# Elementare Grundlagen der Analysis

**Wolfgang Rautenberg**

**Berlin**

Zweite verbesserte Auflage

**Satz und Layout:** Der Autor
**Neufassung vom Juni 2005**

# Vorwort

Die erste Auflage dieses Buches seit geraumer Zeit vergriffen. Daher wurde eine völlig überarbeitete Ausgabe des Buches vorbereitet. Viele Details, vor allem in den ersten acht Abschnitten und im Anhang, wurden geändert.

*Elementare Grundlagen der Analysis* meinen hier mit anderen Akzenten etwa das, was ursprünglich als *Grundlagen der Analysis*, und heute oft als *Aufbau des Zahlensystems* bezeichnet wird. Es geht uns aber nicht nur um die Definition oder Konstruktion der reellen Zahlen nach akademischem Muster, sondern wir wollen diese sogleich in ihrer lebendigen Wirksamkeit vor Augen führen. Der Grenzwertbegriff offenbart kraftvoll seine Wirksamkeit schon unterhalb der Ebene des Differential- und Integralkalküls. Auch dies soll die vorliegende Darstellung sichtbar machen. Besonderer Wert wird auf Eindeutigkeitsfragen gelegt, d.h. bei wichtigen Definitionen und Konstruktionen wird stets die Frage diskutiert, warum diese so und nicht anders zum Erfolg führen.

Wir empfehlen dieses Buch allen Studierenden der Fächer Mathematik und Informatik, als Ergänzungstext zu einer Vorlesung über Analysis oder Mathematik für Informatiker, vor allem aber den Lehramtskandidaten. Es kommen hier einige Dinge zur Sprache, für die in den Universitätskursen über Analysis kaum Zeit zur Verfügung steht – sei es die Analyse der Folge $(1+\frac{1}{n})^n$, seien es Stammbruchentwicklungen, oder CANTORsche Reihen mit ihrer schönen Anwendung eines Irrationalitätsnachweises für die Zahl $\mathbf{e}$, oder seien es grundlegende, auch in den Computerchips implementierte Algorithmen oder einfache Iterationsverfahren der numerischen Analysis. Grundsätzlich wird alles bewiesen, was eines Beweises bedarf. Daher erfordert die Lektüre kein spezielles Vorwissen, aber gewisse Fertigkeiten im logischen Schließen und im Umgang mit der vollständigen Induktion; zumindest aber die Bereitschaft, sich diese bei der Durcharbeitung des Materials anzueignen. $\blacksquare$ markiert das Ende eines Beweises. Um das Selbststudium zu erleichtern, wurden die Lösungen der Übungen am Ende des Buches angegeben. Man sollte aber erst versuchen, diese vor dem Nachblättern selbständig zu lösen.

Auch ein axiomatischer Aufbau der Analysis, ausgehend von den Axiomen eines lückenlosen Körpers, erfordert irgendwann dessen explizite Konstruktion. Dafür gibt es Standardverfahren, etwa die Methode der DEDEKINDschen Schnitte. Eine vom Autor entwickelte und inzwischen ausgereifte Methode wird in den Abschnitten **3 - 5** dieses Buches ausgeführt. Man geht aus von Dezimalzahlen als Schreibfiguren und benutzt zur Definition der Rechenoperationen die Kommaverschiebung (shifting). Darauf beruhen auch wesentlich die internen Prozesse in den elektronischen Rechnern. Der Gewinn ist, dass man von den natürlichen Zahlen ausgehend ohne Umschweife die nichtnegativen reellen Zahlen sichtbar vor Augen hat und mit diesen sogleich rechnen kann.

Diese Methode ist einfach, anschaulich und ökonomisch, und in ihrem Kern auch für eine frühzeitige schulische Behandlung geeignet. Sie ist so präzise wie die klassischen Konstruktionen. Aber sie ist einfacher und ersetzt den akademischen Ballast, der den klassischen Konstruktionen anhaftet, weitestgehend durch elementares Rechnen mit Gleichungen und Ungleichungen. Auch lässt sich zum Beispiel der Divisionsalgorithmus zur Berechnung des reell-wertigen Quotienten ganz elementar, ohne Bezug auf unendliche Reihen, in voller Allgemeinheit rechtfertigen. Nach Abschluss der Konstruktion spielen dezimale und sonstige Darstellungen – wie in der Höheren Analysis durchweg – nur noch eine Nebenrolle, nützlich etwa für numerische Beispiele. Diese stehen in unserem Aufbau *von Anbeginn an* zur Verfügung; man muss diesbezüglich nicht vorgreifen und auf unendliche Reihen vertrösten.

Der größere Teil des Textes ist den elementar-analytischen Anwendungen reeller Zahlen gewidmet und hat einen bescheidenen Hauch von Numerik. Der Zwischenwertsatz wird für LIPSCHITZ-stetige Funktionen bewiesen, weil die elementaren Funktionen nicht nur stetig, sondern auch LIPSCHITZ-stetig sind. Dies vereinfacht sowohl die Beweise und ist auch bedeutsam für die Numerik. Empfehlenswert für eine einprägsame Behandlung numerischer Algorithmen und Beispiele wäre die Ausrüstung mit einem programmierbaren Taschenrechner – alle iterativen Verfahren und numerischen Beispiele in diesem Buch wurden auf einem hochwertigen Taschenrechner programmiert und berechnet.

Dem Ganzen ist in Abschnitt **1** eine kurz gefasste Kulturgeschichte der Zahlen vorangestellt, in deren Mittelpunkt diejenige reeller Zahlen steht. Die Lektüre der einzelnen Abschnitte ist nicht an die gegebene Reihenfolge gebunden. So können **9** und **10** vor **7** gelesen werden, und der Anhang **11** ist gänzlich eigenständig. Indexeinträge haben in der Regel nur eine Seitenzahl, bei Namen meistens den Ort ihres ersten Vorkommens. Das Literaturverzeichnis kann angesichts der weit gefächerten Lehrbuchliteratur nur eine Auswahl darstellen. Negative Zahlen werden erst eingeführt, wenn sie wirklich gebraucht werden. Das erweist sich aus vielen Gründen als vorteilhaft. Aber man muss dieser Vorgehensweise, etwa in einem Kurs, nicht strikt folgen.

Eine Eigenheit der durch CAUCHY und WEIERSTRASS geformten und durch DEDEKIND in gewissem Sinne vollendeten klassischen oder „Reinen“ Analysis ist die bewusste Elimination aller geometrischen oder physikalischen Begriffe aus ihrem Begriffsgefüge. Bis gegen Ende des 19. Jahrhundert berief man sich in Grundfragen auf Konzepte, die der griechischen und der NEWTONschen Quelle der Analysis entstammen. Seitdem wird der Bezug der Analysis zur Geometrie, Physik oder sonstigen Disziplinen in einer zweiten Ebene, der Anwendungsebene, hergestellt.

Dieser Anwendungsbezug und seine Systematik sollten nicht vernachlässigt werden. Deswegen haben wir in **10** einen kleinen aber wichtigen Teil hiervon, nämlich die Verbindung von Zahl und Größe, dem ursprünglichen Grundbegriff der Analysis, systematisch entwickelt. Diese Verbindung erklärt z.B. die historische und sachliche Herkunft der Definitionen arithmetischer Operationen in **4** und **5** besser als die Bruchrechnung. Dass man daraus viel Nutzen ziehen kann, unterstreicht der Isomorphiesatz für lückenlose Größenbereiche, der in seiner Anwendung auf reelle Zahlenbereiche die denkbar natürlichste Erklärung der Exponential- und Logarithmus-Funktion liefert. Ein Nebenprodukt ist die axiomatische Kennzeichnung des Bereichs der reellen Zahlen.

Wir wollen nun kurz erklären, warum die natürlichen Zahlen in einem Anhang gesondert behandelt werden. In einem axiomatischen Aufbau der Analysis werden nebst logischen und kombinatorischen (oder zahlentheoretischen) meist auch mengentheoretische Hilfsmittel verwendet. Diese werden in der Regel unbefangen eingesetzt. Dennoch besteht kein Grund unvermittelt so zu tun, als wüsste man was Mengen, nicht hingegen was natürliche Zahlen sind. Grundlagentheoetisch gesehen sind die letzteren jedenfalls die einfacheren Objekte, obwohl dies nicht unmittelbar auf der Hand liegt.

Bei einer „Konstruktion" der natürlichen Zahlen geht es nach klassischem Verständnis in erster Linie darum, diese mengentheoretisch zu definieren und ihre Eigenschaften, insbesondere die Beweismethode durch Induktion, im Rahmen einer Mengentheorie zu begründen. Es ist bemerkenswert und von großer Bedeutung, dass dies rein prädikatenlogisch möglich ist – dies ist das Fazit des gescheiterten Versuches, die Mathematik rein logisch zu begründen – doch muss dies den Analytiker nicht nicht mehr interessieren als etwa den Zahlentheoretiker. Beide haben sowohl auf der Objekt- als auch auf der Meta-Ebene mit natürlichen Zahlen zu tun und benutzen diese in naiver Weise. Das Thema der Beschaffenheit natürlicher Zahlen ist zwar wesentlich für die Grundlagen der Mathematik als einer zugleich auch philosophisch und erkenntnistheoretisch orientierten Disziplin, betrifft aber die Grundlagen der Analysis nur am Rande. Weder die Herkunft noch das Wesen der natürlichen Zahlen sind für die Grundlagen der Analysis von besonderem Interesse. Wichtig ist nur das Wissen, wie man mit diesen umzugehen und Beweise durch vollständige Induktion zu führen hat.

In jedem Falle hat eine wie immer geartete Konstruktion natürlicher Zahlen einen anderen Charakter als diejenige reeller oder komplexer Zahlen. Die erstere setzt in ihrer Durchführung als auch in einer Diskussion über diese Thematik als Ganzes eine gewisse mathematische Reife voraus, gerade weil es hier zum Teil um subtile Beweise scheinbar offensichtlicher Sachverhalte geht. Wir kommen in **11** darauf zurück. Dort wird zugleich die Theorie des Zählens endlicher Mengen entwickelt, von der ja auch in der Analysis und auch sonstwo ständig und meist unbewusst Gebrauch gemacht wird. Mit den reellen Zahlen war man bis gegen Ende des 19. Jahrhunderts umgegangen „wie im Traum" ([16, Teil II, S. 694]). Mit den Abzählungen endlicher Mengen – wir nennen etwa den Zählsatz in **11.4** – geht man heute noch oft wie im Traume um.

Die Professoren G. Pickert (Gießen), R. Gorenflo und H. Lenz (Berlin) haben mich durch zahlreiche Hinweise bei der Durchsicht des Manuskripts ganz wesentlich unterstützt. Ihnen gilt mein besonderer Dank. Ebenso danke ich den Professoren H. Gericke (Freiburg) und H. Lüneburg (Kaiserslautern). Für technische Unterstützung danke ich Herrn U. Fuchs und den Mitarbeitern des Rechenzentrums am Fachbereich Mathematik/Informatik der Freien Universität.

Berlin, den 9. Juni 2005
Wolfgang Rautenberg

# Inhaltsverzeichnis

# Notation

Wir benutzen die bekannte mengentheoretische Symbolik, aber die Mengentheorie in ihren Inhalten nur in sehr bescheidenem Maße. Wie üblich bezeichnen $M \cup N$, $M \cap N$ und $M\mathbb{N} = x \in M \mid x \notin N$ die *Vereinigung*, den *Durchschnitt*, bzw. die *Differenz* der Mengen $M, N$, und $\subseteq$ die *Inklusion* oder Teilmengenbeziehung. Falls $M \subseteq N$ und $M \neq N$, schreiben wir auch $M \subset N$ (echte Inklusion), aber nur dann, wenn der Umstand $M \neq N$ hervorgehoben werden soll. Die *leere Menge* wird mit $\emptyset$ bezeichnet.

Es sei $M \times N$ die Menge aller geordneten Paare $(a, b)$ mit $a \in M$, $b \in N$, auch das *direkte Produkt* von $M$ und $N$ genannt. Eine Teilmenge von $M \times M$ heißt eine (binäre) Relation auf $M$. Beispiele sind die Ordnungsrelationen in Zahlenbereichen, die im Buch ausführlich behandelt werden.

Üblicher Konvention entsprechend steht $f : M \to N$ abkürzend für „*$f$ ist Funktion* (*Abbildung*) *von $M$ nach $N$*¸“. Ein solches $f$ kann bekanntlich als spezielle Teilmenge von $M \times N$ verstanden werden. Gelegentlich wird auch von der Funktion $f : M \to N$ gesprochen, womit natürlich $f$ gemeint ist. Auch steht z.B. $f : x \mapsto ax$ oft dort, wo eigentlich stehen müsste „Sei $f$ die Funktion $x \mapsto ax$“ (mit $f(x) = ax$). Der Wert $f(x)$ von $f$ an der Stelle $x$ wird auch mit $fx$ bezeichnet, wenn die Deutlichkeit darunter nicht leidet. $f$ ist *injektiv* oder eine *Injektion*, wenn $fx = fy \;\Rightarrow\; x = y$ für alle $x, y \in M$, und *surjektiv*, wenn das Bild $fx \mid x \in M$ von $f$ ganz $N$ ausfüllt. $f$ ist *bijektiv* oder eine *Bijektion*, wenn $f$ injektiv und surjektiv ist. Im bijektiven Falle hat $f$ eine *Umkehrfunktion* $f^{-1} : N \to M$, definiert durch $f^{-1}y = x \;\Leftrightarrow\; fx = y$. Triviales Beispiel einer Bijektion ist die *identische* Abbildung $id_M : x \mapsto x$ von $M$ auf $M$.

Unter einer $n$-stelligen *Operation* auf einer Menge $A$ versteht man eine Abbildung von $A^n$ in $A$. Dabei sei $A^n$ das $n$-fache direkte Produkt von $A$ mit sich selbst. Ist diese nur auf einer Teilmenge von $A^n$ erklärt, spricht man von einer *partiellen* Operation. Prominentes Beispiel ist die Division, die auf einem Zahlenbereich nicht für alle Zahlenpaare sinnvoll erklärt werden kann.

*Terme* sind Zeichenfolgen, die aus Variablen, Konstanten und Operationszeichen in sinnvoller Weise aufgebaut sind. Sinnvoll lässt sich induktiv präzisieren, aber es genügt der Hinweis, dass ein Term $t$ ein wohlbestimmtes Element eines Zahlenbereichs bezeichnet, wenn die vorkommenden Variablen mit Elementen aus eben diesem Bereich belegt worden sind. Für beliebige Terme $s, t$ meint $s \leq t$ stets „$s < t$ oder $s = t$“. Dabei bezeichnet $<$ meist eine Ordnungsrelation auf einer Menge, etwa einem Zahlenbereich. $t \geq s$ sei nur eine andere Schreibweise für $s \leq t$, genauso wie $t > s$ durchweg nur eine andere Schreibweise ist für $s < t$. Es bedeute $s := t$ (außerhalb von Flussdiagrammen wie z.B. in Figur 4), dass der Term $s$ durch den Term $t$ definiert wird.

Wie üblich steht $A \Leftrightarrow B$ häufig für $A$ *genau dann wenn* $B$, und $A \Rightarrow B$ für *wenn* $A$ *so* $B$, insbesondere dann wenn $A, B$ Gleichungen oder Ungleichungen sind. Doch können $A$ und $B$ auch kompliziertere sprachliche Ausdrücke bedeuten.

