

## Einleitung

Wer im Katalog der Mannheimer Universitätsbibliothek das Stichwort *Analysis* eingibt, erhält mehrere Tausend Treffer, wird aber schnell feststellen, daß nur ein Bruchteil der gefundenen Bücher etwas mit Mathematik zu tun hat. Beim großen „Rest“ handelt es sich praktisch ausschließlich um Bücher in englischer Sprache. Dort heißt *analysis* nicht nur Analysis, sondern vor allem auch Analyse, und dieses Wort kommt auch im Deutschen in praktisch jeder Wissenschaft vor, von der Textanalyse über die chemische Analyse und die Analyse politischer oder sozialer Verhältnisse sowie der Arbeit von Analysten an der Börse bis hin zur Psychoanalyse.

Analysis und Analyse sind beide abgeleitet vom griechischen Wort  $\alpha\nu\lambda\upsilon\epsilon\iota\nu$  = auflösen, zerlegen: Bei der chemischen Analyse soll bestimmt werden, welche Stoffe in einem Gemisch enthalten sind, es soll also –zumindest gedanklich – in seine Bestandteile zerlegt werden. Ähnlich soll bei der Analyse wirtschafts- und sozialwissenschaftlicher Zusammenhänge ein (im Allgemeinen komplizierter) Sachverhalt aufgelöst werden in einzelne Fakten, Hintergründe, Entscheidungen und so weiter, und bei der Psychoanalyse soll die Persönlichkeit des Patienten dadurch verstanden werden, daß die Rolle entscheidender Ereignisse aus seinem Leben betrachtet wird.

Auch hier in der Analysis wird es um eine Art von Zerlegung gehen: Wir betrachten Funktionen, mit denen wir z.B. den zeitlichen Verlauf irgendeiner Größe beschreiben wollen. Bei interessanten Größen wie Aktienkursen oder volkswirtschaftlichen Kenngrößen sind diese Funktionen sehr kompliziert; beschränkt man sich jedoch auf ein sehr kurzes Zeitintervall, ist zumindest bei manchen Funktionen mit einem deutlich einfacheren Verhalten zu rechnen. Die Analysis versucht, aus dem Ver-

halten in solchen kleinen Teilintervallen möglichst viele Informationen über die Funktion selbst zu gewinnen. In der Vorlesung *Analysis I* sollen die grundlegenden Werkzeuge für solche Untersuchung vorgestellt werden

Den Umgang mit einem Werkzeug lernt man nur durch dessen Gebrauch: Genauso wie niemand zum Schlosser wird, indem er Bilder von Bohrmaschinen betrachtet und Abhandlungen über das Drehmoment liest, wird niemand, der einfach mathematische Formeln und Sätze auswendig lernt, damit erfolgreich Probleme aus Wirtschafts-, Natur- oder Sozialwissenschaften lösen.

Zum einen beginnen Probleme der wirklichen Welt nie mit Fragestellungen wie „ $f: \mathbb{R}^3 \rightarrow \mathbb{R}$  sei eine mindestens zweimal stetig differenzierbare Funktion“; der Anwender muß sich stets zunächst überlegen, ob sich sein Problem überhaupt mit Mathematik lösen läßt, und wenn ja, mit welcher.

Die Antwort darauf wird selten eindeutig sein: Gelegentlich wird sich dasselbe Problem sowohl mit als auch ohne Mathematik modellieren lassen, und auch wenn man sich für eine mathematische Lösung entscheidet, steht selten eindeutig fest, wie es weiter geht:

Realistische Probleme (im Gegensatz zu Übungs- und Klausuraufgaben) sind fast immer zu komplex für eine vollständige formale Beschreibung. Vor ihrer Übersetzung in Mathematik müssen sie daher vereinfacht werden, wobei aber natürlich die für die jeweilige Problemstellung wesentlichen Aspekte erhalten bleiben müssen. Beispielsweise wird man oft Größen, die ihrem Wesen nach ganzzahlig sind, mit reelle Zahlen modellieren, da die Mathematik der reellen Zahlen erheblich einfacher ist als die der ganzen Zahlen und auch deutlich mehr Methoden zur Verfügung stellt. So sind beispielsweise Geldbeträge stets ganzzahlige Vielfache der kleinsten Währungseinheit, aber wenn es um ein Bruttoinlandsprodukt oder die Umsatzprognose eines Industriebetriebs geht, wird man Veränderungen um einen Cent doch eher als kontinuierlich denn als Sprung empfinden.

Die mathematischen Werkzeuge zur Lösung eines Problems hängen wesentlich ab von solchen Modellierungsentscheidungen: Modellierung durch kontinuierliche Größen verlangt typischerweise analytische Methoden, oft Differentialgleichungen; beim Modellieren mit ganzen

Zahlen ist man in der diskreten Mathematik, wo eher algebraische und zahlentheoretische Methoden im Vordergrund stehen – von denen zumindest ein Teil übrigens auch analytisch ist.

Selbst wenn die Übersetzung eines Problem in Mathematik (oder besser seine Annäherung durch Mathematik) feststeht, gibt es im Allgemeinen verschiedene mathematische Ansätze, die theoretisch allesamt zu korrekten Lösungen führen; in der Praxis kann sich aber der Rechenaufwand zweier Verfahren so stark unterscheiden, daß nur eines der beiden wirklich durchführbar ist, oder aber so, daß zwar beide durchführbar sind, das eine aber erheblich mehr kostet als das andere.

Für einen Anwender geht es somit nicht in erster Linie darum, ob die Voraussetzungen für einen bestimmten mathematischen Satz erfüllt sind: Diese Frage stellt sich erst viel später. Als erstes muß er klären, welche mathematischen Modelle in Frage kommen und welche davon zu Lösungen mit realistischem Aufwand führen. Wer hier erfolgreich sein will, muß vor allem ein Gefühl für die Tragweite und den Aufwand mathematischer Methoden entwickeln.

Dieses Gefühl kann man – wie auch die richtige Technik für den Umgang mit einem Hammer oder einer Feile – nur durch praktische Erfahrung erwerben. Ein wesentlicher Bestandteil dieser Vorlesung sind daher die Übungen, die – im Rahmen dessen, was im ersten Semester möglich ist – die notwendige Erfahrung dazu vermitteln sollen. Hauptziel der Vorlesung ist also, daß Sie praktische Fähigkeiten entwickeln, und das ist nur möglich, indem Sie selbst praktische Erfahrung sammeln. Die Übungen sind somit mindestens genauso wichtig wie die Vorlesung; ohne regelmäßiges Üben ist ein erfolgreicher Besuch dieser Vorlesung praktisch unmöglich.

## Literatur

Es gibt sehr viele Bücher, bei denen *Analysis I* im Titel vorkommt; praktisch jedes davon kann parallel zur Vorlesung verwendet werden. Bei den folgenden beiden Büchern ist die Stoffauswahl und Präsentation ähnlich zu der in der Vorlesung:

KLAUS FRITZSCHE: Grundkurs Analysis I, *Spektrum*, <sup>2</sup>2008

ROLF WALTER: Einführung in die Analysis I, *de Gruyter*, 2007

Sehr knapp findet man alle wesentlichen Teile des Stoffs dargestellt in

OTTO FORSTER: Analysis I, *Springer*, <sup>9</sup>2008

Dazu gibt es auch einen Begleitband mit (teilweise gelösten) Übungsaufgaben:

OTTO FORSTER, RÜDIGER WESSOLY: Übungsbuch zur Analysis I, *Springer*, <sup>4</sup>2008

Wer auch die englische Fachterminologie kennen lernen möchte, kann zum Beispiel in das sehr schön geschriebene Buch

STEPHEN ABBOT: Understanding Analysis, *Springer*, 2001

hineinschauen; für ein Kontrastprogramm zu dieser Vorlesung empfiehlt sich der sehr allgemein und abstrakt geschriebene Klassiker

JEAN DIEUDONNÉ: Grundzüge der modernen Analysis I, *Vieweg*, <sup>3</sup>1986

(Es gibt noch acht weitere Bände, alle auch auf englisch und französisch erhältlich. Wer die alle verstanden hat, verfügt über weit über Bachelor- und Masterniveau hinausgehende Analysiskenntnisse.)

## Kapitel 1 Zahlen

Im Anfang waren die natürlichen Zahlen  $1, 2, 3, \dots$ . Man kann sie addieren und multiplizieren, Subtraktion, Division und Wurzelziehen dagegen sind nicht immer möglich. Aus dem Bestreben, auch solche Operationen möglichst ohne Einschränkungen durchführen zu wollen, entstanden im Laufe der Zeit weitere Zahlbereiche wie die ganzen, rationalen, reellen und komplexen Zahlen. In diesem ersten Kapitel sollen diese kurz vorgestellt werden.

### §1: Mengen

Ein Zahlbereich besteht aus vielen Zahlen; wir sagen, er sei eine *Menge* von Zahlen. Den Begriff „Menge“ verwenden wir dabei genau so, wie ihn GEORG CANTOR 1895 zu Beginn einer Arbeit in den Mathematischen Annalen (Band 46, S. 481) eingeführt hat:

Unter einer 'Menge' verstehen wir jede Zusammenfassung  $M$  von bestimmten wohlunterschiedenen Objecten  $m$  unsrer Anschauung oder unseres Denkens (welche die 'Elemente' von  $M$  genannt werden) zu einem Ganzen.

Wer diese Definition für eher vage hält, hat sicherlich nicht unrecht; in der Tat zeigte CANTOR selbst schon 1897, daß der unkritische Umgang mit diesem Mengenbegriff zu Widersprüchen führen kann. Zu Beginn des zwanzigsten Jahrhunderts entstand deshalb die sogenannte *axiomatische Mengenlehre*, eine auf einem umfangreichen Axiomensystem beruhende präzise Theorie, die allerdings den Nachteil hat, recht formal und unanschaulich zu sein. Da wir Mengen in dieser Vorlesung nur als bequeme Sprechweise verwenden wollen, lohnt sich der mit dem axiomatischen Zugang verbundene Aufwand nicht; hier soll das Wort *Menge* nur im naiven Sinne der CANTORSchen Definition verwendet werden.



GEORG FERDINAND LUDWIG PHILIPP CANTOR (1845–1918) wurde in Sankt Petersburg geboren als Sohn eines dänischen Kaufmanns und einer Russin. 1856 zog die Familie nach Deutschland und CANTOR besuchte zunächst die Realschule, dann die Höhere Gewerbeschule in Darmstadt. Ab 1860 studierte er zunächst Ingenieurwissenschaften an der ETH Zürich, 1862 wechselte er dann zum Mathematikstudium an die Universität Zürich und 1863 nach Berlin, wo er 1867 promovierte. Danach unterrichtete er zunächst an einer Berliner Mädchenschule; 1869 ging er an der Universität Halle, wo er sich gleich darauf habilitierte.

Wie seine Dissertation beschäftigte sich auch seine Habilitationsschrift mit einem Thema aus der Zahlentheorie; danach arbeitete er jedoch vor allem auf dem Gebiet der Analysis. 1872 wurde er außerplanmäßiger Professor in Halle, 1879 erhielt er dort einen Lehrstuhl. Zwischen 1879 und 1884 veröffentlichte er eine Serie von sechs Arbeiten in den Mathematischen Annalen, in denen er die Grundlagen der Mengenlehre darstellte.

Die Objekte unserer Anschauung oder unseres Denkens, die wir zu einem Ganzen zusammenfassen, werden hier meist Zahlen sein. Mit CANTOR benutzen wir geschweifte Klammern, um die Elemente zu einer Menge zusammenzufassen, also zum Beispiel

$$M = \{2, 3, 5, 7\}$$

für die Menge, die aus den Zahlen 2, 3, 5 und 7 besteht. Um auszudrücken, daß ein Objekt  $m$  Element einer Menge  $M$  ist, verwenden wir das einen griechischen  $\varepsilon$  nachempfundene Zeichen „ $\in$ “ und schreiben  $m \in M$ ; in Worten  $m$  Element  $M$  oder auch kurz  $m$  aus  $M$ . Falls  $m$  kein Element von  $M$  ist, wollen wir dies kurz in der Form  $m \notin M$  ausdrücken. Im Falle obiger Menge  $M$  ist also  $5 \in M$ , aber  $11 \notin M$ .

Eine Menge muß nicht notwendigerweise ein Element enthalten; die Menge, die überhaupt kein Element enthält, bezeichnen wir als die *leere Menge* und schreiben sie als  $\emptyset = \{\}$ .

Mengen müssen auch nicht endlich sein; Beispiel einer unendlichen Menge ist etwa die Menge *aller* natürlicher Zahlen. Diese wird üblicherweise mit dem Symbol  $\mathbb{N}$  bezeichnet, d.h.

$$\mathbb{N} = \{1, 2, 3, 4, 5, \dots\}.$$

Oft definieren wir Mengen auch durch Eigenschaften ihrer Elemente; die Menge aller geraden Zahlen können wir beispielsweise schreiben

als

$$G = \{n \in \mathbb{N} \mid n \text{ ist durch } 2 \text{ teilbar}\};$$

das allgemeine Schema ist

$$N = \{n \in M \mid n \text{ hat Eigenschaft } E\}$$

für die Menge aller Elemente  $n$  einer Menge  $M$ , die zusätzlich noch die Eigenschaft  $E$  haben. Auch endliche Mengen lassen sich so definieren; beispielsweise ist

$$\{2, 3, 5, 7\} = \{n \in \mathbb{N} \mid n \text{ ist prim und kleiner als } 10\}.$$

Da Mengen als Zusammenfassungen ihrer Elemente definiert sind, betrachten wir zwei Mengen als gleich, wenn sie dieselben Elemente haben. So ist beispielsweise

$$\{2, 3, 5, 7\} = \{7, 5, 3, 2\} = \{3, 5, 7, 2\};$$

auf die Reihenfolge, in der die Elemente aufgezählt werden, kommt es also nicht an.

Falls jedes Element einer Menge  $N$  auch Element einer anderen Menge  $M$  ist, bezeichnen wir  $N$  als eine *Teilmenge* von  $M$  und schreiben  $N \subseteq M$ . Falls umgekehrt nicht jedes Element von  $M$  bereits in  $N$  liegt, bezeichnen wir  $N$  als eine *echte* Teilmenge von  $M$  und schreiben  $N \subset M$ . So ist beispielsweise  $\{2, 3, 5, 7\} \subset \mathbb{N}$ , denn 2, 3, 5 und 7 sind natürliche Zahlen und selbstverständlich gibt es außer diesen vier noch viele weitere.

Aus zwei Mengen  $M$  und  $N$  lassen sich verschiedene neue Mengen konstruieren; am wichtigsten für uns sind

- Die *Vereinigungsmenge*  $M \cup N$ : Ein Element  $m$  liegt in  $M \cup N$ , wenn es in  $M$  oder in  $N$  (oder in beiden) liegt.
- Der *Durchschnitt*  $M \cap N$ : Ein Element  $m$  liegt in  $M \cap N$ , wenn es sowohl in  $M$  als auch in  $N$  liegt.
- Die *Differenzmenge*  $M \setminus N$ : Ein Element  $m$  liegt in  $M \setminus N$ , wenn es in  $M$ , nicht aber in  $N$  liegt.

Bei mehr als zwei Mengen schreiben wir

$$\bigcup_{i=1}^r M_i = M_1 \cup M_2 \cup \dots \cup M_r$$

für die Vereinigungsmenge, bestehend aus allen Elementen, die in mindestens einer der Mengen  $M_i$  liegen, und

$$\bigcap_{i=1}^r M_i = M_1 \cap M_2 \cap \dots \cap M_r$$

für den Durchschnitt, bestehend aus allen Elementen, die in *jeder* der Mengen  $M_i$  liegen.

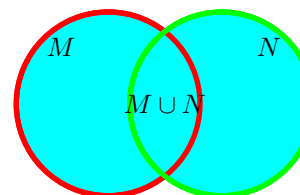
Für die Mengen  $M = \{2, 3, 5, 7\}$  und  $N = \{1, 3, 5, 7, 9\}$  ist also

$$M \cup N = \{1, 2, 3, 5, 7, 9\},$$

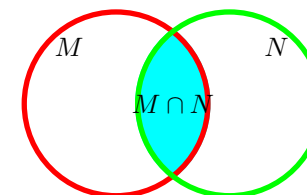
$$M \cap N = \{3, 5, 7\},$$

$$M \setminus N = \{2\} \quad \text{und} \quad N \setminus M = \{1, 9\}.$$

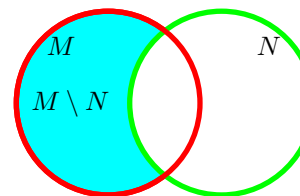
In der folgenden Zeichnung ist  $M$  die Menge aller Punkte im Innern des roten Kreises,  $N$  die im grünen. Vereinigung, Durchschnitt und die beiden Differenzen sind jeweils in cyan eingefärbt:



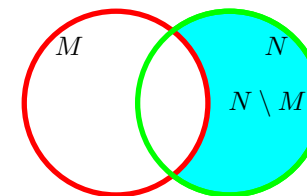
Die Vereinigung  $M \cup N$



Der Durchschnitt  $M \cap N$



Die Differenz  $M \setminus N$



Die Differenz  $N \setminus M$

## §2: Der Körper der rationalen Zahlen

Wenn wir Zahlen nicht nur addieren, sondern auch subtrahieren wollen,

müssen wir negative Zahlen zulassen; wenn wir auch dividieren wollen, brauchen wir Brüche. Während Brüche bereits vor mindestens vier Tausend Jahren den Babyloniern bekannt waren, sind negative Zahlen historisch betrachtet noch ziemlich jung: Erst im sechzehnten Jahrhundert setzte sich langsam durch, daß Zahlen auch negativ sein konnten; als gleichberechtigt mit den positiven Zahlen treten sie wohl erstmalig 1545 in CARDANOS Buch *Ars magna* auf.



GIROLAMO CARDANO (1501–1576) war ein italienischer Mathematiker, Arzt und Naturforscher. Sein Vater war Rechtsanwalt, war aber sehr an Mathematik interessiert. Er diskutierte mit LEONARDO DA VINCI über Geometrie und brachte auch seinem Sohn Mathematik bei. Dieser arbeitete zunächst als Assistent seines Vaters, begann dann aber an der Universität Pavia und später Padua ein Medizinstudium. Nach dem Tod seines Vaters hatte er dessen Vermögen schnell durchgebracht und hielt sich mit Glücksspiel über Wasser, wobei ihm seine guten Kenntnisse der Wahrscheinlichkeitstheorie halfen. 1525 promovierte er in Medizin und eröffnete eine nicht

sonderlich erfolgreiche Praxis in einem kleinen Dorf nahe Padua. 1532 zog er nach Mailand, wo er bei der Piatti Stiftung eine Stelle als Mathematiklehrer bekam. Mehrere seiner Schüler und Kollegen waren auch seine Patienten, und es gelang ihm relativ schnell einen ausgezeichneten Ruf als Arzt aufzubauen. Mathematisch beschäftigte er sich vor allem mit der Lösung von Gleichungen dritten und vierten Grades; 1545 veröffentlichte er sein Buch *Ars Magna*, in dem er die auf TARTAGLIA und CARDANOS Diener FERRARI zurückgehenden Lösungsformeln präsentierte, 1552 erhielt er einen Lehrstuhl für Medizin an der Universität Pavia, beschäftigte sich aber weiterhin auch mit mathematischen und naturwissenschaftlichen Fragen.

Wir wollen hier nicht der historischen Entwicklung folgen, sondern erweitern die natürlichen Zahlen zunächst um die Null zur Menge

$$\mathbb{N}_0 = \{0\} \cup \mathbb{N} = \{0, 1, 2, 3, \dots\}$$

und weiter um die negativen Zahlen zur Menge

$$\mathbb{Z} = \{-3, -2, -1, 0, 1, 2, 3, \dots\}$$

der *ganzen* Zahlen. Als *rationale* Zahl bezeichnen wir einen Bruch der Form

$$\frac{p}{q} \text{ mit } p \in \mathbb{Z} \text{ und } q \in \mathbb{N}.$$

Dadurch, daß wir für den Nenner nur natürliche Zahlen zulassen, ist insbesondere sichergestellt, daß kein Nenner Null auftreten kann.

Wie aus der Schule bekannt ist, können Brüche gekürzt und erweitert werden; zwei Brüche  $\frac{p}{q}$  und  $\frac{r}{s}$  sind also nicht nur dann gleich, wenn  $p = r$  und  $q = s$  ist, sondern bereits dann, wenn es natürliche Zahlen  $n, m$  gibt mit  $np = mr$  und  $nq = ms$ , denn dann ist

$$\frac{p}{q} = \frac{np}{nq} = \frac{mr}{ms} = \frac{r}{s}.$$

Die Menge aller Brüche (mit dieser Gleichheitsdefinition) bezeichnen wir als die Menge  $\mathbb{Q}$  der rationalen Zahlen.

In  $\mathbb{Q}$  lassen sich (abgesehen von der Division durch Null) alle Grundrechenarten unbeschränkt ausführen; wie aus der Schule bekannt sein sollte, ist

$$\frac{p}{q} \pm \frac{r}{s} = \frac{ps \pm qr}{qs}, \quad \frac{p}{q} \cdot \frac{r}{s} = \frac{pr}{qs} \quad \text{und} \quad \frac{p}{q} : \frac{r}{s} = \frac{ps}{qr},$$

wobei man bei den beiden letzten Formeln gegebenenfalls noch mit  $-1$  erweitern muß, um natürliche Zahlen im Nenner zu haben. Für diese Rechenoperationen gelten die „üblichen“ Regeln, die ERNST STEINITZ 1910 durch den Begriff des *Körpers* formalisierte:

**Definition:** Ein *Körper*  $k$  ist eine Menge zusammen mit zwei Rechenoperationen  $+$  und  $\cdot$ , genannt *Addition* und *Multiplikation*, die je zwei Elementen  $a, b \in k$  ein weiteres Element  $a + b$  bzw.  $a \cdot b$  zuordnet, so daß gilt:

I.1) Das Assoziativgesetz der Addition

$$(a + b) + c = a + (b + c) \quad \text{für alle } a, b, c \in k$$

I.2) Es gibt ein Element  $0 \in k$ , so daß gilt

$$a + 0 = 0 + a = a \quad \text{für alle } a \in k$$

I.3) Zu jedem Element  $a \in k$  gibt es ein Element  $a' \in k$ , so daß gilt

$$a + a' = a' + a = 0.$$

I.4) Das Kommutativgesetz der Addition

$$a + b = b + a \quad \text{für alle } a, b \in k$$

## II.1) Das Assoziativgesetz der Multiplikation

$$(a \cdot b) \cdot c = a \cdot (b \cdot c) \quad \text{für alle } a, b, c \in k$$

II.2) Es gibt ein von 0 *verschiedenes* Element  $1 \in k$ , so daß gilt

$$a \cdot 1 = 1 \cdot a = a \quad \text{für alle } a \in k$$

II.3) Zu jedem von 0 verschiedenen Element  $a \in k$  gibt es ein Element  $a'' \in k$ , so daß gilt

$$a \cdot a'' = a'' \cdot a = 1.$$

## II.4) Das Kommutativgesetz der Multiplikation

$$a \cdot b = b \cdot a \quad \text{für alle } a, b \in k$$

## III.) Das Distributivgesetz

$$a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{für alle } a, b, c \in k$$

Das Element  $a'$  aus I.3.) wird üblicherweise als  $-a$  bezeichnet und  $a''$  aus II.3) als  $a^{-1}$ . Statt  $a + (-b)$  schreibt man kurz  $a - b$ , statt  $a \cdot b^{-1}$  entsprechend  $a/b$ . Der Malpunkt wird auch oft weggelassen.



