentheoretisch zu beschreiben.

Wir wollen uns im folgenden von dieser Möglichkeit überzeugen. Dabei machen wir von einigen einfachen Sachverhalten der naiven Mengenlehre Gebrauch. Wir beginnen mit den *Eigenschaften*. Sei zu diesem Zweck M eine Menge von Urelementen oder anderen Objekten, etwa die Menge der reellen Zahlen. E sei eine Eigenschaft über M. Für mathematische Zwecke reicht es nun völlig aus, E zu identifizieren mit der Menge

$$\{r \in M\colon E \textit{ trifft zu auf } r\}$$

derjenigen Elemente von M, die die Eigenschaft E haben. *Den Eigenschaften über M entsprechen so die Teilmengen von M.*

Diese Auffassung hat eine Konsequenz. Z. B. wird dadurch die Eigenschaft über $\mathbb{R}$, Quadrat einer reellen Zahl zu sein, identisch mit der Eigenschaft, nicht negativ zu sein, nämlich gleich der Menge $\{r \in \mathbb{R}\colon r \geqslant 0\}$; denn eine reelle Zahl ist genau dann ein Quadrat, wenn sie nicht negativ ist. Eigenschaften sind jetzt allein durch ihren Umfang, ihre *Extension* bestimmt. Diese *extensionale Auffassung* ist charakteristisch für das mengentheoretische Vorgehen, da ja auch die Mengen allein durch ihre Elemente bestimmt sind. Sie tritt uns in der Mathematik an vielen Stellen entgegen. So wird sie uns etwa bei den Funktionen wieder begegnen: Eine Funktion ist bei gegebenem Definitionsbereich dadurch bestimmt, welche Werte sie den Argumenten zuordnet, und es spielt keine Rolle, *wie* diese Zuordnung definiert wird.

Grundlegend für weitere mengentheoretische Beschreibungsmöglichkeiten ist jetzt eine mengentheoretische Definition von n-Tupeln. Wir beginnen mit dem Fall $n = 2$. Nach K. Kuratowski (1921) definiert man das *geordnete Paar* (a, b) zweier

Objekte a, b mengentheoretisch durch

$$(*) \qquad (a,b) := \{\{a\},\{a,b\}\}.$$

Man weist leicht nach, daß

$$(a,b) = (a',b') \quad \text{genau dann, wenn} \quad a = a' \quad \text{und} \quad b = b'.$$

Diese Äquivalenz ist der einzige Sachverhalt über geordnete Paare, den der Mathematiker wirklich benötigt; die KURATOWSKIsche Festlegung genügt also von daher voll seinen Ansprüchen.

Es sei hier eine Bemerkung angebracht, die grundsätzlich für alle mengentheoretischen Beschreibungen mathematischer Objekte gilt: *Eine mengentheoretische Definition wie* (∗) *verfolgt keine ontologischen Zwecke.* So soll (∗) nicht festlegen, was geordnete Paare *wirklich* sind, sondern nur ein *Modell* für den intuitiven Begriff des geordneten Paares liefern, das den Anforderungen der Mathematik gerecht wird. Dieser „konventionalistische" Standpunkt wird auch durch die Tatsache unterstützt, daß in aller Regel verschiedenartige Definitionen möglich sind, denen man nur schlecht eine ontologische Rangordnung unterschieben könnte. So erfüllt die Definition

$$(a,b) := \{\{\{a\},\emptyset\},\{\{b\}\}\}$$

(WIENER, 1914) den gleichen Zweck wie (∗).

Nachdem geordnete Paare mengentheoretisch definiert sind, lassen sich zwanglos *Tripel* in der Form

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

einführen, dann in ähnlicher Weise Quadrupel, Quintupel usf.

Um den Begriff der zweistelligen – und dann ähnlich der n-stelligen – Relation über einer Menge M mengentheoretisch zu beschreiben, fassen wir eine zweistellige Relation zwischen Elementen von M auf als eine Eigenschaft von geordneten Paaren über M. Setzen wir wie üblich

$$M \times M := \{(a,b)\colon a,b \in M\},$$

so sind also die zweistelligen Relationen über M im Sinne der mengentheoretischen Beschreibung der Eigenschaften gerade die Teilmengen von $M \times M$. Z. B. ist $K := \{(r,s)\colon r,s \in \mathbb{R},\ r < s\}$ die Kleiner-Relation über $\mathbb{R}$, und $2 < 3$ bedeutet, daß $(2,3) \in K$.

Entsprechend kann man in wohlbekannter Weise eine Funktion f von einer Menge M_1 in eine Menge M_2 mengentheoretisch durch ihren Graphen definieren als

$$f = \{(a,f(a))\colon a \in M_1\}.$$

Allgemein ist dann eine Funktion f eine Menge von geordneten Paaren, für die zu jedem Objekt a höchstens ein Objekt b existiert mit $(a,b) \in f$. Die vertraute mathematische Schreibweise „$f\colon M_1 \to M_2$" besagt jetzt, daß $f \subset M_1 \times M_2$ eine Funktion ist, so daß zu jedem $a \in M_1$ ein $b \in M_2$ existiert mit $(a,b) \in f$. Für $a \in M_1$ ist $f(a)$ *das* b mit $(a,b) \in f$.

Ähnlich verfährt man bei höheren Stellenzahlen.

Wir sehen damit exemplarisch, daß es gelingt, die Vielfalt der Objekte, die in einer mathematischen Theorie auftreten, in mengentheoretischer Gestalt systematisch zu beschreiben: Ausgangspunkt ist jeweils ein gewisser Bereich von Urelementen, aus denen die komplizierteren Objekte durch iterierte Mengenbildungsprozesse hervorgehen. So sind einstellige reelle Funktionen Mengen von geordneten Paaren reeller Zahlen. Geordnete Paare reeller Zahlen sind nach (∗) Mengen von Mengen reeller Zahlen. Also sind reelle Funktionen Mengen von Mengen von Mengen reeller Zahlen. Diese Zurückführung der Typenhierarchie auf den Mengenbegriff ermöglicht letztlich die so erfolgreiche mengentheoretische Darstellungsweise in der Mathematik.

Selbstverständlich sind nicht alle Details einer solchen mengentheoretischen Präzisierung für die mathematische „Alltagsarbeit" gleichermaßen wesentlich: Ein Mathematiker benötigt kaum die Definition (∗) des geordneten Paares, und er arbeitet mehr mit dem „dynamischen" intuitiven Funktionsbegriff als mit der eher „statischen" mengentheoretischen Beschreibung. Der Wert einer mengentheoretischen Formulierung mathematischer Begriffe und Sachverhalte offenbart sich also nicht unbedingt bei einem *konsequenten* Gebrauch, er liegt vielmehr in der *Möglichkeit*, sich dieser eleganten und wirksamen Methode dort zu bedienen, wo sie nützlich ist. Mit anderen Worten: Eine mengentheoretische Formulierung soll für den Mathematiker keine Zwangsjacke sein, sondern eine Bereicherung seines methodischen Instrumentariums. Weitere Aspekte diskutieren wir in § 2 und in 3.3.

3. Urelemente als Mengen. Bei den Überlegungen des vorangehenden Abschnitts behalten die Urelemente (Zahlen, Punkte, ...) die Rolle von Atomen; ihre Gestalt bleibt im Dunkeln. Methodologisch braucht das kein Nachteil zu sein: Für die Mathematik ist, wie wir betont haben, die „wahre" Gestalt der Urelemente belanglos. Von der mathematischen Arbeitsweise her gesehen ist ihre Beibehaltung sogar sehr natürlich. Auch bereitet es keine Schwierigkeit, die bislang nur naiv benutzten mengentheoretischen Sachverhalte im Rahmen einer *axiomatischen Mengenlehre mit Urelementen zu präzisieren* – mit dem gleichen Gewinn, den eine präzise Mengenlehre *ohne* Urelemente zu geben vermag (vgl. hierzu die beiden folgenden Paragraphen). Andererseits ist es verlockend, den in 2. eingeschlagenen Weg weiterzugehen und eine mengentheoretische Beschreibung der Urelemente in Angriff zu nehmen, *um so den Mengenbegriff zur alleinigen Grundlage der Mathematik zu machen.*

Es gehört nun zu den großen begrifflichen Leistungen von Mathematik und Mengenlehre, dieses Projekt verwirklicht zu haben. Bahnbrechend hat hier insbesondere DEDEKIND gewirkt; wir haben darauf bereits in der Einleitung genauer hingewiesen. Um ein Beispiel vorzustellen, wollen wir im folgenden kurz auf die *mengentheoretische Definition der natürlichen Zahlen* durch ZERMELO (1908) und VON NEUMANN (1923) eingehen. Wir argumentieren dabei weiterhin intuitiv. Einer Präzisierung im axiomatischen Rahmen wenden wir uns in 2.3 zu.