# Abschnitt 1

ἀεὶ ὁ ἄνθρωπος ἀριθμητίζει[1]
(R. Dedekind)

# Zur Geschichte des Zahlbegriffs

Die Geschichte des Zahlbegriffs ist ein faszinierendes Kapitel der Kulturgeschichte. Der Begriff hat sich im Verlaufe der Zeit mehrfach erweitert. In der griechischen Mathematik hatte das Wort *Zahl* (ἀριθμός) seit Pythagoras (ca. 580 – 500 v.Chr.) eine feste Bedeutung, nämlich die einer positiven natürlichen Zahl als einer Menge von Einheiten, wie in den *Elementen* des Euklid (ca. 300 v.Chr.) definiert; dabei lassen wir die Sonderrolle, welche die 1 als Repräsentant der Einheit hatte, außer acht. Die Bezeichnungen rationale, irrationale, negative, imaginäre und reelle Zahlen kamen erst im Mittelalter auf. Der Erweiterungsprozess verlief über mehrere Etappen, siehe auch [10, 12]. Seit der durch C. F. Gauss (1777 – 1855) vollzogenen Ausräumung der Bedenken gegenüber den von ihm so benannten komplexen Zahlen hat sich der Zahlbegriff in der algebraisch-analytischen Richtung kaum noch verändert. Natürlich hat man sich über den Zahlbegriff häufig geäußert; aber die Frage „Was sind und was sollen die Zahlen?“ vermochte erst gegen Ende des 19. Jahrhunderts, im Zusammenhang mit der Loslösung der Analysis von ihren geometrisch-physikalischen Wurzeln, die Gemüter durchgreifend zu bewegen. Vorher hatte man sich auf den Begriff der messbaren (stetigen) Größe in der geometrisch-physikalischen Welt berufen; reelle Zahl wurde als das definiert, womit man misst, wie bei Stevin, Newton und anderen nachzulesen ist.

G. Cantor verallgemeinerte den ursprünglichen Zahlbegriff in einer anderen Richtung, nämlich zu den transfiniten Kardinal- und Ordinalzahlen, siehe **11.5**. In gewissem Sinne stehen diese der altgriechischen Auffassung von Zahlen als einer Kollektion von Einheiten sogar näher. Jedenfalls sind die reellen Zahlen deswegen nicht reeller, weil sie so heißen. Sie sind auch nicht reeller als die komplexen Zahlen. Denn deren Distanz zu den reellen Zahlen ist aus heutiger Sicht um ein Vielfaches kleiner als die der reellen zu den natürlichen. Komplexe Zahlen lassen sich einfach als Paare reeller Zahlen verstehen, siehe **9.4**. Nicht so einfach ist gewiss der Übergang von der reellen zur komplexen Analysis. Hier gibt es aber höchstens sachliche, aber keine logischen Schwierigkeiten. Diese wohnen dem Begriff der Irrationalzahl bei genauerer Betrachtung grundsätzlich inne, auch wenn die Definition in Abschnitt **3** sehr einfach aussieht.

[1] Der Mensch treibt ewig Arithmetik – Anspielung auf Platons Wort „Gott treibt ewig Geometrie“.

## 1.1 Die ältere Geschichte

Die Vorstellungen über Zahlen in den großen Kulturkreisen des Altertums, bei den Ägyptern, den Sumerern, Babyloniern und den Griechen, den Indern und Chinesen sowie bei den Mayas und Azteken in Amerika sind uns durch Dokumente aus Stein, Knochen, Lehm und Papyrus hinlänglich klar überliefert worden. Was sich außerhalb dieser Kulturkreise oder vor deren Entstehung abgespielt hat, darüber können wir nur Mutmaßungen anstellen. Eine dieser ziemlich verbreiteten und durch Observationen noch lebender so genannter primitiver Kulturen gestützten Mutmaßungen ist, dass der in unserem Sinne noch nicht kultivierte Mensch nur über Worte für „Eins“und „Zwei“ verfügte, alles andere war „mehr“ oder „viel mehr“, ähnlich wie ja auch einige Säugetiere und Vögel über einen anscheinend angeborenen Sinn für die Erfassung von einem oder zwei (bis zu fünf) Gegenständen verfügen. Ein Indiz dafür ist, dass für die Zahlen 1 und 2 in fast allen Sprachen besondere Namen existieren, unterschiedlich zugleich nach ihrer Funktion als Anzahl oder als Ordnungszahl, und dass sich erst bei größeren Zahlen eine zunehmende Regelmäßigkeit sprachlicher Konstruktion ausprägt.

Auch die Geschichte der Zahlbezeichnungen und -benennungen in den verschiedenen Kulturkreisen ist aufschlussreich. Das Bezeichnungsprinzip ist in den antiken Kulturen fast überall dasselbe. Man kann es vielleicht am treffendsten mit dem Wort *Kollektionsprinzip* umschreiben. Kleine Zahlen (bis höchstens 10) wurden meistens durch Striche oder Punkte bezeichnet. Sodann wurde für eine bestimmte Anzahl, z.B. für 10, ein neues Symbol eingeführt. Mehrere dieser Anzahlen wurden wieder zusammengefasst und abermals mit einem neuen Symbol bezeichnet, usw. Typisch in dieser Hinsicht ist das über Jahrtausende verwendete Bezeichnungssystem im alten Ägypten, ebenso das der Griechen und schließlich auch der Römer. In seiner spätmittelalterlichen Form (siehe dazu [23]) wird es auch heute noch verwendet.

Vorteil der auf dem Kollektionsprinzip beruhenden Systeme ist eine unmittelbare Veranschaulichung der jeweils dargestellten Anzahl. Dem aber steht der entscheidende Nachteil gegenüber, dass das Rechnen in diesen Systemen recht kompliziert ist. Die Addition kleinerer Anzahlen ist einfach. Man muss nur die „Zahlhaufen“ zusammenfügen, unter eventueller Verwendung eines neuen Symbols für eine neu entstandene Gruppierungseinheit. Das Multiplizieren hingegen wird schon zu einem Problem, und das Dividieren schließlich ist nur noch eine den Eingeweihten vorbehaltene Kunst. Mit Sicherheit standen den Rechenexperten der damaligen Zeit gewisse technische Hilfsmittel zur Verfügung, z.B. Tafeln und einfache „digitale“ Rechengeräte, ähnlich dem heute an manchen Orten noch gebräuchlichen Abakus, und wahrscheinlich neigen wir nur wegen mangelnder Vertrautheit mit den Rechentechniken der damaligen Zeit dazu, die Schwierigkeiten zu überschätzen. Der präzise Kalender, aufgebaut auf astronomischen Beobachtungen und Berechnungen, die Steuereintreibung und die Landvermessung bei einem komplizierten Bewässerungssystem machen deutlich, dass umfangreiche Rechnungen ständig ausgeführt wurden. In der Regel geschah dies in besonderen Institutionen an den Höfen und Tempeln der Könige, Fürsten und Priester.

Man liest gelegentlich, dass einigen Zahlbezeichnungssystemen in der Vergangenheit eine andere Grundzahl zugrundelag als 10. So wurde bei den Babyloniern und Sumerern bei Bruchzahlen das *sexagesimale* System verwendet, in welchem die Zahl 60 die heutige Rolle der Zahl 10 spielte. In anderen Kulturen war die Zahl 20 eine Art Grundzahl. Die Bezeichnungssysteme des Altertums hatten jedoch noch nicht die wesentlichen Merkmale eines $g$-adischen Positionssystem im heutigen Sinne. Statt von Grunzahlen sollte man besser z.B. von *Kollektionseinheiten* sprechen. Meist wurden und werden in derselben Sprache unterschiedliche Kollektionseinheiten benutzt, je nach Art der Gegenstände, auf die sich die Zählung bezog. Wir beobachten das noch heute etwa an dem aus dem Französischen stammenden Wort *Dutzend*, das nur in bestimmten Zusammenhängen benutzt wird. Sprachreste in den germanischen Sprachen deuten nicht wirklich auf eine frühere Verwendung der 12 als Kollektionseinheit hin. So bedeuteten z.B. die gotischen Namen *àinlif* für 11 und *twalif* für 12 – eng verwandt mit dem deutschen *elf* und *zwölf* und dem englischen *eleven* und *twelve* – ursprünglich etwa „eins gelassen“ bzw. „zwei gelassen“ (nachdem alle Finger zum Zählen verbraucht sind).

Wir wissen recht wenig darüber, ob und in welcher Weise in den vorgriechischen Kulturen – ähnlich wie bei den Griechen – schon ein Nachdenken über den Zahlbegriff stattfand, also eine Philosophie des Zahlbegriffs existierte, wie sie uns von den Pythagoreern (600 – 500 v.Chr.), von Platon (ca. 429 – 348) und anderen überliefert worden ist. Die Griechen unterschieden klar zwischen (natürlichen) Zahlen und *Größen*. Deren Beziehungen wurden durch vier-stellige Relationen, so genannte *Proportionen* beschrieben, aber diese in [8, Buch V] dargestellte, Eudoxos (ca. 408 – 355 v.Chr.) zugeschriebene Proportionenlehre ist keine Lehre von den rationalen Zahlen. Größen waren anschauliche, in der Regel geometrische Objekte wie Strecken- Winkel- oder Flächengrößen (siehe **10.1**). Es war den Griechen bekannt, dass mit Zahlenproportionen (aus natürlichen Zahlen) die Geometrie nicht adäquat zu beschreiben war. Ausdruck dieser Unzulänglichkeit ist, dass die Länge der Diagonalen eines Quadrats sich zur Seitenlänge verhält wie $\sqrt{2}$ zu 1. Euklid liefert einen korrekten Beweis dafür, dass Diagonale und Seite des Quadrats inkommensurabel sind, oder in heutiger Terminologie, dass $\sqrt{2}$ irrational ist. Wahrscheinlich wussten dies schon die Babylonier, die mit unterschiedlichen Näherungen für $\sqrt{2}$ rechneten. Deren Alltagsnäherung – ausgedrückt in Sexagesimalbrüchen – war $1 + \frac{25}{60}$. Sie ist doppelt so genau wie 1,41.

Vielleicht kannten die Babylonier auch folgende Überlegung zur Verbesserung einer rationalen Näherung $r$ für eine Irrationalzahl der Gestalt $\sqrt{n}$: Falls $r < \sqrt{n}$, ist $\sqrt{n} < \frac{n}{r}$; falls aber $\sqrt{n} < r$ so ist $\frac{n}{r} < \sqrt{n}$, so dass $\sqrt{n}$ in jedem Falle zwischen $r$ und $\frac{n}{r}$ liegt. Daher ist das arithmetische Mittel $s = \frac{1}{2}(r + \frac{n}{r})$ sicher eine bessere Näherung für $\sqrt{n}$. Für $n = 2$ und die erste sexagesimale Näherung $r = 1 + \frac{25}{60}$ für $\sqrt{2}$ erhält man nach Abrunden so die den Babyloniern ebenfalls bekannte Näherung $s = 1 + \frac{24}{60} + \frac{51}{60^2} + \frac{10}{60^3}$. Dies ist die beste 3-stellige sexagesimale Näherung; sie ist auf 5 Dezimalstellen genau. Möglicherweise wurde sie aber durch Ausprobieren gefunden. Aus der bloßen Tatsache der Bekanntschaft mit dieser Näherung läßt sich noch nicht folgern, die Babylonier wären im Besitz eines Iterationsverfahrens zur Berechnung von $\sqrt{2}$ gewesen.

Das von den Größen handelnde Buch X der Elemente [8] beginnt mit der folgenden klaren Definition, die wir in **10.9** unverändert übernehmen werden:

> Kommensurabel heißen Größen, die von demselben Maß gemessen werden, und inkommensurabel solche, für die es kein gemeinsames Maß gibt.

Das Phänomen inkommensurabler (oder irrationaler) Streckenverhältnisse in der Geometrie war den Griechen also wohlvertraut, und die Vermutung, dass dies eine ernste Grundlagenkrise der griechischen Mathematik verursachte, ist unbegründet. Statt von inkommensurablen Zahlen hat man von inkommensurablen Größen gesprochen, was durchaus sinnvoll ist. Die These der Pythagoreer „Alles ist Zahl“ war natürlich inzwischen modifiziert worden. Mit den so genannten Proportionen konnten auch inkommensurable Größenverhältnisse näherungsweise bestimmt werden. Man darf daher die Größen- und Proportionenlehre mit Recht als die Urform der Analysis bezeichnen.

Zwar unterschied auch ARCHIMEDES (ca. 287 – 212 v.Chr.) deutlich zwischen Zahl im herkömmlichen Sinne und Messgröße als Verhältnis der gemessenen Größe zur Einheit, aber er hatte anscheinend eine Vorstellung hierüber, die etwa derjenigen über reelle Zahlen (als Maßzahlen für geometrisch-physikalische Größen) in der beginnenden Neuzeit entsprach. Er berechnete auf mathematische Weise nicht nur die Kreiszahl $\boldsymbol{\pi}$ $(= 3{,}141\cdots)$ approximativ zu $\frac{22}{7}$, sondern auch krummlinig begrenzte Flächenstücke der verschiedensten Art. Die griechische Mathematik war zu ihrer Zeit progressiv; in ihr waren auch Erkenntnisse früherer Kulturen des Mittelmeerraumes aufgehoben. Andererseits hat die griechische Zahlkonzeption, eingeordnet in ausgeklügelte philosophische Kategorien, sich später teilweise als Hemmschuh erwiesen. Angesichts des zögerlichen Fortschritts der Mathematik im mittelalterlichen Europa hat man den Eindruck, dass einige der griechischen Dogmen allzu ehrfürchtig beachtet wurden.

Allmählich gelangten neben gebrochenen Zahlen auch negative Zahlen in Gebrauch, die das Lösen von Gleichungen wesentlich vereinfachten. Ein großer Fortschritt, und dies nicht nur im numerischen Bereich, wurde etwa gleichzeitig bewirkt durch eine scheinbare Äußerlichkeit, nämlich die Erfindung des dezimalen Positionssystems.

## 1.2 Die Entwicklung des Dezimalsystems

Die Erklärung der reellen Arithmetik in den Abschnitten **4** und **5** beruht wesentlich auf dem dezimalen Positionssystem. Dieses System stammt in seiner Grundstruktur aus Indien, wurde später von den Arabern übernommen und gelangte um das Jahr 1000 herum über den damals arabisch beherrschten südlichen und westlichen Mittelmeerraum in die italienischen Handelszentren. Das Dezimalsystem wurde mit anderen Bezeichnungen auch in China verwendet und gelangte von dort z.B. nach Japan.

Dieser Prozess vollzog sich unter wechselvollen Umständen. Die Gestalten der Ziffern wurden mehrfach abgewandelt, so dass die im europäischen Raum etwa seit der Erfindung der Buchdruckerkunst standardisierten Ziffern kaum noch Ähnlichkeit haben mit den heutigen im arabischen Raum noch verwendeten Ziffern. Diese äußerlichen Änderungen sind jedoch nicht wesentlich. Vielmehr ist dies die uns heute fast banal erscheinende Erkenntnis, dass natürliche Zahlen mit nur endlich vielen Ziffern bezeichnet werden können, wenn man die Zahlen durch eine *Anordnung* von Ziffern kodiert, unter Einschluss einer Ziffer für die (natürliche) Zahl 0. Dies bedeutet natürlich eine Abkehr von einer Bezeichnungsweise nach dem Kollektionsprinzip.

Trotz klarer Vorteile gegenüber dem in Europa fast überall benutzten römischen Zahlensystem war die Durchsetzung des indisch-arabischen Positionssystems selbst in den progressiven italienischen Handelszentren mit Widerständen verbunden. Dennoch begann sich im 13. Jahrhundert dieses System in Europa zu verbreiten, wobei das 1202 erschienene *Liber abbaci* [22] des LEONARDO von Pisa (auch FIBONACCI genannt, etwa 1170 – 1250) eine erhebliche Rolle gespielt hat. LEONARDO beginnt dieses Werk mit einer musterhaften Erklärung der Rechenalgorithmen im dezimalen Positionssystem.