ERNST STEINITZ (1871–1928) wurde in Schlesien geboren und studierte ab 1890 an den Universitäten Breslau und Berlin. 1894 promovierte er in Breslau, ein Jahr später wurde er Privatdozent an der Technischen Hochschule Berlin-Charlottenburg. 1910 wurde er Professor in Breslau, 1920 in der Universität Kiel. In seinem Buch *Algebraische Theorie der Körper* gab er 1910 die erste Definition eines Körpers und bewies viele Sätze, die noch heute zum Standardstoff jeder Algebra-Vorlesung gehören. Auch die Konstruktion der rationalen Zahlen als Äquivalenzklassen von Paaren ganzer Zahlen geht auf ihn zurück.

Somit ist die Menge  $\mathbb{Q}$  der rationalen Zahlen ein Körper. Dagegen bilden die natürlichen Zahlen keinen Körper: Beispielsweise gibt es dort weder eine Null noch gibt es zu einer natürlichen Zahl  $n$  eine weitere Zahl  $m \in \mathbb{N}$ , so daß  $n + m = 0$  ist. Auch die ganzen Zahlen bilden keinen Körper, denn beispielsweise gibt es für  $z = 5$  keine ganze Zahl  $w$  mit  $zw = 5w = 1$ .

Neben den Grundrechenarten haben wir im Körper der rationalen Zahlen auch noch die Vergleichsoperatoren „>“ und „<“, die wir genau wie die

Grundrechenarten auf die entsprechenden Operationen für ganze Zahlen zurückführen können:

$$\frac{p}{q} > \frac{r}{s} \quad \text{genau dann, wenn} \quad ps > qr \quad \text{und}$$

$$\frac{p}{q} < \frac{r}{s} \quad \text{genau dann, wenn} \quad ps < qr.$$

Um auch die Regeln für den Umgang mit diesen Operatoren zu formalisieren, definieren wir

**Definition:** Ein Körper  $k$  mit Addition „+“ und Multiplikation „·“ heißt *angeordnet*, wenn es eine Relation „>“ gibt, so daß gilt:

1. Für je zwei Elemente  $a, b \in k$  gilt genau eine der drei folgenden Beziehungen:  $a > b$ ,  $b > a$  oder  $a = b$ .
2. Transitivität: Ist  $a > b$  und  $b > c$ , so ist auch  $a > c$ .
3. Ist  $a > b$  und  $c \in k$ , so ist auch  $a + c > b + c$ .
4. Ist  $a > b$  und  $c > 0$ , so ist auch  $ac > bc$ .

Für zwei Elemente  $a, b \in k$  schreiben wir auch  $a < b$ , falls  $b > a$  ist. Außerdem schreiben wir  $a \geq b$ , falls  $a > b$  oder  $a = b$  ist; entsprechend  $a \leq b$ , falls  $a < b$  oder  $a = b$ .

Offensichtlich ist  $\mathbb{Q}$  mit der gerade eingeführten Größerbeziehung ein angeordneter Körper. Wie wir noch sehen werden, kann allerdings nicht jeder Körper angeordnet werden.

### §3: Beweismethoden

Wahrscheinlich haben sich einige Leser gefragt, warum wir die neuen Begriffe *Körper* und *angeordneter Körper* eingeführt haben, nur um einige längst aus der Schule bekannten Rechenregeln zusammenzufassen – und selbst von diesen nur einen kleinen Teil. Wenn wir uns nur mit rationalen Zahlen beschäftigen würden, wäre dies in der Tat ein übertriebener und deshalb überflüssiger Formalismus.

Tatsächlich kennt die Mathematik Körper aller Art, die in den verschiedensten Gebieten Anwendung finden. Für die Analysis ist vor allem der Körper der reellen Zahlen wichtig, mit dem wir uns im nächsten Paragraphen beschäftigen werden, und der auch gleichzeitig unser wichtigstes Beispiel für einen angeordneten Körper sein wird.

Daneben gibt es aber beispielsweise auch endliche Körper: Ein Körper mit 256 Elementen spielt eine wesentliche Rolle bei der Fehlerkorrektur von CDs und DVDs sowie auch beim *Advanced Encryption Standard* AES, mit dem heute ein Großteil der Kommunikation bei (nach derzeitigem Stand) wirklich sicheren Internetverbindungen verschlüsselt wird. Zur Vereinbarung eines Schlüssels für diese Kommunikation wiederum spielen oft endliche Körper eine Rolle, deren Elementanzahl in der Gegend von  $10^{600}$  liegt.

Die Positionsbestimmung per Satellit mit dem *General Positioning System* GPS beruht bekanntlich darauf, daß 24 verschiedene Satelliten ständig die Uhrzeit und ihre Positionsdaten zur Erde funken; der Navigationscomputer berechnet dann aus der Laufzeit der Signale der von ihm empfangenen Satelliten seinen Standort. Alle Satelliten (und je nach Weltgegend zur Erhöhung der Genauigkeit auch noch eine ganze Reihe terrestrischer Sender) funken gleichzeitig und auf derselben Frequenz; daß die Signale trotzdem voneinander getrennt werden können, hängt wesentlich zusammen mit einem endlichen Körper mit 1023 Elementen. Es gibt auch Körper, deren Elemente Funktionen sind und mit denen etwa Computeralgebrasysteme rechnen, um Integrale zu bestimmen: es gibt Zahlkörper, die zum Beispiel für die derzeit schnellsten Faktorisierungsverfahren (und damit Angriffe auf weit verbreitete Kryptoverfahren) eine wichtige Rolle spielen, und so weiter.

Wenn wir nun eine Aussage beweisen und dazu nichts als die Körperaxiome voraussetzen, dann gilt diese Aussage in jedem beliebigen dieser Körper, auch in solchen, die wir gar nicht kennen. Wenn wir dagegen im Beweis alle uns bekannten Eigenschaften der rationalen Zahlen verwenden, dann können wir nur sicher sein, daß die bewiesene Aussage für rationale Zahlen gilt; wenn wir sie später auch für reelle oder komplexe oder sonstige Zahlen brauchen, müssen wir uns einen neuen Beweis überlegen.

Damit sind wir bei einem für die Mathematik extrem wichtigen Konzept angekommen: dem Beweis.

### a) Was ist ein Beweis?

Sowohl in den Naturwissenschaften als auch in den Wirtschafts- und

Sozialwissenschaften gewinnt man Erkenntnisse, indem man hinter Einzelbeobachtungen Gesetzmäßigkeiten vermutet und diese gegebenenfalls durch weitere Experimente oder Studien testet. Da Methoden und Meßinstrumente ständig weiterentwickelt werden, kommt es immer wieder vor, daß eine lange Zeit anerkannte Theorie aufgegeben und durch eine neue ersetzt werden muß.

Aussagen der Mathematik dagegen sind meist von der Art: Unter der Voraussetzung  $A$  gilt die Aussage  $B$ . Der Zusammenhang zwischen den beiden Aussagen  $A$  und  $B$  hängt dabei nicht ab von den Eigenschaften irgendwelcher spezieller Objekte, sondern ist logisch zwangsläufig.

Betrachten wir als Beispiel die Aussage

*Für jede natürliche Zahl  $n$  ist  $n^2 - n + 41$  eine Primzahl.*

Wir können diese Aussage testen, indem wir zunächst die ersten zehn natürlichen Zahlen durchprobieren:

$n$	1	2	3	4	5	6	7	8	9	10
$n^2 - n + 41$	41	43	47	53	61	71	83	97	113	131

Das sind in der Tat alles Primzahlen. Wir können zur Vorsicht noch einige größere  $n$  testen: Für  $n = 32$  etwa erhalten wir die Primzahl 1 033, für  $n = 100$  die Primzahl 9 941, für  $n = 200$  die Primzahl 39 841, für  $n = 5000$  erhalten wir die Zahl 24 995 041, von der man mit einiger Mühe ebenfalls zeigen kann, daß sie prim ist. Um als letztes noch eine „krumme“ Zahl zu testen, nehmen wir die ersten zehn Dezimalstellen von  $\pi$ , d.h.  $n = 3\,141\,592\,654$ ; hier erhalten wir die Zahl 98 696 042 842 862 473 (98 Milliarden 696 Billionen 42 Milliarden 842 Millionen 862 Tausend und 473), von der uns nur noch ein Computer bestätigen kann, daß auch sie prim ist.

Nach den Standards einer experimentellen Wissenschaft haben wir damit recht überzeugende Argumente dafür zusammengetragen, daß obige Behauptung richtig ist. Wir haben allerdings nicht gezeigt, daß alleine schon die Annahme,  $n$  sei eine natürliche Zahl, ausreicht für die Behauptung, daß  $n^2 - n + 41$  notwendigerweise eine Primzahl sein muß, wir haben sie also nicht bewiesen.

In der Tat überlegt man sich leicht, daß obige Behauptung falsch ist: Für  $n = 41$  sind die drei Summanden  $n^2$ ,  $n$  und 41 allesamt durch 41

teilbar, also auch die Summe. Entsprechendes gilt natürlich auch für jedes durch 41 teilbare  $n$ .

Die naheliegende Vermutung, daß dies die einzigen Ausnahmen sein könnten, wird gestützt durch die (mühsame) Überprüfung der Tatsache, daß  $n^2 - n + 41$  für alle  $n < 41$  eine Primzahl ist, aber für  $n = 42$  ist

$$n^2 - n + 41 = n(n - 1) + 41 = 42 \cdot 41 + 41 = 43 \cdot 41$$

wieder durch 41 teilbar, so daß auch das falsch ist.

Selbst sehr viele Beispiele schützen also nicht davor, daß eine Aussage falsch sein kann; gerade in der Zahlentheorie gibt es Fälle, bei denen Milliarden von Fällen überprüft wurden und sich die Aussage dann schließlich doch als falsch herausgestellt hat.

Betrachten wir als nächstes die Aussage *Jede Zehnerzahl ist gerade*. Wir könnten natürlich auch hier zunächst einmal einige Zehnerzahlen wie 20, 70 und 1250 überprüfen, aber eigentlich ist jedem klar, daß wir *nie* ein Gegenbeispiel finden werden. Um das auch zu beweisen, müssen wir nur die verwendeten Begriffe exakt definieren:

Eine natürliche Zahl  $n$  ist eine Zehnerzahl, wenn es eine natürliche Zahl  $m$  gibt mit  $n = 10m$ ; sie ist gerade, wenn es eine natürliche Zahl  $k$  gibt mit  $n = 2k$ . Da  $10 = 2 \cdot 5$  ist, können wir jede Zehnerzahl  $n = 10m$  auch als

$$n = (2 \cdot 5) \cdot m = 2 \cdot (5m) = 2 \cdot k \quad \text{mit} \quad k = 5m$$

schreiben, also ist sie *zwangsläufig* gerade.

Das ist ein wesentlicher Unterschied zur Situation etwa in den Naturwissenschaften: Deren Gesetze sind Hypothesen, die anhand von Beobachtungen und Experimenten aufgestellt werden; man kann nie ausschließen, daß spätere Wissenschaftler etwas finden, was diesen Gesetzen widerspricht. In der Tat führte beispielsweise in der Physik bislang noch jeder wesentliche Fortschritt bei Meßmethoden und -genauigkeit zur Modifikation existierender Gesetze. Selbst das NEWTONSche Gravitationsgesetz, das oft als Inbegriff eines Naturgesetzes gilt, mußte inzwischen durch das EINSTEINSche ersetzt werden, über dessen endgültige Gestalt sich die Experten auch heute noch nicht einig sind.

Verglichen damit sind wir in der Mathematik in einer beneidenswerten Situation: Was die Pythagoräer vor zweieinhalb Tausend Jahren bewiesen haben, ist auch heute noch richtig.

Im Gegenzug muß man freilich zugeben, daß die Aussage *Jede Zehnerzahl ist gerade* deutlich weniger interessant ist als das NEWTONSche oder gar EINSTEINSche Gravitationsgesetz. In der Tat sagte ALBERT EINSTEIN einmal

Insofern sich die Sätze der Mathematik auf die Wirklichkeit beziehen, sind sie nicht sicher, und insofern sie sicher sind beziehen sie sich nicht auf die Wirklichkeit.

Auch HANS GRAUERT (1930–2011), von 1959 bis zu seiner Emeritierung 1988 Professor in Göttingen und einer der bedeutendsten Analytiker des zwanzigsten Jahrhunderts, drückte sich ähnlich aus:

Die Realität ist endlich und deshalb immer unmathematisch. Will man die Mathematik zur Erkenntnis der Realität benutzen, muß man die Realität nach der Mathematik idealisieren.

Trotzdem formulierte EINSTEIN natürlich alle seiner Theorien in der Sprache der Mathematik, und den Mathematiker GRAUERT schreckten diese Aussagen erst recht nicht. Was uns (nicht nur) diese beiden Zitate zeigen, ist vielmehr, daß jede Anwendung mathematischer Methoden aus zwei wesentlich verschiedenen Schritten besteht: Als erstes muß die Wirklichkeit auf ein mathematisches Modell abgebildet werden. Da auch kleine Ausschnitte der Wirklichkeit erheblich komplexer sind als mathematische Theorien, muß es dabei zwangsläufig zu Vereinfachungen kommen, und auch Fehler beispielsweise auf Grund unvollständiger Information lassen sich nie ausschließen.

Im zweiten Schritt, beim mathematischen Modell, liegen klare und präzise Voraussetzungen vor, die zwar vielleicht nicht unbedingt der Realität entsprechen, die aber trotzdem alleiniger Ausgangspunkt des weiteren Vorgehens sind: Ab hier sind alle weiteren Folgerungen zwangsläufig richtig – wenn man einmal davon absieht, daß selbst berühmte Wissenschaftler schon gelegentlich falsche mathematische Beweise veröffentlicht haben und so etwas sicherlich auch in Zukunft immer wieder vorkommen wird. Es gibt eigentlich keinen vernünftigen Grund dafür, daß das Ergebnis für die Anwendungen nützlich sein



muß; daß dies trotzdem fast immer der Fall ist, veranlaßte im Falle der Naturwissenschaften den Physik-Nobelpreisträger EUGENE WIGNER dazu, eine Arbeit mit dem Titel *The Unreasonable Effectiveness of Mathematics in the Natural Sciences* zu veröffentlichen (*Communications in Mathematical Physics* **13** (1960), S. 1–14; auch mehrfach im Internet zu finden).

Eines der Ziele dieser Vorlesung besteht darin, Ihnen zu zeigen, wie ein mathematischer Beweis so geführt wird, daß keine Möglichkeit für einen Irrtum bleibt; die ersten Grundlagen dazu sollen im Rest dieses Paragraphen zusammengestellt werden.

### b) Direkte Beweise

Kehren wir zurück zu einem der zu Beginn dieses Paragraphen erwähnten Probleme: In der Definition eines Körpers ist nur ein Teil der aus der Schule bekannten Rechenregeln für den Umgang mit Zahlen formalisiert. Wenn wir wissen wollen, welche weiteren Regeln wir künftig beispielsweise auch für die (bislang noch gar nicht eingeführten) reellen und komplexen Zahlen sowie auch für endliche Körper verwenden können, müssen wir versuchen, diese Regeln aus den Axiomen abzuleiten, die wir der Definition eines Körpers zu Grunde gelegt haben.

In einfachen Fällen gelingt dies, indem wir einfach bekannte Rechenregeln auf eine Seite der zu beweisenden Formeln anwenden, bis wir sie zur zweiten Seite transformiert haben. Entsprechend können wir auch aus einer Aussage auf eine andere schließen.

Als erstes Beispiel dafür wollen wir beweisen, daß in jedem Körper die Multiplikation mit Null stets auf das Ergebnis Null führt, d.h.  $a \cdot 0 = 0$  für alle  $a \in k$ : Nach dem Distributivgesetz und dem Axiom I.2) ist nämlich

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

Addieren wir nun auf beiden Seiten das Element  $-(a \cdot 0)$ , erhalten wir die Gleichung  $0 = a \cdot 0$  – wobei wir auch noch das Assoziativgesetz I.1) angewandt haben.

Als nächstes wollen wir uns überlegen, daß das Element  $a'$  aus I.3) eindeutig bestimmt ist: Sind  $a'$  und  $a^*$  zwei Elemente eines Körpers  $k$

und gelten für  $a \in k$  die beiden Gleichungen

$$a + a' = a' + a = 0 \quad \text{und} \quad a + a^* = a^* + a = 0,$$

so ist  $a' = a^*$ . Zum Beweis brauchen wir außer den beiden Gleichungen und dem Axiom I.2) nur das Assoziativgesetz der Addition:

$$a' = a' + 0 = a' + (a + a^*) = (a' + a) + a^* = 0 + a^* = a^*$$

Damit folgt nun schnell, daß das Element  $-a$  in jedem Körper auch als  $(-1) \cdot a$  berechnet werden kann: Nach dem Distributivgesetz III), dem Axiom II.2) und dem Kommutativgesetz II.4) ist einerseits

$$a \cdot (1 + (-1)) = a \cdot 1 + a \cdot (-1) = a + (-1) \cdot a,$$

andererseits ist

$$a \cdot (1 + (-1)) = a \cdot 0 = 0.$$

Also ist  $a + (-1) \cdot a = (-1) \cdot a + a = 0$  und somit  $(-1) \cdot a = -a$ .

Mit zu den bekanntesten Rechenregeln aus der Schule zählen die drei binomischen Formeln

$$(a+b)^2 = a^2 + 2ab + b^2, \quad (a-b)^2 = a^2 - 2ab + b^2 \quad \text{und} \quad (a+b)(a-b) = a^2 - b^2.$$

In der Schule werden sie bewiesen via Ausmultiplizieren, d.h. über die Formel

$$(a + b)(c + d) = ac + ad + bc + bd;$$

betrachten wir also zunächst diese.

Nach dem Distributivgesetz III.) aus der Definition eines Körpers gilt für beliebige drei Elemente  $a, b, c$  eines Körpers die Formel  $a(b+c) = ab+ac$ .

Wir haben vier Elemente  $a, b, c, d$ ; betrachten wir an Stelle von  $a$  und  $b$  nur ihre Summe  $a+b$ , haben wir drei Elemente  $a+b, c$  und  $d$ ; nach dem Distributivgesetz ist also

$$(a + b)(c + d) = (a + b)c + (a + b)d.$$

Für den Anfänger erscheint diese Transformation möglicherweise problematisch, weil die Buchstaben  $a, b, c$  im Distributivgesetz und in obiger Formel verschiedene Bedeutung haben; man muß sich klar machen, daß in beiden Formeln die Buchstaben einfach Platzhalter für beliebige Körperelemente sind.

Ein Leser, dem diese Umformung große Schwierigkeiten macht, kann sie in zwei Schritte zerlegen: Wir betrachten vier neue Variablen  $s, t, u, v$  und setzen im Distributivgesetz  $a = s + t$ ,  $b = u$  und  $c = v$ . Dies führt auf die Formel

$$(s + t)(u + v) = (s + t)u + (s + t)v \quad \text{für alle } s, t, u, v \in k .$$

Da die Buchstaben  $s, t, u, v$  nur Platzhalter für Elemente von  $k$  sind, hängt die Richtigkeit einer solchen Formel nicht von der Benennung dieser Platzhalter ab; daher ist diese Aussage gleichbedeutend mit

$$(a + b)(c + d) = (a + b)c + (a + b)d \quad \text{für alle } a, b, c, d \in k .$$

Da wir noch zu interessanten Ergebnissen kommen wollen, werden wir das nur selten so ausführlich hinschreiben; aber die gleiche Idee steckt hinter jeder Umformung, bei der wir eine bekannte Formel anwenden auf Ausdrücke, in denen teilweise dieselben Variablen in anderer Bedeutung vorkommen.

Um die beiden Summanden  $(a + b)c$  und  $(a + b)d$  weiter auszurechnen, brauchen wir ein Distributivgesetz, bei dem der Klammerausdruck auf der linken Seite des Produkts steht; so etwas gibt es nicht unter unseren Körperaxiomen. Wir haben dort aber unter II.4) das Kommutativgesetz der Multiplikation; danach ist

$$(a + b)c = c(a + b) \quad \text{und} \quad (a + b)d = d(a + b) .$$

Die rechten Seiten können wir nach dem vorhandenen Distributivgesetz ausrechnen und erhalten

$$(a + b)c = c(a + b) = ca + cb \quad \text{und} \quad (a + b)d = d(a + b) = da + db .$$

Somit ist  $(a + b)(c + d) = ca + cb + da + db$ .

Wenden wir nun noch auf jeden der vier Summanden auf der rechten Seite das Kommutativgesetz II.4) der Multiplikation und dann das Kommutativgesetz I.4) der Addition auf die beiden mittleren Summanden (und strenggenommen auch einige Assoziativgesetze), haben wir die gewünschte Formel

$$(a + b)(c + d) = ac + ad + bc + bd .$$

Daraus folgen nun leicht die drei binomischen Formeln: Setzen wir  $c = a$  und  $d = b$ , erhalten wir

$$(a + b)^2 = a^2 + ab + ba + b^2 = a^2 + 2ab + b^2 ,$$

wobei wir das Kommutativgesetz der Multiplikation angewandt haben sowie die Formel  $ab + ab = 2ab$ , die aus dem Axiom II.2) sowie dem Distributivgesetz und dem Kommutativgesetz folgt:

$$ab + ab = ab \cdot 1 + ab \cdot 1 = ab \cdot (1 + 1) = ab \cdot 2 = 2 \cdot ab ;$$

die dabei verwendete Gleichung  $1 + 1 = 2$  ist streng genommen die Definition der Zwei.

Es ist klar, daß wir in dieser Vorlesung nie zu interessanten Ergebnissen kommen, wenn wir das ganze Semester über in dieser Ausführlichkeit argumentieren; wer praktische Probleme mit mathematischen Methoden behandeln möchte, muß daher möglichst schnell lernen, seine Beweise so abzukürzen, daß zwar jeder Grundlagen vertraute Leser zumindest im Prinzip in der Lage ist, seine Argumente auf die ganz ausführliche und exakte Form bringen könnte, daß aber seine Präsentation trotzdem so kurz ist, daß nur die wesentlichen Punkte erwähnt werden. Was die „wesentlichen“ Punkte sind, hängt natürlich stark vom Umfeld ab: In einer Vorlesung *Analysis I* werden Dinge ausführlich behandelt, die in einer Vorlesung für Fünftsemester stillschweigend als bekannt vorausgesetzt werden; in einer Bachelor-Arbeit wird man vielleicht noch gelegentlich kurz sagen *Wie aus der Analysis bekannt, . . .*; spätestens in einer Masterarbeit wird man Sätze aus der *Analysis I* anwenden, ohne sie zu erwähnen.