ZERMELO setzt der Reihe nach

$$0 := \emptyset, \qquad 1 := \{\emptyset\}, \qquad 2 := \{\{\emptyset\}\}, \ldots;$$

VON NEUMANN definiert

$$0 := \emptyset, \qquad 1 := \{\emptyset\}, \qquad 2 := \{\emptyset, \{\emptyset\}\}, \ldots$$

und allgemein

$$n + 1 := n \cup \{n\}.$$

Seine Vorgehensweise hat gegenüber der ZERMELOschen den technischen Vorteil, daß jede Zahl gerade die Menge der vorangehenden Zahlen ist und daß daher die $<$-Beziehung mit der $\in$-Beziehung zusammenfällt. Vom kardinalen Standpunkt aus sind die VON NEUMANNschen Zahlen natürliche Maßstäbe für endliche Mächtigkeiten – enthält doch die Zahl n genau n Elemente! (Diese Eigenschaft hat auch eine verwandte Definition CANTORS (1895; vgl. [2, S. 289 f.]).) Schließlich läßt sich die VON NEUMANNsche Zahlenreihe leicht gemäß

$$0, 1, 2, \ldots, \omega := \{0, 1, 2, \ldots\},\ \omega + 1 := \omega \cup \{\omega\},$$

$$\omega + 2 := (\omega + 1) \cup \{\omega + 1\}, \ldots,\ \omega + \omega := \{1, 2, \ldots, \omega, \omega + 1, \omega + 2, \ldots\}, \ldots$$

ins Transfinite fortsetzen; auch für die sich so ergebenden *Ordinalzahlen* fallen $<$- und $\in$-Beziehung zusammen.

Definiert man nun für die mengentheoretischen Zahlen auf kanonische Weise eine Nachfolgerfunktion ($n \mapsto n \cup \{n\}$ (im Sinne von 1.2 genauer: $\{(n, n \cup \{n\}) : n \in \omega\}$) bei VON NEUMANN, $n \mapsto \{n\}$ bei ZERMELO), so läßt sich ohne Mühe zeigen, daß die PEANOschen Axiome gelten. Da diese Axiome alle Eigenschaften garantieren, die der Mathematiker von den natürlichen Zahlen benötigt, liefern also die mengentheoretischen Definitionen von VON NEUMANN und ZERMELO (zusammen mit der jeweiligen Nachfolgerfunktion) adäquate Modelle. Selbstverständlich – und das sei an dieser Stelle noch einmal wiederholt – kann eine solche Definition uns nicht sagen, was natürliche Zahlen *wirklich* sind!

Es bereitet jetzt keine Schwierigkeit mehr, auch die arithmetischen Operationen und die weiteren Zahlen (ganze, rationale, reelle, komplexe) in geeigneter Weise mengentheoretisch zu definieren. Man braucht nur einem der üblichen Wege zum Aufbau der Zahlbereiche zu folgen.

Wir stehen damit an einem Punkt, wo es möglich ist, die anfangs schier erdrückend erscheinende Vielfalt mathematischer Objekte allein auf den Mengenbegriff zu reduzieren. Wenn wir etwa von der VON NEUMANNschen Definition der natürlichen Zahlen ausgehen, so pflanzt sich die Eigenschaft, eine Menge zu sein, von der Null über die natürlichen Zahlen, die ganzen Zahlen etc. auf alle mathematischen Objekte fort. Sie alle sind dann Mengen; an ihrem Anfang steht, gleichsam als einziges Urelement, die leere Menge. Etwas prägnanter können wir sagen, daß sich das Universum der mathematischen Objekte allein mit Hilfe der Mengenbildung „aus dem Nichts" erschaffen läßt.

§ 2. Axiomensysteme der Mengenlehre

Bislang haben wir nur im Rahmen einer *intuitiven* Mengenlehre eingesehen, daß wir mit Hilfe des Mengenbegriffs allein zu einer Darstellung der Mathematik gelangen können. Die Tragweite, die der Mengenlehre damit zukommt, fordert zu einer gründlichen Analyse des Mengenbegriffs heraus und zu einer Präzisierung unseres

Vorgehens. Dies kann am ehesten durch die Aufstellung eines *Axiomensystems der Mengenlehre* erfolgen. Wir wollen in diesem Paragraphen einige Systeme schildern, und zwar Systeme für eine Mengenlehre ohne Urelemente. Dabei werden wir auch auf Schwierigkeiten eingehen, die bei der Geburt solcher Systeme Pate gestanden haben und letztlich in einer inkorrekten Mengenvorstellung wurzeln. Mit der Aufstellung hinreichend starker mengentheoretischer Axiomensysteme gelingt es, die Gegenstände der gesamten Mathematik auf eine einheitliche axiomatische Basis zu gründen.

1. Die Russellsche Antinomie. Gottlob FREGE (1848–1925), einer der Väter der mathematischen Logik, gab im ersten Band seiner *Grundgesetze der Arithmetik* [7] ein Axiomensystem für die CANTORsche Mengenlehre an. Sein Ziel war es, die Mathematik logisch-mengentheoretisch zu begründen. Eines seiner Axiome präzisiert die Vorstellung von Mengen als Umfänge oder Extensionen von Eigenschaften, wie man sie aus dem CANTORschen Mengenbild herauslesen konnte und wie sie auch DEDEKIND öfter benutzt hat. Es lautet in heutiger Sprechweise:

Fregesches Komprehensionsaxiom (von lat. *comprehensio* = das Zusammenfassen). Zu jeder Eigenschaft E existiert die *Menge*

$$M_E := \{x\colon x \textit{ ist Menge und } E \textit{ trifft zu auf } x\}.$$

Im Sommer 1901 entdeckte Bertrand RUSSELL (1872–1970) die Inkonsistenz des Komprehensionsaxioms: Wählt man als E die Eigenschaft, nicht Element von sich selbst zu sein, so liefert Komprehension die *Menge*

$$M_R := \{x\colon x \textit{ ist Menge und } x \notin x\}.$$

Für diese gilt offenbar

$$M_R \in M_R \Leftrightarrow M_R \text{ ist Menge und } M_R \notin M_R.$$

Da M_R eine Menge ist, erhält man

$$M_R \in M_R \Leftrightarrow M_R \notin M_R,$$

also einen Widerspruch.

Bereits einige Wochen zuvor hatte ZERMELO diese Antinomie dem Philosophen E. HUSSERL mitgeteilt. Ein entsprechender schriftlicher Vermerk HUSSERLS fand sich in dessen Nachlaß. Allem Anschein nach hat ZERMELO seiner Entdeckung zunächst keine große Bedeutung beigelegt. Waren doch auf naiver Ebene bereits andere Antinomien bekannt. So z. B. die Antinomie von BURALI-FORTI (1897, scharf gefaßt von RUSSELL 1903): Der in 1.3 angedeuteten Bildung der VON NEUMANNschen Ordinalzahlen, bei der die $<$-Beziehung mit der $\in$-Beziehung zusammenfällt, kann man entnehmen, daß die Menge Ω aller Ordinalzahlen – ihre Existenz unterstellt! – ähnlich wie ω oder $\omega + \omega$ wieder eine Ordinalzahl ist. Also gilt $\Omega \in \Omega$ im Widerspruch dazu, daß eine Ordinalzahl nicht kleiner sein kann als sie selbst (oder auch im Widerspruch zum Fundierungsaxiom; vgl. 2.).

CANTOR nannte solche „gefährlichen" Mengen, wie M_R oder Ω, *absolut unendliche* oder *inkonsistente Vielheiten* [2, S. 443 f.]. Sie waren für ihn keine

Mengen im eigentlichen Sinn; FREGE ist also mit seinem Komprehensionsaxiom wesentlich über den von CANTOR naiv abgesteckten Rahmen hinausgegangen.