Das Prinzip des Dezimalsystems, nämlich jede positive natürliche Zahl $k$ als endliche Ziffernfolge $z_0 z_1 \cdots z_n$ aus den Dezimalziffern 0 bis 9 darzustellen, ist allgemein bekannt. Es sind $z_0, \ldots, z_n$ diejenigen eindeutig bestimmten Ziffern mit $z_0 \neq 0$, so dass

$$k = z_0 \cdot g^n + z_1 \cdot g^{n-1} + \ldots + z_{n-1} \cdot g + z_n\,,$$

wobei im vorliegenden Falle $g = 10$ ist. Nun hätte man anstatt 10 hier jede beliebige andere natürliche Zahl $g \geq 2$ als Grundzahl wählen können. Man spricht dann vom *g-adischen System*; siehe hierzu auch **11.6**. Für $g > 10$ braucht man entsprechend mehr, für $g < 10$ weniger Ziffern [2)] Das so genannte *Binärsystem* ($g = 2$, auch *Dualsystem* genannt) hat G. W. LEIBNIZ (1646 – 1716) sehr geschätzt; es diente ihm als Beispiel für seine These, dass das innere Wesen der Dinge von göttlicher Harmonie geprägt ist. Das System hat nur die Ziffern $0, 1$ (die *Binärziffern*) und besitzt große praktische Bedeutung, denn die meisten digitalen Computer rechnen binär. Für $g = 8$ spricht man vom *Oktalsystem*. Hier entfallen einfach die Ziffern 8 und 9. In der folgenden Tabelle sind die natürlichen Zahlen von 1 beginnend in der oberen Reihe im Dezimalsystem, und darunter im Binär- bzw. Oktalsystem geschrieben.

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $g := 10$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | $\cdots$ |
| $g := 2$ | 1 | 10 | 11 | 100 | 101 | 110 | 111 | 1000 | 1001 | 1010 | 1011 | 1100 | $\cdots$ |
| $g := 8$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 10 | 11 | 12 | 13 | 14 | $\cdots$ |

Hierbei erhebt sich natürlich die von historischen Betrachtungen ganz unabhängige Frage, ob eine andere Grundzahl als 10 Vorteile im Alltagsleben hätte. Diese Frage muss bejaht werden, und zwar zugunsten des Oktalsystems und nicht des oft zitierten Duodezimalsystems (Basis 12). Wesentliche Vorteile des Oktalsystems sind folgende:

---

[2)] Die Länge $\ell_g k$ der Zifferfolge von $k$ in $g$-adischer Darstellung nimmt entsprechend zu. Nach (17) in **7.4** ist $\ell_g k$ für größere $k$ ungefähr gleich $\frac{\ell_{10} k}{\log g}$. Das ist für $g = 8$ nur das etwa 1,1-fache der dezimalen Länge. Die ersten 500 natürlichen Zahlen haben eine mittlere dezimale Ziffernlänge von 2,78, und eine mittlere oktale Ziffernlänge von $\frac{\Sigma_{k<500}\ell_8 k}{500} = 2{,}856$. Das ist ein Zuwachs von weniger als 3%.

- Das Einmaleins würde sich um fast die Hälfte (über 40verkürzen.
- Interpolationen bei Messvorgängen oder bei der Benutzung von Tabellen wären erheblich schneller auszuführen.
- Die Herstellung von Tastaturen und Messeinrichtungen (Schreib- und Rechenmaschinen, Waagen, Gewichtssätze und Skalen aller Art) würde sich verbilligen und ihre Handhabung vereinfacht.
- Die Ziffern $0, \ldots, 7$ des Oktalsystems werden durch die acht Tripel $000, \ldots, 111$ aus Binärziffern umkehrbar eindeutig kodiert. Deshalb erfolgt die Umrechnung einer reellen Zahl vom Oktal- ins Binärsystem *ziffernweise* in Dreierblöcken, und umgekehrt. So ist $2{,}4375 = 2{,}34_8 = 10{,}011100_2$ (weil $2 = 010_2$, $3 = 011_2$ und $4 = 100_2$). Die von kleinen und großen Rechnern ständig vorgenommenen Umrechnungen vom Dezimal- ins Dualsystem und umgekehrt würden praktisch entfallen, und damit entfiele zugleich eine permanente Quelle von Rundungsfehlern.
- Die Kapazität der Binärspeicher für Zahlen in den elektronischen Rechnern wäre viel besser auslastbar, weil in einer sagen wir 16-Bit Zelle mit einem Bitplatz für das Vorzeichen (siehe **9.4**) jede der $2^{16} - 1 = 65\,535$ ganzen Zahlen $\pm n$ mit $0 \leq n \leq 2^{15} - 1 = 32\,767 = 77777_8$ Platz fände. Das ist für 4-stellige ganzzahlige Dezimalzahlen mit Vorzeichen zu viel, für 5-stellige zu wenig.

Nur am Rande sei ein im Zeitalter der Computer natürlich unerheblicher Vorteil erwähnt: Die 2-te Rundung der Kreiszahl $\pi$ ist 3,14. Im Oktalsystem ist diese $3{,}11_8$. Sie ist nicht nur für Handrechnungen bequemer, sondern ihr relativer Fehler ist mit rund $0{,}031 der dezimalen Rundung (ca.\ 0{,}051$

Nun, es lässt sich auch in einer Welt des Dezimalsystems leben. Hier wie dort lässt sich nämlich die Zifferndarstellung natürlicher Zahlen in der bekannten, anschaulichen Weise auch auf Zifferndarstellungen nichtganzer Zahlen erweitern, was Anlass gibt zu den Dezimalzahlen oder Kommazahlen, auch Dezimalbrüche genannt. Diese sind etwa im 15. Jahrhundert an verschiedenen Orten des damals schon riesigen Verbreitungsraumes der indisch-arabischen Ziffern aufgetaucht. Von erheblicher Bedeutung für die Popularisierung der Dezimalbrüche in Westeuropa war vermutlich das Rechenbuch *De Thiende* von S. STEVIN (1548 – 1620), in welchem z.B. die Zahl 27,847 in der Weise 27(0)8(1)4(2)7(3) notiert wird. In STEVINs Büchlein finden sich auch die bekannten, noch heute benutzten Algorithmen für Addition und Multiplikation abbrechender Dezimalzahlen, die in Abschnitt **4** des vorliegenden Buches zum Fundament der Dezimalzahlarithmetik gemacht werden. Die Schreibweise STEVINs ist zur Erläuterung dieser Algorithmen anhand von Beispielen ausgezeichnet geeignet. Es ist nicht klar erkennbar, ob STEVIN die Dezimalbrüche unabhängig von Vorläufern erfunden hat, die es nachweislich mindestens 100 Jahre früher gegeben hat. Jedenfalls erkennt und verdeutlicht er klar die Vorteile eines allgemeinen dezimalen Maßsystems. In der Überzeugung, dass sich dieses früher oder später durchsetzen werde, schreibt er in [34] einleitend:

> Ob nun hierdurch kostbare, nicht käufliche Zeit gewonnen werden wird... überlasse ich gern Ihrem Urteil. Nun könnte mir jemand sagen, dass viele Dinge sich auf den ersten Blick oft besonders gut anlassen, aber wenn man sie durchführen will, so kann man nichts damit ausrichten, ebenso wie es bei den Neuerungen der Revolutionäre oft zugeht, welche im kleinen gut sind, aber im großen nichts taugen. Denen antworten wir, dass solch Zweifel hier keinesfalls bestehen kann, weil es... nun täglich in der Praxis genug erprobt wird, nämlich durch verschiedene erfahrene Landmesser hier in Holland, denen wir es erklärt haben ...

Vermutlich sind die Dezimalbrüche im Prozess des Experimentierens mit Rechenalgorithmen oder in der Praxis des Messens eher zufällig entstanden, obwohl eine dezimale Maßskala noch zu Ende des 18. Jahrhunderts weitgehend unbekannt war. Sie wurde erst nach der Französischen Revolution eingeführt und den kontinentalen Europäern durch Napoleon sozusagen zwangsverordnet.

Die Dezimalbrüche wurden und werden noch heute unterschiedlich notiert. Bezeichnungen wie $2|47$ oder $2\underline{47}$ oder 2(0)4(1)7(2) für 2,47 waren im Gebrauch. Heute ist neben dem Dezimalkomma der angelsächsische Dezimalpunkt die am weitesten verbreitete Schreibweise und es wäre vorteilhaft, wenn diese sich global durchsetzen würde.

Für Addition, Subtraktion und Multiplikation natürlicher Zahlen im Dezimalsystem wurden schon im frühen Mittelalter von verschiedenen Autoren im wesentlichen dieselben Verfahren empfohlen, die wir heute kennen. Hauptschwierigkeit war das Dividieren; hier gab es viele phantasievolle Methoden, deren Hauptprobleme insgesamt alle darauf hinaus liefen, mit den „Resten“ fertig zu werden; denn die Dezimalbrüche kamen ja erst viel später, nämlich im 17. Jahrhundert, allmählich in Gebrauch. Leonardo von Pisa löst z.B. die Gleichung $x^3 + 2x^2 + 10x = 20$ mit dem sehr exakten Näherungswert

$$1.22^{\mathrm{I}}7^{\mathrm{II}}42^{\mathrm{III}}33^{\mathrm{IV}}4^{\mathrm{V}}40^{\mathrm{VI}},$$

einer bequemen, leicht durchschaubaren Schreibweise für $1+\frac{22}{60}+\frac{7}{60^2}+\frac{42}{60^3}+\frac{33}{60^4}+\frac{4}{60^5}+\frac{40}{60^6}$ (bei Leonardo geschrieben als „unum et minuta .XXII. et secunda .VII. et tertia .XLII. et quarta .XXXIII. et quinta .IIII. et sexta .XL.“). Uns erscheint kurios, dass die natürliche Zahlen dezimal, Bruchzahlen aber zugleich sexagesimal dargestellt wurden, auch wenn die Zahl 60 besonders günstig ist hinsichtlich der relativen Anzahl ihrer Teiler. Dazu muss man aber bedenken, dass Bruchzahlen als wesentlich verschieden von natürlichen angesehen wurden, und nicht einmal Zahlen genannt wurden.

Es ist hier nicht der Platz darzustellen, welche ausgeklügelten Methoden erforderlich gewesen sind, um in diesem gemischten System Multiplikations- und Divisionsaufgaben zu lösen. Ein anderer Autor aus dem 16. Jahrhundert berechnet z.B. die Quadratwurzel aus 10 wie folgt: Er multipliziert zunächst 10 mit dem Wert $10^6$ und findet, dass $3162^2$ die am nächsten bei $10^7 = 10 \cdot 10^6$ liegende Quadratzahl ist, also $\sqrt{10^7} \approx 3162$ ($\approx$ meint *ungefähr gleich*). Er hätte nun schreiben können $\sqrt{10} \approx 3{,}162$, denn es ist

$$\sqrt{10} = \frac{\sqrt{10^7}}{\sqrt{10^6}} \approx \frac{3162}{10^3} = 3 + \frac{162}{1000}.$$

Statt dessen rechnet er $\frac{162}{1000}$ mit dem $g$-adischen Algorithmus aus **8.2** für $g = 60$ in einen Sexagesimalbruch um, $\frac{162}{1000} = 9^{\mathrm{I}}43^{\mathrm{II}}12^{\mathrm{III}}$, und erhält so $\sqrt{10} \approx 3.9^{\mathrm{I}}43^{\mathrm{II}}12^{\mathrm{III}}$. Dieses Ergebnis ist übrigens nicht die beste 3-stellige Sexagesimalbruchnäherung für $\sqrt{10}$. Vielmehr ist dies die Zahl $3.9^{\mathrm{I}}44^{\mathrm{II}}12^{\mathrm{III}} = 3{,}1622\,777\cdots$, die sich erst in der 7. Dezimalen von $\sqrt{10} = 3{,}1622\,776\cdots$ unterscheidet. Siehe hierzu auch **8.4**.

Reste des Sexagesimalsystems für Brüche haben sich bis in die heutige Zeit erhalten, insbesondere in der Winkelteilung und der Zeitmessung. Dezimale Winkelteilung schafft hier nur zusätzliche Verwirrung, denn Gewohnheiten sind nun einmal mächtig.

Die Geschichte der dezimalen Arithmetik ist ein Musterbeispiel dafür, wie wichtig scheinbare Äußerlichkeiten – im vorliegenden Falle die dezimale Darstellung reeller Zahlen – für den Fortschritt einer Wissenschaft sein können. Diese Darstellung hat mit Sicherheit geholfen, Schwierigkeiten mit den Irrationalzahlen als weniger gravierend zu empfinden. Schon die dezimale Darstellung natürlicher Zahlen war ein gewaltiger Fortschritt. Sie waren nun der Öffentlichkeit zugänglich und nicht mehr Statussymbol einer geistigen Elite. Die Griechen hatten fraglos das zahlentheoretische Wissen, um eine dezimale oder andere g-adische Darstellung natürlicher Zahlen einwandfrei zu begründen. Der EUKLIDische Algorithmus zur Bestimmung des größten gemeinsamen Teilers benutzt nämlich dasselbe Hilfsmittel, die so genannte Division mit Rest, siehe dazu **11.6**. Aber sie haben diesen Weg anscheinend nicht gesucht und auch nicht zufällig gefunden. Am Fehlen der Null lag es sicher nicht, denn durch eine Modifikation der $g$-adischen Darstellung kommt man auch ohne die Ziffer 0 ausgezeichnet zurecht, Übung 11.7. Im übrigen benutzten die Griechen und andere ein Symbol für die Null auch bei der Kennzeichnung fehlender Glieder in der sexagesimalen Bruchdarstellung.

## 1.3 Die neuere Geschichte

Eine Antwort auf die Frage nach dem *Wesen* des Zahlbegriffs, also dergestalt „Was sind die Zahlen“, hängt nicht unerheblich von dem abstrakten Begriffsgefüge ab, über das der Mensch einer jeweiligen Epoche unter Einbeziehung aller Erfahrungen der Vergangenheit verfügt. Die Maßstäbe wissenschaftlicher Strenge sind einem beständigen Wandel unterworfen, und sie werden fortfahren sich zu wandeln. Man darf sich z.B. nicht darüber wundern, dass GAUSS in mehreren der von ihm erstmals geführten Beweise des Fundamentalsatzes der Algebra bedenkenlos Zwischenwertargumente für stetige Funktionen benutzte, obwohl bekanntlich doch der Stetigkeitsbegriff später erst präzisiert und noch später die Notwendigkeit des Beweises von Zwischenwertargumenten allgemeine Anerkennung fand. Mit einem Wort, seit den Zeiten von GAUSS haben sich unsere Ansichten über das, was einer Begründung bedarf, erheblich gewandelt.

NEWTON und auch LEIBNIZ verstanden unter einer reellen Zahl das Verhältnis zweier Größen gleicher Art, von denen eine als Einheit betrachtet wurde und noch im 19. Jahrhundert wurden Irrationalzahlen etwa als Größenverhältnisse von Strecken erklärt. Die Klärung des Begriffs einer reellen Zahl nach modernen Maßstäben ist wesentlich das

Werk R. Dedekinds (1831 – 1916), obwohl auch andere, wie etwa B. Bolzano (1781 – 1848), A. Cauchy (1789 – 1857), K. Weierstrass (1815 – 1897), G. Cantor (1845 – 1918) und D. Hilbert (1862 – 1943) in diesem Zusammenhang genannt werden müssen. Der Aufbau des Systems der reellen Zahlen nach modernen Maßstäben begrifflicher Strenge ist Gegenstand der Abschnitte **3 - 5** dieses Buches und muss daher an dieser Stelle nicht erläutert werden.