Wie im wirklichen Leben muß man also auch bei der Anwendung mathematischer Methoden einen Entwicklungsprozess durchlaufen: Anfänglich sind nur ganz ausführliche Beweise verständlich, im Verlauf des mathematischen Erwachsenwerdens bekommen dann aber auch heute sehr kryptisch erscheinende Argumente einen klar erkennbaren Sinn.

In dieser Vorlesung werden extrem ausführliche Beweise wie der obige nur in der ersten Vorlesungswoche und nur auf dem ersten Übungsblatt vorkommen; danach werden die Argumente immer kürzer, wobei freilich immer die wesentlichen Ideen erkennbar bleiben müssen.

Der Beweis der zweiten und dritten binomischen Formel soll zur Illustration des Gesagten auf dem Niveau der zweiten Vorlesungswoche durchgeführt werden: Wir setzen in der Formel für das Ausmultiplizieren

$c$  durch  $a$  und  $d$  durch  $-b$  dann ist

$$(a+b)(a-b) = a^2 - ab + ba - b^2 = a^2 - b^2,$$

die dritte binomische Formel. Für die zweite müssen wir  $b$  durch  $-b$  ersetzen und  $c = a, d = -b$  setzen; dann ist

$$(a-b)^2 = (a-b)(a-b) = a^2 - ab - ba + b^2 = a^2 - 2ab + b^2.$$

### c) Beweis durch Widerspruch

Manchmal läßt sich eine Aussage nur schwer direkt beweisen, ihr Gegenteil aber leicht widerlegen. Als einfaches Beispiel betrachten wir die Aussage

*$k$  sei ein angeordneter Körper und  $a \in k$ . Ist  $a > 0$ , so ist  $-a < 0$ , und für  $a < 0$  ist  $-a > 0$ .*

Wir nehmen an, diese Aussage sei falsch, es gebe also z.B. ein  $a \in k$  mit  $a > 0$ , für das  $-a$  nicht kleiner als Null ist. Da  $-a \neq 0$ , muß dann nach der ersten Eigenschaft aus der Definition eines angeordneten Körpers  $-a > 0$  sein, und nach der zweiten dieser Eigenschaften ist

$$0 = a + (-a) > 0 + 0 = 0, \quad \text{also} \quad 0 > 0.$$

Das ist natürlich Unsinn, also führt die Annahme,  $-a$  sei nicht kleiner als Null zu einem Widerspruch. Damit muß  $-a < 0$  sein. Genauso folgt auch, daß für  $a < 0$  gelten muß  $-a > 0$ .

Damit können wir nun beispielsweise beweisen, daß in jedem angeordneten Körper das Quadrat eines Elements  $a \neq 0$  größer als Null sein muß: Ist  $a \neq 0$ , so ist entweder  $a > 0$  oder  $a < 0$ . Im ersten Fall können wir einfach die Ungleichung  $a > 0$  nach der dritten Regel aus der Definition eines angeordneten Körpers mit  $a$  multiplizieren und erhalten die gewünschte Ungleichung

$$a^2 = a \cdot a > 0 \cdot a = 0.$$

Ist  $a < 0$ , so ist  $-a > 0$  und, wie wir oben gesehen haben, ist in jedem Körper  $(-a) = (-1) \cdot a$ . Somit ist

$$(-a)^2 = ((-1) \cdot a) \cdot ((-1) \cdot a) = (-1) \cdot (-1) \cdot a \cdot a = a^2.$$

Da wir bereits wissen, daß das Quadrat der positiven Zahl  $-a$  positiv ist, folgt damit die gewünschte Aussage  $a^2 > 0$ .

Wir sind bislang von den natürlichen zu den rationalen Zahlen fortgeschritten, weil wir alle Grundrechenarten ausführen wollten; wenn wir quadratische Gleichungen lösen wollen, müssen wir aber auch Wurzeln ziehen können. Das ist aber im Körper der rationalen Zahlen leider nicht immer möglich: Als nächstes Beispiel eines Beweises durch Widerspruch (den bereits vor rund zweieinhalb Jahrtausenden die Pythagoräer kannten), wollen wir uns überlegen, daß es keine rationale Zahl gibt, deren Quadrat gleich zwei ist:

Angenommen, es gäbe einen Bruch mit Quadrat zwei. Dann gäbe es auch einen entsprechenden gekürzten Bruch  $p/q$ . Dessen Zähler und Nenner würden die Gleichung  $p^2 = 2q^2$  erfüllen, also wäre  $p^2$  eine gerade Zahl. Dann wäre aber auch  $p$  gerade, denn das Quadrat einer ungeraden Zahl ist ungerade. Wenn  $p$  eine gerade Zahl wäre, müßte aber  $p^2$  durch vier teilbar sein, so daß auch  $q^2 = p^2/2$  eine gerade Zahl wäre, also müßte auch  $q$  eine gerade Zahl sein. Das widerspricht der Tatsache, daß  $p/q$  als gekürzter Bruch vorausgesetzt war. Also kann es keinen Bruch geben, dessen Quadrat gleich zwei ist, die Gleichung  $x^2 = 2$  hat also keine rationale Lösung.

Ein Beweis durch Widerspruch zeigt, daß das Gegenteil der zu beweisenden Aussage nicht richtig sein kann; er zeigt aber leider nicht, *warum* die Aussage inhaltlich betrachtet richtig sein muß. Da wir die Aussagen, die wir beweisen, auch verstehen wollen, ist ein inhaltlicher Beweis sicherlich vorzuziehen; oft gibt es aber – wie im vorliegenden Beispiel – keine Alternative.

### d) Vollständige Induktion

Unter *Induktion* versteht man in der Wissenschaftstheorie die Methode, auf Grund von Einzelbeobachtungen zu einer allgemeinen Aussage zu kommen. Wie wir bereits gesehen haben, ist das in der Mathematik nicht immer ein erfolgversprechender Ansatz, auch wenn heute immer mehr mathematische Aussagen durch systematisches Probieren – meist mit Computern – gefunden werden: Will man sicher sein, daß eine so erhaltene Aussage richtig ist, muß man sie zunächst beweisen.

Das Prinzip der *vollständigen* Induktion ist anwendbar auf Sätze, die sich auf irgendeine Weise auffassen lassen als Aussagen über eine natürliche

Zahl, also z.B. die Aussage

*Für jede natürliche Zahl  $n$  ist die Summe der ersten  $n$  natürlichen Zahlen gleich  $\frac{1}{2}n(n+1)$ .*

Klassische Induktion würde bedeuten, daß wir diese Aussage für verschiedene Werte von  $n$  überprüfen:

$$1 = 1 = \frac{1 \cdot 2}{2}, \quad 1 + 2 = 3 = \frac{2 \cdot 3}{2}, \quad 1 + 2 + 3 = 6 = \frac{3 \cdot 4}{2}, \quad \dots$$

Bei der *vollständigen* Induktion zeigen wir die Aussage zunächst für  $n = 1$ . Danach betrachten wir sie für  $n = 2$  unter der Voraussetzung, daß wir sie für  $n = 1$  bereits kennen:

$$1 + 2 = \frac{1 \cdot 2}{2} + 2 = \frac{2 + 4}{2} = \frac{6}{2} = \frac{2 \cdot 3}{2}.$$

Danach betrachten wir sie für  $n = 3$  unter der Voraussetzung, daß wir sie für  $n = 2$  bereits kennen:

$$1 + 2 + 3 = \frac{2 \cdot 3}{2} + 3 = \frac{6 + 6}{2} = \frac{12}{2} = \frac{3 \cdot 4}{2},$$

und so weiter.

In dieser Form erscheint dies deutlich umständlicher und unübersichtlicher als der naive und direkte Zugang. Vorteilhaft wird dieser Ansatz erst, wenn man die Schlüsse von 1 auf 2, von 2 auf 3, von 3 auf 4 usw. zusammenfasst zu einem einzigen Schluß der Form: Wenn die Aussage für eine natürliche Zahl  $n$  gilt, dann gilt sie auch für  $n + 1$ .

Im vorliegenden Fall ist dies möglich. Um die Argumente kurz und übersichtlich zu halten, führen wir dazu erst eine Abkürzung für Summen ein: Ist  $a_i$  irgendein Ausdruck, der von einer natürlichen Zahl  $i$  abhängt und der für alle  $i$  zwischen den beiden Zahlen  $n$  und  $m$  definiert ist, so schreiben wir

$$\sum_{i=n}^m a_i \stackrel{\text{def}}{=} a_n + a_{n+1} + a_{n+2} + \dots + a_{m-1} + a_m,$$

also zum Beispiel

$$\sum_{i=1}^n i = 1 + 2 + \dots + (n-1) + n \quad \text{und} \quad \sum_{i=1}^n i^2 = 1^2 + 2^2 + \dots + (n-1)^2 + n^2.$$

In dieser Schreibweise wird die zu beweisende Formel zu

$$\sum_{i=1}^n i = \frac{n(n+1)}{2}.$$

Wir nehmen an, diese Formel gelte für eine feste Zahl  $n$ , und wir wollen zeigen, daß sie dann auch für  $n + 1$  gilt. Dazu schreiben wir

$$\begin{aligned} \sum_{i=1}^{n+1} i &= \sum_{i=1}^n i + (n+1) = \frac{n(n+1)}{2} + (n+1) = \frac{n(n+1) + 2(n+1)}{2} \\ &= \frac{(n+2)(n+1)}{2} = \frac{(n+1)(n+2)}{2}, \end{aligned}$$

und das ist genau die gewünschte Formel für  $n + 1$ .

Damit haben wir ein Verfahren gefunden, um aus der Richtigkeit der Formel für eine Zahl  $n$  auf deren Richtigkeit für  $n + 1$  zu schließen. Da wir außerdem wissen, daß die Formel für  $n = 1$  gilt, können wir sie so nacheinander für  $n = 2, n = 3, n = 4$  und so weiter beweisen, insgesamt also für *jede* natürliche Zahl  $n$ . Diese Vorgehensweise bezeichnet man als das

*Prinzip der vollständigen Induktion:* Falls eine Aussage  $A(n)$  über eine natürliche Zahl  $n$  für  $n = 1$  gilt und falls für jede natürliche Zahl  $n$  aus ihrer Richtigkeit von  $A(n)$  die von  $A(n + 1)$  folgt, dann gilt die Aussage für jede natürliche Zahl  $n$ .

Den Beweis der Aussage  $A(1)$  bezeichnet man als den *Induktionsanfang*; den Beweis von  $A(n + 1)$  unter der Voraussetzung oder *Induktionsannahme*, daß  $A(n)$  gilt, als den *Induktionsschritt*.

Das Prinzip der vollständigen Induktion ist sehr universell anwendbar, liefert allerdings nicht immer den bestmöglichen Beweis: Die Summe  $S_n$  der ersten  $n$  Zahlen hätten wir auch direkt berechnen können, indem wir sie einmal vorwärts und einmal rückwärts aufsummieren:

$$\begin{aligned} S_n &= 1 + 2 + \dots + (n-1) + n \\ S_n &= n + (n-1) + \dots + 2 + 1 \end{aligned}$$

Der  $i$ -te Summand rechts ist in der ersten Zeile  $i$ , in der zweiten  $n + 1 - i$ . Addieren wir die beiden Gleichungen, so erhalten wir rechts daher eine

Summe von  $n$  Summanden, die allesamt gleich  $n + 1$  sind. Somit ist

$$2S_n = n(n + 1) \quad \text{und} \quad S_n = \frac{n(n + 1)}{2}.$$

Bei diesem direkten Beweis wird sofort klar, *warum* die Summe der ersten  $n$  natürlichen Zahlen gleich  $\frac{1}{2}n(n + 1)$  ist; beim Beweis via vollständige Induktion wird nur klar, *daß* diese Formel gilt. Dafür geht freilich der Beweis via vollständige Induktion stur nach Schema F, während man für den direkten Beweis erst auf die Idee kommen muß, die Summe einmal vorwärts und einmal rückwärts hinzuschreiben.

In leicht modifizierter Form kann das Prinzip der vollständigen Induktion auch verwendet werden, um Aussagen  $A(n)$  zu beweisen, die erst ab einer gewissen natürlichen Zahl  $a$  gelten: Dazu müssen wir die Aussage beim Induktionsanfang statt für eins für die Zahl  $a$  beweisen, und beim Induktionsschritt müssen wir nur beweisen, daß für jedes  $n \geq a$  gilt: Aus der Richtigkeit von  $A(n)$  folgt die von  $A(n + 1)$ .

Als Beispiel dazu betrachten wir die BERNOULLISCHE Ungleichung: In einem angeordneten Körper gilt

$$(1 + x)^n > 1 + nx \quad \text{für alle } n \geq 2 \text{ und alle } x > -1 \text{ außer } x = 0.$$

Für  $n = 1$  hätten wir hier die offensichtlich falsche Aussage  $1 + x > 1 + x$ , die Ungleichung kann also nicht für alle  $n \in \mathbb{N}$  gelten.

Für  $n = 2$  ist nach der ersten binomischen Formel

$$(1 + x)^2 = 1 + 2x + x^2,$$

und das ist sogar für alle  $x \neq 0$  größer als  $1 + 2x$ , denn wie wir bereits wissen, ist das Quadrat einer jeden von Null verschiedenen Zahl echt größer als Null. Dies nehmen wir als Induktionsanfang.

Für den Induktionsschritt nehmen wir an, die Gleichung sei für irgendein  $n \geq 2$  erfüllt; wir müssen zeigen, daß sie dann auch für  $n + 1$  gilt. Dazu schreiben wir  $(1 + x)^{n+1} = (1 + x)^n \cdot (1 + x)$  und wissen, daß der erste Faktor  $(1 + x)^n$  auf der rechten Seite nach der Induktionsannahme größer ist als  $1 + nx$ . Außerdem wissen wir, daß  $x > -1$ , also  $1 + x > 0$  ist. Daher können wir die Ungleichung  $(1 + x)^n > 1 + nx$  mit  $(1 + x)$  multiplizieren und erhalten die neue Ungleichung

$$(1 + x)^{n+1} > (1 + nx)(1 + x) = 1 + (n + 1)x + nx^2.$$

Dies wiederum ist größer als  $1 + (n + 1)x$ , denn  $x^2$  und  $n$  sind beide positiv, also auch  $nx^2$ .

Nach dem Prinzip der vollständigen Induktion ist damit die Ungleichung von JACOB BERNOULLI bewiesen.



JACOB BERNOULLI (1654–1705) war der erste Mathematiker aus der weitverzweigten BERNOULLI-Familie. Auf Druck seiner Eltern mußte er sich an der Universität Basel zwar für Philosophie und Theologie einschreiben und erhielt auch die entsprechenden Abschlüsse, aber er verwendete einen Großteil seiner Zeit für das Studium der Mathematik. Auf anschließenden Reisen nach Frankreich, die Niederlande und England kam er mit vielen führenden Mathematikern seiner Zeit in Kontakt. 1683 kehrte er nach Basel zurück und lehrte an der dortigen Universität Mechanik, theoretische Physik und Mathematik.

#### §4: Von den rationalen zu den reellen Zahlen

Im Körper der rationalen Zahlen können wir zwar alle Grundrechenarten unbeschränkt ausführen (außer natürlich der Division durch Null); wie wir gerade gesehen haben, können wir aber beispielsweise nicht die Wurzel aus zwei ziehen. Unser nächstes Ziel ist die Erweiterung der rationalen Zahlen zu einer Menge, in der wir eine Lösung der Gleichung  $x^2 = 2$  finden können. Dabei wollen wir das bislang Erreichte nicht aufgeben, auch die neue Menge soll also ein angeordneter Körper sein, so daß insbesondere alle im letzten Paragraphen bewiesenen Rechenregeln weiterhin gelten.

Fragen wir einen Taschenrechner, der mit zehn Ziffern rechnet, was  $\sqrt{2}$  ist, erhalten wir die Antwort  $x = 1,414213562$ . Daß diese Antwort nicht richtig sein kann, sehen wir, wenn wir  $x^2$  von Hand oder mit einem Computer in voller Genauigkeit berechnen:

$$x^2 = 1,999999998944727844.$$

Das unterscheidet sich zwar nur wenig von der Zwei, ist aber definitiv ungleich zwei – was im übrigen auch schon ohne Rechnung klar war, denn  $x$  ist schließlich eine rationale Zahl, und wir wissen, daß das Quadrat einer rationalen Zahl nie zwei sein kann.

Nicht nur mit Taschenrechnern oder Computern können wir uns rationale Zahlen verschaffen, deren Quadrat sich nur wenig von der Zwei unterscheidet; die babylonischen Mathematiker wußten schon vor rund vier Tausend Jahren, wie sie beliebig genaue Näherungslösungen konstruieren konnten. Heute wird ihr Verfahren benannt nach HERON von Alexandria, der es etwa zwei Tausend Jahre später in einem Lehrbuch präsentierte.

Die Idee ist einfach: Ist  $x^2 = 2$ , so ist  $x = 2/x$ . Ist aber  $x^2 < 2$  und  $x$  positiv, so ist  $2/x > x$  und  $(2/x)^2 > 2$ ; im Falle  $x^2 > 2$  ist entsprechend  $2/x < x$  und  $(2/x)^2 < 2$ . Da also eine der beiden Zahlen  $x$  und  $2/x$  ein zu kleines Quadrat hat, die andere ein zu großes, sollte das Quadrat ihres Mittelwerts  $y$  näher bei der Zwei liegen als die Quadrate der beiden Ausgangszahlen.

Mit dem Mittelwert  $y$  können wieder genauso verfahren: Eine der beiden Zahlen  $y^2$  und  $4/y^2$  ist größer als 2, die andere kleiner; der Mittelwert aus  $y$  und  $2/y$  sollte also eine bessere Näherung sein als  $y$ , und so weiter.

Da uns bei diesem Verfahren schnell die Buchstaben ausgehen, bezeichnen wir die so konstruierten Zahlen nicht mit  $x, y, z, \dots$ , sondern mit  $x_0, x_1, x_2, x_3, \dots$ ; auf diese Weise können wir das Verfahren beliebig oft anwenden, ohne uns ständig neue Buchstaben überlegen zu müssen.

Wir gehen aus von einer positiven Zahl aus einem angeordneten Körper  $k$ . Im Augenblick kennen wir erst einen solchen Körper, nämlich den Körper  $\mathbb{Q}$  der rationalen Zahlen, aber wie wir bald sehen werden, funktioniert der Algorithmus auch für reelle Zahlen – die zumindest den Babyloniern noch nicht bekannt waren.

### Algorithmus von Heron

Gegeben ist eine Zahl  $a > 0$  aus dem angeordneten Körper  $k$ ; konstruiert werden Näherungslösungen  $x_i, i = 1, 2, 3, \dots$  der Gleichung  $x^2 = a$ .

Wir gehen aus von irgendeiner positiven Zahl  $x_0 \in k$ , z.B.  $x_0 = 1$ , und konstruieren daraus sukzessive neue Zahlen gemäß der Vorschrift

$$x_i = \frac{1}{2} \left( x_{i-1} + \frac{a}{x_{i-1}} \right) \quad \text{für } i = 1, 2, 3, \dots$$

Eine solche Vorgehensweise bezeichnet man in der Mathematik als *Rekursion*: Wir haben keine geschlossene Formel für  $x_i$ , sondern können  $x_i$  nur berechnen, wenn wir  $x_{i-1}$  kennen, zu dessen Berechnung wir wiederum  $x_{i-2}$  brauchen, und so weiter. Da wir  $x_0$  kennen oder besser gesagt willkürlich festlegen, können wir auf diese Weise nacheinander beliebig viele  $x_i$  bestimmen.

Wenn wir dies beispielsweise auf  $a = 2$  anwenden und mit  $x_0 = 1$  anfangen, erhalten wir

$$x_1 = \frac{1}{2} \left( 1 + \frac{2}{1} \right) = \frac{3}{2}$$

$$x_2 = \frac{1}{2} \left( \frac{3}{2} + \frac{2 \cdot 2}{3} \right) = \frac{17}{12}$$

$$x_3 = \frac{1}{2} \left( \frac{17}{12} + \frac{2 \cdot 12}{17} \right) = \frac{577}{408},$$

und so weiter. Um die Qualität der Näherungslösung besser beurteilen zu können, berechnen wir die  $x_i^2$  in (näherungsweise) Dezimaldarstellung; dies führt auf folgende Tabelle:

$i$	$x_i$	$x_i^2$
0	1	1
1	$\frac{3}{2}$	2,25
2	$\frac{17}{12}$	2,00694 44444 ...
3	$\frac{577}{408}$	2,00000 60073 04882 ...
4	$\frac{665857}{470832}$	2,00000 00000 04510 ...
5	$\frac{88673108897}{627013566048}$	2,00000 00000 00000 00000 00025 ...

Schon  $x_3^2$  unterscheidet sich also erst in der sechsten Nachkommastelle von zwei, bei  $x_4^2$  haben wir bereits elf Nullen nach dem Komma und bei  $x_5^2$  sogar 23. Läßt man einen Computer weiterrechnen, stellt sich heraus, daß sich  $x_{10}^2$  erst ab der 784. Stelle von der Zwei unterscheidet.

Zumindest in diesem Fall funktioniert der Algorithmus also sehr gut. Wir wollen sehen, was wir allgemein darüber sagen können.

Für  $x_0$  können wir eine beliebige positive Zahl wählen; je nachdem, wie groß diese ist, kann  $x_0^2$  kleiner oder größer als  $a$  sein. Falls es ein Element  $x \in k$  gibt mit  $x^2 = a$  und wir mit einem positiven solchen Element starten, ist sogar  $x_0^2 = a$ .

Wir wollen uns überlegen, daß danach aber, d.h. für  $i \geq 1$ , alle  $x_i^2 \geq a$  sind. Dazu schreiben wir  $x_i = \frac{1}{2}(x_{i-1} + a/x_{i-1})$  und berechnen nach den binomischen Formeln

$$\begin{aligned} x_i^2 - a &= \frac{1}{4} \left( x_{i-1} + \frac{a}{x_{i-1}} \right)^2 - a = \frac{1}{4} \left( x_{i-1}^2 + 2a + \frac{a^2}{x_{i-1}^2} \right) - a \\ &= \frac{1}{4} \left( x_{i-1}^2 - 2a + \frac{a^2}{x_{i-1}^2} \right) = \frac{1}{4} \left( x_{i-1} - \frac{a}{x_{i-1}} \right)^2. \end{aligned}$$

Da Quadrate stets größer oder gleich Null sind, ist damit  $x_i^2 \geq a$  für alle  $i \geq 1$ .