Die Entdeckung der RUSSELLschen Antinomie rief die Gegner der Mengenlehre auf den Plan, die in den infinitären mathematischen und mengentheoretischen Konzepten die Wurzel für solche Widersprüche sahen und die sich auf das konstruktiv Kontrollierbare zurückziehen wollten. Einer der Wegbereiter dieser Auffassung und darin ein Gegenspieler DEDEKINDS und insbesondere CANTORS ist Leopold KRONECKER (1823–1891). Ein Zitat aus dem Jahre 1886 (vgl. [13, S. 336]) mag das illustrieren.

„... selbst der allgemeine Begriff einer unendlichen Reihe ... ist ... nur mit dem Vorbehalte zulässig, daß in jedem speziellen Falle auf Grund des arithmetischen Bildungsgesetzes der Glieder ... gewisse Voraussetzungen als erfüllt nachgewiesen werden, welche die Reihen wie endliche Ausdrücke anzuwenden gestatten, und welche also das Hinausgehen über den Begriff einer *endlichen* Reihe eigentlich unnötig machen."

Die unterschiedlichen erkenntnistheoretischen Standpunkte CANTORS und KRONECKERS haben nicht nur zu wissenschaftlichen Kontroversen geführt, sondern auch die persönlichen Beziehungen zwischen beiden stark belastet, ein Umstand, unter dem CANTOR sehr gelitten hat.

Exponent der kritisch-konstruktiven Haltung wurde in der Folgezeit der holländische Mathematiker L. E. J. BROUWER (1881–1966). Die von ihm vertretene Richtung ist heute als *Intuitionismus* bekannt [9].

Auf der anderen Seite versuchten zahlreiche Mathematiker, unter ihnen auch RUSSELL und ZERMELO, durch eine Revision der sich in den FREGEschen Axiomen niederschlagenden Vorstellung über den Mengenbegriff zu einer widerspruchsfreien Axiomatisierung der von CANTOR eröffneten Möglichkeiten zu kommen. Einer der entschiedensten geistigen Führer dieser Richtung wurde David HILBERT (1862–1943); vgl. [10].

Im folgenden sollen die bekanntesten Axiomensysteme kurz vorgestellt werden. Sie gelten heute bei den Mengentheoretikern als widerspruchsfrei. Ein Widerspruchsfreiheitsbeweis, wie er bis in die zwanziger Jahre unseres Jahrhunderts hinein noch für möglich gehalten wurde, kann nach einem von K. GÖDEL (1931) stammenden Satz der mathematischen Logik selbst mit Hilfsmitteln von der methodischen Stärke der Mengenlehre nicht erbracht werden (vgl. z. B. [5, S. 226 ff.]). Auf gewisse inhaltliche Argumente, die für die Widerspruchsfreiheit sprechen, werden wir in 3.1 eingehen.

2. Zermelosche und Zermelo–Fraenkelsche Mengenlehre. Im Jahre 1908 schuf Ernst ZERMELO (1871–1953) ein wegweisendes Axiomensystem [21]. Mit einer später durch A. FRAENKEL und Th. SKOLEM vorgenommenen Ergänzung stellt es bis heute das wichtigste System dar. Unverkennbar ist der Einfluß DEDEKINDS. ZERMELO beschreibt sein Unterfangen so:

Angesichts der RUSSELLschen Antinomie „bleibt ... nichts anderes übrig, als ..., ausgehend von der historisch bestehenden ‚Mengenlehre' die Prinzipien aufzusuchen, welche zur Begründung dieser mathematischen Disziplin erforderlich sind ... in der Weise ..., daß man die Prinzipien einmal eng genug einschränkt, um alle Widersprüche auszuschließen, gleichzeitig aber auch weit genug ausdehnt, um alles Wertvolle dieser Lehre beizubehalten".

Wir geben im folgenden die ZERMELOschen Axiome in einer heute üblichen, leicht modifizierten Form an. Inhaltlich gesehen beschreiben sie ein „Universum" von Mengen; Urelemente treten nicht auf.

Existenzaxiom, **Ex**. *Es gibt eine Menge.*

(ZERMELO fordert stattdessen, jedoch äquivalent zu **Ex** auf der Basis der übrigen Axiome, die Existenz der leeren Menge.)

Extensionalitätsaxiom, **Ext**. *Zwei Mengen, die die gleichen Mengen als Elemente besitzen, sind gleich.*

Ext spiegelt die extensionale Auffassung vom Mengenbegriff, der zufolge eine Menge allein durch ihre Elemente bestimmt ist.

Aussonderungsaxiom, **Aus**. *Zu jeder Eigenschaft E von Mengen und zu jeder Menge x existiert eine Menge y, die genau aus den Elementen von x besteht, welche die Eigenschaft E haben.* Die Menge y ist nach **Ext** eindeutig bestimmt. Schreibweise: $y = \{z \in x\colon E \text{ trifft zu auf } z\}$.

Aus übernimmt die Rolle des FREGEschen Komprehensionsaxioms. Doch anders als bei FREGE werden Komprehensionen auf den Rahmen bereits vorgegebener Mengen eingeschränkt. In dieser Vorsichtsmaßnahme ZERMELOS kommt sein Bestreben zum Ausdruck, mit den Axiomen nicht ein gleichsam fertiges Universum von Mengen zu beschreiben, sondern sich an einem Aufbau des Universums „von unten her" zu orientieren. Dem entspricht auch der Charakter der meisten folgenden Axiome: Sie sagen aus, wie man aus bereits vorhandenen Mengen neue gewinnen kann. Man kann übrigens leicht feststellen, daß sich – zumindest ad hoc – der RUSSELLsche Widerspruch im ZERMELOschen System nicht mehr nachvollziehen läßt.

Welche Eigenschaften sind in **Aus** zugelassen? ZERMELO denkt an besonders „konkrete" Eigenschaften, die er *definit* nennt, jedoch ohne dafür eine befriedigende Präzisierung angegeben zu haben. Eine schärfere Festlegung erfolgt durch den norwegischen Mathematiker und Logiker Thoralf SKOLEM (1922): Zugelassen werden diejenigen Eigenschaften, die sich in der prädikatenlogischen Sprache erster Stufe (vgl. [5]) formulieren lassen. Dabei darf nur $\in$ als nichtlogisches Symbol verwandt werden, das Gleichheitszeichen ist erlaubt, und die Variablen laufen über Mengen. Eine Reihe von Beispielen folgt in 3.; doch sei ein besonders einfaches Beispiel bereits hier angeführt: Die Eigenschaft E, die auf eine Menge z genau dann zutrifft, wenn $z \neq z$ ist, genügt der SKOLEMschen Forderung. Wählen wir jetzt eine Menge x_0, deren Existenz durch **Ex** verbürgt wird, so sichert **Aus** die Existenz der Menge $\{z \in x_0\colon z \neq z\}$, also die Existenz der leeren Menge $\emptyset$ (die ja aufgrund von **Ext** eindeutig bestimmt ist).

Paarmengenaxiom, **Paar**. *Zu je zwei Mengen x, y gibt es eine* (nach **Ext** dann *die*) *Menge z, welche genau x und y als Elemente besitzt.*

Wir schreiben $z = \{x, y\}$ und für $\{x, x\}$, die *Einermenge* von x, abkürzend $\{x\}$. Nach **Ext** ist stets $\{x, y\} = \{y, x\}$; $\{x, y\}$ hat also nicht die Eigenschaft eines *geordneten* Paares von x und y.

Vereinigungsmengenaxiom, **U-Ax**. *Zu jeder Menge X* (die man sich in diesem Zusammenhang am besten als ein „System" von Mengen vorstellt) *gibt es die Menge Y der Elemente der Elemente von X.*

Schreibweise: $Y = \bigcup X$, in der Mathematik häufiger $Y = \bigcup_{x \in X} x$. Z. B. ist $\bigcup \emptyset = \emptyset$, $\bigcup \{x\} = x$.

Potenzmengenaxiom, **Pot**. *Zu jeder Menge x gibt es die Menge y aller Teilmengen von x*, die sogenannte *Potenzmenge* von x.

Schreibweise: $y = Pot(x)$.