Jedem Schritt einer Erweiterung des Zahlbegriffs ging ein erster Schritt voraus, nämlich das formale Operieren mit Termen wie z.B. $7 - 12$, $2 + \sqrt{3}$, $5 + \sqrt{-15}$ [3)], verbunden mit ernsten Auseinandersetzungen über die Frage, ob solchen Rechengrößen nun eine reale Existenz zukommt oder ob sie nur fiktive oder imaginäre, in unserer Einbildung existierende Größen darstellen. Die begriffliche Erfassung irrationaler Größen und ihre Einordnung in das Zahlsystem war zweifellos der entscheidende und schwierigste Schritt. Lassen wir zum Problem der irrationalen Zahlen einen bekannten Mathematiker des ausgehenden Mittelalters zu Worte kommen, M. Stifel (1486 – 1567). In seinem Hauptwerk [35] schreibt er:

> Mit Recht wird bei den irrationalen Zahlen darüber disputiert, ob sie wahre Zahlen sind oder nur fiktive. Denn bei Beweisen an geometrischen Figuren haben die irrationalen Zahlen noch Erfolg, wo uns die rationalen im Stich lassen, und sie beweisen genau das, was die rationalen Zahlen nicht beweisen konnten. Wir werden also veranlasst, ja gezwungen zuzugeben, dass sie in Wahrheit existieren, nämlich auf Grund ihrer Wirkungen, die wir als wirklich, gewiss, und feststehend empfinden.
>
> Aber andere Gründe veranlassen uns zu der entgegengesetzten Behauptung, dass wir nämlich bestreiten müssen, dass die irrationalen Zahlen Zahlen sind. Nämlich wenn wir versuchen, sie der Zählung zu unterwerfen und sie mit rationalen Zahlen in ein Verhältnis zu setzen, dann finden wir, dass sie uns fortwährend entweichen, so dass keine von ihnen sich genau erfassen lässt... Es kann aber nicht etwas eine wahre Zahl genannt werden, bei dem es keine Genauigkeit gibt und was zu wahren Zahlen kein bekanntes Verhältnis hat. So wie eine unendliche Zahl keine Zahl ist, so ist eine irrationale Zahl keine wahre Zahl, weil sie sozusagen unter einem Nebel der Unendlichkeit verborgen ist; doch ist das Verhältnis einer irrationalen Zahl zu einer rationalen nicht weniger unbestimmt als das einer unendlichen zu einer endlichen. Ferner: wenn die irrationalen Zahlen wahre Zahlen wären, dann wären sie entweder ganze oder gebrochene. Gebrochene Zahlen nenne ich diejenigen, die aus einem Zähler und einem Nenner bestehen, und zwar so, dass diese zwischen zwei aufeinanderfolgende ganze Zahlen fallen, wie z.B. $8\frac{7}{9}$ zwischen 8 und 9 fällt.

Die Vorstellung über Irrationalzahlen beschränkte sich bei Stifel überdies nur auf Wurzelausdrücke wie z.B. $\sqrt[3]{2} + \sqrt{5}$. Mit Beginn der Neuzeit, vor allem seit Erfindung der Logarithmen und der Differential- und Integralrechnung, hat sich ein allgemeinerer als nur durch Wurzelausdrücke oder geometrische Verhältnisse beschreibbarer Begriff

[3)] zuerst aufgetreten in [4] und durch $5 \cdot \tilde{p} \cdot R \cdot \tilde{m} \cdot 15$ bezeichnet, $\tilde{p} =$ plus, $\tilde{m} =$ minus, $R =$ Radix.

der Irrationalzahl als unentbehrlich erwiesen. Diese speziellen Zahlen bilden nur ein Tröpfchen im riesigen Meer aller Irrationalzahlen.

Wegen der Dichtheit der rationalen in der Menge $\mathbb{R}$ aller reellen Zahlen ist anschaulich klar, dass eine Irrationalzahl $a$ durch das Paar $(U_a, V_a)$ eindeutig bestimmt ist, wobei $U_a$ die Menge aller rationalen Zahlen $\leq a$ und $V_a$ diejenige aller rationalen Zahlen $> a$ bezeichne. Diesen natürlich schon lange bekannten Umstand hat DEDEKIND in [5] benutzt, um $a$ als eben dieses Paar (von ihm *Schnitt* genannt) zu definieren [4]. Auf der Menge dieser Schnitte lassen sich dann Rechenoperationen erklären, und das Resultat ist der Körper der reellen Zahlen (auch die Bezeichnung *Körper* stammt von DEDEKIND). Dies war aus historischer Sicht die erste, in allen Details begrifflich klare Konstruktion des Körpers der reellen Zahlen. Übrigens könnte man statt von Schnitten rationaler Zahlen z.B. auch von Schnitten abbrechender Dezimalzahlen ausgehen; man benötigt nicht unbedingt alle rationalen Zahlen zur Ausführung der Konstruktion.

Im Anschluss an DEDEKIND wurden von anderen Autoren ähnliche mengentheoretische Konstruktionen vorgeschlagen, z.B. von CANTOR dasjenige mittels Fundamentalfolgen (CAUCHY-Folgen) und von P. BACHMANN das Verfahren der Intervallschachtelungen. Schließlich wurde – im wesentlichen von HILBERT in [17] – auch eine vollständige axiomatische Charakterisierung des Bereichs der reellen Zahlen angegeben. Diese wurde später in einzelnen Punkten verfeinert, siehe auch **10.5**. In [37] wurde erstmals auch die additive Gruppe der reellen Zahlen bis auf Isomorphie eindeutig gekennzeichnet, wodurch sich ein allseitig befriedigendes Verständnis des logarithmischen Rechnens ergibt. Dies wird in **10.6** im einzelnen erläutert.

Die Analysis hat in der 2. Hälfte des 20. Jahrhunderts noch einmal eine wesentliche Bereicherung ihrer Grundlagen durch die Entwicklung der so genannten *Nichtstandard-Analysis* durch A. ROBINSON erfahren. Diese hat schon heute nicht allein nur theoretisches Interesse insofern, als durch sie die alte Idee von LEIBNIZ von den Differentialen als unendlich kleinen Zahlen in überraschender Weise Wirklichkeit geworden ist, sondern viele Phänomene aus dem Bereich der Molekularphysik und der Quantenfeldtheorie lassen sich mit ihrer Hilfe besser modellieren als mit Mitteln der traditionellen Analysis und Funktionalanalysis. Grundlage der Nichtstandard-Analysis ist die Existenz von Modellen einer jeden formalisierten oder formalisierbaren Theorie, welche ein ausgezeichnetes Modell der Theorie, meist das Standardmodell genannt, unter Erhalt aller dort gültigen Aussagen erweitern. Das betrifft insbesondere die Theorie der reellen Zahlen unter Einschluss eines ausreichenden Fragments der Mengenlehre. Darauf können wir im Rahmen dieser Darstellung aber nicht eingehen. Siehe [20] oder [32].

---

[4] genauer, $a$ wird durch diesen Schnitt *hervorgebracht*, wie DEDEKIND sich ausdrückte. DEDEKIND war der Meinung, Zahlen seien eine freie Schöpfung des menschlichen Geistes und es bedürfe keiner direkten Identifizierung von $a$ mit dem Schnitt $(U_a, V_a)$. Diese beträfe konsequenterweise dann auch rationale Schnitte, wodurch die rationalen Zahlen eine neue Qualität erhielten. Ob nun $a$ mit $(U_a, V_a)$ identifiziert wird oder nicht, ist im Prinzip unwesentlich; denn weder aus der einen noch aus der anderen Verfahrensweise lässt sich ein zusätzlicher Gewinn ziehen.

# Abschnitt 2

# Natürliche Zahlen und Rechenregeln für nichtnegative Zahlen

Mit den einfachsten Eigenschaften der natürlichen Zahlen $0, 1, 2, \ldots$ ist jedermann vertraut. Daher wollen wir uns an dieser Stelle nicht mit Erläuterungen darüber aufhalten, was diese Zahlen sind und woher die üblichen, in **2.1** genannten Rechenregeln kommen. Einzelheiten hierüber findet man im Anhang (Abschnitt **11**). Man lernt frühzeitig, wie natürliche Zahlen $n, m$ im Dezimalsystem der Größe nach verglichen, addiert, multipliziert und – falls dies möglich ist – voneinander subtrahiert werden. Zu den Elementarkenntnissen im weiteren Sinne gehört auch eine gewisse Bekanntschaft mit Primzahlen, mit elementaren Teilbarkeitseigenschaften, der Potenz $n^m$, und dem überaus wichtigen Beweisverfahren durch vollständige Induktion.

$\mathbb{N}$ bezeichnet überall die Menge der natürlichen Zahlen zu denen wir auch die Zahl 0 rechnen. Die Buchstaben $n, m, i, j, k$ werden ausschließlich als Variable für natürliche Zahlen verwendet. Daran sollte man sich stets erinnern, weil Zusätze wie $n \in \mathbb{N}$ in der Regel unterbleiben. Die auf $n$ unmittelbar folgende natürliche Zahl $n + 1$ werde mit $n^+$ bezeichnet und heißt der *Nachfolger* von $n$. Eine von 0 verschiedene natürliche Zahl heiße *positiv.* $\mathbb{N}_+$ bezeichne die Menge der positiven natürlichen Zahlen.

In diesem Abschnitt befassen wir uns mit denjenigen arithmetischen Eigenschaften oder Rechenregeln des Zahlenbereichs $\mathbb{N}$, welche bei Erweiterung von $\mathbb{N}$ zu anderen Bereichen nichtnegativer Zahlen in jedem Falle ihre Gültigkeit bewahren sollen, insbesondere für die Bereiche der endlichen Dezimalzahlen und der nichtnegativen reellen Zahlen. Einige dieser Eigenschaften werden besonders hervorgehoben und in Gestalt eines Axiomensystems zusammengestellt, aus welchem alle übrigen interessierenden arithmetischen Eigenschaften hergeleitet werden. Aus dem Axiomensystem ergeben sich zwangsläufig auch gewisse Eigenschaften der nachträglich definierten Subtraktion und der Potenz. Von allen arithmetischen Eigenschaften sollen so viele wie möglich auch nach Einführung negativer Zahlen ihre Gültigkeit bewahren. Vorerst lassen wir die negativen Zahlen aus Gründen der Übersichtlichkeit jedoch aus dem Spiel.

## 2.1 Grundrechenregeln für nichtnegative Zahlen

Mit bloßen Fertigkeiten in der Handhabung des Zahlenrechnens kann man noch keine Mathematik betreiben oder anwenden. Man muss die Rechengesetze kennen, die dieser Handhabung zugrundeliegen. Die im folgendem Axiomensystem zusammengestellten Rechengesetze oder Rechenregeln gelten sämtlich im Bereich der natürlichen Zahlen. Sie sind im Ergebnis des Umgangs mit diesen Zahlen formuliert worden und insofern empirischer Herkunft. Nicht aufgeführt sind hierbei Gesetze von rein logischem (d.h. allgemeingültigem) Charakter, wie z.B. $a = b \ \Rightarrow\ a + c = b + c$.

Die angegebenen Axiome sind charakteristisch für alle Bereiche nichtnegativer Zahlen. Für Zahlbereiche unter Einschluss negativer Zahlen sind diese geringfügig zu modifizieren. Dadurch gelangt man zu den Rechengesetzen für geordnete Ringe, deren Vorteil vor allem der Wegfall hinderlicher Einschränkungen hinsichtlich der Subtraktion ist.

**Axiome des Rechnens mit nichtnegativen Zahlen**

| | | |
|---|---|---|
| $\mathsf{N}^+ :\ a + 0 = a$ | $\mathsf{N}^\times :\ a \cdot 1 = a$ | (Neutralitätsgesetze), |
| $\mathsf{K}^+ :\ a + b = b + a$ | $\mathsf{K}^\times :\ a \cdot b = b \cdot a$ | (Kommutativgesetze), |
| $\mathsf{A}^+ :\ a + (b + c) = (a + b) + c$ | $\mathsf{A}^\times :\ a \cdot (b \cdot c) = (a \cdot b) \cdot c$ | (Assoziativgesetze), |
| $\mathsf{D} :\ (a + b) \cdot c = a \cdot c + b \cdot c$ | | (Distributivgesetz), |
| $\mathsf{V} :$ *Entweder* $a = b$ *oder* $a < b$ *oder* $b < a$ | | (Vergleichbarkeit), |
| $\mathsf{E} :$ *Es gibt ein* $d \neq 0$ *mit* $a + d = b$ *genau dann, wenn* $a < b$ | | |
| $\mathsf{F} :\ a \cdot b \neq 0$ *für* $a, b \neq 0$ | | (Nullteilerfreiheit), |
| $\mathsf{G} :\ 0 \neq 1.$ | | |

Diese Schreibweisen sind genau genommen Abkürzungen für gewisse *Aussagen*. So lautet etwa $\mathsf{K}^+$ ausführlich *Für alle* $a, b :\ a + b = b + a$. Wir verwenden auch übliche Konventionen der Klammerersparnis, z.B. bei der Niederschrift des Terms $a{\cdot}c{+}b{\cdot}c$ in Axiom $\mathsf{D}$ die Verabredung, dass der Malpunkt stärker bindet als +. Den Gepflogenheiten folgend wird der Malpunkt künftig nur dann geschrieben, wenn dies der Deutlichkeit dienlich ist. Axiom $\mathsf{E}$ beschreibt die Einschränkung für die Existenz einer von 0 verschiedenen Lösung für Gleichungen der Gestalt $a + x = b$.

Im Bereich der natürlichen Zahlen gelten weitere spezifische Gesetze, etwa dass zu jeder Zahl eine unmittelbar nachfolgende gibt, sowie insbesondere das schon erwähnte Beweisprinzip der vollständigen Induktion. Anders als diese Besonderheiten behalten jedoch die angegebenen Axiome und damit auch die daraus durch logische Folgerung herleitbaren Rechenregeln für alle Bereiche nichtnegativer Zahlen ihre Gültigkeit.

Unter einem *Bereich* verstehe man eine nichtleere Menge $A$ – genannt die *Trägermenge* – mit gewissen auf $A$ erklärten Operationen und Relationen, sowie gewissen ausgezeichneten Elementen aus $A$. Sind $+, \cdot$ solche Operationen, $<$ eine Relation, und $0, 1$ ausgezeichnete Elemente, und gelten dort alle oben aufgelisteten Axiome, so heiße dieser ein *Elementar-Bereich* oder kurz ein *$\mathcal{E}$-Bereich*. Diesen Namen wählen wir, weil obiges Axiomensystem und seine Folgerungen genau die Rechengesetze umfasst, die man im mathematischen Elementarunterricht, bewusst oder unbewusst noch vor Einführung der negativen Zahlen erlernt. Das prominenteste Beispiel eines $\mathcal{E}$-Bereichs ist $\mathbb{N}$. Weitere Beispiele, darunter die Bereiche der nichtnegativen rationalen bzw. reellen Zahlen, werden noch konstruiert. In algebraischer Terminologie handelt es sich um *natürlich geordnete kommutative Halbringe mit* 0 *und* 1. Letztere sind bei Weglassen der Null genau die Positivbereiche geordneter kommutativer (nullteilerfreier) Ringe, siehe Abschnitt **9**. Doch ist der zuletzt genannte Fakt vorerst gänzlich belanglos.

Wir unterscheiden i.a. nicht zwischen der Bezeichnung eines $\mathcal{E}$-Bereichs und seiner Trägermenge $A$. Stets sei $A_+ = a \in A \mid a > 0$. Hier und anderswo ist $a > 0$ wie verabredet nur eine andere Schreibweise für $0 < a$.

## 2.2 Folgerungen aus den Grundrechenregeln

Es werden nun aus den angegebenen Axiomen weitere Rechengesetze als logische Folgerungen hergeleitet. Zunächst sei vermerkt, dass aufgrund der Assoziativgesetze $\mathsf{A}^+$ und $\mathsf{A}^\times$ Klammern innerhalb von Termen, die entweder nur $+$ oder nur $\cdot$ enthalten, nicht geschrieben werden müssen. Auch kommt es wegen $\mathsf{K}^+$ und $\mathsf{K}^\times$ auf die Reihenfolge der Summanden bzw. Faktoren nicht an. Es gilt $c(a+b) = ca + cb$ nach $\mathsf{K}^\times$ und $\mathsf{D}$.