Als Näherungsausdrücke für die Quadratwurzel von  $a$  sind die  $x_i$  ab  $x_1$  also stets zu groß – es sei denn, wir hätten ein  $x_i$  mit  $x_i^2 = a$ , was zumindest in interessanten Fällen nur selten vorkommen wird. Somit ist  $x_j$  genau dann eine bessere Näherung als  $x_i$ , wenn  $x_j < x_i$  ist.

Wenn uns das Verfahren immer bessere Näherungsausdrücke liefert, sollte also für alle  $i \geq 1$  gelten:  $x_{i+1} < x_i$ . Auch das können wir leicht direkt nachrechnen:

$$x_i - x_{i+1} = x_i - \frac{1}{2} \left( x_i + \frac{a}{x_i} \right) = \frac{1}{2} \left( x_i - \frac{a}{x_i} \right) \geq 0,$$

denn wegen  $x_i^2 > a$  ist  $x_i > a/x_i$ .

Aus  $x_{i+1} < x_i$  folgt, daß  $a/x_{i+1} > a/x_i$  ist; daher haben für  $i \geq 1$  die Ungleichungskette

$$\frac{a}{x_i} < \frac{a}{x_{i+1}} < x_{i+1} < x_i,$$

und wenn es ein Element  $x$  gibt mit  $x^2 = a$ , dann ist auch

$$\frac{a}{x_i} < \frac{a}{x_{i+1}} < x < x_{i+1} < x_i;$$

wir können  $x$  also sowohl von oben als auch von unten immer enger eingrenzen und erhalten somit immer bessere Näherungslösungen. Das besagt aber noch nicht, daß diese Näherungslösungen auch beliebig gut werden müssen, denn es könnte ja sein, daß der Abstand zwischen  $a/x_i$  und  $x_i$  eine gewisse Schranke nie unterschreiten kann.

Um diese Frage zu untersuchen, vergleichen wir die Differenz zwischen  $x_i$  und  $a/x_i$  mit der zwischen  $x_{i+1}$  und  $a/x_{i+1}$ :

$$\begin{aligned} x_i - \frac{a}{x_i} &= \frac{x_i^2 - a}{x_i} \quad \text{und} \\ x_{i+1} - \frac{a}{x_{i+1}} &= \frac{1}{2} \left( x_i + \frac{a}{x_i} \right) - \frac{2a}{x_i + a/x_i} = \frac{x_i^2 + a}{2x_i} - \frac{2ax_i}{x_i^2 + a} \\ &= \frac{(x_i^2 + a)^2 - 4ax_i^2}{2x_i(x_i^2 + a)} = \frac{(x_i^4 + 2ax_i^2 + a^2) - 4ax_i^2}{2x_i(x_i^2 + a)} = \frac{(x_i^4 - 2ax_i^2 + a^2)}{2x_i(x_i^2 + a)} \\ &= \frac{(x_i^2 - a)^2}{2x_i(x_i^2 + a)} = \frac{x_i^2 - a}{x_i} \cdot \frac{x_i^2 - a}{2(x_i^2 + a)} \\ &< \frac{(x_i^2 - a)}{x_i} \cdot \frac{1}{2} = \frac{1}{2} \left( x_i - \frac{a}{x_i} \right), \end{aligned}$$

denn da  $x_i^2$  größer ist als  $a$ , ist insbesondere

$$0 < \frac{x_i^2 - a}{2(x_i^2 + a)} < \frac{1}{2}.$$

Die Differenz zwischen  $x_{i+1}$  und  $a/x_{i+1}$  ist also höchstens halb so groß wie die zwischen  $x_i$  und  $a/x_i$ . Tatsächlich verringert sie sich nicht nur im obigen Beispiel sogar noch sehr viel dramatischer, aber solche Details sind im Augenblick noch zweitrangig; wichtig ist zunächst vor allem die prinzipielle Vorgehensweise.

Wenn wir diese Abschätzung iterieren, erhalten wir die Ungleichungskette

$$\begin{aligned} x_{i+1} - \frac{a}{x_{i+1}} &< \frac{1}{2} \left( x_i - \frac{a}{x_i} \right) < \frac{1}{4} \left( x_{i-1} - \frac{a}{x_{i-1}} \right) \\ &< \frac{1}{8} \left( x_{i-2} - \frac{a}{x_{i-2}} \right) < \dots < \frac{1}{2^i} \left( x_1 - \frac{a}{x_1} \right). \end{aligned}$$

In der Klammer ganz rechts steht eine feste Zahl, die nicht von  $i$  abhängt. Diese Zahl wird durch  $2^i$  dividiert, und nach den uns bekannten Rechenregeln wissen wir, daß wir eine rationale Zahl durch fortgesetztes Halbieren beliebig klein machen können.

Wenn wir allerdings *nur* die Axiome eines angeordneten Körpers voraussetzen, dann finden wir dort nichts, womit wir eine derartige Aussage beweisen könnten. In der Tat gibt es angeordnete Körper mit Elementen, die größer sind als jede Zweierpotenz und auch größer als jede ganze Zahl; da diese Körper hier in der Vorlesung keine Rolle spielen (und auch sonst nicht übermäßig wichtig sind), möchte ich nicht weiter darauf eingehen.

Wichtig für uns ist nur, daß wir für einen beliebigen angeordneten Körper *nicht* beweisen können, daß sich obere und untere Annäherung an die Lösung von  $x^2 = a$  beim HERON-Verfahren beliebig nahe kommen; wenn wir das zeigen wollen, brauchen wir zusätzliche Eigenschaften.

Das Ziel dieses Paragraphen besteht, wie die Überschrift sagt, darin, aus den rationalen Zahlen die reellen zu konstruieren, und solange wir uns nur in den rationalen Zahlen bewegen, können wir eine feste Zahl beliebig klein machen, indem wir sie durch immer größere Zweierpotenzen dividieren.

Wir wissen nun also, daß beispielsweise die Gleichung  $x^2 = 2$  zwar keine Lösung im Körper  $\mathbb{Q}$  der rationalen Zahlen hat, wir wissen aber auch, daß wir rationale Zahlen finden können, deren Quadrat der Zwei beliebig nahe kommt. Wir wollen die rationalen Zahlen erweitern zu einem größeren Körper, in dem wir eine Lösung dieser Gleichung (und hoffentlich auch vieler anderer in  $\mathbb{Q}$  unlösbarer Gleichungen) haben.

Für eine solche Erweiterung müssen wir festlegen, was die „neuen“ Zahlen sein sollen. Die natürliche Zahlen sind uns wohlbekannt; wenn wir sie aber definieren müßten, bliebe uns kaum etwas anderes übrig, als so vorzugehen, wie der Zahlbegriff wohl auch historisch entstand: Die Zahlen kommen vom Abzählen, zwei ist also einfach eine Abkürzung für  $1 + 1$ , drei für  $1 + 1 + 1$  und so weiter. Rationale Zahlen sind Brüche, und auch das sind eigentlich Rechenvorschriften: Der Bruch  $\frac{1595703}{9042317}$  bezeichnet das Ergebnis der Division von 159570 durch 9042317, und

wenn wir das mit genügender Sorgfalt ausrechnen, kommen wir zum Ergebnis, daß wir diese Zahl auch besser als  $\frac{3}{17}$  schreiben können.

Bei den nun einzuführenden reellen Zahlen geht es – wie wir im Laufe der Vorlesung immer wieder sehen werden – um deutlich kompliziertere Objekte, aber wie bei allen Zahlen wollen wir auch damit rechnen und wir wollen Ergebnisse, die wir je nach Anwendung in Euro und Cent oder in Metern und Zentimetern ausdrücken können. Dafür sind Eingrenzungen, wie sie uns das Verfahren von HERON liefert, durchaus geeignet; das einzige Problem ist, daß wir wie bei den Brüchen dieselbe Zahl auf verschiedene Weise darstellen können: HERON liefert uns schließlich für *jeden* positiven Startwert Annäherungen an  $\sqrt{a}$ , aber natürlich hängen die konkreten Zahlen ab vom Startwert.

Wir formalisieren zunächst die Annäherung durch obere und untere Schranken:

**Definition:**  $k$  sei ein angeordneter Körper und  $a, b$  seien zwei Elemente von  $k$ .

- a) Das abgeschlossene Intervall  $[a, b]$  ist die Menge aller  $x \in k$  mit  $a \leq x \leq b$ .
- b) Das offene Intervall  $(a, b)$  ist die Menge aller  $x \in k$  mit  $a < x < b$ .
- c) Die halboffenen Intervalle sind definiert als

$$[a, b) = \{x \in k \mid a \leq x < b\} \quad \text{und} \quad (a, b] = \{x \in k \mid a < x \leq b\}.$$

Mit dem Verfahren von HERON finden wir somit immer kleinere Intervalle  $[a/x_i, x_i]$ . Um eine solche Formulierung präzise zu fassen, müssen wir noch erklären, was „immer kleiner“ bedeuten soll.

Anschaulich ist das zwar völlig klar, aber bei komplizierten mathematischen Sachverhalten kann uns aber unsere Anschauung gelegentlich auch in die Irre führen. Dies müßten beispielsweise die Mathematiker des späten siebzehnten sowie des achtzehnten Jahrhunderts schmerzlich erfahren: Gemeinsam mit den Physikern (zu denen ein Großteil von ihnen selbst gehörte) war es ihnen gerade erst gelungen war, mit der neu geschaffenen Differential- und Integralrechnung die Natur in einem bis dahin nicht bekannten Maße zu erklären und überprüfbare Vorhersagen zu machen, von denen ihre Vorgänger nicht einmal zu träumen gewagt



hatten. Genau wie mancher heutige Wirtschaftsguru glaubten sie, die Welt vollständig im Griff zu haben und ihren Ablauf wie den eines Uhrwerks erklären zu können. Für spekulierende Philosophen und in Dogmen verhaftete Theologen hatten sie nur Verachtung übrig.

Deren Rache kam postwendend. Teilweise waren ihre Entgegnungen die üblichen, damals unter Gelehrten aller Art sehr beliebten Polemiken, teilweise auch in literarischer Form. Auch heute noch oft gelesen (und auch in deutscher Übersetzung erhältlich) sind einige Romane des irischen Schriftstellers und Geistlichen JONATHAN SWIFT (1667-1745; seit 1713 Dekan der Dubliner St.-Patricks-Kathedrale). So ist etwa der dritte Band von *Gulliver's Travels* eine stellenweise sehr boshafte Satire, die sich zwar grundsätzlich gegen alles Moderne richtet, in einigen Teilen aber auch speziell die Mathematiker (Reise nach Lapute) oder die mathematisch arbeitenden Naturwissenschaftler und Ingenieure (Reise nach Balnibarbi) aufs Korn nimmt.

Viel schmerzhafter waren allerdings einige rein wissenschaftliche Streitschriften, die berühmteste darunter *The Analyst; or, a Discourse Adressed to an Infidel Mathematician. Wherein it is examined whether the Object, Principles, and Inferences of the modern Analysis are more distinctly conceived, or more evidently deduced, then Religious Mysteries and Points of Faith*, die GEORGE BERKELEY 1734 veröffentlichte. Daß sie auch heute noch gelesen wird, sieht man unter anderem daran, daß sie mehrfach als Volltext im Internet zu finden ist.



GEORGE BERKELEY (1685–1753) wurde im irischen Kilkenny geboren und besuchte (wie JONATHAN SWIFT) zunächst das dortige Kilkenny College, die wohl berühmteste Privatschule Irlands. Im Alter von nur 15 Jahren begann er mit der Studium von Mathematik, Logik, Latein, Griechisch, Hebräisch, Französisch, Philosophie und schließlich auch Theologie am Trinity College der Universität Dublin, wo er 1704 seinen B.A. erwarb. Gleich darauf veröffentlichte er ein elementares Lehrbuch der Arithmetiks, unterrichtete aber ab 1707 Theologie. Philosophisch war er ein Vertreter des Empirismus, wonach nur unsere Wahrnehmungen real sind; eine davon unabhängige Realität gibt es nicht. 1734

wurde er Bischof von CLOYNE in der irischen Grafschaft Cork.

BERKELEY verwendete für seine mathematischen Beispielen im *Analyst* und in anderen Schriften genau die Art von Schlußfolgerungen, mit denen die Analytiker damals arbeiteten und in denen oft von beliebig kleinen oder gar unendlich kleinen Größen die Rede war; was er damit bewies waren aber entweder offensichtlich unsinnige Sätze oder er berechnete denselben Flächeninhalt mit verschiedenen Methoden, die sich allesamt nicht von den damals üblichen mathematischen Vorgehensweisen unterscheiden ließen, die aber trotzdem zu lauter verschiedenen Ergebnissen für genau dieselbe Fläche führten. Auch wenn zumindest in den heute noch bekannten Werken von Mathematikern dieser Zeit kaum falsche Behauptungen zu finden sind, war damit doch gezeigt, daß ihre Beweise nicht exakt waren, sondern nur dank guter Intuition zu richtigen Ergebnissen führten.

Gegen Intuition in der Mathematik ist selbstverständlich nichts einzuwenden; kein wirklich interessantes mathematisches Ergebnis wurde je ohne Intuition bewiesen, und zumindest soweit ich weiß haben Erbsenzähler, die immer nur eine Rechenvorschrift an die andere reihen, noch nie einen auch nur moderat nichttrivialen Satz bewiesen.

Die Mathematik als exakte Wissenschaft hat freilich den Anspruch, daß alle ihre Aussagen zumindest im Prinzip auf einige wenige vorher festgelegte Annahmen, die Axiome zurückführbar sind. Das Ideal jeglicher Mathematik waren selbst noch im siebzehnten und achtzehnten Jahrhundert die *Elemente* von EUKLID, die auch von Philosophen und Theologen als Urbeispiel jeglicher exakter Wissenschaft betrachtet wurden. (Erst Anfang des zwanzigsten Jahrhunderts entdeckten Mathematiker, daß EUKLID auch noch einige (wenige) Annahmen benutzte, die er nirgends explizit voraussetzte.)

Somit drängte die Frage, wie auch Probleme der für die Naturwissenschaften so wichtigen Analysis mit der gleichen Strenge bewiesen werden konnten wie die der EUKLIDischen Geometrie. Wie BERKELEY gezeigt hatte, waren die Methoden der damaligen Analytiker nicht mit denen der griechischen Geometer vergleichbar; die Argumentation mit beliebig oder unendlich kleinen Größen kann zu willkürlichen Schlußfolgerungen führen, und BERKELEY fragt bezüglich dieser Größen polemisch: *May we not call them the Ghosts of departed Quantities?*

Erst um 1800 fand CAUCHY einen mögliche Ausweg aus dem Dilemma: Er führte die Grundlagen der Analysis zurück auf die Theorie algebraischer Ungleichungen und konnte damit die analytischen Begriffe in einer Weise definieren, die klare Beweise im Stile der griechischen Geometrie erlaubte. Einziger Nachteil seiner Methode ist, daß er damit etwas in die Mathematik einführte, was noch heute der Schrecken vieler Anfänger ist: Die berühmten griechischen Buchstaben  $\varepsilon$  (epsilon) und  $\delta$  (delta). Wir brauchen hier zunächst nur das  $\varepsilon$  und definieren *beliebig klein* so, daß für jede vorgegebene Schranke  $\varepsilon$  die Folgenglieder schließlich kleiner als  $\varepsilon$  werden. In diesem Satz ist noch nicht ganz klar, was „schließlich“ bedeuten soll; die einzig sinnvolle exakte Interpretation ist aber wohl die, daß es eine Zahl  $n_0$  geben muß (die natürlich von  $\varepsilon$  abhängt), so daß der Betrag aller Folgenglieder  $a_n$  für  $n \geq n_0$  kleiner als  $\varepsilon$  sein soll. Den Betrag definieren wir dabei auf die übliche Art:

**Definition:** Der Betrag eines Elements  $x \in k$  eines angeordneten Körpers ist

$$|x| = \begin{cases} x & \text{falls } x \geq 0 \\ -x & \text{falls } x < 0 \end{cases}.$$

Damit haben wir alles zusammen, was wir zur präzisen Definition des anschaulichen Konzepts „wird beliebig klein“ brauchen:

**Definition:** Eine Folge  $(a_n)_{n \in \mathbb{N}}$  von Elementen  $a_n$  eines angeordneten Körpers  $k$  heißt *Nullfolge*, wenn es zu jedem  $\varepsilon > 0$  aus  $k$  ein  $n_0 \in \mathbb{N}$  gibt, so daß  $|a_n| < \varepsilon$  für alle  $n \geq n_0$ .



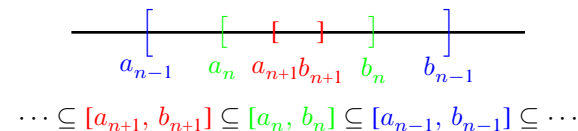
Baron AUGUSTIN LOUIS CAUCHY (1789–1857) stellte als erster durch die exakte Definition von Begriffen wie *Konvergenz* und *Stetigkeit* die Analysis auf ein sicheres Fundament. In insgesamt 789 Arbeiten beschäftigte er sich u.a. auch mit komplexer Analysis, Variationsrechnung, Differentialgleichungen, FOURIER-Analysis, Permutationsgruppen, der Diagonalisierung von Matrizen und der theoretischen Mechanik. Als überzeugter Royalist hatte er häufig Schwierigkeiten mit den damaligen Regierungen; er lebte daher mehrere Jahre im Exil in Turin und später in Prag, wo er (mit sehr mäßigem Erfolg) den französischen Thronfolger unterrichtete.

Typisches Beispiel einer Nullfolge ist etwa die Folge mit  $a_n = 1/n$ : Hier ist  $|a_n| = |1/n| = 1/n$  genau dann kleiner als  $\varepsilon$ , wenn  $n > 1/\varepsilon$  ist; als  $n_0$  können wir also jede natürliche Zahl nehmen, die größer ist als  $1/\varepsilon$ .

Mit diesem Begriff ist nun auch klar, wie die Forderung, daß Intervalle  $[a_n, b_n]$  beliebig klein werden, präzise formuliert werden kann:

**Definition:** Eine Intervallschachtelung über einem angeordneten Körper  $k$  ist eine Folge von Intervallen  $([a_n, b_n])_{n \in \mathbb{N}}$  mit den folgenden beiden Eigenschaften:

1.  $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$  für alle  $n \in \mathbb{N}$
2. Die Folge der  $(b_n - a_n)_{n \in \mathbb{N}}$  ist eine Nullfolge.



Genau wie wir die rationalen Zahlen durch Quotienten ganzer Zahlen (mit positivem Nenner) definiert haben, können wir nun reelle Zahlen definieren durch Intervallschachtelungen über  $\mathbb{Q}$ . Genau wie eine rationale Zahl in vielfältiger Weise als Quotient ganzer Zahlen darstellbar ist, läßt sich auch eine reelle Zahl durch die verschiedensten Intervallschachtelungen definieren: Für die Quadratwurzel haben wir ja gesehen, daß uns das HERON-Verfahren für *jeden* Startwert  $x_0 > 0$  eine Intervallschachtelung liefert, von der wir intuitiv sagen würden, daß sie sich auf die Quadratwurzel zusammenzieht. Wir müssen nun nur noch formal definieren, wann zwei Intervallschachtelungen dieselbe reelle Zahl definieren sollen. Da die Intervalle immer kleiner werden, sollten wir erwarten, daß es bei zwei Intervallschachtelungen zur selben reellen Zahl auch die Intervallenden immer mehr aufeinander zugehen, und genau so wollen wir die reellen Zahlen auch definieren:

**Definition:** Reelle Zahlen sind gegeben durch Intervallschachtelungen über dem Körper  $\mathbb{Q}$  der rationalen Zahlen. Dabei sollen zwei Intervallschachtelungen  $([a_n, b_n])_{n \in \mathbb{N}}$  und  $([c_n, d_n])_{n \in \mathbb{N}}$  genau dann dieselbe reelle Zahl definieren, wenn die Folgen  $(c_n - a_n)_{n \in \mathbb{N}}$  und

$(d_n - b_n)_{n \in \mathbb{N}}$  Nullfolgen sind. Die Menge aller reeller Zahlen bezeichnen wir mit  $\mathbb{R}$ .

Damit sind reelle Zahlen definiert, wenn auch die Definition etwas seltsam erscheint. Wir werden aber im Verlauf der Vorlesung immer wieder sehen, daß wir die reellen Zahlen nicht wirklich in den Griff bekommen können; sie sind eine Idealisierung der Wirklichkeit, und eine Rechenvorschrift, mit der wir sie sukzessive immer besser annähern können ist leider das Beste, was wir an Konkretisierung erwarten können. Bei praktischen Problemen werden wir nie über ein meist relativ kleines  $n$  hinauskommen, und trotzdem zeigt jahrtausendelange Erfahrung der Natur- und Ingenieurwissenschaften und auch mehr als hundert Jahre Erfahrung der Wirtschaftswissenschaften, daß die reellen Zahlen ein besseres Beschreibungsmodell der Wirklichkeit liefern als beispielsweise die rationalen Zahlen.

Die rationalen Zahlen können wir leicht als Teilmenge der reellen Zahlen interpretieren: Für eine rationale Zahl  $q \in \mathbb{Q}$  nehmen wir einfach die Intervallschachtelung, deren sämtliche Intervalle  $[a_n, b_n]$  gleich dem Intervall  $[q, q]$  sind.

Unabdingbare Voraussetzung für die Nützlichkeit der reellen Zahlen ist natürlich, daß wir damit rechnen können. Da wir sie durch Intervallschachtelungen definiert haben, liegt es nahe, daß wir zunächst versuchen, mit Intervallen rationaler Zahlen zu rechnen. Tatsächlich sind auch Intervalle reeller Zahlen nützlich für manche Anwendungen; daher wollen wir gleich mit Intervallen in beliebigen angeordneten Körpern rechnen, wobei wir hoffen, daß sich die reellen Zahlen als Beispiel eines solchen angeordneten Körpers herausstellen werden.

Reelle Zahlen sind durch Intervallschachtelungen gegeben; daher liegt es nahe, zunächst Rechenoperationen für Intervalle zu definieren. Dies hat durchaus auch eine praktische Bedeutung, denn viele Daten aus der realen Welt lassen sich oft nicht mit beliebiger Genauigkeit ermitteln; verlässliche obere und untere Schranken lassen sich aber trotzdem oft finden. In diesem Fall wissen wir nur, daß der gesuchte Wert zwischen diesen beiden Schranken liegt, d.h. wir kennen ein Intervall, in dem er mit Sicherheit liegt.

Meist wird man in so einer Situation einen Schätzwert nehmen, z.B. den

Mittelwert zwischen oberer und unterer Schranke, und damit rechnen in der Hoffnung, daß dieselbe Rechnung mit einem anderen Wert aus dem Intervall auf ein ähnliches Ergebnis geführt hätte. Je umfangreicher und komplizierter die Rechnung ist, desto fragwürdiger wird diese Annahme.