Die bisherigen Axiome sind erfüllt im Bereich derjenigen *endlichen* Mengen, die man aus der leeren Menge jeweils durch endlich viele Übergänge der Gestalt $x \mapsto \{x\}$, $x, y \mapsto x \cup y$ erhält. Es fehlt also noch ein Axiom, das die Existenz unendlicher Mengen sicherstellt, um die Kraft der Mengenlehre im Transfiniten zu entfalten.

Unendlichkeitsaxiom, **Inf**. *Es gibt eine induktive Menge*, das heißt, eine Menge, die $\emptyset$ enthält und mit jedem z auch $z \cup \{z\}$. (Da „$\cup$" noch nicht eingeführt ist, steht hier $z \cup \{z\}$ als Abkürzung für eine Menge, deren Elemente genau die Elemente von z und z selbst sind.)

Anschaulich gesehen enthält eine induktive Menge auf jeden Fall die von Neumannschen natürlichen Zahlen $\emptyset$, $\{\emptyset\}$, $\{\emptyset, \{\emptyset\}\}$, usf. Wie wir in 3. sehen werden, wird **Inf** entscheidend gebraucht, um die Existenz der Menge ω dieser Zahlen zu beweisen.

Wir vervollständigen das Zermelosche Axiomensystem mit dem

Auswahlaxiom, **AC** (= **A**xiom of **C**hoice). *Zu jeder Menge gibt es eine Auswahlfunktion.* Dabei heißt eine auf einer Menge X definierte Funktion eine *Auswahlfunktion* zu X, wenn $f(y) \in y$ für alle $y \in X$, $y \neq \emptyset$.

Bei dieser Formulierung stoßen wir zunächst auf eine Schwierigkeit: Das Zermelosche Axiomensystem ist konzipiert als ein Axiomensystem für Mengen, in dem als Grundbegriff nur die $\in$-Beziehung zwischen Mengen auftritt. Dieser Intention trägt ja auch die Skolemsche Präzisierung der in **Aus** zugelassenen Eigenschaften Rechnung. In der von uns gewählten anschaulichen Formulierung von **AC** kommt dagegen der Funktionsbegriff vor. (Vgl. dazu die äquivalente Formulierung **AC'** in 3.2.) Man kann jedoch Funktionen mengentheoretisch beschreiben, wie wir das in 1.3 intuitiv angedeutet haben, und dann in **AC** den Gebrauch des Funktionsbegriffs mit Hilfe seiner mengentheoretischen Definition eliminieren. Die obige Formulierung von **AC** kann dann als eine besser verständliche Abkürzung für das so entstehende Axiom aufgefaßt werden.

Wie wir im folgenden Abschnitt exemplarisch darlegen werden, reicht das ZERMELOsche Axiomensystem aus, um praktisch alle mengentheoretischen Sachverhalte abzuleiten, die der Mathematiker benötigt. Nur selten, und zwar dort, wo der mengentheoretische Rahmen außergewöhnlich stark „strapaziert" wird (z. B. im Zusammenhang mit der Definition der CONWAYspiele und CONWAYzahlen in Kap. 12, 2.3 bzw. 7.1) und in der Mengenlehre selbst benötigt man einige weitere Axiome: das *Fundierungsaxiom*, **Fund**, und das *Ersetzungsaxiom*, **Ers**.

Fund (VON NEUMANN 1925, in der folgenden Formulierung nach ZERMELO 1930) schließt die Existenz pathologischer Mengen x mit $x \in x$ aus oder auch absteigende $\in$-Ketten der Gestalt $\ldots x_2 \in x_1 \in x_0$. Es besagt, daß jede nicht-leere Menge x ein *$\in$-minimales* Element hat, das heißt, ein Element y mit $x \cap y = \emptyset$: *Zu jeder Menge $x \neq \emptyset$ gibt es ein $y \in x$, das mit x kein Element gemeinsam hat.*

Um z. B. einzusehen, daß mit **Fund** stets $x \notin x$ gilt, bilde man nach **Paar** zu vorgegebenem x die Menge $\{x\}$. Da x das einzige Element von $\{x\}$ ist, muß x $\in$-minimales Element von $\{x\}$ sein, insbesondere ist also $x \notin x$.

Ers (MIRIMANOFF 1917, FRAENKEL 1922, SKOLEM 1923) besagt intuitiv: Ersetzt man die Elemente einer Menge „auf vernünftige Weise" durch andere, so entsteht wieder eine Menge. Formaler: *R sei eine zweistellige Beziehung zwischen Mengen, so daß zu jeder Menge x höchstens eine Menge y existiert mit xRy. Dann gibt es zu jeder Menge X die Menge $\{y$: es gibt ein $x \in X$ mit $xRy\}$.*

Dabei sind, ähnlich wie bei **Aus**, solche Beziehungen R zugelassen, die sich in der prädikatenlogischen Sprache erster Stufe mit $\in$ beschreiben lassen. Beispiele sind etwa die Beziehungen

$$x R y :\Leftrightarrow x = y; \qquad x R y :\Leftrightarrow y = \{x\}.$$

Mit der letzteren bekommen wir aus **Ers** sofort, daß die Einermengen der Elemente einer Menge wieder eine Menge bilden. (Man kann statt mit **Ers** auch mit **Aus** und **Pot** argumentieren.)

Ers ist ebenfalls ein Spezialfall des FREGEschen Komprehensionsaxioms. Es ist stärker als **Aus**; **Aus** kann auf der Basis der restlichen ZERMELOschen Axiome aus **Ers** bewiesen werden.

Das sogenannte *Zermelo-Fraenkelsche Axiomensystem* **ZF** (ohne Auswahlaxiom) umfaßt die bisher formulierten Axiome mit Ausnahme von **AC**. Durch dessen Hinzunahme entsteht das System **ZFC**, das heute den meisten mengentheoretischen Betrachtungen zugrunde liegt.

3. Einige Folgerungen. Wir wollen nun an einigen elementaren Beispielen darlegen, daß bereits die ZERMELOschen Axiome ausreichen, um – bis auf wenige Ausnahmen – diejenigen Sachverhalte abzuleiten, die der Mathematiker für ein mengentheoretisches Vorgehen benötigt. Großenteils handelt es sich dabei um Resultate, die wir in 1.2 und 1.3 bereits *intuitiv* benutzt haben.

(a) Die *leere Menge* $\emptyset$: Ihre Existenz haben wir in 2. bewiesen.

(b) BOOLE*sche Kombinationen.* Es seien Mengen x, y vorgegeben.

Nach **Aus** existiert die Menge $\{z \in x\colon z \in y\}$, also der *Durchschnitt* $x \cap y$ von x und y.

Nach **Paar** existiert zunächst die Menge $\{x, y\}$ und nach $\bigcup$-**Ax** dann die Menge $\bigcup \{x, y\}$. Sie besteht aus den Elementen von x und aus den Elementen von y, ist also die *Vereinigung* $x \cup y$ von x und y.

Nach **Aus** existiert auch die Menge $\{z \in x\colon z \notin y\}$, also die *mengentheoretische Differenz* $x \setminus y$ von x und y.

(c) *Verallgemeinerter Durchschnitt.* Zu jeder nicht-leeren Menge X, die wir uns in diesem Zusammenhang als „Mengensystem" denken, existiert der *Durchschnitt* $\bigcap X = \bigcap_{y \in X} y$ von X; man erhält ihn mit **Aus** in der Form

$$\bigcap X = \{z \in \bigcup X\colon z \in y \text{ für alle } y \in X\}.$$

Es ist jetzt nicht schwer, die bekannten Gesetzmäßigkeiten für $\cap, \cup, \setminus, \bigcap, \bigcup$ abzuleiten.

(d) *Geordnete Paare und kartesische Produkte.* Zu zwei Mengen x, y liefert dreimalige Anwendung von **Paar** das *geordnete Paar* $(x, y) = \{\{x\}, \{x, y\}\}$. Die Existenz des kartesischen Produktes $x \times y = \{(u, v)\colon u \in x \text{ und } v \in y\}$ erhalten wir folgendermaßen: Ist $u \in x$, so ist $\{u\} \subset x$, also $\{u\} \subset x \cup y$ und daher $\{u\} \in Pot(x \cup y)$. Ist weiter $v \in y$, ergibt sich entsprechend $\{u, v\} \in Pot(x \cup y)$. Da $(u, v) = \{\{u\}, \{u, v\}\}$, gilt also $(u, v) \in Pot(Pot(x \cup y))$. Daher ist

$$x \times y = \{z \in Pot(Pot(x \cup y))\colon \text{Es gibt } u \in x \text{ und } v \in y \text{ mit } z = (u, v)\},$$

und die Existenz dieser Menge ergibt sich mit **Aus.** (Die Bedingung „$z = (u, v)$" läßt sich leicht durch $\in$ allein ausdrücken!)