Ferner ist $b < a + b$, vorausgesetzt $a \neq 0$. Denn weil $b + a = a + b$, existiert ein $d \neq 0$ mit $b + d = a + b$, und zwar $d = a$, so dass $b < a + b$ gemäß $\mathsf{E}$. Das ergibt speziell $0 < a + 0 = a$ für $a \neq 0$. Mit anderen Worten, 0 ist das kleinste Element, kurzum es ist $0 < a$ für alle $a \neq 0$. In jedem $\mathcal{E}$-Bereich gilt demnach

$$\mathsf{H}: \quad 0 \leq a.$$

Ferner: $a + b = 0$ impliziert $a = b = 0$. Denn wäre z.B. $b \neq 0$, so ergäbe sich $a < 0$ nach Axiom $\mathsf{E}$. Dies aber ist unmöglich; denn nach $\mathsf{H}$ ist $0 \leq a$ und damit wird $a < 0$ durch $\mathsf{V}$ ausgeschlossen. Hiermit erhalten wir für $<$ nun leicht die Eigenschaft der *Transitivität*

$$\mathsf{T}: \quad \textit{Wenn } a < b \textit{ und } b < c, \textit{ so } a < c.$$

Denn sei $a < b$ und $b < c$, so dass $a + d = b$ und $b + e = c$ für gewisse $d, e \neq 0$ gemäß $\mathsf{E}$. Dann erhält man $a + d + e = c$, und hieraus $a < c$ nach $\mathsf{E}$, weil mit $e \neq 0$ immer auch $d + e \neq 0$. Wesentlich aus $\mathsf{E}$ folgen auch die so genannten *Monotoniegesetze*

$\mathsf{M}^+: \quad a < b \Rightarrow a + c < b + c, \qquad \mathsf{M}^\times: \quad a < b \Rightarrow ac < bc \quad (c > 0)$ [1]. Denn sei $a < b$, also $a + d = b$ für ein $d \neq 0$. Dann folgt $a + c + d = b + c$ und mit $\mathsf{E}$ $a + c < b + c$. Das beweist $\mathsf{M}^+$. Ferner ergibt $a + d = b$ offenbar $ac + dc = bc$ gemäß $\mathsf{D}$. Ist nun auch $c \neq 0$, so gilt gemäß $\mathsf{F}$ auch $dc \neq 0$, also $ac < bc$, wiederum nach $\mathsf{E}$. Das beweist $\mathsf{M}^\times$. Gelegentlich ist es nützlich, die folgenden offensichtlich beweisbaren Abschwächungen von $\mathsf{M}^+$ und $\mathsf{M}^\times$ zur Verfügung zu haben:

[1] Im Interesse eines übersichtlichen Schriftbildes stehen Zusatzvoraussetzungen oft in Klammern, anstelle eines Textes wie „falls $c > 0$“ im vorliegenden Falle.

$$a \leq b \;\Rightarrow\; a+c \leq b+c \quad ; \quad a \leq b \;\Rightarrow\; ac \leq bc, \quad \text{falls } c > 0.$$

Ist $A$ ein beliebiger $\mathcal{E}$-Bereich, und hat $f : A \to A$ die Eigenschaft $a < b \;\Rightarrow\; fa < fb$ für alle $a, b \in A$, so folgt allein aufgrund von $\mathsf{V}$ schon die Umkehrung hiervon; also $fa < fb \;\Rightarrow\; a < b$. Denn $a \not< b \Rightarrow b \leq a \Rightarrow fb \leq fa \Rightarrow fa \not< fb$. Also gilt die Umkehrung von $\mathsf{M}^+$: $a+c < b+c \;\Rightarrow\; a < b$, und analog von $\mathsf{M}^\times$. Ähnlich folgt die

*Streichungsregel*: $\quad c+a = c+b \;\Rightarrow\; a = b.$

Denn für $a \neq b$ gilt $a < b$ oder $b < a$ gemäß $\mathsf{V}$, was mit $\mathsf{M}^+$ in beiden Fällen zu einem Widerspruch zu $c+a = c+b$ führt. Mit einem analogen Schluss aus $\mathsf{M}^\times$ die

*Kürzungsregel*: $\quad ca = cb \;\Rightarrow\; a = b, \quad (c \neq 0).$

Als nächstes zeigen wir $a0 = 0a = 0$ für alle $a$. $\mathsf{N}^+$, $\mathsf{N}^\times$, sowie $\mathsf{K}^\times$ und $\mathsf{D}$ ergeben nämlich $a + 0 = a = 1a = (1+0)a = 1a + 0a = a + 0a$. Also $0a = 0$ nach der Streichungsregel.

Mit $n$ bezeichnet man in einem $\mathcal{E}$-Bereich $A$ zugleich auch das $n$-Vielfache von $1 \in A$, also $n = 1 + \ldots + 1$ mit genau $n$ Summanden. Man kann nämlich unschwer zeigen, dass die Gesamtheit der 1-Vielfachen ein isomorphes Bild des $\mathcal{E}$-Bereichs $\mathbb{N}$ ist. Für die Ordnung folgt dies leicht durch wiederholte Anwendung von $\mathsf{M}^+$, ausgehend von $0 < 1$. Denn diese ergibt $0 < 1 < 1+1 < 1+1+1 < \ldots$ Auch hat $n + m$ eine unzweideutige Bedeutung, ob man $n, m$ nun als Elemente von $\mathbb{N}$ oder als 1-Vielfache von $A$ auffasst. Dasselbe gilt für $m \cdot n$. Deswegen darf die natürliche Zahl $n$ mit dem $n$-Vielfachen von 1 identifiziert werden. Mit anderen Worten, man darf ohne Einschränkung der Allgemeinheit davon ausgehen, dass $\mathbb{N}$ Teilbereich eines jeden $\mathcal{E}$-Bereichs ist.

Seien $a, b$ Elemente eines $\mathcal{E}$-Bereichs, $b \leq a$. Dann gibt es ein $d$ mit $b + d = a$ (für $a = b$ wähle man $d = 0$). Aus $b + d = b + e = a$ folgt $d = e$ mit der Streichungsregel, also gibt es genau ein $d$ mit $b + d = a$. Dieses $d$ wird mit $a - b$ bezeichnet und heißt die *Differenz* von $a$ und $b$. Es gilt definitionsgemäß also $(a - b) + b = b + (a - b) = a$, sowie

$$(1) \quad a - b = d \;\Leftrightarrow\; b + d = a \;\Leftrightarrow\; d + b = a \quad (a \geq b).$$

Mit dieser auch *Subtraktion* genannten, nur partiell erklärten Operation lässt sich wie gewohnt rechnen – nur muss die Definiertheit der vorkommenden Terme gesichert sein. Zum Beispiel ist $a(b - c) = ab - ac$ falls $b \geq c$; denn dann ist auch $ab \geq ac$ und die Behauptung folgt mit (1) und $\mathsf{D}$ aus $ac + a(b - c) = a(c + (b - c)) = ab$. Ähnlich folgt

$$(2) \quad (a - c) + (c - b) = a - b \qquad (a \geq c \geq b).$$

Auch diese Gleichung ist vorerst nur unter der in Klammern gesetzten Bedingung sinnvoll. Sie ergibt sich mit (1) aus $(a - c) + (c - b) + b = (a - c) + c = a$.

Es sei abschließend vermerkt, dass Axiom $\mathsf{G}$ weder ein logisches ist noch folgt es aus den übrigen Axiomen. Man könnte es höchstens abschwächen zu *Es gibt Elemente* $a \neq b$, weil dann $a < b$ oder $b < a$ nach $\mathsf{V}$, also existiert ein $d \neq 0$ gemäß $\mathsf{E}$. Wäre dann $1 = 0$, ergäbe sich der Widerspruch $d = d1 = d0 = 0$.

## 2.3 Die Potenz

Sei $A$ ein beliebiger $\mathcal{E}$-Bereich und $a \in A$. Dann wird $a^n$ wie folgt rekursiv erklärt:

$$a^0 = 1 \quad ; \quad a^{n+1} = a^n \cdot a,$$

so dass $a^1 = a^0 \cdot a = a$, $a^2 = a^1 \cdot a = a \cdot a$, usw. Insbesondere ist $0^0 = 1$. Dies ist eine bequeme Konvention. Es gilt dann $1 + c + c^2 + \ldots + c^n = \sum_{i \leq n} c^i$ ohne zusätzliche Verabredungen offenbar auch für $n = c = 0$. Für $n \neq 0$ hingegen ist stets $0^n = 0$.

Es gelten die folgenden Regeln der Potenzrechnung in $A$ mit Exponenten aus $\mathbb{N}$:

$$\begin{array}{lll}
\mathsf{P}^+ :\ a^{m+n} = a^m a^n, & \mathsf{P}^\times :\ a^{mn} = (a^m)^n, & \mathsf{P}_\times :\ (ab)^n = a^n b^n, \\
\mathsf{P}^< :\ n < m \Rightarrow a^n < a^m \quad (a > 1), & \mathsf{P}_< :\ a < b \Rightarrow a^n < b^n \quad (n \neq 0), & \\
\mathsf{P}^= :\ a^n = a^m \Rightarrow n = m \quad (a \neq 0, 1), & \mathsf{P}_= :\ a^n = b^n \Rightarrow a = b \quad (n \neq 0). &
\end{array}$$

$\mathsf{P}^+, \mathsf{P}^\times, \mathsf{P}_\times$ beweist man meist durch vollständige Induktion. Es geht aber auch mit folgendem einfachen Satz, der die Induktion vorverlegt und oft deutlicher erkennen lässt, warum gewisse Identitäten gelten. $A$ kann hierin eine beliebige Menge sein.

**Satz 2.1.** *Seien $F : A \times \mathbb{N} \to A$, $f : \mathbb{N} \to A$ und $g : \mathbb{N} \to A$ Funktionen derart, dass*

$$(*) \quad f0 = g0 \quad ; \quad f(n+1) = F(fn, n) \quad ; \quad g(n+1) = F(gn, n), \quad \textit{für alle } n \in \mathbb{N}.$$

*Dann sind $f$ und $g$ identisch.*

**Beweis.** Wir zeigen $fn = gn$ durch Induktion über $n$. Es ist $f0 = g0$. Sei $fn = gn$ angenommen. $(*)$ liefert dann $f(n+1) = F(fn, n) = F(gn, n) = g(n+1)$; also ist auch $f(n+1) = g(n+1)$. Folglich gilt $fn = gn$ für alle $n$, d.h. $f = g$. ∎

Analoges gilt auch, wenn im Satz $\mathbb{N}$ durch $\mathbb{N}_+$ und $f0 = g0$ durch $f1 = g1$ ersetzt wird. Kurzum, es ist $f = g$, falls $f$ und $g$ dieselben Anfangswerte haben und derselben Rekursionsgleichung genügen. Um die Beweiskraft dieses Satzes an einem simplen Beispiel zu verdeutlichen, bezeichne $fn$ die Summe der ersten $n$ ungeraden Zahlen, so dass $f(n+1) = fn + 2n + 1 = F(fn, n)$ mit $F(x, n) = x + 2n + 1$. Ferner sei $gn = n^2$. Dann ist auch $g1 = 1$ und wegen $g(n+1) = (n+1)^2 = gn + 2n + 1$ genügen $f$ und $g$ derselben Rekursionsgleichung. Also $f = g$, d.h. $1 = 1^2$, $1 + 3 = 2^2$, $1 + 3 + 5 = 3^2$ usw.

Zum Nachweis von $\mathsf{P}^+$ betrachte man $f : \mathbb{N} \to A$ mit $fn = a^{m+n}$ und $g : \mathbb{N} \to A$ mit $gn = a^m \cdot a^n$, wobei $A$ ein $\mathcal{E}$-Bereich ist und $a \in A$ sowie $m \in \mathbb{N}$ beliebig aber fest gewählt seien. Dann gilt gewiss $f0 = a^m = g0$, und nach den Definitionen ist

$$f(n+1) = a^{m+n} \cdot a = f(n) \cdot a \quad ; \quad g(n+1) = a^m a^{n+1} = a^m a^n a = g(n) \cdot a.$$

Daher genügen $f$ und $g$ der Voraussetzung $(*)$ von Satz 2.1, mit $F(x, n) = x \cdot a$. Also $f = g$ und $\mathsf{P}^+$ ist bewiesen. Genauso zeigt man $\mathsf{P}^\times$ und $\mathsf{P}_\times$ durch den Nachweis, dass $f : n \mapsto (a^m)^n$ und $g : n \mapsto a^{mn}$ bzw. $f : n \mapsto (ab)^n$ und $g : n \mapsto a^n b^n$ die Bedingung $(*)$ für geeignetes $F$ erfüllen. Mit derselben Methode ergibt sich auch die nützliche Formel

$$(3) \quad 1 - c^{n+1} = (1 - c)(1 + c + c^2 + \ldots + c^n) \qquad (c \leq 1).$$

Denn für $f : n \mapsto 1 - c^{n+1}$ und $g : n \mapsto (1 - c)(1 + c + \ldots + c^n)$ ist sicher $f0 = g0$, und

$$\begin{aligned} f(n+1) &= 1 - c^{n+2} = (1-c^{n+1}) + (c^{n+1} - c^{n+2}) &&= fn + (1-c)c^{n+1}, \\ g(n+1) &= (1-c)(1+c+c^2+\ldots+c^n+c^{n+1}) &&= gn + (1-c)c^{n+1}. \end{aligned}$$

Damit gilt $(*)$ in Satz 2.1 mit $F(x,n) = x + (1-c)c^{n+1}$, und (3) ist bewiesen.

Wichtig und in allen $\mathcal{E}$-Bereichen gültig ist ferner die so genannte *binomische Formel*

$$(4) \qquad (a+b)^n = \textstyle\sum_{k\leq n} \binom{n}{k} a^{n-k} b^k = a^n + \binom{n}{1} a^{n-1} b + \ldots + \binom{n}{n-1} a b^{n-1} + b^n.$$

Diese verallgemeinert die Gleichung $(a+b)^2 = a^2 + 2ab + b^2$, welche für $n = 2$ durch (4) mit erfasst wird. Hier sind die so genannten *Binomialkoeffizienten* $\binom{n}{k}$ für jedes Paar $(n,k)$ mit $n \geq k$ definierte natürliche Zahlen, die wie folgt rekursiv erklärt sind:

$$\binom{n}{0} = \binom{n}{n} = 1 \quad ; \quad \binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1} \quad (k < n).$$

$$\begin{array}{ccccccccc} &&&& 1 &&&& \\ &&& 1 && 1 &&& \\ && 1 && 2 && 1 && \\ & 1 && 3 && 3 && 1 & \\ 1 && 4 && 6 && 4 && 1 \\ &&&& \vdots &&&& \end{array}$$

PASCAL*sches Dreieck*

Deren Entstehung lässt sich mit dem berühmten PASCALschen Dreieck leicht veranschaulichen. Die 1 an der Spitze gilt als 0-te Zeile und es sei $\binom{0}{0} = 1$. In der $n$-ten Zeile des Dreiecks für $n > 0$ mit den Randzahlen $\binom{n}{0} = 1$ und $\binom{n}{n} = 1$ findet man die $\binom{n}{k}$ für $k = 1, \ldots, n-1$ durch Addition der beiden jeweils darüber stehenden Zahlen. So ist $\binom{2}{1} = \binom{1}{0} + \binom{1}{1} = 1 + 1 = 2$ und $\binom{4}{3} = 3 + 3 = 6$. Induktion zeigt $\binom{n}{1} = n = \binom{n}{n-1}$ für alle $n \geq 1$. Dies erkennt man auch mit einem Blick auf das PASCALsche Dreieck. Es ist sinnvoll, $\binom{n}{k} = 0$ zu setzen für $n < k$. Dann ist $\binom{n}{k}$ für alle $n, k$ definiert.

## 2.4 Übungen

1. In den klammerfreien Termen dieser Übung sei Linksklammerung gemeint (im Zweifelsfalle auch sonstwo). So meint $a + b - c$ stets $(a + b) - c$. Man zeige:

   $$(5)\ a + (b - c) = a + b - c \quad (b \geq c), \quad (6)\ a \geq b + c \ \Leftrightarrow\ a - b \geq c \quad (a \geq b),$$
   $$(7)\ a + c \geq b \ \Leftrightarrow\ a \geq b - c \quad (b \geq c), \quad (8)\ a - (b + c) = a - b - c \quad (a \geq b + c),$$
   $$(9)\ a - (b - c) = a + c - b \quad (a + c \geq b \geq c).$$

   Man beachte, (6),(7) garantieren die Wohldefiniertheit aller Terme in (8),(9).

2. Man beweise die im Text formulierten Eigenschaften $\mathsf{P}^<, \mathsf{P}_<, \mathsf{P}^=$ und $\mathsf{P}_=$.

3. Man bestätige in $\mathcal{E}$-Bereichen $(a-b)^2 = a^2 + b^2 - 2ab$ $(a \geq b)$, und $a^2 + b^2 \geq 2ab$.

4. Man bestätige durch Induktion über $n \geq 2$ folgende Behauptungen:

   (10) $\quad (1+a)^n > 1 + na \quad (a > 0$, BERNOULLI*sche Ungleichung*$)$.