Will man zu sicheren Aussagen über das Ergebnis kommen, muß man von vorn herein mit Intervallen rechnen; das Ergebnis ist dann zwar auch nur ein Intervall, und wenn man Pech hat sogar ein recht großes, aber man hat immerhin die Sicherheit, daß der gesuchte Wert auf jeden Fall dort liegt.

Deshalb entstanden schon vor rund fünfzig Jahren Programmiersprachen, wie Triplex ALGOL oder Pascal XSC, die auch mit Intervallen rechnen können; heute gibt es sowohl für C als auch die Computeralgebrasysteme entsprechende Klassenbibliotheken, und derzeit arbeitet eine IEEE *working group* an einem internationalen Standard IEEE P1788 für Intervallarithmetik.

Ist  $*$  irgendeine Rechenoperation und sind  $[a, b]$ ,  $[c, d]$  zwei Intervalle, so soll das Intervall  $[a, b] * [c, d]$  für jedes  $x \in [a, b]$  und jedes  $y \in [c, d]$  das Rechenergebnis  $x * y$  enthalten.

Im Falle der Addition bedeutet dies, daß wir Schranken suchen für alle möglichen Summen  $x + y$  mit  $a \leq x \leq b$  und  $c \leq y \leq d$ . Nach den Rechenregeln für angeordnete Körper können wir die beiden Ungleichungen  $a \leq x$  und  $c \leq y$  addieren zu  $a + c \leq x + y$  und entsprechend auch an der oberen Grenze; hier drängt sich die Definition

$$[a, b] + [c, d] \stackrel{\text{def}}{=} [a + c, b + d]$$

also förmlich auf. (Beim numerischen Rechnen, wo Rundungsfehler zu befürchten sind, müßte man zusätzlich noch darauf achten, daß bei der Berechnung der unteren Schranke  $a + c$  im Zweifelsfall abgerundet wird, für  $b + d$  aber aufgerundet. Diese Probleme wollen wir aber hier in der Analysis I noch völlig ausklammern; wir gehen davon aus, daß wir alle Rechenoperationen zumindest im Prinzip exakt ausführen können.)

Auch bei der Subtraktion gibt es keine Probleme: Ist  $c \leq y \leq d$ , so ist  $-d \leq -y \leq -c$ ; deshalb setzen wir

$$[a, b] - [c, d] \stackrel{\text{def}}{=} [a - d, b - c].$$

Bei der Multiplikation lassen sich zumindest für positive Zahlen  $a, c, x, y$  wieder die beiden Ungleichungen  $a \leq x$  und  $c \leq y$  kombinieren zu  $ac \leq xy$ ; sobald aber negative Zahlen ins Spiel kommen, gilt es eine ganze Reihe von Fällen zu unterscheiden. Da es uns nur ums Prinzip geht, nicht um eine möglichst effiziente praktische Implementierung, können wir uns aus der Affäre ziehen, indem wir einfach alle vier möglichen Produkte einer Schranke des ersten Intervalls mit einer Schranke des zweiten Intervalls berechnen und dann den kleinsten der vier Werte als untere, den größten als obere Schranke des Intervalls nehmen:

$$[a, b] \cdot [c, d] \stackrel{\text{def}}{=} [\min(ac, ad, bc, bd), \max(ac, ad, bc, bd)],$$

wobei allgemein  $\min(a_1, a_2, \dots, a_n)$  die kleinste der Zahlen  $a_1, \dots, a_n$  bezeichnen soll, das *Minimum* also, und  $\max(a_1, a_2, \dots, a_n)$  die größte, das *Maximum*.

Bleibt noch das Problem der Division. Da Division durch Null in einem Körper nicht sinnvoll erklärt werden kann, darf bei der Division von Intervallen offensichtlich der Divisor nicht die Null enthalten. Somit können wir nur durch Intervalle dividieren, bei denen entweder beide Schranken positiv oder beide negativ sind. In diesem enthält Intervall  $[\frac{1}{d}, \frac{1}{c}]$  alle Inversen der Zahlen aus  $[c, d]$ ; daher ist ähnlich zum Fall der Multiplikation

$$[a, b]/[c, d] = \left[ \min\left(\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}\right), \max\left(\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}\right) \right].$$

Damit wissen wir, wie man mit Intervallen rechnet; reelle Zahlen sind aber durch Intervallschachtelungen definiert. Auch hier bietet sich eine offensichtliche Lösung an: Wollen wir zwei reelle Zahlen, die durch Intervallschachtelungen  $([a_n, b_n])_{n \in \mathbb{N}}$  und  $([c_n, d_n])_{n \in \mathbb{N}}$  definiert sind, miteinander verknüpfen, so verknüpfen wir einfach für jedes  $n$  die Intervalle  $[a_n, b_n]$  und  $[c_n, d_n]$  miteinander. Dabei gibt es einige Probleme. Am offensichtlichsten ist das mit der Division: Selbstverständlich kann eine von Null verschiedene reelle Zahl auch durch eine Intervallschachtelung definiert werden, bei der zumindest einige der Intervalle eine negative untere und eine positive obere Grenze haben.

Wir wollen uns überlegen, daß dies höchstens für endlich viele Intervalle am Anfang gelten kann, genauer:

*Für jede Intervallschachtelung  $([a_n, b_n])$ , die sich nicht auf die Null zusammenzieht, gibt es einen Index  $n_0$ , so daß entweder  $a_n$  und  $b_n$  beide positiv sind für alle  $n \geq n_0$  oder aber  $a_n$  und  $b_n$  beide negativ sind für alle  $n \geq n_0$ .*

Offensichtlich reicht es, wenn wir einen einzigen Index  $n_0$  finden, für den  $a_{n_0}$  und  $b_{n_0}$  beide negativ oder beide positiv sind: Da alle folgenden Intervalle in  $[a_{n_0}, b_{n_0}]$  liegen, gelten dann dieselben Ungleichungen auch für alle  $n \geq n_0$ .

Wir beweisen dies durch Widerspruch: Falls es keinen solchen Index  $n_0$  gibt, ist  $a_n \leq 0 \leq b_n$  für alle  $n \in \mathbb{N}$ .

Nach Definition einer Intervallschachtelung ist die Folge der  $b_n - a_n$  eine Nullfolge, d.h. für jedes  $\varepsilon > 0$  gibt es ein  $n_0$ , so daß  $b_n - a_n \leq \varepsilon$  ist für alle  $n \geq n_0$ . Nach unseren Annahmen ist dann also

$$b_n - a_n = |b_n| + |a_n| < \varepsilon \quad \text{für alle } n \geq n_0.$$

Dann ist aber auch  $|b_n| < \varepsilon$  und  $|a_n| < \varepsilon$  für alle  $n \geq n_0$ , denn die Summe zweier Beträge ist größer oder gleich jedem einzelnen. Damit sind  $(a_n)_{n \in \mathbb{N}}$  und  $(b_n)_{n \in \mathbb{N}}$  Nullfolgen, die Intervallschachtelung zieht sich also auf die Null zusammen, im Widerspruch zur Voraussetzung.

Somit muß es einen Index  $n_0$  geben, ab dem entweder alle Intervallgrenzen positiv oder alle negativ sind, womit die Behauptung bewiesen wäre. ■

Die Intervalle  $[a_n, b_n]$  mit  $n \geq n_0$  bilden natürlich auch eine Intervallschachtelung und definieren dieselbe reelle Zahl; bei der Division können (und müssen) wir daher einfach den Divisor beschreiben durch eine Intervallschachtelungen aus Intervallen, von denen keines die Null enthält.

Ein zweites Problem besteht darin zu zeigen, daß die so definierten Folgen von Intervallen überhaupt Intervallschachtelungen sind.

Sind also beispielsweise  $([a_n, b_n])_{n \in \mathbb{N}}$  und  $([c_n, d_n])_{n \in \mathbb{N}}$  zwei Intervallschachtelungen, ist dann auch  $([a_n + c_n, b_n + d_n])_{n \in \mathbb{N}}$  eine Intervallschachtelung?

Wir müssen zwei Bedingungen überprüfen: Als erstes muß für alle  $n \in \mathbb{N}$  gelten:

$$[a_{n+1} + c_{n+1}, b_{n+1} + d_{n+1}] \subseteq [a_n + c_n, b_n + d_n].$$

Da wir von zwei Intervallschachtelungen ausgegangen sind, wissen wir, daß auf jeden Fall

$$[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n] \quad \text{und} \quad [c_{n+1}, d_{n+1}] \subseteq [c_n, d_n]$$

ist, wir haben also die Ungleichungen

$$a_n \leq a_{n+1} \leq b_{n+1} \leq b_n \quad \text{und} \quad c_n \leq c_{n+1} \leq d_{n+1} \leq d_n$$

für alle  $n \in \mathbb{N}$ . In einem angeordneten Körper können wir Ungleichungen addieren; daher ist auch

$$a_n + c_n \leq a_{n+1} + c_{n+1} \leq b_{n+1} + d_{n+1} \leq b_n + d_n,$$

und das ist gleichbedeutend mit der behaupteten Inklusion.

Als zweites müssen wir zeigen, daß die Differenz der Intervallgrenzen eine Nullfolge bildet. Diese Differenz ist beim  $n$ -ten Intervall

$$(b_n + d_n) - (a_n + c_n) = (b_n - a_n) + (d_n - c_n).$$

Wir wissen, daß  $(b_n - a_n)_{n \in \mathbb{N}}$  und  $(d_n - c_n)_{n \in \mathbb{N}}$  beides Nullfolgen sind, und wir wollen wissen, ob dies auch für ihre gliedweise Summe gilt. Diese Frage ist auch unabhängig vom aktuellen Beweis interessant; wir formulieren die Antwort daher als einen eigenen *Hilfssatz* oder – wie man in der Mathematik meistens sagt – als *Lemma*:

**Lemma:** Sind  $(a_n)_{n \in \mathbb{N}}$  und  $(b_n)_{n \in \mathbb{N}}$  beide Nullfolgen, so ist auch  $(a_n + b_n)_{n \in \mathbb{N}}$  eine Nullfolge.

*Beweis:* Wir müssen zeigen, daß es zu jedem  $\varepsilon > 0$  ein  $n_0 \in \mathbb{N}$  gibt, so daß  $|a_n + b_n| \leq \varepsilon$  ist. Wir wissen, daß  $(a_n)_{n \in \mathbb{N}}$  und  $(b_n)_{n \in \mathbb{N}}$  Nullfolgen sind; wenn wir aber die Definition direkt einsetzen, folgt nichts brauchbares, da wir ja noch addieren müssen.

Stattdessen gehen wir aus von  $\varepsilon/2$ ; auch dazu gibt es ein  $n_1 \in \mathbb{N}$ , so daß  $|a_n| < \varepsilon/2$  für alle  $n \geq n_1$  und ein  $n_2 \in \mathbb{N}$ , so daß  $|b_n| < \varepsilon/2$  für alle  $n \geq n_2$ . Bezeichnet  $n_0 = \max(n_1, n_2)$  die größere der beiden Zahlen, so ist also für alle  $n \geq n_0$

$$|a_n| + |b_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Was uns interessiert ist nicht  $|a_n| + |b_n|$ , sondern der Betrag von  $a_n + b_n$ . Wir werden bald allgemeine Regeln kennen lernen, wie man solche Ausdrücke miteinander in Verbindung bringt; hier soll ein kurzes *ad hoc* Argument genügen: Ist  $a_n \geq 0$  und  $b_n \geq 0$ , sind die beiden Zahlen gleich, genauso wenn  $a_n \leq 0$  und  $b_n \leq 0$ . Ist aber  $a_n > 0$  und  $b_n < 0$ , so ist  $b_n < a_n + b_n < a_n$ , also ist  $|a_n + b_n|$  kleiner als das Maximum von  $|a_n|$  und  $|b_n|$ . Dieses wiederum ist kleiner als die Summe der beiden Beträge. Entsprechend können wir argumentieren, wenn  $a_n < 0 < b_n$  ist, so daß in allen Fällen gilt  $|a_n + b_n| \leq |a_n| + |b_n| \leq \varepsilon$ . Damit haben wir nachgewiesen, daß auch  $(a_n + b_n)_{n \in \mathbb{N}}$  eine Nullfolge ist. ■

Wenden wir dieses Lemma speziell an auf die Differenzen zwischen den Grenzen von Intervallschachtelungen, schließt dies den Nachweis ab, daß für zwei Intervallschachtelungen  $([a_n, b_n])_{n \in \mathbb{N}}$  und  $([c_n, d_n])_{n \in \mathbb{N}}$  auch  $([a_n, b_n] + [c_n, d_n])_{n \in \mathbb{N}}$  eine Intervallschachtelung ist.

Um zu zeigen, daß auch  $([a_n, b_n] - [c_n, d_n])_{n \in \mathbb{N}}$  eine Intervallschachtelung ist, können wir die obige Rechnung fast wörtlich wiederholen; alternativ können wir auch zeigen, daß mit  $([c_n, d_n])_{n \in \mathbb{N}}$  auch  $([-d_n, -c_n])_{n \in \mathbb{N}}$  eine Intervallschachtelung ist, und die Folge dann umschreiben als  $([a_n, b_n] + [-d_n, -c_n])_{n \in \mathbb{N}}$ . Einzelheiten seien den interessierten Lesern als Übungsaufgaben überlassen.

Deutlich unangenehmer wird die Sache bei Produkten, denn falls Intervalle auftreten, deren Grenzen verschiedene Vorzeichen haben, sind hier viele Fallunterscheidungen notwendig. Nun haben wir aber gesehen, daß jede Intervallschachtelung, die sich nicht auf die Null zusammenzieht, ab einem gewissen Index nur noch Intervalle enthält, bei denen entweder beide Intervallgrenzen positiv oder aber beide negativ sind. Wenn wir uns auf Intervallschachtelungen begrenzen, bei denen dies für *alle* Intervalle gilt, wird die Situation einfacher; beispielhaft möchte ich den Fall betrachten, daß  $([a_n, b_n])_{n \in \mathbb{N}}$  und  $([c_n, d_n])_{n \in \mathbb{N}}$  Intervallschachtelungen sind, für die alle Intervallgrenzen positiv sind. Dann ist

$$[a_n, b_n] \cdot [c_n, d_n] = [a_n c_n, b_n d_n].$$

Wir müssen als erstes zeigen, daß  $[a_{n+1}, b_{n+1}] \cdot [c_{n+1}, d_{n+1}]$  für alle  $n$  in diesem Intervall liegt, daß also gilt

$$a_n c_n \leq a_{n+1} c_{n+1} \leq b_{n+1} d_{n+1} \leq b_n d_n.$$

Dies folgt schnell aus der Tatsache, daß für positive Zahlen  $a, b, c, d$  gilt: Ist  $a < b$  und  $c < d$ , so ist auch  $ac < bd$ , wir können Ungleichungen also genauso miteinander multiplizieren, wie wir sie oben addiert haben. Dann müssen wir noch zeigen, daß die Folge der Intervalllängen  $(b_n d_n - a_n c_n)_{n \in \mathbb{N}}$  eine Nullfolge ist. Wir wissen, daß  $(b_n - a_n)_{n \in \mathbb{N}}$  und  $(d_n - c_n)_{n \in \mathbb{N}}$  Nullfolgen sind, also sollten wir die irgendwie ins Spiel bringen. Dazu schreiben wir

$$b_n d_n - a_n c_n = b_n (d_n - c_n) + c_n (b_n - a_n).$$

Nun sei irgendein  $\varepsilon > 0$  vorgegeben, und wir wollen ein  $n_0$  finden, so daß der Betrag dieses Ausdrucks für alle  $n \geq n_0$  kleiner ist als  $\varepsilon$ .

Da alle hier stehenden Terme positiv sind, können wir auf Betragsstriche verzichten und die rechte Seite der Gleichung so abschätzen, wie sie dasteht. Da  $b_n$  die obere Grenze eines Intervalls ist, das in  $[a_1, b_1]$  liegt, ist  $b_n \leq b_1$ . Genauso ist auch  $d_n \leq d_1$  und  $c_n \leq d_n$ , also  $c_n \leq d_1$ . Somit ist

$$b_n d_n - a_n c_n \leq b_1 (d_n - c_n) + d_1 (b_n - a_n)$$

mit positiven Zahlen  $b_1$  und  $d_1$ .

Da  $(d_n - c_n)_{n \in \mathbb{N}}$  und  $(b_n - a_n)_{n \in \mathbb{N}}$  Nullfolgen sind, können wir natürliche Zahlen  $n_1$  und  $n_2$  finden, so daß

$$d_n - c_n < \frac{\varepsilon}{2b_1} \text{ für alle } n \geq n_1 \text{ und } b_n - a_n < \frac{\varepsilon}{2d_1} \text{ für alle } n \geq n_2.$$

Für  $n \geq n_0 = \max(n_1, n_2)$  gilt dann

$$b_n d_n - a_n c_n \leq b_1 (d_n - c_n) + d_1 (b_n - a_n) < b_1 \cdot \frac{\varepsilon}{2b_1} + d_1 \cdot \frac{\varepsilon}{2d_1} = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Damit haben wir auch hier eine Nullfolge.

Die anderen Fälle gehen ähnlich; da wir in dieser Vorlesung noch zu wesentlich interessanteren Ergebnissen kommen wollen, sei auf Einzelheiten dazu wie auch zur Division verzichtet.

Auch auf ein weiteres technisches Problem möchte ich nur kurz eingehen: Eine reelle Zahl läßt sich durch viele verschiedene Intervallschachtelungen darstellen. Angenommen, wir haben  $x$  dargestellt sowohl durch die Intervallschachtelung  $([a_n, b_n])_{n \in \mathbb{N}}$  als auch durch  $([c_n, d_n])_{n \in \mathbb{N}}$ . Entsprechend sei eine reelle Zahl  $y$  gegeben durch

zwei Intervallschachtelungen  $([p_n, q_n])_{n \in \mathbb{N}}$  und  $([r_n, s_n])_{n \in \mathbb{N}}$ . Dann können wir  $x + y$  einmal berechnen über die Intervallschachtelung  $([a_n, b_n] + [p_n, q_n])_{n \in \mathbb{N}}$ , aber auch über  $([c_n, d_n] + [r_n, s_n])_{n \in \mathbb{N}}$ . Unsere Definition der Addition ist nur sinnvoll, wenn beide Intervallschachtelungen dieselbe reelle Zahl definieren.

Um dies nachzuweisen, müssen wir zeigen, daß die Differenzen der unteren wie auch der oberen Intervallgrenzen Nullfolgen sind, daß also

$$((a_n + p_n) - (c_n + r_n))_{n \in \mathbb{N}} \quad \text{und} \quad ((b_n + q_n) - (d_n + s_n))_{n \in \mathbb{N}}$$

Nullfolgen sind. Das ist aber einfach:

$$(a_n + p_n) - (c_n + r_n) = (a_n - c_n) + (p_n - r_n) \quad \text{und}$$

$$(b_n + q_n) - (d_n + s_n) = (b_n - d_n) + (q_n - s_n).$$

Da  $([a_n, b_n])_{n \in \mathbb{N}}$  und  $([c_n, d_n])_{n \in \mathbb{N}}$  dieselbe reelle Zahl darstellen, sind die Folgen  $(a_n - c_n)_{n \in \mathbb{N}}$  und  $(b_n - d_n)_{n \in \mathbb{N}}$  Nullfolgen, entsprechend auch  $(p_n - r_n)_{n \in \mathbb{N}}$  und  $(q_n - s_n)_{n \in \mathbb{N}}$ . Somit folgt die Behauptung einfach aus dem obigen Lemma, wonach die Summe zweier Nullfolgen wieder eine Nullfolge ist. Ähnlich zeigt man auch entsprechende Ergebnisse für die anderen Grundrechenarten.

Damit haben wir den Körper  $\mathbb{Q}$  der rationalen Zahlen erweitert zu einer Menge  $\mathbb{R}$ , auf der wir ebenfalls alle Rechenoperationen durchführen können. Nun stellt sich natürlich die Frage, ob auch  $\mathbb{R}$  ein Körper ist. Dies ist in der Tat der Fall:

**Satz:** Die reellen Zahlen mit den oben definierten Rechenoperationen bilden einen Körper.

Beim *Beweis* möchte ich mich auf eine kurze Darstellung der wesentlichen Punkte beschränken. Bei den Assoziativ- und Kommutativgesetzen sowie dem Distributivgesetz gibt es keine Probleme, denn die gelten sogar schon für das Rechnen mit Intervallen: Beispielsweise ist

$$[a_n, b_n] + [c_n, d_n] = [a_n + c_n, b_n + d_n]$$

$$= [c_n + a_n, d_n + b_n] = [c_n, d_n] + [a_n, b_n],$$

wir können die Aussage also zurückführen auf das entsprechende Gesetz für rationale Zahlen. Genauso sieht es aus mit der Existenz von Null-

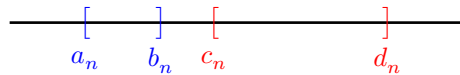
und Einselement: Wenn wir diese definieren durch Intervallschachtelungen, bei denen sämtliche Intervalle gleich  $[0, 0]$  beziehungsweise  $[1, 1]$  sind, haben wir auch hier entsprechende Aussagen schon für Intervalle. Bei der Existenz von Inversen geht das allerdings nicht: Ist etwa die reelle Zahl  $x$  gegeben durch die Intervallschachtelung  $([a_n, b_n])_{n \in \mathbb{N}}$ , so können wir  $-x$  definieren durch die Intervallschachtelung  $([-b_n, -a_n])_{n \in \mathbb{N}}$ . Die Summe

$$[a_n, b_n] + [-b_n, -a_n] = [a_n - b_n, b_n - a_n]$$

ist aber nur dann das Intervall  $[0, 0]$ , wenn  $a_n = b_n$  ist. Trotzdem definiert die Intervallschachtelung  $([a_n - b_n, b_n - a_n])_{n \in \mathbb{N}}$  die Null, denn nach Definition einer Intervallschachtelung ist ja  $(b_n - a_n)_{n \in \mathbb{N}}$  eine Nullfolge, und daraus folgt sofort, daß auch  $(a_n - b_n)_{n \in \mathbb{N}}$  eine ist. Mit nur geringfügig größerem Aufwand beweist man auch die Existenz von multiplikativen Inversen. ■

Was uns jetzt noch fehlt, ist die Ordnungsrelation: Wir wollen  $\mathbb{R}$  zu einem angeordneten Körper machen. Auch das geht leicht mit Hilfe von Intervallschachtelungen:

**Definition:** Die reellen Zahlen  $x$  und  $y$  seien gegeben durch die Intervallschachtelungen  $([a_n, b_n])_{n \in \mathbb{N}}$  und  $([c_n, d_n])_{n \in \mathbb{N}}$ . Dann ist  $x < y$  genau dann, wenn es einen Index  $n$  gibt, für den  $b_n < c_n$  ist.