Eine Ausdehnung dieser Betrachtungen auf höhere Stellenzahlen bereitet keine Schwierigkeiten, ebensowenig die Herleitung der grundlegenden Eigenschaften von Relationen und Funktionen, die man dazu wie in 1.2 mengentheoretisch definiert.

(e) *Natürliche Zahlen.* Anschaulich ist die Menge $\omega = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \ldots\}$ der VON NEUMANNschen natürlichen Zahlen die kleinste Menge, die $\emptyset$ enthält und mit jedem z auch $z \cup \{z\}$, also im Sinne der bei **Inf** eingeführten Terminologie die kleinste induktive Menge. Ist y_0 irgendeine induktive Menge, können wir ω ohne Benutzung von „Pünktchen" – zunächst noch intuitiv – definieren als

$$\omega = \bigcap \{y \subset y_0\colon y \text{ induktiv}\}.$$

In dieser Gestalt nun läßt sich die Existenz von ω auf der Basis der ZERMELOschen Axiome *beweisen*: Zunächst existiert nach **Inf** eine induktive Menge, etwa y_0. Man sieht leicht, daß die Induktivität eine im Aussonderungsaxiom zugelassene Eigenschaft ist. Somit existiert mit $Pot(y_0)$ auch $\{y \subset y_0\colon y \text{ induktiv}\} = \{y \in Pot(y_0)\colon y \text{ induktiv}\}$ und nach (c) dann ω selbst.

Leicht ergibt sich, daß ω induktiv ist. Daher ist mit $n \in \omega$ stets $n \cup \{n\} \in \omega$, und wir können die Nachfolgerfunktion σ über ω ähnlich wie das kartesische Produkt in (d) gewinnen als

$$\begin{aligned}\sigma &= \{(n, n \cup \{n\})\colon n \in \omega\} \\ &= \{z \in Pot(Pot(\omega))\colon \text{Es gibt } n \in \omega \text{ mit } z = (n, n \cup \{n\})\}.\end{aligned}$$

(Mit **Ers** ist der Beweis für die Existenz von σ noch einfacher!)

Einem Nachweis der gewünschten arithmetischen Eigenschaften von ω und σ steht nun nichts mehr im Wege. Das Induktionsaxiom z. B. wird trivial. Seine Aussage

> *Trifft eine Eigenschaft von natürlichen Zahlen auf* 0 *zu und vererbt sie sich stets von einer natürlichen Zahl n auf den Nachfolger* $n + 1$, *so trifft sie auf alle natürlichen Zahlen zu*

besagt für ω und σ in mengentheoretischer Formulierung

> *Enthält eine Teilmenge von* ω *das Element* $\emptyset$ *und mit jedem z auch* $z \cup \{z\}$, *so ist sie gleich* ω, das heißt, *jede induktive Teilmenge von* ω *ist gleich* ω.

Das aber folgt sofort aus der Definition von ω.

Schließlich lassen sich jetzt auch Mächtigkeitsbegriffe exakt definieren: Zwei Mengen x, y heißen *gleichmächtig*, wenn es eine bijektive Funktion von x auf y gibt. Eine Menge ist *endlich*, wenn sie gleichmächtig zu einem Element von ω ist; sie heißt *abzählbar unendlich*, falls sie gleichmächtig zu ω ist, und *überabzählbar*, falls sie weder endlich noch abzählbar unendlich ist. Man kann leicht zeigen, daß zwei verschiedene Elemente von ω nicht gleichmächtig sind. Die VON NEUMANNschen natürlichen Zahlen repräsentieren also in eindeutiger Weise die endlichen Mengen und heißen in diesem Sinn auch die *endlichen Kardinalzahlen*. Die Menge ω ist die kleinste unendliche Kardinalzahl.

4. Mengenlehre mit Klassen. Um die RUSSELLsche Antinomie zu vermeiden, hat ZERMELO das FREGEsche Komprehensionsaxiom zum Aussonderungsaxiom abgeschwächt. Seine Intention ging dahin, das Universum der Mengen „von unten" aufzubauen und keine Mengenbildung „quer durch das Universum" zuzulassen. Doch kann man die intuitiv so einleuchtende Vorstellung, die hinter dem Komprehensionsaxiom steht, beibehalten, wenn man einige Vorsichtsmaßnahmen einbaut. Man verlangt nicht, daß die Komprehension von Mengen wieder zu *Mengen* führt, sondern gesteht zu, daß dabei auch „zu große Mengen", „Unmengen" entstehen können, CANTORS absolut unendliche Vielheiten. Zur sprachlichen Unterscheidung nennt man die durch Komprehensionen mit Mengen entstehenden Objekte *Klassen*. So spricht man von der Klasse V aller Mengen (*Allklasse*),

$$V = \{x\colon x \text{ ist Menge}\},$$

oder auch von der *Klasse aller Gruppen* als der Klasse aller Mengen der Gestalt (x, f), wobei x eine nicht-leere Menge ist und $f\colon x \times x \to x$ eine Funktion, die die Gruppenaxiome erfüllt.

Klassen sind also ihrer Herkunft nach Umfänge oder Extensionen von Eigenschaften für Mengen. Daher sind zwar ihre Elemente Mengen, aber sie selbst brauchen keine Mengen zu sein. Diese Terminologie deckt sich nicht ganz mit dem *mathematischen* Sprachgebrauch: In der Mathematik spricht man zuweilen auch dort von *Klassen*, wo von vornherein eigentlich *Mengen* gemeint sind (Äquivalenz*klassen*, Rest*klassen*).

Das revidierte Komprehensionsaxiom sichert zu jeder Eigenschaft E von Mengen (die gewissen Bedingungen genügt; s. u.) die Existenz der *Klasse*

$$K_E = \{x\colon x \textit{ ist Menge, und } E \textit{ trifft zu auf } x\}.$$

Die Abweichung vom mathematischen Sprachgebrauch wird in gewisser Weise dadurch aufgehoben, daß aufgrund dieses Axioms jede Menge eine Klasse ist; denn eine Menge x läßt sich schreiben als $x = \{z\colon z \textit{ ist Menge und } z \in x\}$. Die Umkehrung gilt nicht allgemein. Denn für die „RUSSELLsche Klasse"

$$K_R = \{x\colon x \textit{ ist Menge und } x \notin x\}$$

gewinnt man mit

$$K_R \in K_R \Leftrightarrow K_R \textit{ ist Menge und } K_R \notin K_R$$

sofort, daß K_R keine Menge ist; sonst gälte ja $K_R \in K_R$ genau dann, wenn $K_R \notin K_R$. K_R ist damit eine *echte Klasse*. Diese Überlegung zeigt zugleich, wie man durch die Unterscheidung von Mengen und (echten) Klassen dem Widerspruch in der RUSSELLschen Antinomie entkommen kann.

Auch die anderen oben erwähnten Klassen sind echte Klassen: Wäre die Allklasse V eine Menge, so nach **Aus** auch die „RUSSELLsche Klasse" $K_R = \{x \in V\colon x \notin x\}$, und es ergäbe sich ein Widerspruch, nämlich die RUSSELLsche Antinomie. Unter Benutzung des Fundierungsaxioms ist der Beweis noch einfacher: Wäre V eine Menge, so gälte $V \in V$ im Widerspruch zu **Fund.** – Im Fall der Gruppen argumentiert man so: Da jede nicht-leere Menge Trägermenge einer Gruppe sein kann, ergibt sich für die Klasse G aller Gruppen „durch Nachrechnen", daß $(\bigcup\bigcup G) \cup \{\emptyset\} = V$. Wäre also G eine Menge, so auch V. – Übrigens bilden auch die Ordinalzahlen eine echte Klasse; denn andernfalls würde man auf die Antinomie von BURALI-FORTI geführt (vgl. 1.3).