   (11) $\quad 1 < (1-a)^n + na \leq 1 + \binom{n}{2} a^2 \quad (0 < a < 1)$.

   (12) $\quad (1-a)^n > 1 - na \quad (0 < na \leq 1)$.

5. Man beweise durch Induktion über $n$ oder mit Satz 2.1 die Formel (4).

# Abschnitt 3

# Reelle Zahlen und ihre Anordnung

Endliche oder abbrechende Dezimalzahlen, auch Kommazahlen genannt, sind uns von Kindheit an vertraut. Gleichungen wie zum Beispiel $1{,}23\,m = 123\,cm$ haben einen unmittelbar anschaulichen Sinn. Auch lernt man frühzeitig, wie man solche Zahlen addiert, multipliziert und der Größe nach vergleicht.

Nun ist bekannt, dass sich z.B. bei den so genannten nicht aufgehenden Divisionen und anderen Rechenverfahren mit endlichen Dezimalzahlen nichtabbrechende Dezimalfolgen ergeben, die sich nach dem Komma unbegrenzt fortsetzen. Dabei stellt sich heraus, dass gewisse dieser Folgen als Ergebnis solcher Operationen nicht erscheinen, nämlich diejenigen mit einem „Neunerende“. Diese werden wir bei unserem Vorgehen als unzulässig ausschließen, und zwar nicht nur weil man sie im Prinzip nicht benötigt, sondern weil sie unser intuitives Bild über die dichte Anordung reeller Zahlen stören würden, siehe **3.2**. Alle übrigen Dezimalfolgen nennen wir kurzerhand reelle Zahlen.

Mit einer bloßen Benennung ist es natürlich nicht getan. Worauf es nämlich ankommt ist der Nachweis, dass man mit diesen Zahlen wie gewohnt rechnen kann und dass sie eben alle jene Eigenschaften besitzen, welche die reellen Zahlen auszeichnen. Diesen Nachweis werden wir erbringen, und zwar ohne uns auf unendliche Reihen zu berufen, ja nicht einmal auf die Bruchrechnung. Diese wird sozusagen übersprungen. In diesem Abschnitt befassen wir uns zunächst mit der Anordnung dieser Zahlen. Gerechnet wird erst in den Abschnitten **4** und **5**. Wegen des anschaulichen Charakters der Thematik werden hier von Anfang an gleich die nichtabbrechenden Dezimalfolgen in die Betrachtung mit einbezogen. Dadurch lässt sich auch die Sonderrolle der Dezimalfolgen mit einem Neunerende sehr gut veranschaulichen.

Anders als in der geordneten Menge der endlichen Dezimalzahlen gilt in der geordneten Menge aller Dezimalzahlen der Satz von der oberen Grenze, der sehr einfach beweisbar ist. Dieser garantiert nicht nur die unbeschränkte Ausführbarkeit der Division – von der Division durch 0 abgesehen – sondern auch die Existenz von Funktionen wie z.B. der Exponentialfunktion, und macht dadurch Analysis überhaupt erst möglich.

## 3.1 Reelle Zahlen als Dezimalfolgen

Eine *Folge* von Elementen einer Menge $M$ sei eine Abbildung, welche jedem $n \in \mathbb{N}$ ein Element $a_n \in M$ zuordnet. Eine solche Folge werde durch $\langle a_n \rangle$, oder z.B. auch $\langle a_i \rangle$, bezeichnet. Gelegentlich betrachten wir auch durch $\langle a_n \rangle_{n \geq k}$ bezeichnete Folgen mit dem Definitionsbereich $n \in \mathbb{N} \mid n \geq k$, sowie auch *endliche Folgen* $\langle a_0, \dots, a_n \rangle$ oder $\langle a_i \rangle_{i \leq n}$ der Länge $n + 1$. Es sind dies Abbildungen von $0, \dots, n$ in eine Menge.

Unter einer *Dezimalfolge* sei nun eine Folge $\langle z_n \rangle$ verstanden, so dass $z_i$ für jedes $i > 0$ eine Dezimalziffer ist, die eine natürliche Zahl zwischen 0 und 9 einschließlich bezeichnet (welche man sich mit der entsprechenden Ziffer identifiziert denken kann); hingegen darf $z_0$ eine beliebige natürliche Zahl in Dezimaldarstellung sein, also eine nicht mit 0 beginnende endliche Ziffernfolge, oder aber $z_0 = 0$. Die Folge $\langle z_n \rangle$ werde wie gewohnt als $z_0{,}z_1 z_2 \cdots$ notiert. Ist $a = z_0{,}z_1 z_2 \cdots$, darf $z_n$ der Deutlichkeit halber auch mit $z_n^a$ bezeichnet werden. $z_n$ $(= z_n^a)$ wird für $n > 0$ die *$n$-te Dezimale* (oder *Kommastelle*) von $a$ genannt. $n$ heiße der *Index* (genauer, *Stellenindex*) von $z_n$. Die natürliche Zahl $z_0$, der *ganzzahlige Teil von $a$*, werde mit $\operatorname{Int} a$ bezeichnet (gelesen *int $a$*). Für $a = 13{,}746 \cdots$ z.B. ist $\operatorname{Int} a = z_0 = 13$, $z_1 = 7$ die erste Dezimale, usw.

Wenn im folgenden von einer Dezimalzahl $a = z_0{,}z_1 \cdots z_n zzz \cdots$ die Rede ist, so ist stets gemeint, dass $z_i = z$ für *alle* $i > n$. Falls $a$ von der Gestalt $z_0{,}z_1 \cdots z_n 000 \cdots$ ist, also $z_i = 0$ für alle $i > n$, heiße $a = z_0{,}z_1 z_2 \cdots$ eine *endliche* oder *abbrechende Dezimalzahl der Stellenzahl $n$*. Diese denken wir uns mit der im anschaulichen Sinne endlichen Dezimalfolge $\langle z_i \rangle_{i \leq n}$ identifiziert, welche wir in gewohnter Weise als $z_0{,}z_1 \cdots z_n$ schreiben. Doch soll die Gleichung $a = z_0{,}z_1 \cdots z_n$ nicht etwa zum Ausdruck bringen, dass $n$ die kleinste Stellenzahl von $a$, also $z_n \neq 0$ ist. Die kleinste Stellenzahl von $a$ ist vielmehr das kleinste $m$ mit $z_{m+i} = 0$ für alle $i > 0$. Demnach sind z.B. sowohl 3,14 als auch 3,140 Bezeichnungen für die Dezimalzahl $3{,}14000 \cdots$, und diese Zahl hat die Stellenzahlen 2, 3, ... Aufgrund dieser Verabredung haben je zwei endliche Dezimalzahlen sicher eine gemeinsame Stellenzahl, und natürlich mehr als eine.

Es bezeichne $\mathbb{E}$ die Menge aller endlichen Dezimalzahlen. Die Zahl $z_0{,}000 \cdots$ der Stellenzahl 0 identifizieren wir mit der natürlichen Zahl $z_0$. Man kann dies auch so ausdrücken, dass $z_0{,}000 \cdots$ als neue Bezeichnung für $z_0 \in \mathbb{N}$ verwendet werden darf. Insbesondere ist $0{,}000 \cdots = 0$. Die natürlichen Zahlen gehören daher zu $\mathbb{E}$. Sie sind in diesem Zusammenhang als die endlichen Dezimalzahlen der Stellenzahl 0 gekennzeichnet.

**Bemerkung.** In diesem Stadium der Theorie hat die Schreibweise $z_0{,}z_1 \cdots z_n$ noch nichts mit einer Darstellung $z_0{,}z_1 \cdots z_n = z_0 + z_1 \cdot \frac{1}{10^1} + \ldots + z_n \cdot \frac{1}{10^n}$ zu tun. Erst nachdem die Terme $\frac{1}{10^n}$ und die Rechenoperationen auf $\mathbb{E}$ erklärt wurden, ist die Frage nach der eben erwähnten Identität sinnvoll. Vorläufig wissen wir jedoch weder was $\frac{1}{10}$ ist, noch was zum Beispiel die Summe $3 + \frac{1}{10}$ bedeutet. Wir vermeiden deshalb auch die Bezeichnung Dezimalbruch, denn Dezimalzahlen sind gemäß obiger Auffassung gewisse Folgen natürlicher Zahlen und keine Brüche im herkömmlichen Sinne.

Eine Dezimalfolge der Form $z_0{,}z_1 \cdots z_n 999 \cdots$ mit einem so genannten Neunerende heiße *unzulässig*; alle übrigen Dezimalfolgen seien *zulässig* genannt. Der Grund für diese Unterscheidung wird in **3.2** erklärt; sie kann später nach einer geeigneten Verabredung wieder aufgehoben werden. Es bezeichne $\mathbb{D}$ die Menge aller zulässigen Dezimalfolgen, die wir fortan einfach *Dezimalzahlen*, oder auch einfach nur (nichtnegative) *reelle Zahlen* nennen. Dazu gehören insbesondere die endlichen Dezimalzahlen, kurz $\mathbb{E} \subseteq \mathbb{D}$. Eine von 0 verschiedene Zahl aus $\mathbb{D}$ heiße *positiv.* $\mathbb{E}_+$ und $\mathbb{D}_+$ bezeichnen die Mengen der positiven Elemente von $\mathbb{E}$, bzw. von $\mathbb{D}$.

Wir haben den Objekten unserer Betrachtung damit im voraus die Bezeichnung *reelle Zahlen* zugestanden. Dies wird in **3.3** und den Abschnitten **4** und **5** gerechtfertigt. In diesem Sinne seien Funktionen $f : X \to \mathbb{D}$ mit $X \subseteq \mathbb{D}$ auch *reelle Funktionen* genannt. Nach Einführung der negativen Zahlen werden diese Bezeichnungen natürlich entsprechend erweitert.

Ist $a = z_0{,}z_1z_2\cdots$, so heiße $z_0{,}z_1\cdots z_n$ die *n-te kanonische Näherung* oder kurz die *n-te Näherung* von $a$ Diese werde fortan mit $a_{.n}$ bezeichnet. Offenbar ist $a \in \mathbb{E}$ genau dann, wenn es ein $n$ gibt mit $a = a_{.n}$. Denn $a = a_{.n} = z_0{,}z_1\cdots z_n$ heißt nichts anderes als $z_{n+i} = 0$ für $i > 0$. Speziell ist $a \in \mathbb{N}$ genau dann, wenn $a = a_{.0}$ $(= \operatorname{Int} a)$.

**Beispiel.** Für $a = 3{,}14000\cdots$ ist $a_{.0} = 3$, $a_{.1} = 3{,}1$ und $a_{.2} = a_{.3} = \ldots = a = 3{,}14$.

## 3.2 Die Anordnung der Dezimalzahlen

Allgemein versteht man unter einer *Anordnung* oder *Ordnung* einer Menge $M$ eine Relation $<$ auf $M$ mit den Eigenschaften $\mathsf{T}$ und $\mathsf{V}$ der Transitivität und der Vergleichbarkeit, siehe **2.1**. $\mathsf{V}$ lässt sich hierbei auch ersetzen durch die beiden Forderungen $\mathsf{I}$ : $a \not< a$ (*Irreflexivität*), sowie $\mathsf{V}'$: $a < b$ *oder* $a = b$ *oder* $b > a$, einer Abschwächung von $\mathsf{V}$. Wegen $\mathsf{I}$ schließen sich die Fälle $a < b$, $b < a$, $a = b$ nämlich gegenseitig aus. Wäre z.B. $a < b$ und $b < a$, folgt $a < a$ nach $\mathsf{T}$, ein Widerspruch zu $\mathsf{I}$. Dass anderseits $\mathsf{V}$ auch **I** impliziert, ist klar: wegen $a = a$ entfällt die Möglichkeit $a < a$.

Wir gehen aus von der üblichen (lexikographischen) Anordnung der Dezimalfolgen: Sind $a = z_0{,}z_1z_2\cdots$ und $a' = z'_0{,}z'_1z'_2\cdots$ voneinander verschieden und ist $z_0 \neq z'_0$, so soll $a < a'$ bzw. $a' > a$ sein, je nachdem ob $z_0 < z'_0$ oder $z_0 > z'_0$. Wenn aber $z_0 = z'_0$, so richten wir uns nach den Werten von $z_1$ und $z'_1$ und allgemein nach dem kleinsten Index $n$, so dass $z_n \neq z'_n$. Wir fassen dies zusammen in der folgenden

**Definition.** Für $a = z_0{,}z_1z_2\cdots$ und $a' = z'_0{,}z'_1z'_2\cdots$ sei $a < a'$ genau dann, wenn $a \neq a'$ und wenn $z_n < z'_n$ für den kleinsten Index $n$ mit $z_n \neq z'_n$.

Offenbar ist $a_{.n} \leq a$ für alle $n$. Ist $a < b$, gibt es offenbar ein $n$ mit $a < b_{.n}$, und auch umgekehrt. So ist $2{,}6398\cdots < 2{,}640197\cdots$, weil bereits $2{,}6398\cdots < 2{,}64$. Es ist klar, dass $<$ die Eigenschaften $\mathsf{V}$ und $\mathsf{T}$ hat, also eine Ordnung ist, insbesondere auf $\mathbb{D}$. Ferner ist klar, dass sich für natürliche Zahlen nichts anderes ergibt, als die dort bereits vorhandene Anordnung. Anders als in $\mathbb{N}$ gibt es jedoch kein kleinstes positives

Element in $\mathbb{D}$. Dies erkennt man leicht durch Betrachtung der folgenden Zahlenreihe, die im weiteren eine wichtige Rolle spielen wird:

$$\varepsilon_0 = 1; \quad \varepsilon_1 = 0{,}1; \quad \varepsilon_2 = 0{,}01; \quad \varepsilon_3 = 0{,}001 \text{ usw.}$$

Es ist unmittelbar klar, dass $\varepsilon_0 > \varepsilon_1 > \ldots > 0$ und dass zu jedem $a \in \mathbb{D}$ ein $n$ existiert mit $\varepsilon_n < a$. Offenbar ist $\varepsilon_n$ unter allen von 0 verschiedenen endlichen Dezimalzahlen der Stellenzahl $n$ gerade die kleinste.

Nunmehr kann erklärt werden, warum die Dezimalfolgen $a = z_0{,}z_1 \cdots z_n999\cdots$ eine Sonderrolle in den Betrachtungen spielen und zu unzulässigen Zahlen degradiert wurden. Es gibt nämlich keine Dezimalfolge, die im Sinne der erklärten Anordnung zwischen $a = z_0{,}z_1 \cdots z_n999\cdots$ und $a' = z_0{,}z_1 \cdots z_n^+$ [1)] liegt, wobei $z_n \neq 9$. Aus $a \leq x \leq a'$ folgt nämlich $x = a$ oder $x = a'$ für alle $x \in \mathbb{D}$. Man verdeutliche sich dies an Beispielen, etwa $a = 3{,}4999\cdots$ und $a' = 3{,}5000\cdots$. Um der anschaulich klaren Forderung gerecht zu werden, dass zwischen zwei reellen Zahlen mindestens eine weitere liegt, ist man genötigt, eine der beiden folgenden Alternativen zu wählen: Entweder man schließt die Dezimalfolge $a$ von den Betrachtungen aus – d.h. man erklärt sie für unzulässig – oder man identifiziert die beiden Dezimalfolgen $a$ und $a'$, d.h. man sieht sie als Bezeichnungen für dasselbe Objekt an. Es gibt gewisse technische, aber keine sachlichen Gründe, warum wir uns für die Wahl der ersten Alternative entschließen. Prinzipiell könnte man die endlichen Dezimalzahlen sogar ganz aus dem Spiel zu lassen; denn jede reelle Zahl hat eine eindeutige nichtabbrechende Darstellung wie in **8.2** gezeigt wird.