Natürlich gilt diese Relation dann auch für alle auf  $n$  folgenden Indizes, denn das  $(n+1)$ -te Intervall liegt ja vollständig im  $n$ -ten.

**Satz:** Mit dieser Ordnungsrelation ist  $\mathbb{R}$  ein angeordneter Körper.

*Beweis:* Wir müssen als erstes zeigen, daß für zwei beliebige reelle Zahlen  $x, y$  stets genau eine der drei Relationen  $x < y$ ,  $x = y$  oder  $y < x$  gilt:

$x$  und  $y$  seien gegeben durch die Intervallschachtelungen  $([a_n, b_n])_{n \in \mathbb{N}}$  und  $([c_n, d_n])_{n \in \mathbb{N}}$ . Falls für irgendein  $n$  gilt  $b_n < c_n$ , ist  $x < y$ , falls  $d_n < a_n$  ist  $y < x$ . Beides gleichzeitig kann nicht gelten, denn wenn

eine dieser Relationen für ein  $n$  gilt, gilt sie auch für alle folgenden; gäbe es also für jede der beiden Relationen einen Index, so gäbe es auch ein  $n$ , für das beides gilt, d.h.  $d_n < a_n \leq b_n < c_n \leq d_n$ . Damit wäre die rationale Zahl  $d_n$  echt kleiner als sie selbst, was natürlich absurd ist.

Falls es kein  $n$  gibt, für das eine der beiden Relationen erfüllt ist, gelten für alle  $n$  die beiden Ungleichungen  $b_n \geq c_n$  und  $d_n \geq a_n$ . Wir müssen zeigen, daß dann  $x = y$  ist. Dazu bilden wir die Differenz  $x - y$ ; sie ist gegeben durch die Intervallschachtelung

$$([a_n, b_n] - [c_n, d_n])_{n \in \mathbb{N}} = ([a_n - d_n, b_n - c_n])_{n \in \mathbb{N}}.$$

Nach unserer Annahme ist  $b_n - c_n \geq 0$  für alle  $n$  und  $a_n - d_n \leq 0$  für alle  $n$ , die linke Intervallgrenze ist also stets kleiner oder gleich Null, die rechte größer oder gleich Null. Wie wir oben gesehen haben, ist das nur möglich, wenn sich die Intervallschachtelung auf die Null zusammenzieht, d.h.  $x - y = 0$  und  $x = y$ .

Zur Transitivität haben wir drei Elemente  $x, y, z$ , gegeben durch Intervallschachtelungen  $([a_n, b_n])_{n \in \mathbb{N}}$ ,  $([c_n, d_n])_{n \in \mathbb{N}}$  und  $([e_n, f_n])_{n \in \mathbb{N}}$ . Wir müssen zeigen: Ist  $x < y$  und  $y < z$ , so ist auch  $x < z$ .

Da  $x < y$ , gibt es ein  $n$ , so daß  $b_n < c_n$  ist; entsprechend gibt es wegen  $y < z$  ein  $m$ , so daß  $d_m < e_m$  ist. Wenn eine solche Ungleichung für einen Index gilt, gilt sie auch für alle folgenden; für jede Zahl  $k$  größer oder gleich dem Maximum von  $n$  und  $m$  ist also sowohl  $b_k < c_k$  also auch  $d_k < e_k$ . Dann ist aber auch  $b_k < c_k \leq d_k < e_k$ , also  $b_k < e_k$ . Das wiederum bedeutet, daß  $x < z$  ist.

Als nächstes muß gezeigt werden: Ist  $x < y$ , so gilt auch für jedes  $z \in \mathbb{R}$  die Ungleichung  $x + z < y + z$ .

$x, y$  und  $z$  seien wieder gegeben durch die Intervallschachtelungen  $([a_n, b_n])_{n \in \mathbb{N}}$ ,  $([c_n, d_n])_{n \in \mathbb{N}}$  und  $([e_n, f_n])_{n \in \mathbb{N}}$ . Wegen  $x < y$  gibt es ein  $n_0 \in \mathbb{N}$ , so daß  $b_n < c_n$  ist für alle  $n \geq n_0$ .

Die Intervallschachtelungen zu  $x + z$  und  $y + z$  sind

$$([a_n + e_n, b_n + f_n])_{n \in \mathbb{N}} \quad \text{und} \quad ([c_n + e_n, d_n + f_n])_{n \in \mathbb{N}}.$$

Für  $n \geq n_0$  ist  $b_n + f_n < c_n + f_n$ ; wir brauchen aber ein  $n$ , für das  $b_n + f_n < c_n + e_n$  ist. Falls es kein solches  $n$  gäbe, wäre  $b_n + f_n \geq c_n + e_n$  für alle  $n \in \mathbb{N}$ , also  $c_n - b_n \leq f_n - e_n$  für alle  $n \in \mathbb{N}$ .

Die Folge der  $f_n - e_n$  ist aber nach Definition einer Intervallschachtelung eine Nullfolge; für jedes  $\varepsilon > 0$  gäbe es also eine Zahl  $n$ , für die insbesondere  $f_n - e_n < \varepsilon$  wäre. Für jedes  $n \geq n_0$  ist aber  $c_n \geq c_{n_0}$  und  $b_n \leq b_{n_0}$ , also  $c_n - b_n \geq c_{n_0} - b_{n_0}$ , wir hätten also die Ungleichung

$$c_{n_0} - b_{n_0} \leq c_n - b_n \leq f_n - e_n < \varepsilon.$$

Da  $b_{n_0} < c_{n_0}$  ist, steht links eine feste positive Zahl, zu der wir leicht ein  $\varepsilon > 0$  finden können, für das diese Ungleichung nicht gilt, z.B.  $\varepsilon = \frac{1}{2}(c_{n_0} - b_{n_0})$ . Somit ist unsere Annahme falsch; es gibt also ein  $n$ , für das  $b_n + f_n < c_n + e_n$  ist, d.h.  $x + z < y + z$ .

Ähnlich folgt auch das noch fehlende Axiom, wonach Ungleichungen mit positiven Zahlen multipliziert werden können. ■

Damit können wir in den reellen Zahlen so rechnen, wie wir es gewohnt sind und können alle Rechenregeln, die in angeordneten Körpern gelten anwenden, ohne uns um Intervallschachtelungen kümmern zu müssen.

Beispielsweise wissen wir, daß es in  $\mathbb{R}$  zu jeder positiven Zahl  $a > 0$  genau eine positive reelle Zahl  $x > 0$  gibt mit  $x^2 = a$ ; wir bezeichnen sie natürlich wie gewohnt mit  $\sqrt{a}$ . Mit dieser Zahl können wir nun unter Ausnutzung aller Körperaxiome rechnen.

Als erstes Beispiel dafür wollen wir das Quadrat von  $x = \sqrt{3} + \sqrt{5}$  ausrechnen: Da  $\mathbb{R}$  ein Körper ist, gelten insbesondere auch die binomischen Formeln, also ist

$$(\sqrt{3} + \sqrt{5})^2 = 3 + 2\sqrt{3}\sqrt{5} + 5 = 8 + 2\sqrt{3}\sqrt{5}.$$

Das wollen wir natürlich weiter vereinfachen zu  $8 + 2\sqrt{15}$ , und das ist auch in der Tat möglich, denn es gilt

**Lemma:** Für zwei positive reelle Zahlen  $a, b$  ist  $\sqrt{ab} = \sqrt{a}\sqrt{b}$ .

*Beweis:* Nach dem Kommutativ- und Assoziativgesetz der Multiplikation ist

$$(\sqrt{a}\sqrt{b})^2 = \sqrt{a}\sqrt{b}\sqrt{a}\sqrt{b} = \sqrt{a}\sqrt{a}\sqrt{b}\sqrt{b} = ab.$$

$\sqrt{a}\sqrt{b}$  ist also eine Lösung der Gleichung  $x^2 = ab$ . Da  $\sqrt{a}$  und  $\sqrt{b}$  positiv sind, ist es eine positive Lösung, und die einzige positive Lösung von  $x^2 = ab$  ist  $x = \sqrt{ab}$ . ■

Auch komplizierte Ausdrücke können wir berechnen, z.B.  $\frac{\sqrt{3}+1}{\sqrt{3}-1}$ .

Hier ist das Hauptproblem zunächst einmal der Nenner, und den können wir in Ausdrücken wie diesem vereinfachen über die dritte binomische Formel:  $(\sqrt{3}+1)(\sqrt{3}-1) = 3 - 1 = 2$ . Wenn wir den Bruch mit  $\sqrt{3}+1$  erweitern, erhalten wir also

$$\frac{\sqrt{3}+1}{\sqrt{3}-1} = \frac{(\sqrt{3}+1)^2}{(\sqrt{3}-1)(\sqrt{3}+1)} = \frac{3+2\sqrt{3}+1}{3-1} = \frac{4+2\sqrt{3}}{2} = 2 + \sqrt{3},$$

was sicherlich ein sehr viel angenehmerer Ausdruck ist.

## §5: Komplexe Zahlen

Wir alle wissen, wie man quadratische Gleichungen löst; das entsprechende Verfahren war – genau wie das heute nach HERON benannte Verfahren zur Berechnung der Quadratwurzel – bereits vor vier Tausend Jahren den babylonischen Mathematikern bekannt.

Der Trick, den sie zur Lösung der Gleichung  $x^2 + px + q = 0$  benutzten, ist die sogenannte quadratische Ergänzung: Wir schreiben die Gleichung um als

$$\left(x + \frac{p}{2}\right)^2 - \frac{p^2}{4} + q = 0 \quad \text{oder} \quad \left(x + \frac{p}{2}\right)^2 = \frac{p^2}{4} - q$$

und können sie nun durch einfaches Wurzelziehen lösen – falls die Wurzel existiert.

Wenn  $p$  und  $q$  reelle Zahlen sind und wir auch Lösungen  $x \in \mathbb{R}$  suchen, haben wir im vorigen Paragraphen gesehen, daß jede positive Zahl genau eine positive Quadratwurzel hat; da  $(-1) \cdot (-1) = 1$  ist, ist natürlich auch deren Negatives eine Zahl mit gleichem Quadrat. Falls also die sogenannte *Diskriminante*

$$\Delta \stackrel{\text{def}}{=} \frac{p^2}{4} - q$$

positiv ist, haben wir im reellen Fall die beiden Lösungen

$$x + \frac{p}{2} = \pm \sqrt{\frac{p^2}{4} - q} \quad \text{oder} \quad x = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}.$$

Für  $\Delta = 0$  gibt es offensichtlich nur die eine Lösung  $x = -\frac{1}{2}p$ .



Ist aber  $\Delta < 0$ , so können wir keine reelle Zahl mit Quadrat  $\Delta$  finden: Wir wissen schließlich, daß in einem angeordneten Körper alle Quadrate größer oder gleich Null sind.

Das ist keine wirklich neue Situation für uns: Wir haben zu Beginn dieses Kapitels die natürlichen Zahlen betrachtet, erweiterten diese, um Subtrahieren zu können, zu den ganzen Zahlen, diese dann, um Divisionen zu ermöglichen, zum Körper  $\mathbb{Q}$  der rationalen Zahlen, und weil wir dort die Gleichung  $x^2 = 2$  nicht lösen konnten, haben wir den nochmals erweitert zum Körper  $\mathbb{R}$  der reellen Zahlen. Offensichtlich ist nun wieder einmal eine Erweiterung fällig.

Wir brauchen im wesentlichen nur eine einzige neue Zahl: Die *imaginäre Einheit*  $i$ , deren Quadrat gleich minus eins sein soll. Wenn wir die gewohnten Rechenregeln beibehalten können, läßt sich dann auch für jede negative reelle Zahl  $-a$  (mit  $a > 0$ ) eine Lösung der Gleichung  $x^2 = -a$  finden, denn

$$(i\sqrt{a})^2 = (i\sqrt{a}) \cdot (i\sqrt{a}) = i^2 \cdot (\sqrt{a})^2 = -a.$$

Natürlich ist auch  $-i\sqrt{a}$  eine Lösung, aber das ist gegenüber dem Reellen keine neue Situation.

Wir suchen somit nach einem Körper, der einerseits die reellen Zahlen enthält, andererseits aber auch ein neues Element  $i$ , die imaginäre Einheit, für die  $i^2 = -1$  ist.

Da man in einem Körper beliebig addieren und multiplizieren kann, muß dieser Körper für je zwei reelle Zahlen  $a, b \in \mathbb{R}$  auch das Element  $a+bi$  enthalten. Wir wollen sehen, daß wir mit (zunächst noch formalen) Summen dieser Art auch tatsächlich auskommen:

**Definition:** a) Die Menge der komplexen Zahlen ist

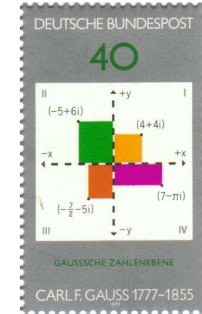
$$\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}.$$

b) Für  $c = a + ib \in \mathbb{C}$  bezeichnen wir  $a$  als den *Realteil* von  $c$  und  $b$  als den *Imaginärteil* von  $c$ ; in Formeln:  $a = \Re c$  und  $b = \Im c$ .

Eine komplexe Zahl ist somit gegeben durch *zwei* reelle Zahlen  $a$  und  $b$ , den Realteil  $a$  und den Imaginärteil  $b$ . Eine Zahl der Form  $a + 0i$  identifizieren wir mit der reellen Zahl  $a$ ; Zahlen der Form  $0 + bi$  oder kurz  $bi$

bezeichnen wir als (rein) imaginär, weil sie nach Auffassung der Mathematiker des 18. Jahrhunderts „imaginär“, d.h. nur in der Vorstellung vorhanden waren. Der Name *komplexe* Zahl schließlich kommt daher, daß eine solche Zahl aus einer reellen und einer imaginären Zahl zusammengesetzt ist.

Genau wie man die reellen Zahlen oft durch eine Gerade veranschaulicht, kann man also die Menge der komplexen Zahlen veranschaulichen durch eine Ebene, deren Punkt  $(a, b)$  bezüglich eines vorgegebenen kartesischen Koordinatensystems die komplexe Zahl  $a+ib$  darstellen soll. Man redet hier von der GAUSSschen Zahlenebene, deren Idee auf einer Briefmarke zum 200. Geburtstag von GAUSS veranschaulicht ist.



CARL FRIEDRICH GAUSS (1777–1855) leistete wesentliche Beiträge zur Zahlentheorie, zur nichteuklidischen Geometrie, zur Funktionentheorie, zur Differentialgeometrie und Kartographie, zur Fehlerrechnung und Statistik, zur Astronomie und Geophysik usw. Als Direktor der Göttinger Sternwarte baute er zusammen mit dem Physiker Weber den ersten Telegraphen. Er leitete die erste Vermessung und Kartierung des Königreichs Hannover und zeitweise auch den Witwenfond der Universität Göttingen; seine hierbei gewonnene Erfahrung nutzte er für erfolgreiche Spekulationen mit Aktien. Sätze und Verfahren von GAUSS sind noch heute allgegenwärtig in der Mathematik, Physik und Statistik.

Wenn wir die Menge  $\mathbb{C}$  zu einem Körper machen wollen, müssen wir zunächst Rechenoperationen definieren. Durch die Körperaxiome und die Tatsache, daß  $\mathbb{R}$  ein Teilkörper von  $\mathbb{C}$  sein soll, ist dabei eigentlich alles vorgegeben: Wenn ein Assoziativgesetz der Addition und ein Distributivgesetz gelten soll, muß

$$(a + bi) \pm (c + di) = (a \pm c) + (b \pm d)i$$

sein, und Multiplizieren heißt einfach Ausmultiplizieren nach dem Distributiv- und Kommutativgesetz:

$$(a + bi)(c + di) = ac + a \cdot di + bi \cdot c + bd \cdot i^2 = (ac - bd) + (bc + ad)i.$$

Für die Division können wir den gleichen Trick anwenden wie am Ende des letzten Paragraphen bei der Division durch reelle Wurzel­ausdrücke: Nach der dritten binomischen Formel (oder der gerade hingeschriebenen Definition des Produkts) ist

$$(c + di)(c - di) = c^2 - d^2 \cdot i^2 = c^2 + d^2.$$

$c^2 + d^2$  ist für alle komplexen Zahlen mit  $c + di \neq 0$  positiv, insbesondere also ungleich Null; daher können wir Brüche mit Nenner  $c + di \neq 0$  berechnen als

$$\frac{a + bi}{c + di} = \frac{(a + bi)(c - di)}{(c + di)(c - di)} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2} i.$$

**Satz:** Die Menge  $\mathbb{C}$  mit den oben definierten Verknüpfungen ist ein Körper.

*Beweis:* Eigentlich müßten wir alle Körperaxiome einzeln überprüfen; wie bei den reellen Zahlen wäre das recht langweilig und uninstruktiv. Um nicht gar zuviel Papier zu produzieren, möchte ich mich daher im Sinne des Umweltschutzes auf die interessantesten beschränken.

Völlig uninteressant sind die Axiome, die sich mit der Addition in  $\mathbb{C}$  befassen: Da die Addition komponentenweise für Realteil und Imaginärteil definiert ist, folgen alle Axiome sofort aus den entsprechenden Axiomen für  $\mathbb{R}$ . Das Neutralelement bezüglich der Addition ist natürlich  $0 + 0i$ , und  $-(a + bi) = (-a) + (-b)i$  ist das Negative von  $a + ib$ .

Die Forderungen an die Multiplikation sind weniger offensichtlich. Unmittelbar einsichtig ist das Kommutativgesetz; das Assoziativgesetz dagegen ist eine eher unangenehme sture Nachrechnung:

$$\begin{aligned} ((a + bi)(c + di))(e + fi) &= ((ac - bd) + (ad + bc)i)(e + fi) \\ &= ((ac - bd)e - (ad + bc)f) + ((ac - bd)f + (ad + bc)e)i \\ &= (ace - bde - adf - bcf) + (acf - bdf + ade + bce)i \end{aligned}$$

und

$$\begin{aligned} (a + bi)((c + di)(e + fi)) &= (a + bi)(ce - df) + (cf + de)i \\ &= (a(ce - df) - b(cf + de)) + (a(cf + de) + b(ce - df))i \\ &= (ace - adf - bcf - bde) + (acf + ade + bce - bdf)i \end{aligned}$$

Die Kommutativität der Addition in  $\mathbb{R}$  zeigt, daß beide Ausdrücke denselben Realteil und denselben Imaginärteil haben, also gleich sind.

Neutralelement bezüglich der Multiplikation kann, wenn alles Sinn haben soll, nur  $1 + 0i$  sein, und in der Tat rechnet man ohne Schwierigkeiten nach, daß jedes Element  $a + bi$  sowohl bei Links- wie auch bei Rechtsmultiplikation hiermit sich selbst liefert.

Bleibt die Existenz eines multiplikativen Inversen; so wie wir oben die Division komplexer Zahlen definiert haben, ist schnell klar, daß  $1/(a+bi)$  invers zu  $a + bi$  ist:

$$\begin{aligned} \frac{1}{a + bi} \cdot (a + bi) &= \left( \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2} i \right) \cdot (a + bi) \\ &= \frac{a^2 + b^2}{a^2 + b^2} + \frac{ab - ba}{a^2 + b^2} i = 1. \end{aligned}$$

Das noch verbleibende Distributivgesetz sind wieder Rechnerei bis zum Abwinken, mit der ein interessierter Leser keinerlei Schwierigkeit haben sollte. ■

Damit haben wir unser Ziel erreicht, einen Körper zu finden, der die reellen Zahlen enthält und zusätzlich auch noch Quadratwurzeln aller reeller Zahlen unabhängig von deren Vorzeichen.

Wie sieht es aus mit Quadratwurzeln *komplexer* Zahlen? Hat die Gleichung  $z^2 = c$  für jede komplexe Zahl  $c$  eine Lösung?

Wir schreiben  $c = a + bi$  und machen einen Ansatz  $z = x + yi$ , wobei nun  $a, b$  bekannte und  $x, y$  gesuchte *reelle* Zahlen sind. Die Gleichung  $z^2 = c$  oder  $(x + yi)^2 = a + bi$  wird ausmultipliziert zu den beiden Gleichungen  $x^2 - y^2 = a$  und  $2xy = b$  für Real- und Imaginärteil. Daher ist

$$\begin{aligned} (x - yi)^2 &= (x^2 - y^2) - 2xyi = a - bi, \text{ also} \\ (x + iy)^2(x - iy)^2 &= (a + ib)(a - ib) = a^2 + b^2. \end{aligned}$$

Nun ist aber  $(x + iy)(x - iy) = x^2 + y^2 \geq 0$ , also ist  $(x^2 + y^2)^2 = a^2 + b^2$  und  $x^2 + y^2 = \sqrt{a^2 + b^2}$ . Zusammen mit der obigen Gleichung  $x^2 - y^2 = a$  haben wir somit das Gleichungssystem

$$x^2 + y^2 = \sqrt{a^2 + b^2} \quad \text{und} \quad x^2 - y^2 = a.$$

Addition bzw. Subtraktion dieser beiden Gleichungen führt auf

$$2x^2 = \sqrt{a^2 + b^2} + a \quad \text{und} \quad 2y^2 = \sqrt{a^2 + b^2} - a,$$

also ist

$$x = \pm \frac{\sqrt{2}}{2} \sqrt{\sqrt{a^2 + b^2} + a} \quad \text{und} \quad y = \pm \frac{\sqrt{2}}{2} \sqrt{\sqrt{a^2 + b^2} - a},$$

wobei die beiden Vorzeichen so gewählt werden müssen, das  $xy$  dasselbe Vorzeichen wie  $b$  hat, denn schließlich ist  $2xy = b$ .

Nachdem wir Quadratwurzeln aus komplexen Zahlen ziehen können, ist klar, daß wir mit der üblichen Methode der quadratischen Ergänzung oder nach der üblichen Formel auch quadratische Gleichungen mit komplexen Koeffizienten lösen können.

Gelegentlich können wir sogar kompliziertere Gleichungen auf ähnliche Weise lösen: Wenn wir beispielsweise alle  $z \in \mathbb{C}$  suchen, für die  $z^3 = 1$  ist, schreiben wir wieder  $z = x + iy$  und berechnen

$$z^3 = (x + iy)^3 = x^3 + 3x^2 \cdot iy + 3x \cdot (iy)^2 + (iy)^3 = x^3 - 3xy^2 + (3x^2y - y^3)i.$$

Dieser Ausdruck ist genau dann gleich eins, wenn

$$x^3 - 3xy^2 = 1 \quad \text{und} \quad 3x^2y - y^3 = y(3x^2 - y^2) = 0$$

ist, und die zweite Gleichung ist genau dann erfüllt, wenn  $y = 0$  oder  $y^2 = 3x^2$  ist.

$y = 0$  in die erste Gleichung eingesetzt führt auf  $x^3 = 1$ , also (da  $x$  eine reelle Zahl ist)  $x = 1$ . Die Wurzel  $x + iy = 1$  ist in diesem Fall also einfach die wohlbekanntere reelle Kubikwurzel.