Bei Axiomatisierungen einer Mengenlehre mit Klassen muß neben dem Umgang mit Mengen natürlich auch der Umgang mit Klassen und das Wechselspiel zwischen Klassen und Mengen axiomatisch geregelt werden. Wesentlich ist dabei das revidierte Komprehensionsaxiom. Ähnlich wie bei **Aus** und **Ers** werden nur elementar definierbare Eigenschaften zur Komprehension zugelassen. Die Mengenaxiome (etwa die von **ZFC**) werden teilweise modifiziert. So wird **Ext** jetzt allgemein für Klassen ausgesprochen. **Aus** besagt einfach, daß der Durchschnitt einer Menge mit einer Klasse wieder eine Menge ist. Die wichtigsten Systeme sind:

(i) die **NBG**-*Mengenlehre*, fußend auf Arbeiten von VON NEUMANN (1925 ff.) und aufgebaut wesentlich durch P. BERNAYS und durch K. GÖDEL (1937 ff.);

(ii) die KELLEY-MORSE-Mengenlehre (H. WANG (1949) und A. P. MORSE (1939 ff.), bekannt geworden durch den Anhang in J. L. KELLEYS Lehrbuch der Topologie [12]); sie unterscheidet sich von der **NBG**-Mengenlehre durch liberalere Bedingungen an die Definierbarkeit der Eigenschaften E im Komprehensionsaxiom.

Eine Mengenlehre mit Klassen erweist sich in der Mathematik dort als vorteilhaft, wo echte Klassen Gegenstand der Untersuchung werden, wie z. B. in der Kategorientheorie. Bis zu einem gewissen Grad handelt es sich dabei jedoch nur um Vorteile sprachlicher Natur. Für eine eingehende Diskussion verweisen wir auf LEVY [16].

§ 3. Einige metamathematische Aspekte

Was haben wir mit einer Axiomatisierung der Mengenlehre erreicht? Sicherlich haben wir eine präzise formulierte Basis für mengentheoretische Betrachtungen geschaffen und damit auch für die Mengenlehre den Standard einer axiomatisch aufgebauten deduktiven Theorie gewonnen. Gewarnt durch die Inkonsistenz des FREGEschen Axiomensystems, müssen wir uns allerdings die Frage stellen, ob die heute benutzten mengentheoretischen Axiomensysteme wirklich widerspruchsfrei sind. Bereits in 2.1 haben wir mit einem Verweis auf den sogenannten *zweiten Gödelschen Unvollständigkeitssatz* feststellen müssen, daß wir uns nicht mit einem Beweis von der Widerspruchsfreiheit überzeugen können. Wir können also höchstens intuitive Argumente vorbringen. Solche Argumente diskutieren wir in 1. Dabei klammern wir das Auswahlaxiom zeitweilig aus; ihm wenden wir uns ausführlich in 2. zu. In 3. streifen wir eine auch für die Mathematik nützliche methodologische Möglichkeit: Auf der Basis präziser mengentheoretischer Axiomensysteme läßt sich unter Umständen die Unlösbarkeit von Problemen aus Mathematik und Mengenlehre exakt beweisen.

1. Die von Neumannsche Hierarchie. Wir wählen für unsere Betrachtungen das Axiomensystem **ZFC**. Im Hinblick auf seine Widerspruchsfreiheit können wir zunächst darauf verweisen, daß ein jahrzehntelanger intensiver Umgang mit ihm nicht zu Widersprüchen geführt hat. (Abschwächend müssen wir allerdings eingestehen, daß wir schon morgen mit einem Widerspruch konfrontiert sein könnten.)

Weiter können wir vorbringen, daß die einzelnen **ZFC**-Axiome einsichtige Eigenschaften des intuitiven Mengenbegriffs spiegeln. Doch hierbei kann nicht ausgeschlossen werden, daß die *Gesamtheit* dieser *im einzelnen* einsichtigen Axiome zu Unverträglichkeiten führt. Zudem erscheint das System bei oberflächlicher Betrachtung vielleicht allzu sehr durch Einzelaspekte bestimmt und durch Zufälligkeiten geprägt.

Gegen die letzten Einwände gibt es jedoch ein *inhaltlich* überzeugendes Argument zumindest für **ZF**: der sogenannte *kumulativ-hierarchische* Charakter des Universums aller Mengen. Um ihn herauszuarbeiten, betrachten wir die sogenannten *von Neumannschen Stufen* V_α für die Ordinalzahlen $\alpha = 0, 1, 2, \ldots, \omega, \omega + 1, \ldots, \omega + \omega, \ldots$ (vgl. 1.3). Sie sind induktiv definiert durch

$$V_0 := \emptyset,$$

$$V_1 := Pot(\emptyset) = \{\emptyset\},$$

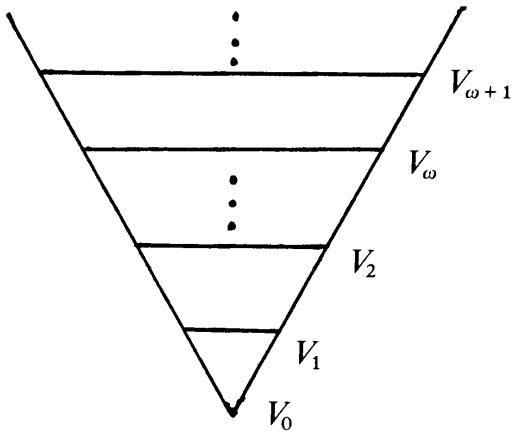
allgemein

$$V_{\alpha+1} := Pot(V_\alpha),$$

und für sogenannte *Limeszahlen*, wie ω oder $\omega + \omega$, die keinen unmittelbaren ordinalen Vorgänger haben, setzt man V_α gleich der Vereinigung aller vorangehenden V_β:

$$(*) \qquad V_\alpha = \bigcup_{\beta \in \alpha} V_\beta.$$

Die VON NEUMANNschen Stufen entstehen also dadurch, daß man, ausgehend von der leeren Menge, die Potenzmengenbildung über alle Ordinalzahlen iteriert und dabei an den Limeszahlstellen die Vereinigung gemäß (*) bildet. Sie konstituieren die *von Neumannsche Hierarchie.* Diese Hierarchie ist *kommutativ* in dem Sinn, daß jedes V_α Teilmenge aller späteren V_β ist:



Man kann nun in **ZF** beweisen, daß jede Menge Element eines V_α ist, das heißt, daß die VON NEUMANNsche Hierarchie das Universum der Mengen ausschöpft. Man kann darüber hinaus nach D. SCOTT (vgl. [4, S. 141 ff.]) sogar zeigen, daß **ZF** in einem präzisen Sinn gerade so stark ist, daß es einen kumulativ-hierarchischen Aufbau des Mengenuniversums sicherstellt. *Die Einsichtigkeit, die man einem solchen Aufbau zubilligen kann, überträgt sich damit auch auf das System* **ZF**.

2. Das Auswahlaxiom. Das Auswahlaxiom wird als ein Schlußprinzip zuerst um die Jahrhundertwende von den Italienern G. PEANO, R. BETTAZZI und B. LEVI erwähnt und kritisiert. Es hat aber bereits vorher Anwendung gefunden, so in mengentheoretischem Kontext bereits bei CANTOR und DEDEKIND. Die erste explizite Formulierung findet sich dann bei ZERMELO (1904) in der Form aus 2.2:

AC. *Zu jeder Menge gibt es eine Auswahlfunktion.* Äquivalent dazu: *Ist X eine nicht-leere Menge von nicht-leeren Mengen, so ist das direkte Produkt der Elemente von X nicht-leer* (denn es besteht ja gerade aus den Funktionen $f: X \to \bigcup X$ mit $f(x) \in x$ für $x \in X$, das heißt, aus den Auswahlfunktionen zu X).

Für sein Axiomensystem aus dem Jahre 1908 benutzt ZERMELO eine andere Form, deren Äquivalenz mit **AC** im gleichen Jahr durch B. RUSSELL gezeigt wurde:

AC′. *Ist X eine Menge zueinander disjunkter nicht-leerer Mengen, so gibt es eine Menge, die mit jedem Element von X genau ein Element gemeinsam hat.* Das heißt, *jede Äquivalenzrelation besitzt ein Repräsentantensystem.*

Zu „**AC** $\Rightarrow$ **AC′**“: Ist X eine Menge zueinander disjunkter nicht-leerer Mengen, so ist das Bild einer Auswahlfunktion zu X eine Menge mit den in **AC′** verlangten Eigenschaften.