Obige anschauliche Betrachtung macht nun deutlich, dass $\mathbb{D}$ hinsichtlich der Relation $<$ dicht geordnet ist, d.h. zwischen je zwei derartigen Zahlen liegt noch wenigstens eine, und damit unendlich viele weitere. Ganz allgemein heißt eine geordnete Menge *dicht geordnet*, wenn sie mindestens zwei Elemente enthält und zwischen je zwei Elementen noch ein weiteres Element liegt. Die Dichtheit der Anordnung von $\mathbb{D}$ lässt sich nun sogar zu folgender Aussage verschärfen, welche zugleich die Dichtheit von $\mathbb{E}$ offenbart:

**Satz 3.1.** *Zwischen je zwei Elementen $a, b \in \mathbb{D}$ mit $a < b$ liegt wenigstens ein $c \in \mathbb{E}$, d.h. es ist $a < c < b$. Mit anderen Worten, $\mathbb{E}$ liegt dicht in $\mathbb{D}$.*

**Beweis.** Sei $a = z_0{,}z_1z_2\cdots < b$ und $k$ kleinster Index mit $a < b_{.k}$. Falls $b_{.k} < b$, wähle man $c = b_{.k}$ womit die Behauptung offenbar erfüllt ist. Falls aber $b_{.k} = b$, gibt es wegen der Zulässigkeit von $a$ sicher ein $n > k$ mit $z_n \neq 9$. Dann gilt für $c = z_0{,}z_1 \cdots z_k \cdots z_n^+$ sicher $a < c$, aber auch $c < b_{.k} = b$. ■

Es sei dem Anfänger geraten, diesen Beweis anhand von Beispielen zu verfolgen. Sei etwa $a = 3{,}49996\cdots$ und $b = 3{,}5001\cdots$, so dass $a < b_{.1}$. Eine gemäß Beweisvorschrift konstruierte Zahl ist $c = 3{,}5$. Ist hingegen $b$ selbst gleich 3,5 wähle man $c = 3{,}49997$.

[1)] Die Nachfolgeroperation soll sich im Falle $n > 0$ nur auf die letzte Ziffer beziehen. Da wir nicht strikt zwischen einer Ziffer als Symbol und der durch sie bezeichneten natürlichen Zahl unterscheiden, ist der Nachfolger $z^+$ von $z < 9$ wieder eine Ziffer, so dass $z_0{,}z_1 \cdots z_n^+$ im Falle $n > 0$ nur für $z_n \neq 9$ wohlerklärt ist. Zum Beispiel ist $0{,}98^+ = 0{,}99$. Für $n = 0$ sei $z_0{,}z_1 \cdots z_n^+ = z_0 + 1$.

## 3.3 Der Satz von der oberen Grenze

Dieser Satz ist grundlegend für die gesamte Theorie der reellen Zahlen und Funktionen. Zu seiner Formulierung in Satz 3.2 werden einige Begriffe benötigt, die wir gleich für eine beliebige geordnete Menge $M$ formulieren, obwohl sie zunächst nur für den Spezialfall der geordneten Mengen $\mathbb{E}$ und $\mathbb{D}$ wichtig sind.

**Definition.** Eine Teilmenge $X \subseteq M$ einer geordneten Menge $M$ heißt *nach oben beschränkt*, wenn ein $b \in M$ existiert mit $x \leq b$ für alle $x \in X$. Jedes derartige $b$ heißt eine *obere Schranke* für $X$. Ein Element $s \in M$ heißt *obere Grenze* oder *Supremum* für $X \subseteq M$, symbolisch $s = \sup X$, wenn $s$ eine kleinste obere Schranke von $X$ ist, d.h. es gelten (a) $s$ ist obere Schranke von $X$, und (b) $s \leq b$ für jede obere Schranke $b$ von $X$. (b) ist wegen Eigenschaft $\vee$ gleichwertig zu (b'): Zu jedem $u < s$ gibt es ein $x \in X$ mit $u < x$, d.h. man kommt $s$ von links mit Elementen aus $X$ beliebig nahe.

**Beispiel 1.** Sei $a \in \mathbb{D}$ und $X = a_{.0}, a_{.1}, a_{.2}, \ldots$ die durch $a$ beschränkte Menge der kanonischen Näherungen von $a$. Wir behaupten $\sup X = a$. Ist nämlich $u < a$ beliebig gewählt, so ist $u < a_{.n}$ für hinreichend großes $n$ gemäß Definition von $<$, so dass Bedingung (b') erfüllt ist. Insbesondere ist $\sup 0{,}3\,;\, 0{,}33\,;\ldots = 0{,}333\cdots$.

**Beispiel 2.** . $X = z_0{,}9\,;\, z_0{,}99\,;\, z_0{,}999\,;\ldots$ ist offenbar beschränkt durch $z_0 + 1$ und es gilt $\sup X = z_0 + 1$. Denn gewiss ist $z_0 + 1$ obere Schranke von $X$. Sei nun $u < z_0 + 1$. Falls $\operatorname{Int} u < z_0$ ist sicher $u < z_0$. Aber auch im Falle $\operatorname{Int} u = z_0$ gibt es – weil $u$ kein Neunerende hat – offenbar ein $n$ mit $u < z_0, \underbrace{9\cdots 9}_{n}$. Allgemeiner sei $X \subseteq \mathbb{D}$, $n \geq 0$, $s = z_0{,}z_1 \cdots z_n^+$ obere Schranke von $X$, und zu jedem $m$ existiere ein $x \in X$ mit $z_0{,}z_1 \cdots z_n \underbrace{9\cdots 9}_{m} < x$. Dann ist $\sup X = s$. Denn offenbar gibt es ein hinreichend großes $k$ mit $u < z_0{,}z_1 \cdots z_n \underbrace{9\cdots 9}_{k}$ und damit ist auch $u < x$ für ein gewisses $x \in X$.

Völlig analog definiert man die Begriffe nach *unten beschränkt, untere Schranke* und *untere Grenze (Infimum)*. Eine eventuell existierende untere Grenze für $X$ wird mit $\inf X$ bezeichnet. Eine *beschränkte* Teilmenge von $M$ ist eine solche, die zugleich nach oben und nach unten beschränkt ist. Eine nach oben beschränkte Teilmenge $X$ von $M$ muss eine obere Grenze besitzen. Wenn sie aber eine solche hat, so ist diese eindeutig bestimmt. Denn sind $a, b$ obere Grenzen für $X$, so ist $a \leq b$, weil $a$ kleinste obere Schranke ist. Aus analogem Grunde ist auch $b \leq a$, also $a = b$. Obere und untere Grenze von $X$ können, aber müssen nicht zu $X$ gehören.

Jede Teilmenge von $\mathbb{D}$ ist nach unten beschränkt, z.B. durch 0. Daher bedeutet für $\mathbb{D}$ beschränkt dasselbe wie nach oben beschränkt. Die leere Teilmenge von $\mathbb{D}$ hat außerdem gerade das Supremum 0. Deswegen ist die Voraussetzung $X \neq \emptyset$ in Satz 3.2 unten eigentlich überflüssig! Hingegen hat $\emptyset$ in $\mathbb{D}$ kein Infimum. Das liegt natürlich daran, dass $\mathbb{D}$ ein kleinstes aber kein größtes Element hat. In einer geordneten Menge ohne kleinstes und größtes Element hat $\emptyset$ weder ein Supremum noch ein Infimum.

Es lässt sich sehr einfach zeigen, dass nicht jede beschränkte Teilmenge von $\mathbb{E}$ ein Supremum in $\mathbb{E}$ besitzt. Damit lässt sich viel einfacher als anhand von $\mathbb{Q}$ (siehe **6**) die Existenz dichter, lückenhafter Mengen verdeutlichen. Hat nämlich $X \subseteq \mathbb{E}$ ein Supremum $s$ bereits in $\mathbb{E}$, so ist $s$ auch Supremum für $X$ in $\mathbb{D}$, wie aus der Dichtheit von $\mathbb{E}$ in $\mathbb{D}$ sofort folgt. Hätte also z.B. $X = 0{,}3\,;\, 0{,}33\,;\, \dots$ ein Supremum $s \in \mathbb{E}$, so wäre $s$ auch Supremum für $X$ in $\mathbb{D}$. Das aber ist ausgeschlossen, denn $\sup X = 0{,}333\cdots$ liegt nicht in $\mathbb{E}$. Anders sind die Verhältnisse in $\mathbb{D}$. Denn hier gilt der

**Satz 3.2.** *Jede nichtleere beschränkte Teilmenge $X$ von $\mathbb{D}$ besitzt ein Supremum.*

**Beweis.** Entweder hat $X$ ein Supremum in $\mathbb{E}$ oder nicht. Im ersten Fall ist nichts zu beweisen. Es liege also der zweite Fall vor. Wir werden jetzt die Ziffern einer nichtabbrechenden Dezimalzahl $s = z_0{,}z_1z_2\cdots$ schrittweise konstruieren und beweisen $s = \sup X$.

Da $X$ nichtleer und beschränkt ist, gibt es gewiss eine größte natürliche Zahl $z_0$, so dass gerade noch $z_0 \leq x$ für wenigstens ein $x \in X$. Seien nun $z_0, \dots, z_n$ schon definiert, und zwar so, dass $z_0{,}z_1\cdots z_n \leq x$ für ein gewisses $x \in X$. Dann sei $z_{n+1}$ die größte Ziffer, so dass noch $z_0{,}z_1\cdots z_{n+1} \leq x$ für wenigstens ein $x \in X$. Weil gemäß Voraussetzung $z_0{,}z_1\cdots z_n0 \leq x$ für ein $x \in X$, ist $z_{n+1}$ wohlbestimmt, und falls $z_{n+1} < 9$, ist $z_0{,}z_1\cdots z_n^+$ sogar obere Schranke von $X$. Damit ist $s = z_0{,}z_1z_2\cdots$ wohlerklärt. Als erstes zeigen wir, dass $s$ zulässig ist. Wäre $s = z_0{,}z_1\cdots z_n999\cdots$ für den Fall $n > 0$, wäre nach Beispiel 2 $z_0{,}z_1\cdots z_n^+$ Supremum von $X$, im Widerspruch dazu, dass wir uns im zweiten Fall befinden. Als nächstes zeigen wir, $s$ ist obere Schranke für $X$. Sei $a \in X$. Falls $a = s$ ist auch $a \leq s$. Falls $a \neq s$, sei $k$ kleinster Index mit $z_k \neq z_k^a$ (= $k$-te Ziffer von $a$). Aufgrund der Bestimmung von $z_k$ folgt $z_k^a \leq z_k$ also $z_k^a < z_k$, daher $a < s$. Schließlich ist noch zu bestätigen, dass $s$ wirklich $s = \sup X$. Das aber ist klar, denn für jedes $n$ gibt es gemäß Konstruktion ein $x \in X$ mit $z_0{,}z_1\cdots z_n \leq x$. Damit ist der Satz bewiesen. ■

Man kann die Fallunterscheidung in diesem Beweis auch vermeiden, indem man die Konstruktion eines unzulässigen $s$ in Kauf nimmt und $s$ sodann durch die entsprechende zulässige Dezimalzahl ersetzt. Jedenfalls liefert dieser Beweis ein Verfahren zur schrittweisen Berechnung von $\sup X$ für konkret gegebenes $X$. So folgt nach Definition der Multiplikation in **5** leicht, dass $s^2 = a$, wobei $s = \sup x \in \mathbb{E} \mid x^2 \leq a$. Deshalb kann man $s = \sqrt{a}$ für explizit gegebenes $a \in \mathbb{D}$ auch ohne ein spezielles Verfahren mit Satz 3.2 näherungsweise und mit beliebiger Genauigkeit notfalls auch von Hand ausrechnen. Der Satz ist aber ein mächtiges Existenzprinzip eben auch dann, wenn man bestenfalls nur weiß, dass $X$ beschränkt ist. Dazu das subtile und lehrreiche Beispiel unten. Sind $M$ und $N$ geordnete Mengen, so heißt $f\colon M \to N$ *monoton wachsend* oder einfach *wachsend*, wenn $x < y \;\Rightarrow\; fx \leq fy$, für alle $x, y \in M$. Teilmengen von $M$ der Gestalt $x \in M \mid u \leq x \leq v$ mit $u < v$ heißen *abgeschlossene Intervalle* oder einfach *Intervalle*. Ist $c$ ein Wert mit $fc = c$, so heißt $c$ ein *Fixpunkt* von $f$.

**Beispiel 3.** Sei $I$ ein Intervall in $\mathbb{D}$ und $f\colon I \to I$ wachsend. Wir behaupten, $f$ hat einen Fixpunkt in $I$. Und zwar ist $s = \sup X$ ein solcher, wobei $X := x \in I \mid x \leq fx$. Mit $X \subseteq I$ ist auch $s \in I$ also ist $fs$ wohlerklärt. $x \in X$, also $x \leq fx$, impliziert $fx \in X$, weil auch $fx \leq f(fx)$. Aus $x \in X$ folgt daher $x \leq fx \leq s$ und so $fx \leq fs$. Also ist $fs$ obere Schranke für $X$ und deshalb $s \leq fs$. Daher ist auch $s \in X$, also $fs \in X$ und somit $fs \leq s$. Insgesamt ergibt sich $s = fs$, d.h. $s$ ist in der Tat ein Fixpunkt von $f$.

Von vielen Anwendungen des Satzes 3.2 nennen wir die Erklärungsmöglichkeit von Potenzen $b^x$ mit beliebiger Basis $b > 0$ und irrationalen Exponenten $x$ in **7.3**. Diese Konstruktion ist im Grunde viel harmloser als die des letzten Beispiels. Denn $b^x$ wird durch eine konstruktive Anwendung von Satz 3.2 definiert. Immer wenn die Basis $b$ und der Exponent $x$ explizit gegeben sind, ist auch $b^x$ explizit berechenbar. Hingegen ist ein Fixpunkt von $f: I \to I$ i.a. nur berechenbar, wenn $f$ über die bloße Monotonie und Berechenbarkeit hinaus weitere Zusatzeigenschaften hat, siehe hierzu **8.4**.

## 3.4 Lückenlosigkeit der reellen Zahlen

Es ist nützlich, sich den Satz von der oberen Grenze auf unterschiedliche Weise zu veranschaulichen, z.B. durch die Lückenlosigkeit der Ordnung von $\mathbb{D}$. Dazu die folgenden ganz elementaren, auf Dedekind [5] zurückgehenden Definitionen.

Unter einem *Schnitt* einer geordneten Menge $M$ versteht man ein Paar nichtleerer Teilmengen $U, V$ von $M$ derart, dass

$$(1)\ \ U \cup V = M, \quad (2)\ \ U \cap V = \emptyset, \quad (3)\ a \in U, b \in V \ \Rightarrow\ a < b.$$

$U$ heißt die *Unterklasse* und $V$ die *Oberklasse* des Schnitts $(U, V)$. Offenbar gibt es nur die folgenden drei Typen von Schnitten:

I. $U$ hat ein größtes und $V$ ein kleinstes Element,
II. $U$ hat kein größtes und $V$ kein kleinstes Element,
III. $U$ hat ein größtes und $V$ kein kleinstes Element, oder umgekehrt.

Schnitte vom Typ I heißen *Sprünge*, solche vom Typ II *Lücken*, und Schnitte vom Typ III seien Dedekind*sche Schnitte* genannt. Figur 1 veranschaulicht die drei Typen. Jedem Schnitt $(U, V)$ vom Typ III ist eindeutig ein so genanntes *Schnittelement* zugeordnet, womit das größte Element von $U$ oder das kleinste Element von $V$ gemeint ist. Dieses ist offensichtlich sowohl Supremum von $U$ als auch Infimum von $V$.