Setzen wir  $y^2 = 3x^2$  in die erste Gleichung ein, erhalten wir die Gleichung  $-8x^3 = 1$  mit der Lösung  $x = -\frac{1}{2}$ . Wegen  $y^2 = 3x^2$  gibt es für  $y$  die beiden Möglichkeiten  $\pm \frac{1}{2}\sqrt{3}$ , die dritten Wurzeln der Eins sind also

$$1, \quad \rho = -\frac{1}{2} + \frac{1}{2}\sqrt{3}i \quad \text{und} \quad \bar{\rho} = -\frac{1}{2} - \frac{1}{2}\sqrt{3}i.$$

Damit ist klar, daß jede von Null verschiedene reelle Zahl  $a$  drei Kubikwurzeln hat: Die reelle Wurzel  $\sqrt[3]{a}$  sowie die beiden komplexen Wurzeln  $\rho \sqrt[3]{a}$  und  $\bar{\rho} \sqrt[3]{a}$ .

Für die Lösungen der Gleichung  $z^4 = 1$  müssen wir nicht so lange rechnen: Wenn  $z^4 = 1$  ist, muß  $z^2 = \pm 1$  sein.  $z^2 = 1$  hat die beiden

reellen Lösungen  $z = 1$  und  $z = -1$ ; die Lösungen von  $z^2 = -1$  sind  $z = i$  und  $z = -i$ .

Auch die Lösungen von  $z^4 = -1$  lassen sich leicht berechnen: Aus  $z^4 = -1$  folgt  $z^2 = \pm i$ , und die obigen Formeln für die Quadratwurzeln einer komplexen Zahl zeigen, daß dann

$$z = \frac{\sqrt{2}}{2} (\pm 1 \pm i)$$

sein muß, wobei alle vier Vorzeichenkombinationen möglich sind.

Damit können wir also eine ganze Reihe von Gleichungen lösen, für die es im Reellen keine Lösung gibt; tatsächlich gilt sogar der

**Fundamentalsatz der Algebra:** Jedes Polynom

$$f(x) = x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_2x^2 + a_1x + a_0$$

mit komplexen Koeffizienten  $a_i$  vom Grad  $n \geq 1$  hat mindestens eine komplexe Nullstelle.

Der Beweis dieses Satzes ist mit den Mitteln einer Vorlesung *Analysis I* nicht möglich: Wir wissen zwar bereits, daß er für quadratische Gleichungen gilt, und im Falle reeller Koeffizienten zeigt eine Kurvendiskussion, daß es für Gleichungen ungeraden Grades mindestens eine Nullstelle geben muß. Für beliebige Polynome gibt es aber leider keine so einfachen Argumente.

Mit Methoden aus der Algebra zeigte GAUSS 1815, daß eine geschickte Kombination der beiden gerade betrachteten Spezialfälle zu einem Beweis des allgemeinen Satzes führt; die dazu benötigte Algebra steht aber erst gegen Ende einer *Algebra I*-Vorlesung zur Verfügung. Meist wird der Satz auch dort nicht bewiesen, denn es gibt inzwischen einen einfacheren Beweis mit Mitteln der *Funktionentheorie*, d.h. der Analysis komplexer Veränderlicher. Eine dritter Beweis schließlich arbeitet mit Methoden der algebraischen Topologie; auch hier folgt der Satz ziemlich schnell aus Standardsätzen dieses Gebiets.

Wenn wir den Fundamentalsatz der Algebra glauben, können wir ihn aber leicht verschärfen:

**Satz:** Zu jedem Polynom

$$f(x) = x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_2x^2 + a_1x + a_0$$

mit komplexen Koeffizienten  $a_i$  gibt es  $n$  (nicht notwendigerweise verschiedene) komplexe Zahlen  $c_1, \dots, c_n$ , so daß

$$f(x) = (x - c_1)(x - c_2) \cdots (x - c_n)$$

ist. Insbesondere ist  $f(x)$  genau dann gleich Null, wenn  $x$  eine dieser Zahlen  $c_i$  ist.

*Beweis* durch vollständige Induktion: Für  $n = 1$  ist  $f(x) = x + a_0$ , also können wir einfach  $c_1 = -a_0$  setzen, und der Induktionsanfang ist bewiesen.

Für den Induktionsschritt nehmen wir an, der Satz sei für einen festen Grad  $n$  bewiesen, und betrachten ein Polynom

$$g(x) = x^{n+1} + b_nx^n + b_{n-1}x^{n-1} + \dots + b_2x^2 + b_1x + b_0$$

mit komplexen Koeffizienten  $b_i$ . Nach dem Fundamentalsatz der Algebra gibt es mindestens eine komplexe Nullstelle. Wir wählen eine dieser Nullstellen aus, nennen sie  $c_{n+1}$  und dividieren das Polynom  $g(x)$  mit Rest durch das Polynom  $(x - c_{n+1})$ :

$$g(x) : (x - c_{n+1}) = f(x) \text{ Rest } r(x).$$

Da wir durch ein lineares Polynom dividiert haben, ist der Rest ein konstantes Polynom, hängt also nicht von  $x$  ab, und wir können auch einfach  $r(x) = r$  schreiben mit einer komplexen Zahl  $r$ . Somit ist

$$g(x) = (x - c_{n+1})f(x) + r.$$

Setzen wir hier  $x = c_{n+1}$ , wird die Gleichung zu  $0 = 0 + r$ , d.h.  $r = 0$  und  $g(x) = (x - c_{n+1})f(x)$ .

Da  $g(x)$  ein Polynom vom Grad  $n + 1$  war, muß  $f(x)$  Grad  $n$  haben, und wie Ausmultiplizieren zeigt, kann der Koeffizient von  $x^n$  nur eins sein. Nach Induktionsvoraussetzung gibt es daher komplexe Zahlen  $c_1, \dots, c_n$ , so daß  $f(x) = (x - c_1) \cdots (x - c_n)$  ist. Somit ist

$$g(x) = (x - c_{n+1})f(x) = (x - c_1) \cdots (x - c_n)(x - c_{n+1}),$$

wie behauptet. ■

Damit haben wir einen Körper gefunden, in dem wir nicht nur alle Grundrechenarten (außer Division durch Null) ausführen können, sondern in dem wir sogar beliebige Polynomgleichungen lösen können. Trotzdem werden wir uns im Rest der Vorlesung fast immer auf die reellen Zahlen beschränken.

Der Grund ist einfach: Bei unseren bisherigen Zahlbereichserweiterungen ist es uns immer gelungen, alle wesentlichen Eigenschaften des kleineren Bereichs auf den größeren zu übertragen. Beim Übergang von den reellen zu den komplexen Zahlen konnten wir aber nur die Körperaxiome übertragen, nicht mehr die eines angeordneten Körpers:

Wäre  $\mathbb{C}$  bezüglich irgendeiner Ordnungsrelation ein angeordneter Körper, so wäre  $i^2 = -1 > 0$ , da in einem angeordneten Körper jedes Quadrat eines von Null verschiedenen Elements positiv ist. Somit wäre  $0 = (-1) + 1 = i^2 + 1^2 > 0$ , was absurd ist. Es gibt also keine mit den Rechenoperationen verträgliche Ordnungsrelation auf den komplexen Zahlen, so daß viele Methoden der (reellen) Analysis nicht auf komplexe Zahlen anwendbar sind. Trotzdem gibt es natürlich Beziehungen zwischen der reellen und der komplexen Analysis; an einigen wenigen Stellen werden die uns auch in dieser Vorlesung das Leben erleichtern.

Eine einzige Konstruktion für angeordnete Körper können wir auch auf komplexe Zahlen übertragen: den Betrag. Wir können den Betrag einer komplexen Zahl  $c$  natürlich nicht definieren als  $c$  für  $c \geq 0$  und  $-c$  sonst, aber wir können definieren

**Definition:** a) Der *Betrag*  $|c|$  der komplexen Zahl  $a + bi$  ist

$$|c| = \sqrt{a^2 + b^2} = \sqrt{(a + bi)(a - bi)}.$$

b)  $\bar{c} = a - bi$  heißt die zu  $c$  *konjugiert komplexe* Zahl.

Der Betrag einer komplexen Zahl ist also eine *reelle* Zahl, denn unter der Wurzel steht ja eine Summe zweier Quadrate reeller Zahlen. Er ist stets größer oder gleich Null und verschwindet wie im Reellen genau dann, wenn bereits  $c = 0$  ist, denn  $a^2 + b^2$  kann für reelle Zahlen  $a, b$  nur verschwinden für  $a = b = 0$ .

Im nächsten Kapitel werden wir eine anschauliche, geometrische Interpretation des Betrags einer komplexen Zahl kennenlernen.

## §6: Die Mächtigkeit einer Menge

Wir haben im bisherigen Verlauf der Vorlesung unseren Zahlbereich immer mehr erweitert, von den natürlichen zu den ganzen Zahlen, von den ganzen zu den rationalen und den reellen, und schließlich im letzten Paragraphen noch zu den komplexen Zahlen. Natürlich ist in der Inklusionskette

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

jede Menge eine echte Teilmenge der folgenden. Wenn wir es mit endlichen Mengen zu tun hätten, wäre somit klar, daß die Elementanzahl von  $\mathbb{Z}$  größer wäre als die von  $\mathbb{N}$ , aber kleiner als die von  $\mathbb{Q}$ , und so weiter.

Um zu sehen, warum wir bei unendlichen Mengen nicht so einfach argumentieren können, wollen wir die Menge  $G$  der geraden natürlichen Zahlen vergleichen mit der Menge  $\mathbb{N}$  aller natürlichen Zahlen. Natürlich ist  $G$  eine echte Teilmenge von  $\mathbb{N}$ ; es gibt schließlich auch noch ungerade Zahlen. Da auf jede ungerade Zahl eine gerade folgt und umgekehrt, liegt die Aussage nahe, daß es genauso viele gerade wie ungerade Zahlen gibt und daß  $\mathbb{N}$  somit doppelt so viele Elemente enthält wie  $G$ .

Nun sind  $G$  und  $N$  aber unendliche Mengen; was ist das Doppelte von unendlich?

Wie problematisch die Situation wirklich ist, sehen wir daran, daß wir jede gerade Zahl  $g \in G$  schreiben können als  $g = 2n$  mit  $n \in \mathbb{N}$ ; bei dieser Sicht der Dinge sollte es also genauso viele gerade wie natürliche Zahlen geben, eben unendlich viele. Können wir dann nur unterscheiden zwischen endlichen und unendlichen Mengen und müssen sagen, daß alle unendlichen Mengen gleich viele Elemente enthalten?

CANTOR fand einen Weg, um auch zwischen unendlichen Mengen noch weiter zu differenzieren:

**Definition:** a) Eine Abbildung  $f: M \rightarrow N$  zwischen zwei Mengen  $M, N$  ist eine Vorschrift, die jedem Element  $m \in M$  ein eindeutig bestimmtes Element  $f(m) \in N$  zuordnet.

b) Eine Abbildung  $f: M \rightarrow N$  heißt *injektiv*, wenn verschiedene Elemente von  $M$  auf verschiedene Elemente von  $N$  abgebildet werden,

wenn also  $f(m) = f(n)$  nur dann gilt, wenn auch  $m = n$  ist.

c)  $f$  heißt *surjektiv*, wenn es zu jedem  $n \in N$  ein Urbild  $m \in M$  gibt, d.h. ein Element mit  $f(m) = n$ .

d)  $f$  heißt *bijektiv*, wenn  $f$  sowohl injektiv als auch surjektiv ist.

e) Zwei Mengen  $M, N$  heißen *gleichmächtig*, wenn es eine bijektive Abbildung  $f: M \rightarrow N$  gibt.

f) Eine Menge  $M$  heißt *abzählbar*, wenn es eine bijektive Abbildung  $f: T \rightarrow M$  einer Teilmenge  $T \subseteq \mathbb{N}$  auf  $M$  gibt; andernfalls heißt sie *überabzählbar*.

In diesem Sinne ist beispielsweise die Abbildung  $f: \mathbb{N} \rightarrow \mathbb{Z}$ , die jede natürliche Zahl auf sich selbst abbildet, injektiv, aber nicht surjektiv, da die Null und die negativen Zahlen keine Urbilder haben. Dagegen ist die Abbildung

$$f: \begin{cases} \mathbb{Z} \rightarrow \mathbb{N}_0 \\ z \mapsto |z| = \begin{cases} z & \text{falls } z \geq 0 \\ -z & \text{falls } z < 0 \end{cases} \end{cases}$$

surjektiv, aber nicht injektiv, da beispielsweise die Eins sowohl  $+1$  als auch  $-1$  als Urbild hat. Beispiele von abzählbaren Mengen sind zunächst die endlichen Mengen, denn natürlich können wir zu einer Menge  $M$  mit  $n$  Elementen eine bijektive Abbildung

$$\{1, 2, 3, \dots, n\} \rightarrow M$$

finden, die die Elemente von  $M$  zählt. Beispiel einer unendlichen abzählbaren Menge sind etwa die geraden Zahlen; wir können sie abzählen durch die Abbildung  $f: \mathbb{N} \rightarrow G$  mit  $f(n) = 2n$ . Somit sind also die Mengen  $\mathbb{N}$  und  $G$  gleichmächtig, obwohl  $G$  eine echte Teilmenge von  $\mathbb{N}$  ist. Das erscheint auf den ersten Blick seltsam, aber wir müssen uns daran gewöhnen, daß beim Umgang mit dem Unendlichen immer wieder Phänomene auftreten, die zumindest auf den ersten Blick paradox erscheinen. Es ist eine charakteristische Eigenschaft unendlicher Mengen, daß sie echte Teilmengen haben, die gleichmächtig zur Gesamtmenge sind.

Wir können auch eine bijektive Abbildung von  $\mathbb{N}$  nach  $\mathbb{Z}$  finden, indem wir die ganzen Zahlen wie folgt anordnen

$$0, 1, -1, 2, -2, 3, -3, 4, -4, \dots$$

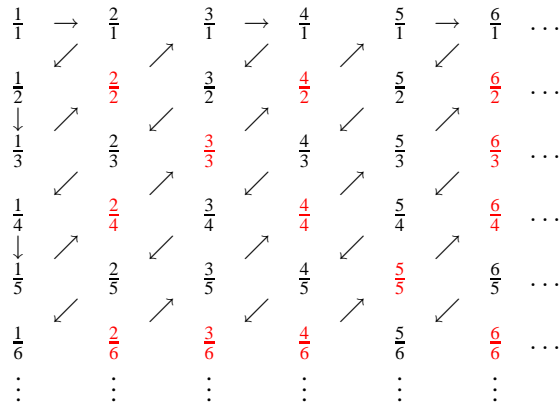
und der natürlichen Zahl  $n$  das  $n$ -te Element dieser Liste zuordnen. In Formeln ausgedrückt haben wir dann die Abbildung

$$f: \begin{cases} \mathbb{N} \rightarrow \mathbb{Z} \\ n \mapsto f(n) = \begin{cases} \frac{n}{2} & \text{falls } n \text{ gerade} \\ \frac{1-n}{2} & \text{falls } n \text{ ungerade} \end{cases} \end{cases}$$

Somit ist auch  $\mathbb{Z}$  abzählbar.

Betrachten wir als nächstes die rationalen Zahlen! Da zwei rationale Zahlen beliebig nahe beieinander liegen können, fällt es schwer, sich hier eine Abzählung vorzustellen. Wenn wir allerdings daran denken, daß eine rationale Zahl gegeben ist durch ihren Zähler und ihren Nenner, läßt sich doch leicht eine konstruieren. Sie geht auf CANTOR zurück und wird heute als erstes CANTORSches Diagonalverfahren bezeichnet.

Er zeigt zunächst, daß die Menge  $\mathbb{Q}^+$  aller positiver rationaler Zahlen gleichmächtig ist zu  $\mathbb{N}$ , indem er die positiven rationalen Zahlen in einer zweidimensionalen Anordnung aufschreibt, wobei an der Position  $(i, j)$  der Bruch  $i/j$  steht. Dann geht er mäanderförmig durch diese Anordnung:



Dabei kommt natürlich jeder Bruch mehrfach vor: Beispielsweise ist  $\frac{1}{2} = \frac{2}{4} = \frac{3}{6}$  und so weiter. CANTOR streicht deshalb alle Brüche, die schon einmal vorgekommen sind (in der obigen Abbildung sind das die rot eingezeichneten) und erhält so eine Folge, in der jede positive

rationale Zahl genau einmal vorkommt:

$$\frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{1}{3}, \frac{3}{1}, \frac{4}{1}, \frac{3}{2}, \frac{2}{3}, \frac{1}{4}, \frac{5}{1}, \frac{6}{1}, \frac{5}{2}, \frac{4}{3}, \frac{3}{4}, \frac{2}{5}, \frac{1}{6}, \dots$$

Die Abbildung, die der natürlichen Zahl  $n$  das  $n$ -te Glied dieser Folge zuordnet, ist offensichtlich eine Bijektion  $g$  von  $\mathbb{N}$  auf die Menge aller positiver rationaler Zahlen. Sie läßt sich zwar nicht durch eine kurzen Formel ausdrücken, aber für jede Zahl  $n$  läßt sich eindeutig bestimmen, auf welchen Bruch sie abgebildet wird.

Diese Bijektion läßt sich, wenn man die negativen Brüche entsprechend auflistet, fortsetzen zu einer Bijektion  $\mathbb{Z} \rightarrow \mathbb{Q}$ , die eine positive Zahl  $z$  auf  $g(z)$  abbildet, die Null auf die Null, und eine negative Zahl  $z$  auf  $-g(-z)$ : Für negatives  $z$  ist  $-z$  positiv,  $g(-z)$  ein positiver Bruch und  $-g(-z)$  dessen Negatives. Somit sind  $\mathbb{Q}$  und  $\mathbb{Z}$  gleichmächtig, also auch  $\mathbb{Q}$  und  $\mathbb{N}$ .

Um zu sehen, daß nicht jede unendliche Menge gleichmächtig zu  $\mathbb{N}$  ist, betrachten wir die Menge  $P$  aller Teilmengen von  $\mathbb{N}$ . Wir wollen uns überlegen, daß es keine bijektive Abbildung  $f: \mathbb{N} \rightarrow P$  geben kann. Der Beweis geht wieder auf CANTOR zurück; man spricht heute vom zweiten CANTORSchen Diagonalverfahren.

Da sich am Aufwand nichts ändert, beweisen wir gleich allgemeiner:

**Lemma:** Ist  $P$  die Menge aller Teilmengen einer Menge  $M$ , so kann es keine bijektive Abbildung  $M \rightarrow P$  geben;  $M$  und  $P$  sind also nicht gleichmächtig.

Zum *Beweis* betrachten wir eine Abbildung  $f: M \rightarrow P$ , z.B. die Abbildung mit  $f(m) = \{m\}$  oder irgendeine andere. Für jedes Element  $m \in M$  ist  $f(m)$  ein Element von  $P$ , also eine Teilmenge von  $M$ . Wir können uns daher fragen, welche Elemente  $m \in M$  in der Menge  $f(m)$  liegen. Dazu betrachten wir die Menge

$$T = \{m \in M \mid m \notin f(m)\}.$$

$T$  ist auf jeden Fall eine Teilmenge von  $M$ . Falls es ein Element  $t \in M$  gäbe mit  $f(t) = T$ , so könnte  $t$  nicht in  $T = f(t)$  liegen, denn  $T$  besteht ja genau aus den Elementen  $m \in M$ , die nicht in  $f(m)$  liegen. Wenn aber  $t$  nicht in  $f(t)$  liegt, dann liegt  $t$  nach Definition von  $T = f(t)$  in dieser

Menge. Sowohl die Annahme  $t \in T$  als auch die Annahme  $t \notin f(t)$  führen also zu Widersprüchen. Somit kann es kein  $t \in M$  geben mit  $f(t) = T$ . Daher kann  $f$  nicht surjektiv und erst recht nicht bijektiv sein. ■

Wie sieht es aus mit den reellen Zahlen?

Da wir sie via Intervallschachtelung beliebig genau durch rationale Zahlen approximieren können, sollten wir nicht erwarten, daß es sehr viel mehr reelle als rationale Zahlen gibt; wie CANTOR gezeigt hat, sind aber die reellen Zahlen *nicht* abzählbar. Zum Beweis geht er genauso vor wie bei der Menge aller Teilmengen von  $\mathbb{N}$ : Angenommen, die reellen Zahlen wären abzählbar, es gäbe also eine bijektive Abbildung  $f: \mathbb{N} \rightarrow \mathbb{R}$ . Wir wollen durch eine Intervallschachtelung eine reelle Zahl  $x$  konstruieren, die nicht im Bild von  $f$  liegen kann und so zeigen, daß  $f$  im Widerspruch zur Annahme doch nicht surjektiv ist.

Wir starten mit irgendeinem Intervall  $[a_1, b_1]$  mit rationalen Grenzen  $a_1, b_1 \in \mathbb{Q}$ , das die reelle Zahl  $f(1)$  nicht enthält. Darin können wir ein Teilintervall  $[a_2, b_2]$  mit höchstens halber Länge finden, das  $f(2)$  nicht enthält: Falls  $f(2)$  nicht gerade gleich der Intervallmitte  $\frac{1}{2}(a_1 + b_1)$  ist, können wir einfach das Intervall halbieren und mindestens eines der beiden Intervalle

$$\left[ a_1, \frac{a_1 + b_1}{2} \right] \quad \text{oder} \quad \left[ \frac{a_1 + b_1}{2}, b_1 \right]$$

auswählen, und für  $f(2) = \frac{1}{2}(a_1 + b_1)$  können wir zum Beispiel das linke Drittel

$$\left[ a_1, a_1 + \frac{b_1 - a_1}{3} \right]$$

nehmen.

Entsprechend konstruieren wir rekursiv weitere Intervalle: Innerhalb eines jeden Intervalls  $[a_n, b_n]$  wählen wir ein Teilintervall  $[a_{n+1}, b_{n+1}]$  von höchstens halber Länge, das  $f(n+1)$  nicht enthält.

Damit haben wir offensichtlich eine Intervallschachtelung: Die Bedingung  $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$  ist nach Konstruktion erfüllt, und da die Intervalllänge für jedes neue  $n$  mindestens halbiert wird, ist auch klar, daß  $(b_n - a_n)_{n \in \mathbb{N}}$  eine Nullfolge ist. Somit definiert  $([a_n, b_n])_{n \in \mathbb{N}}$  eine reelle Zahl  $x$ .

Nach unserer Annahme, es gäbe eine bijektive Abbildung  $f: \mathbb{N} \rightarrow \mathbb{R}$ , müßte es ein  $n \in \mathbb{N}$  geben mit  $x = f(n)$ . Da unsere Intervalle so konstruiert sind, daß  $f(n)$  nicht im Intervall  $[a_n, b_n]$  liegt, kann das aber unmöglich der Fall sein: Schließlich liegt die durch eine Intervallschachtelung definierte reelle Zahl in *jedem* der Intervalle. Daher kann es keine bijektive Abbildung  $f: \mathbb{N} \rightarrow \mathbb{R}$  geben, die reellen Zahlen sind also im Gegensatz zu den rationalen Zahlen *nicht* abzählbar.