Zu „**AC′** $\Rightarrow$ **AC**“: Sei X eine Menge, o. B. d. A. $\emptyset \notin X$. Über $\{(y, z)\colon y \in X \text{ und } z \in y\}$ werde eine Äquivalenzrelation $\sim$ definiert gemäß

$$(y, z) \sim (y', z') : \text{genau dann, wenn } y = y'.$$

Sei – nach **AC′** – S ein Repräsentantensystem von $\sim$. Dann ist S (Graph einer) Auswahlfunktion zu X. □

Es ist heute eine Fülle von Aussagen bekannt, die auf der Basis von **ZF** zu **AC** äquivalent sind. Für die Mathematik bedeutsam ist ein Lemma, das auf F. HAUSDORFF (1909, 1914) zurückgeht und durch Arbeiten M. ZORNS (1935) allgemein unter Mathematikern bekannt wurde:

ZORNsches Lemma. *Eine halbgeordnete Menge, in der jede linear geordnete Teilmenge eine obere Schranke hat, besitzt (mindestens) ein maximales Element.*

Das Auswahlaxiom erscheint in der Formulierung **AC** intuitiv sehr einsichtig. Dennoch hat es in Mengenlehre und Mathematik zu Kontroversen Anlaß gegeben. Lehrreich (und erfrischend!) sind in diesem Zusammenhang die Ausführungen ZERMELOS in [20]. Wir erwähnen einige Punkte, die die Sonderstellung des Auswahlaxioms beleuchten und die kritische Einstellung ihm gegenüber erklärlich machen.

(a) „*Inkonstruktivität*". Die **ZF**-Axiome sind so formuliert oder lassen sich – was **Inf** betrifft – so formulieren, daß die Mengen, deren Existenz in den Axiomen gefordert wird (die Paarmenge bei **Paar**, die Potenzmenge bei **Pot**, ω beim umformulierten Unendlichkeitsaxiom etc.), *explizit* unter Benutzung etwaiger Ausgangsmengen *definiert werden können.* Beim Auswahlaxiom ist das *nicht* der Fall; **AC** verlangt nicht die Existenz von (in einem vernünftigen Sinn) *definierbaren* Auswahlfunktionen, **AC′** nicht die Existenz von *definierbaren* Repräsentantensystemen. Man kann sich diesen (nachweisbar unvermeidlichen) „inkonstruktiven" Charakter des Auswahlaxioms exemplarisch vor Augen führen, wenn man etwa versucht, auf $Pot(\mathbb{R})$ eine Auswahlfunktion zu definieren, oder wenn man für die Äquivalenzrelation auf $\mathbb{R}$, gemäß der genau die Zahlen mit rationaler Differenz äquivalent sind, ein Repräsentantensystem definieren möchte.

Solche Repräsentantensysteme liefern Beispiele von Mengen reeller Zahlen, die nicht LEBESGUE-meßbar sind (G. VITALI 1905). Man macht hierbei wesentlich von **AC** Gebrauch; denn man kann nach SOLOVAY [17] umgekehrt zeigen, daß eine schwächere Form des Auswahlaxioms, die für Analysis und Maßtheorie ausreicht, mit der Forderung verträglich ist, daß jede Teilmenge von $\mathbb{R}$ LEBESGUE-meßbar ist.

Die „Inkonstruktivität" von **AC** hat zur Folge, daß Beweise, die wesentlich das Auswahlaxiom oder das ZORNsche Lemma benutzen, in einem weiten Sinn ebenfalls völlig „inkonstruktiv" sind. Zum Beispiel liefert der übliche Beweis dafür, daß jeder Vektorraum eine Basis besitzt, keine Andeutung, wie eine solche Basis im Einzelfall aussehen könnte. So garantiert etwa **ZFC** nicht die Existenz einer definierbaren HAMELbasis, das heißt, einer definierbaren Basis von $\mathbb{R}$ als Vektorraum über $\mathbb{Q}$.

(b) „*Paradoxien*". **AC** hat – im Zusammenwirken mit den übrigen Axiomen! – einige paradox anmutende Konsequenzen. Wir erwähnen das *Kugelparadoxon* von TARSKI und BANACH (1924): Die massive Einheitskugel läßt sich so in endlich viele Teile zerlegen, daß aus diesen Teilen zwei neue Einheitskugeln zusammengesetzt

werden können. (Natürlich sind die Teile nicht meßbar, so daß man die Ausgangskugel sicherlich nicht entsprechend zersägen kann!)

Um die Diskussion voranzubringen, hat man – ähnlich wie beim Parallelenaxiom in der euklidischen Geometrie – Untersuchungen darüber angestellt, ob das Auswahlaxiom vielleicht aus den übrigen Axiomen beweisbar oder widerlegbar sei. Beides trifft nicht zu (COHEN 1963 bzw. GÖDEL 1938). Dabei sei die Widerspruchsfreiheit von **ZF** unterstellt. Insbesondere wird also durch die Hinzunahme von **AC** die Widerspruchsfreiheit von **ZF** nicht zerstört. (Sonst wäre **AC** in **ZF** widerlegbar.) Diese beweistheoretische Rechtfertigung stärkt auch die Mathematik in ihrem Urteil: Sie hat sich, wie vielfältige und weittragende Anwendungen des ZORNschen Lemmas und anderer Äquivalente in den verschiedensten Disziplinen zeigen, *für* das Auswahlaxiom entschieden.

In einigen Fällen erweist sich übrigens das Auswahlaxiom als entbehrlich. So kann man ohne **AC** zeigen, daß jede *endliche* Menge eine Auswahlfunktion besitzt. Für abzählbare Mengen oder Mengen von abzählbaren Mengen gilt das im allgemeinen nicht mehr. Auf Mengen von Mengen natürlicher Zahlen kann eine Auswahlfunktion dadurch definiert werden, daß man aus jeder Menge von Zahlen die kleinste Zahl wählt, und dies läßt sich ohne **AC** bewerkstelligen. Man kann überdies allgemein zeigen (und zwar unter Benutzung der GÖDELschen konstruktiblen Mengen), daß **AC** zum Beweis *zahlentheoretischer* Sätze an keiner Stelle benötigt wird. Nahezu erschöpfende Auskunft über das Auswahlaxiom erteilt [11].

3. Unabhängigkeitsbeweise. Es gibt in der Mathematik eine Reihe von Problemen, die trotz intensiven Bemühens bislang noch nicht gelöst werden konnten. So z. B. die FERMATsche Vermutung, der zufolge für $n \geqslant 3$ und positive natürliche Zahlen a, b, c stets $a^n + b^n \neq c^n$. Der Mißerfolg in der Bearbeitung solcher Probleme kann trivial bedingt sein: Lösungen sind jederzeit möglich, aber einfach noch nicht gefunden worden. Daneben kommen aber auch tiefere Gründe in Frage, etwa Gründe *komplexitätstheoretischer* Natur. So können zwar Lösungen existieren, aber jeder Lösungsweg ist viel zu lang, als daß er in absehbarer Zeit zum Ziele führen könnte. Überlegungen dieser Art werden insbesondere durch neuere Ergebnisse der *Komplexitätstheorie* nahegelegt (vgl. etwa [18, S. 127 ff.]). Schließlich kann die Ursache *prinzipieller* Natur sein in dem Sinne, daß eine Lösung grundsätzlich nicht möglich ist.

Entsprechende Ergebnisse haben zur Voraussetzung, daß eine methodische Basis für die Mathematik vorliegt, auf die man sich beziehen kann. In den Axiomensystemen der Mengenlehre, wie z. B. **ZFC**, stehen uns solche Basen zur Verfügung. In der Tat hat man mittlerweile starke Methoden entwickeln können, um Unbeweisbarkeitsresultate zu erzielen. Im wesentlichen handelt es sich dabei um die bereits mehrfach erwähnten Methoden der konstruktiblen Mengen und des Erzwingens (Forcing). Eine Darstellung findet sich in [14]. Die bislang angeführten Unbeweisbarkeitsresultate – z. B. im vorangehenden Abschnitt – bedienen sich durchweg dieser Techniken.

Einer der ersten wesentlichen Erfolge ist der in der Einleitung erwähnte Unabhängigkeitsbeweis für die Kontinuumshypothese. CANTOR hat sich immer wieder um einen Beweis der Hypothese bemüht. Mehrfach hat er sich zuversichtlich geäußert [2, S. 192, 244], zumal er Teilergebnisse erzielen konnte, so z. B. den

Beweis für offene und für abgeschlossene Zahlenmengen. HILBERT stellte das Kontinuumproblem an die erste Stelle einer Liste von dreiundzwanzig ihm zukunftsträchtig erscheinenden offenen Fragen, über die er im Jahre 1900 auf dem internationalen Mathematikerkongreß in Paris vortrug. Würde doch ein Beweis der Kontinuumshypothese zeigen, daß das Kontinuum die kleinste überabzählbare Mächtigkeit besitzt, und damit helfen, die Kluft zum Abzählbaren zu überbrücken. Der Unabhängigkeitsbeweis zeigt uns, daß CANTOR scheitern mußte!

Ein weiteres Beispiel für eine unabhängige Aussage ist die sogenannte *Souslinsche Hypothese*, der zufolge sich die Ordnung der reellen Zahlen dadurch charakterisieren läßt, daß sie dicht ohne erstes und letztes Element und vollständig ist und daß sie zudem kein überabzählbares System zueinander disjunkter offener Intervalle zuläßt.

Man kann bislang nicht ausschließen, daß auch die FERMATsche Vermutung unabhängig von **ZFC** ist. Doch anders als bei der Kontinuumshypothese oder der SOUSLINschen Hypothese würde ein Unabhängigkeitsbeweis zugleich die Richtigkeit bestätigen. Ist nämlich die FERMATsche Vermutung falsch, so existiert ein Gegenbeispiel, das sich auf der Basis von **ZFC** „durch Nachrechnen" als Gegenbeispiel verifizieren läßt. Unabhängigkeit kann also nur dann vorliegen, wenn die Vermutung wahr ist. Eine ähnliche Schlußweise erlauben alle Aussagen über natürliche Zahlen, die aus einer Reihe von Allquantoren bestehen, denen ein quantorenfreier Kern, wie z. B. eine diophantische (Un-)Gleichung, folgt. Und ähnlich kann man auch bei Aussagen schließen, die zu solchen Aussagen äquivalent sind. Dazu gehören z. B. die GOLDBACHsche Vermutung („*Jede gerade Zahl* $\geqslant 4$ *ist Summe zweier Primzahlen*") und – trotz ihres analytischen Aussehens – auch die RIEMANNsche Vermutung.

Epilog

Wie wir exemplarisch gesehen haben, reichen die gegenwärtigen Axiomensysteme der Mengenlehre aus, um mengentheoretische Modelle für die Objekte der Mathematik anzugeben und den methodischen Umgang mit ihnen nachzuvollziehen. Ein darauf gegründetes Verständnis der Mathematik ist nicht nur hilfreich für eine Klärung ihrer Begriffe, es öffnet auch die Tür zu einem reichen Vorrat an mengentheoretischen Methoden, und es schafft eine einheitliche axiomatische Basis für die Mathematik.

Wie tragfähig ist diese Basis? Ihre Widerspruchsfreiheit ist nicht beweisbar; wir können nur intuitive Argumente für sie ins Feld führen, wie etwa die Natürlichkeit der VON NEUMANNschen kumulativen Hierarchie. Unterstellen wir wie bisher, sie *sei* widerspruchsfrei. Wie weit trägt sie dann? Wir haben Grenzen kennengelernt, etwa die Unabhängigkeit von Kontinuumshypothese und SOUSLINscher Hypothese. Es ließen sich hier noch viele andere Beispiele anführen.

Auch diese *Unvollständigkeit* mengentheoretischer Axiomensysteme ist nach einem Satz von GÖDEL unausweichlich (vgl. [5, S. 226 ff.]). Wir können sie nur mildern, indem wir etwa konkret versuchen, **ZFC** um intuitiv einsichtige Axiome zu erweitern. Eine Fülle von entsprechenden Vorschlägen ist in diesem Zusammen-

hang bereits diskutiert worden. Noch haben sich keine allseits überzeugenden neuen Prinzipien herausgeschält. Angesichts dieser Situation könnte man übereinkommen, verschiedene einsichtige oder methodisch interessante, vielleicht gar miteinander unverträgliche Erweiterungen nebeneinander zu dulden. Die Geometrie hat uns gezeigt, wie fruchtbar eine solche Entwicklung sein kann.

Eine radikalere Vorgehensweise könnte es sich zum Ziele setzen, bei der Grundlegung der Mathematik die Mengenlehre CANTORscher Prägung überhaupt abzulösen. Ernsthafte und interessante Versuche liegen bereits vor; so etwa von kategorieller Seite oder in der sogenannten *alternativen Mengenlehre* [19], die sich an den Bedürfnissen der Nichtstandardanalysis orientiert. Wie weit es ihnen gelingen wird, der CANTORschen Mengenlehre, die ja gerade in voller Frische in das zweite Jahrhundert ihrer Existenz eingetreten ist, zur Konkurrenz zu gereichen, wird wohl erst die (fernere?) Zukunft entscheiden.

Literatur

[1] BOLZANO, B.: Paradoxien des Unendlichen. Leipzig 1851
[2] CANTOR, G.: Gesammelte Abhandlungen mathematischen und philosophischen Inhalts. Herausgegeben von E. Zermelo. Berlin 1933
[3] DEDEKIND, R.: Gesammelte mathematische Werke. Herausgegeben von R. Fricke, E. Noether, Ö. Ore. Braunschweig 1932
[4] EBBINGHAUS, H.-D.: Einführung in die Mengenlehre. Darmstadt 1976
[5] EBBINGHAUS, H.-D., FLUM, J., THOMAS, W.: Einführung in die Mathematische Logik. Darmstadt 1978
[6] FELSCHER, W.: Naive Mengen und abstrakte Zahlen I, III. Mannheim 1978/9
[7] FREGE, G.: Grundgesetze der Arithmetik I, II. Jena 1893/1903
[8] HALMOS, P. R.: Naive Mengenlehre. Göttingen 1968
[9] HEYTING, A.: Intuitionism. An Introduction. Amsterdam 1956
[10] HILBERT, D.: Über das Unendliche. Math. Ann. 95 (1926)
[11] JECH, J.: The Axiom of Choice. Amsterdam 1973
[12] KELLEY, J. L.: General Topology. Princeton 1955
[13] KRONECKER, L.: Über einige Anwendungen der Modulsysteme und elementare algebraische Fragen. J. Reine Angew. Math. 99 (1886)
[14] KUNEN, K.: Set Theory. An Introduction to Independence Proofs. Amsterdam 1980
[15] LEVY, A.: Basic Set Theory. Heidelberg 1979
[16] LEVY, A.: The Role of Classes in Set Theory. In: Sets and Classes (herausgegeben von G. H. Müller). Amsterdam 1976
[17] SOLOVAY, R. M.: A Model of Set Theory in Which Every Set of Reals is Lebesgue-Measurable. Ann. Math. 92 (1970)
[18] SPECKER, E., V. STRASSEN: Komplexität von Entscheidungsproblemen. Ein Seminar. Heidelberg 1976
[19] VOPĚNKA, P.: Mathematics in the Alternative Set Theory. Leipzig 1979
[20] ZERMELO, E.: Neuer Beweis für die Möglichkeit einer Wohlordnung. Math. Ann. 65 (1908)
[21] ZERMELO, E.: Untersuchungen über die Grundlagen der Mengenlehre. I. Math. Ann. 65 (1908)

Namenverzeichnis

Sachverzeichnis

Porträts berühmter Mathematiker

Leonhard EULER (1707 - 1783)

Carl Friedrich GAUSS (1777 - 1855)

William Rowan HAMILTON
(1805 - 1865)

Richard DEDEKIND (1831 - 1916)

Federzeichnungen von Martina Koecher

Georg CANTOR (1845 - 1918)

Ferdinand Georg FROBENIUS
(1849 - 1917)

Heinz HOPF (1894 - 1971)

Abraham ROBINSON (1918 - 1972)