Somit kennen wir nun zwei überabzählbare Mengen, die Menge aller Teilmengen von  $\mathbb{N}$  und die Menge der reellen Zahlen. Da stellt sich natürlich die Frage, ob diese beiden Mengen gleichmächtig sind. Es ist zwar etwas umständlich, aber nicht schwer, zu zeigen, daß dies in der Tat der Fall ist. Für diese Vorlesung ist das jedoch ohne größere Bedeutung, so daß hier auf einen Beweis verzichtet sei.

Als nächstes können wir mit CANTOR fragen, ob *jede* überabzählbare Teilmenge von  $\mathbb{R}$  schon die gleiche Mächtigkeit wie  $\mathbb{R}$  hat, oder ob es auch noch Zwischenstufen gibt.

Nach unseren Erfahrungen mit  $\mathbb{N}$  und den positiven geraden Zahlen sollte ziemlich klar sein, daß beispielsweise die Intervalle  $[0, 1]$  und  $[0, 2]$  gleichmächtig sind; die Abbildung

$$\begin{cases} [0, 1] \rightarrow [0, 2] \\ x \mapsto 2x \end{cases}$$

ist offensichtlich bijektiv. Abgeschlossene Intervalle verschiedener (positiver) Länge sind also gleichmächtig. Da die Abbildung

$$\begin{cases} \{x \in \mathbb{R} \mid x \geq 0\} \rightarrow (0, 1] \\ x \mapsto \frac{1}{1+x^2} \end{cases}$$

offensichtlich auch bijektiv ist, ändert sich auch durch eine unendliche Länge nichts an der Mächtigkeit, also spricht doch einiges dafür, daß alle überabzählbaren Teilmengen von  $\mathbb{R}$  gleichmächtig mit  $\mathbb{R}$  sind.

CANTOR stellte dies 1878 als eine Vermutung auf, die sogenannte *Kontinuumshypothese*. Er konnte sie aber nicht beweisen, und auch später gelang es keinem Mathematiker, sie zu beweisen oder zu widerlegen.

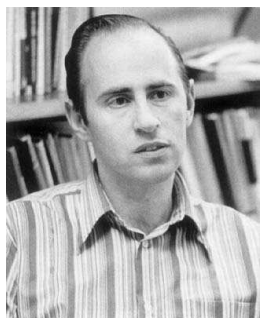
1938 konnte KURT GÖDEL schließlich zeigen, daß sich die Kontinuumshypothese nicht widerlegen läßt, wenn man nur die üblichen Axiome der Mengenlehre und der reellen Zahlen voraussetzt; es ist also insbesondere nicht möglich, ein explizites Gegenbeispiel zu konstruieren.

1963 bewies dann allerdings PAUL COHEN, daß sie sich aus diesen Axiomen auch nicht beweisen läßt: Die Kontinuumshypothese ist weder wahr noch falsch, sondern unentscheidbar! Genauer ausgedrückt: Wenn die Axiome der Mengenlehre widerspruchsfrei sind, dann bleiben sie auch widerspruchsfrei, wenn wir die Kontinuumshypothese als neues Axiom hinzufügen; sie bleiben aber auch dann widerspruchsfrei, wenn wir das Gegenteil der Kontinuumshypothese dazu nehmen.



KURT GÖDEL (1906–1978) wurde im damaligen Österreich-Ungarn in Brünn geboren; heute heißt der Ort Brno und liegt in der Tschechischen Republik. In der Schule interessierte er sich vor allem für Mathematik und Sprachen; 1923, zum Ende seiner Schulzeit beherrschte er bereits den gesamten Standardstoff der Universitätsmathematik. An der Universität Wien studierte er Mathematik und Theoretische Physik und spezialisierte sich schließlich auf Logik. 1931 bewies er seinen Unvollständigkeitssatz, wonach kein Axiomensystem die Theorie der natürlichen Zahlen vollständig beschreiben kann. 1940 emigrierte er nach USA und lebte

bis zu seinem Tod in Princeton. Dort lernte er 1942 ALBERT EINSTEIN (1879–1955) kennen und bewies unter anderem, daß es in dessen allgemeiner Relativitätstheorie geschlossene Weltlinien geben kann, entlang derer man theoretisch in die Vergangenheit gelangen kann.



PAUL JOSEPH COHEN (1934–2007) wurde in New Jersey geboren, studierte an der Universität von Chicago und promovierte dort über FOURIER-Reihen. Nach kurzen Aufenthalten in Rochester, am Massachusetts Institute of Technology und in Princeton kam er 1961 nach Stanford, wo er bis zu seiner Emeritierung im Jahre 2004 (und darüber hinaus) lehrte. 1966 erhielt er für seine Arbeiten zur Logik die Fields Medal, den damals bedeutendsten Preis der Mathematik. Außer über Logik und Mengenlehre arbeitete er auch auf dem Gebiet der Analysis, insbesondere der Theorie partieller Differentialgleichungen, sowie der Maßtheorie.

Wir haben somit die freie Auswahl, ob wir uns für eine Mathematik mit oder ohne Kontinuumshypothese entscheiden wollen. Logiker untersuchen beide Optionen, den meisten Mathematikern aber ist die Frage schlicht gleichgültig, da von den wirklich relevanten Sätzen der Analysis keiner davon abhängt, ob man die Kontinuumshypothese voraussetzt oder nicht.

## §7: Gleitkommazahlen

Im vorigen Paragraphen haben wir gesehen, daß die Menge der reellen Zahlen nicht einmal abzählbar ist; trotzdem rechnen wir ständig mit reellen Zahlen, und meistens benutzen wir dazu einen Taschenrechner oder einen Computer. Auch für einen Leser, der nicht die geringste Ahnung von der Funktionsweise eines Taschenrechners oder Computers hat, sollte klar sein, daß dort nur endlich viele Zahlen dargestellt werden können. Dasselbe gilt natürlich auch für unsere Gehirne.

Wir beschreiben also die Wirklichkeit schon seit über zwei Jahrtausenden mit Hilfe einer überabzählbaren Menge, können damit auch erstaunlich korrekte Vorhersagen machen, und trotzdem können wir weder mit Papier und Bleistift noch mit Taschenrechner oder Computer mehr als endlich viele Zahlen verarbeiten.

Für diese Erfolge ist in erster Linie die Numerische Mathematik verantwortlich; seit etwa fünfzig Jahren gibt es zumindest teilweise auch noch die Computeralgebra als Alternative. Deren Verfahrenn erfordern zwar meist erheblich größeren Rechenaufwand, dafür aber können sie bei manchen Problemen sogar exakte Ergebnisse liefern.

*Numerik* und *Computeralgebra* sind Vorlesungen für höhere Semester und damit kein Teil der *Analysis I*. Da aber auch Studenten der *Analysis I* routinemäßig Taschenrechner und Computer benutzen, möchte ich schon hier zumindest kurz die Prinzipien des numerischen Rechnens erläutern.

Das Rechnen mit reellen Zahlen per Computer folgt völlig anderen Regeln als denen, die wir vom Körper der reellen Zahlen gewohnt sind. Zunächst einmal müssen wir uns notgedrungen auf eine endliche Teilmenge von  $\mathbb{R}$  beschränken; in der Numerik sind dies traditionellerweise die sogenannten Gleitkommazahlen, die je nach Programmiersprache



als Datentyp **real** oder **float** bezeichnet werden; meist gibt es auch noch einen Typ **double** für das Rechnen mit doppelter Genauigkeit.

Eine Gleitkommazahl wird dargestellt in der Form  $x = \pm m \cdot b^{\pm e}$ , wobei die *Mantisse*  $m$  je nach Konvention zwischen 0 und 1 oder 1 und 10 liegt und der *Exponent*  $e$  eine ganze Zahl aus einem gewissen vorgegebenen Bereich ist. Die Basis  $b$  ist in heutigen Computern gleich zwei; in einigen älteren Mainframe Computern sowie in vielen Taschenrechnern wird auch  $b = 10$  verwendet.

Praktisch alle heute gebräuchliche CPUs für Computer richten sich beim Format für  $m$  und  $e$  nach dem IEEE-Standard 754 von 1985. Hier ist  $b = 2$ , und einfach genaue Zahlen werden in einem Wort aus 32 Bit gespeichert. Das erste dieser Bits steht für das Vorzeichen, null für positive, eins für negative Zahlen. Danach folgen acht Bit für den Exponenten  $e$  und 23 Bit für die Mantisse  $m$ .

Die acht Exponentenbit können interpretiert werden als eine ganze Zahl  $n$  zwischen 0 und 255; wenn  $n$  keinen der beiden Extremwerte 0 und 255 annimmt, wird das gesamte Bitmuster interpretiert als die Gleitkommazahl (Mantisse im Zweiersystem, d.h.  $m_i \in \{0, 1\}$ )

$$\pm 1, m_1 \dots m_{23} \times 2^{n-127}.$$

Die Zahlen, die in obiger Form dargestellt werden können, liegen somit zwischen  $2^{-126} \approx 1,175 \cdot 10^{-37}$  und  $(2 - 2^{-23}) \cdot 2^{127} \approx 3,403 \cdot 10^{38}$ . Das führende Bit der Mantisse ist stets gleich eins (sogenannte normalisierte Darstellung) und wird deshalb gleich gar nicht erst abgespeichert. Der Grund liegt natürlich darin, daß man ein führendes Bit null durch Erniedrigung des Exponenten zum Verschwinden bringen kann – es sei denn, man hat bereits den niedrigstmöglichen Exponenten  $n = 0$ , entsprechend  $e = -127$ .

Für  $n = 0$  gilt daher eine andere Konvention: Jetzt wird die Zahl interpretiert als

$$\pm 0, m_1 \dots m_{23} \times 2^{-126};$$

man hat somit einen (unter Numerikern nicht unumstrittenen) *Unterlaufbereich* aus sogenannten *subnormalen* Zahlen, in dem mit immer weniger geltenden Ziffern Zahlen auch noch positive Werte bis hinunter

zu  $2^{-23} \times 2^{-126} = 2^{-149} \approx 1,401 \cdot 10^{-44}$  dargestellt werden können, dazu natürlich die Null, bei der sämtliche 32 Bit gleich null sind.

Auch der andere Extremwert  $n = 255$  hat eine Sonderbedeutung: Falls alle 23 Mantissenbit gleich null sind, steht dies je nach Vorzeichenbit für  $\pm\infty$ , andernfalls für NAN (*not a number*), d.h. das Ergebnis einer illegalen Rechenoperation wie  $\sqrt{-1}$  oder  $0/0$ . Das Ergebnis von  $1/0$  dagegen ist nicht NAN, sondern  $+\infty$ , und  $-1/0 = -\infty$ .

Doppeltgenaue Gleitkommazahlen werden entsprechend dargestellt; hier stehen insgesamt 64 Bit zur Verfügung, eines für das Vorzeichen, elf für den Exponenten und 52 für die Mantisse. Durch die elf Exponentenbit können ganze Zahlen  $n$  zwischen null und 2047 dargestellt werden; abgesehen von den beiden Extremfällen entspricht dies dem Exponenten  $e = n - 1023$ , der somit Werte zwischen  $-1022$  und  $1023$  annehmen kann.

Der Exponent  $e$  sorgt dafür, daß Zahlen aus einem relativ großen Bereich dargestellt werden können, er hat aber auch zur Folge, daß die Dichte der darstellbaren Zahlen in den verschiedenen Größenordnung stark variiert: Am dichtesten liegen die Zahlen in der Umgebung der Null, und mit steigendem Betrag werden die Abstände benachbarter Zahlen immer größer.

Um dies anschaulich zu sehen, betrachten wir ein IEEE-ähnliches Gleitkommasystem mit nur sieben Bit, einem für das Vorzeichen und je drei für Exponent und Mantisse. Das folgende Bild zeigt die Verteilung der darin darstellbaren Zahlen:



Um ein Gefühl dafür zu bekommen, was dies für das praktische Rechnen mit Gleitkommazahlen bedeutet, betrachten wir ein analoges System mit der uns besser vertrauten Dezimaldarstellung von Zahlen (für die es einen eigenen IEEE-Standard 854 von 1987 gibt), und zwar nehmen wir an, daß wir eine dreistellige dezimale Mantisse haben und Exponenten zwischen  $-3$  und  $3$ . Da es bei einer von zwei verschiedenen Basis keine Möglichkeit gibt, bei einer normalisierten Mantisse die erste Ziffer einzusparen, schreiben wir die Zahlen in der Form  $\pm 0, m_1 m_2 m_3 \cdot 10^e$ .

Zunächst einmal ist klar, daß die Summe zweier Gleitkommazahlen aus diesem System nicht wieder als Gleitkommazahl in diesem System darstellbar sein muß: Ein einfaches Gegenbeispiel wäre die Addition der größten darstellbaren Zahl  $0,999 \cdot 10^3 = 999$  zu  $5 = 0,5 \cdot 10^1$ : Natürlich ist das Ergebnis 1004 nicht mehr im System darstellbar. Der IEEE-Standard sieht vor, daß in so einem Fall eine *overflow*-Bedingung gesetzt wird und das Ergebnis gleich  $+\infty$  wird. Wenn man (wie es die meisten Compiler standardmäßig tun) die *overflow*-Bedingung ignoriert und mit dem Ergebnis  $+\infty$  weiterrechnet, kann dies zu akzeptablen Ergebnissen führen: Beispielsweise wäre die Rundung von  $1/(999 + 5)$  auf die Null für viele Anwendungen kein gar zu großer Fehler, auch wenn es dafür in unserem System die sehr viel genauere Darstellung  $0,996 \cdot 10^{-3}$  gibt. Spätestens wenn man das Ergebnis mit 999 multipliziert, um den Wert von  $999/(999 + 5)$  zu berechnen, sind die Konsequenzen aber katastrophal: Nun bekommen wir eine Null anstelle von  $0,996 \cdot 10^0$ . Auch wenn wir anschließend 500 subtrahieren, sind die Konsequenzen katastrophal:  $\infty - 500 = \infty$ , aber  $(999 + 5) - 500 = 504$  ist eine Zahl, die sich in unserem System sogar exakt darstellen ließe!

Selbst ohne Bereichsüberschreitung kann es Probleme geben: Beispielsweise ist

$$123 + 0,0456 = 0,123 \cdot 10^3 + 0,456 \cdot 10^{-1} = 123,0456$$

mit einer nur dreistelligen Mantisse nicht exakt darstellbar. Hier sieht der Standard vor, daß das Ergebnis zu einer darstellbaren Zahl gerundet wird, wobei mehrere Rundungsvorschriften zur Auswahl stehen. Voreingestellt ist üblicherweise eine Rundung zur nächsten Maschinenzahl; wer etwas anderes möchte, muß dies durch spezielle Bits in einem Prozessorstatusregister spezifizieren. Im Beispiel würde man also  $123 + 0,0456 = 123$  oder (bei Rundung nach oben) 124 setzen und dabei zwangsläufig einen Rundungsfehler machen.

Wegen solcher unvermeidlicher Rundungsfehler gilt das Assoziativgesetz selbst dann nicht, wenn keine Bereichsüberschreitung auftritt: Bei Rundung zur nächsten Maschinenzahl ist beispielsweise

$$(0,456 \cdot 10^0 + 0,3 \cdot 10^{-3}) + 0,4 \cdot 10^{-3} = 0,456 \cdot 10^0 + 0,4 \cdot 10^{-3} = 0,456 \cdot 10^0,$$

aber

$$0,456 \cdot 10^0 + (0,3 \cdot 10^{-3} + 0,4 \cdot 10^{-3}) = 0,456 \cdot 10^0 + 0,7 \cdot 10^{-3} = 0,457 \cdot 10^0.$$

Ein mathematischer Algorithmus, dessen Korrektheit unter Voraussetzung der Körperaxiome bewiesen wurde, muß daher bei Gleitkomma-rechnung kein korrektes oder auch nur annähernd korrektes Ergebnis mehr liefern – ein Problem, das keinesfalls nur theoretische Bedeutung hat und schon oft zu als korrekt anerkannten unsinnigen Lösungen führte. Ein häufig zitierter Spruch des lange in Karlsruhe und in Freiburg lehrenden Numerikers KARL NICKEL (1924–2009), dessen Originalquelle ich leider nirgends finden konnte, sagt:

Der (naive) Anfänger **glaubt** an jede einzelne Ziffer

Der (erfahrene) Programmierer **vertraut** auf die Hälfte der Stellen

Der (wissende) Pessimist **mißtraut** sogar dem Vorzeichen.

Natürlich kennt die numerischen Mathematik auch Methoden, bei denen man *sicher* sein kann, daß sie ein im Rahmen einer vorgegebenen Schranke korrekte Ergebnisse liefern; als Beispiel haben dafür haben wir bereits die Intervallarithmetik kennengelernt. Bei naiven numerischen Ansätzen ist aber oft genug der Mißerfolg vorprogrammiert.

Eine elementare Einführung in den Standard IEEE 754 und seine Anwendung beim numerischen Rechnen bietet

MICHAEL L. OVERTON: Numerical Computing with IEEE Floating Point Arithmetic, *SIAM*, 2001

Für weitere Einzelheiten sei auf die Vorlesung Numerik verwiesen.

## §8: Wie real sind die reellen Zahlen?

Wir haben die reellen Zahlen abstrakt über Intervallschachtelungen definiert und sind dabei auf eine überabzählbare Menge gekommen. Diese benutzen wir zur Lösung praktischer Probleme, wobei wir im allgemeinen mit Computern oder Taschenrechnern arbeiten, deren Zahlbereich nur eine endliche Teilmenge von  $\mathbb{R}$  ist und die selbst damit nicht exakt rechnen können.

Die Mathematik hat den Anspruch, sichere Ergebnisse zu liefern; die kann uns das rundungsfehlerbehaftete Rechnen mit Gleitkommazahlen höchstens dann garantieren, wenn wir uns auf Intervalle als Ergebnisse

beschränken. In der Praxis werden diese dann allerdings oft so lang, daß sie uns kaum noch nützliche Information liefern.

Manche Probleme sind sogar grundsätzlich unlösbar: Wie DANIEL RICHARDSON 1969 zeigte, können wir nicht einmal immer entscheiden, ob eine durch einen relativ einfachen Ausdruck gegebene reelle Zahl verschwindet oder nicht. Mit dem, was wir heute wissen, können wir seinen Satz so formulieren:

**Satz von Richardson:** Es gibt kein Verfahren, das in endlich vielen Schritten entscheidet, ob ein beliebig vorgegebener Ausdruck bestehend aus rationalen Zahlen,  $\pi$ , einer Variablen  $x$  sowie den Funktionen  $+$ ,  $\cdot$ , Sinus und Betrag gleich Null ist.

DANIEL RICHARDSON wurde 1941 in Chicago geboren. Er studierte Mathematik in New York und in San Francisco, wo er 1962 seinen Bachelor bekam. Danach ging er an die Universität von Bristol und promovierte dort 1965 in mathematischer Logik. Nach verschiedenen ein- bis zweijährigen Tätigkeiten an Universitäten und in der Industrie wurde er 1988 zunächst Lecturer, später Senior Lecturer, an der Universität von Bath. <http://people.bath.ac.uk/masdr/>

Ein vollständiger Beweis des Satzes von RICHARDSON wäre fast eine eigene einsemestrige Vorlesung und würde uns weit vom eigentlichen Stoff der Analysis I wegführen, da er wesentlich auf Argumenten aus der Logik sowie der Zahlentheorie beruht. Interessenten finden alle Einzelheiten sowie verschiedene Anwendungen im Buch

YURI V. MATIYASEVICH: Hilbert's Tenth Problem, *MIT Press*, 1993

## §9: Zusammenfassung

In diesem Kapitel ging es darum, den Zahlbereich ausgehend von den natürlichen Zahlen so zu erweitern, daß immer mehr Gleichungen lösbar werden:

- Um Gleichungen der Form  $a + x = b$  stets lösen zu können, erweiterten wir die natürlichen Zahlen zu den ganzen Zahlen.
- Um auch die Gleichung  $ax = b$  mit  $a \neq 0$  lösen zu können, erweiterten wir die ganzen Zahlen zu den rationalen Zahlen. Dort können wir alle Grundrechenarten (außer der Division durch Null) unbeschränkt ausführen. Zahlbereiche mit dieser Eigenschaft definierten wir anhand einiger offensichtlicher, aus der Schule bekannten

Rechenregeln als *Körper*. Zusätzlich haben wir für rationale Zahlen eine Ordnungsrelation, die mit den Rechenoperationen kompatibel ist; das führte uns auf das Konzept des angeordneten Körpers. Beim Nachweis von Eigenschaften, die für alle Körper gelten, lernten wir Beweisprinzipien kennen wie den Beweis durch Widerspruch oder durch vollständige Induktion.

- In den rationalen Zahlen sind Gleichungen wie  $x^2 = 2$  nicht lösbar. Mit dem bereits den Babyloniern vor vier Tausend Jahren bekannten Algorithmus von HERON konnten wir die Gleichung aber in beliebig guter Näherung lösen; dies führte uns auf den Begriff der Intervallschachtelung über den wir reelle Zahlen definierten, die ebenfalls einen angeordneten Körper bilden.
- Selbst anschaulich klare Konzepte wie „beliebig klein“ müssen auf dem Umweg über algebraische Ungleichungen definiert werden, bevor man präzise damit argumentieren kann. Anders läßt sich die Gefahr von Fehlschlüssen nicht vermeiden.
- Die Gleichung  $x^2 = a$  für  $a < 0$  kann in einem angeordneten Körper nicht gelöst werden, da dort das Quadrat eines jeden Elements größer oder gleich Null ist. Indem wir die imaginäre Einheit  $i = \sqrt{-1}$  einführten, konnten wir immerhin noch einen Körper definieren, den der komplexen Zahlen, in dem alle Elemente Quadratwurzeln haben und über dem sogar jedes nichtkonstante Polynom Nullstellen hat und in Linearfaktoren zerfällt.
- Die reellen wie auch die komplexen Zahlen bilden überabzählbare Mengen, in denen wir viele Gleichungen nur näherungsweise lösen können. Sie sind eine Idealisierung der Wirklichkeit, die sich in der Vergangenheit als extrem nützlich erwiesen hat. Trotzdem gibt es Probleme, die nicht entscheidbar sind.
- Numerisches Rechnen ist zwar die übliche Weise, wie man in der Praxis mit reellen Zahlen umgeht; bei naivem Umgang mit Rechenregeln können dabei aber leicht unsinnige Ergebnisse produziert werden. Die numerische Mathematik hat jedoch für die meisten praktischen Probleme Algorithmen entwickelt, die zu sinnvollen Ergebnissen führen.