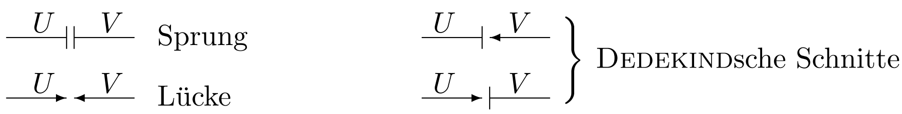



*Fig. 1 Schnitt-Typen geordneter Mengen*

Sei $M$ eine dicht geordnete Menge, d.h. $M$ hat keine Sprünge und enthält mehr als ein Element. Hat in einem solchen Falle $M$ auch keine Lücken, d.h. ist $M$ *lückenlos geordnet*, kurz, ist $M$ lückenlos, so verbleiben offenbar nur Dedekindsche Schnitte. Umgekehrt ist eine geordnete Menge $M$ mit nur Dedekindschen Schnitten natürlich lückenlos. $M$ erfüllt in diesem Falle auch den Satz von der oberen Grenze (d.h. Satz 3.2 mit $M$ für $\mathbb{D}$). Denn sei $X \subseteq M$ nichtleer und beschränkt, $V$ die Menge der oberen Schranken, und $U := M\mathtt{V}$. Dann ist das Schnittelement $s$ des Schnitts $(U, V)$ gerade das Supremum von $M$, wie leicht zu erkennen ist. Gilt umgekehrt in $M$ der Satz von der oberen Grenze, so ist $M$ auch lückenlos; denn das Supremum der Unterklasse eines Schnittes ist offenbar dessen Schnittelement. Damit haben wir folgenden Satz bewiesen:

**Satz 3.3.** *Für eine dicht geordnete Menge $M$ sind gleichwertig*

(i) *$M$ ist lückenlos geordnet,*
(ii) *Jeder Schnitt von $M$ ist ein* DEDEKIND*scher,*
(iii) *In $M$ gilt der Satz von der oberen Grenze.*

## 3.5 Übungen

1. Es bezeichne $f$ jeweils eine der der Bedingung $f1 = 1$ genügenden Funktionen
$$n \mapsto 1 + 2 + \ldots + n, \quad n \mapsto \binom{n+1}{2}, \quad n \mapsto \tfrac{n(n+1)}{2} \qquad (n > 0).$$
Man zeige, diese genügen alle der Rekursionsgleichung $f(n+1) = fn + n + 1$ und sind nach Satz 2.1 daher identisch. Kurzum, $\sum_{0<i\leq n} i = \binom{n+1}{2} = \frac{n(n+1)}{2}$. Diese Gleichungen gelten übrigens auch für $n = 0$.

2. Der *Abstand* $|a - b|$ zweier Elemente eines $\mathcal{E}$-Bereichs sei wie folgt erklärt: Es sei $|a - b| = a - b$ für $a \leq b$ und $|a - b| = b - a$ sonst, d.h. $a < b$. Offenbar gelten dann $|a - b| = |b - a|$, sowie $|a - b| = 0 \Leftrightarrow a = b$. Man beweise die so genannte *Dreiecksungleichung* $|a - b| \leq |a - c| + |b - c|$. Diese Bezeichnung ist konsistent mit der in **9.2** eingeführten Notation $|a|$ (absoluter Betrag).

3. Ein $\mathcal{E}$-Bereich $A$ heißt *archimedisch* (geordnet), wenn es zu allen $a, b \in A_+$ ein $n$ gibt mit $na > b$. Sicher ist $\mathbb{N}$ archimedisch. Man beweise, dass in einem archimedischen $\mathcal{E}$-Bereich $A$ zu allen $a, b \in A_+$ mit $a > 1$ ein $n$ mit $a^n > b$ existiert.

4. Man beweise, ein lückenlos geordneter $\mathcal{E}$-Bereich $A$ ist archimedisch (die Umkehrung gilt i.a. nicht wie das Beispiel des $\mathcal{E}$-Bereichs $\mathbb{E}$ zeigen wird).

5. Seien $M, G$ geordnete Mengen. $f : M \to G$ heißt *strikt wachsend* oder *ordnungstreu*, wenn $a < b \Rightarrow fa < fb$, für alle $a, b \in M$. Sei nun $f : \mathbb{E} \to G$ wachsend und $G$ lückenlos (wie z.B. $G = \mathbb{D}$). Die Funktion $\bar{f} : \mathbb{D} \to G$ sei erklärt durch
$$\bar{f}a = \sup f(a_{.n}) \mid n \in \mathbb{N}$$
und heiße die *natürliche Fortsetzung* von $f$ auf $\mathbb{D}$. Man beweise

   (a) $\bar{f}$ ist Fortsetzung von $f$, d.h. $\bar{f}a = fa$ für alle $a \in \mathbb{E}$,
   (b) $\bar{f}$ ist wachsend, und mit $f$ ist auch $\bar{f}$ strikt wachsend.

# Abschnitt 4

# Arithmetik der abbrechenden Dezimalzahlen

Vor einer Erklärung der Rechenoperationen für beliebige Dezimalzahlen definieren wir diese zuerst im Bereich $\mathbb{E}$ der endlichen oder abbrechenden Dezimalzahlen. Dies ist mehr als nur eine technische Vorbereitung auf die reelle Arithmetik. Das Rechnen mit endlichen Dezimalzahlen verdient schon deswegen spezielle Aufmerksamkeit, weil wir damit im Alltag ständig zu tun haben. Auch rechnen sowohl Taschenrechner als auch Hochleistungscomputer mit endlichen Dezimalzahlen [1)].

Bei der Summierung der Dezimalzahlen $z_0{,}z_1 \cdots z_n$ und $z'_0{,}z'_1 \cdots z'_m$ verfährt man bekanntlich wie folgt: Zuerst wird durch „Anhängen von Nullen“ die Stellenzahl angeglichen; kurz, es darf hierbei $m = n$ angenommen werden. Sodann summiert man die natürlichen Zahlen $z_0 z_1 \cdots z_n$ und $z'_0 z'_1 \cdots z'_n$ – d.h. man denkt sich die Kommata vorübergehend entfernt – und trennt in dieser Summe die letzten $n$ Stellen durch das Komma ab. Ist hingegen das Produkt von $z_0{,}z_1 \cdots z_n$ und $z'_0{,}z'_1 \cdots z'_m$ zu bilden, wird $z_0 z_1 \cdots z_n$ mit $z'_0 z'_1 \cdots z'_m$ multipliziert, und in diesem Resultat werden sodann $n + m$ Stellen durch das Komma abgetrennt.

Diese einfachen Kommasetzungsregeln lassen sich leicht formal fassen. Auf diese Weise wird nicht nur die Ausführung der Rechenoperationen mit Dezimalzahlen auf diejenige mit natürlichen Zahlen zurückgeführt, sondern es lassen sich auch die Rechengesetze auf eine verblüffend einfache Weise beweisen. Rationale Zahlen oder auch nur eine spezielle Form der Bruchrechung treten hierbei nicht in Erscheinung. Es zeigt sich, dass entgegen einer landläufigen Meinung die Dezimalzahlarithmetik keineswegs der Bruchrechnung als eines theoretischen Fundaments bedarf. Was man hier benötigt ist lediglich die Gewissheit, dass die mehr als 1000 Jahre alten Rechenverfahren für Summe und Produkt natürlicher Zahlen im Dezimalsystem tatsächlich das gewünschte Ergebnis liefern, und dies setzen wir selbstverständlich voraus.

[1)] genau genommen, falls deren Chips speziell für kaufmännisches Rechnen ausgelegt sind; sonst wird intern mit Binärzahlen gerechnet, was aber keinen äußerlich erkennbaren Unterschied ausmacht.

## 4.1 Rechnen mit endlichen Dezimalzahlen

Zuerst führen wir, und zwar gleich für beliebige Dezimalzahlen, die Operationen $\xrightarrow{n}$ und $\xleftarrow{n}$ der $n$-fachen *Kommaverschiebung nach rechts* beziehungsweise *nach links* ein. Sei $a \in \mathbb{D}$. Dann entstehe $a\xrightarrow{n}$ aus $a$ dadurch, dass in $a$ das Komma um $n$ Stellen nach rechts verschoben wird. Analog sei $a\xleftarrow{n}$ das Resultat der Kommaverschiebung in $a$ um $n$ Stellen nach links. Abbrechende Dezimalzahlen werden durch die Operationen $\xrightarrow{n}, \xleftarrow{n}$ wieder in abbrechende, nichtabbrechende in nichtabbrechende überführt. Eine $n$-stellige Dezimalzahl wird durch $n$-fache Kommaverschiebung nach rechts offenbar in eine natürliche Zahl überführt. So erhält man unter Beachtung von $\varepsilon_n = 0,\underbrace{0\cdots01}_{n}$

$$25{,}3\xrightarrow{2} = 2530 \quad ; \quad 25{,}3\xleftarrow{3} = 0{,}0253 \quad ; \quad 1\xrightarrow{n} = 10^n \quad ; \quad 1\xleftarrow{n} = \varepsilon_n \quad ; \quad \varepsilon_n\xrightarrow{n} = 1.$$

Es ist offensichtlich und sei nur nebenbei erwähnt, dass jedes $a \in \mathbb{E}_+$ sich in der Weise $a = z_0{,}z_1\cdots z_n\xrightarrow{k}$ oder $a = z_0{,}z_1\cdots z_n\xleftarrow{k}$ für geeignetes $k$ so schreiben lässt, dass $z_0$ eine Dezimalziffer $\neq 0$ ist, die so genannte *Gleitkommadarstellung*. Z.B. ist $27 = 2{,}7\xrightarrow{1}$. Auf dem Monitor oder gegebenenfalls auch im Taschenrechner-Display erscheint $a$ im ersten Falle als $z_0{,}z_1\cdots z_n\mathrm{E}k$, im zweiten als $z_0{,}z_1\cdots z_n\mathrm{E}{-k}$. Auch rechnerintern wird das Exponentensymbol E lediglich als Kennzeichnung einer Kommaverschiebung (*shifting*) interpretiert. Diesbezüglich gelten offenbar die folgenden Regeln:

$$\mathsf{S}_0: \; a\xrightarrow{0} = a\xleftarrow{0} = a,$$

$$\mathsf{S}_1: \; a\xrightarrow{n}\xrightarrow{m} = a\xrightarrow{n+m} \quad ; \quad a\xleftarrow{n}\xleftarrow{m} = a\xleftarrow{n+m},$$

$$\mathsf{S}_2: \; a\xleftarrow{m}\xrightarrow{n} = a\xrightarrow{n}\xleftarrow{m} = \begin{cases} a\xrightarrow{n-m} & \text{falls } n \geq m, \\ a\xleftarrow{m-n} & \text{falls } m \geq n, \end{cases}$$

$$\mathsf{S}_3: \; a\xrightarrow{n} < b\xrightarrow{n} \Leftrightarrow a < b \Leftrightarrow a\xleftarrow{m} < b\xleftarrow{m} \quad ; \quad a\xrightarrow{n} = b\xrightarrow{n} \Leftrightarrow a = b \Leftrightarrow a\xleftarrow{m} = b\xleftarrow{m},$$

$$\mathsf{S}_4: \; a\xrightarrow{n} = a \cdot 10^n \text{ für natürliche Zahlen } a.$$

Wer $\mathsf{S}_1 - \mathsf{S}_4$ nicht für offensichtlich hält, kann auch streng induktiv vorgehen und muss nur beachten, dass $\xrightarrow{1}$ eine Bijektion von $\mathbb{E}$ ist mit der Umkehroperation $\xleftarrow{1}$, und dass $\xrightarrow{n}$ die $n$-fache Wiederholung von $\xrightarrow{1}$ darstellt. $\mathsf{S}_1$ liefert $a\xrightarrow{n}\xrightarrow{m} = a\xrightarrow{n+m} = a\xrightarrow{m+n} = a\xrightarrow{m}\xrightarrow{n}$, und analog ist $a\xleftarrow{n}\xleftarrow{m} = a\xleftarrow{m}\xleftarrow{n}$. $\mathsf{S}_4$ ergibt sich aus der dezimalen Darstellung natürlicher Zahlen und liefert die nützliche, später wesentlich verallgemeinerte Gleichung

$$\mathsf{S}_5: \; (a+b)\xrightarrow{n} = a\xrightarrow{n} + b\xrightarrow{n} \text{ für natürliche Zahlen } a, b.$$

Es wird nun die Addition in $\mathbb{E}$ wie folgt erklärt:

**Definition.** Die Summe $a + b$ von $a, b \in \mathbb{E}$ sei erklärt durch $(a\xrightarrow{n} + b\xrightarrow{n})\xleftarrow{n}$, wobei $n$ eine beliebige gewählte gemeinsame Stellenzahl von $a$ und $b$ ist.

Danach ist z.B. $3{,}14 + 8{,}7 = (314 + 870)\xleftarrow{2} = 1184\xleftarrow{2} = 11{,}84$. Diese Definition ist nichts anderes als eine präzise Formulierung des üblichen Additionsverfahrens, das im Prinzip auch in den Chips elektronischer Rechner „verdrahtet“ ist. Der Dezimalpunkt bei Eingabe der Operanden bewirkt intern nur eine entsprechende Shift-Operation.

Die angegebene Definition hängt nur scheinbar von $n$ ab. Wählt man einmal $n$, ein andermal $n'$ in der geforderten Weise und ist etwa $n' = n + k$, so ergibt sich

$$\begin{aligned}
(a\overset{n'}{\rightarrow} + b\overset{n'}{\rightarrow})\overset{n'}{\leftarrow} &= (a\overset{n}{\rightarrow}\overset{k}{\rightarrow} + b\overset{n}{\rightarrow}\overset{k}{\rightarrow})\overset{k}{\leftarrow}\overset{n}{\leftarrow} && (\text{gemäß } \mathsf{S}_1)\\
&= (a\overset{n}{\rightarrow} + b\overset{n}{\rightarrow})\overset{k}{\rightarrow}\overset{k}{\leftarrow}\overset{n}{\leftarrow} && (\text{gemäß } \mathsf{S}_5,\ \text{weil } a\overset{n}{\rightarrow}, b\overset{n}{\rightarrow} \in \mathbb{N})\\
&= (a\overset{n}{\rightarrow} + b\overset{n}{\rightarrow})\overset{n}{\leftarrow} && (\text{weil } c\overset{k}{\rightarrow}\overset{k}{\leftarrow} = c \text{ gemäß } \mathsf{S}_2).
\end{aligned}$$

**Definition.** Das Produkt $a \cdot b$ von $a, b \in \mathbb{E}$ sei erklärt durch $(a\overset{n}{\rightarrow} \cdot b\overset{m}{\rightarrow})\overset{n+m}{\longleftarrow}$, wobei $n$ und $m$ Stellenzahlen von $a$ bzw. $b$ seien.

**Beispiele.** $2{,}31 \cdot 1{,}2 = (2{,}31\overset{2}{\rightarrow} \cdot 1{,}2\overset{1}{\rightarrow})\overset{3}{\leftarrow} = (231 \cdot 12)\overset{3}{\leftarrow} = 2772\overset{3}{\leftarrow} = 2{,}772$. Ferner ist $a \cdot 10^n = (a\overset{m}{\rightarrow} \cdot 10^n\overset{0}{\rightarrow})\overset{m}{\leftarrow} = (a\overset{m}{\rightarrow}\overset{n}{\rightarrow})\overset{m}{\leftarrow} = a\overset{n}{\rightarrow}$ für $m$-stelliges $a$, speziell $10\varepsilon_{n+1} = \varepsilon_n$. Analog ist $a \cdot \varepsilon_n = (a\overset{m}{\rightarrow} \cdot \varepsilon_n\overset{n}{\rightarrow})\overset{m+n}{\longleftarrow} = (a\overset{m}{\rightarrow} \cdot 1)\overset{m+n}{\longleftarrow} = a\overset{m}{\rightarrow}\overset{m+n}{\longleftarrow} = a\overset{n}{\leftarrow}$. Daher lässt sich zu jedem $a \in \mathbb{E}$ und jedem $b \in \mathbb{D}_+$ offenbar ein $n$ so wählen, dass $a\varepsilon_n \leq b$. Ferner ist $(ab)\overset{n}{\rightarrow} = a \cdot b \cdot 10^n = a(b\overset{n}{\rightarrow})$; es ist also gleichgültig, wie $ab\overset{n}{\rightarrow}$ gelesen wird.

Auch das Produkt hängt von der Wahl der Stellenzahlen von $a$ und $b$ nicht ab. Wichtig vor allem ist, dass die Einschränkungen der auf diese Weise erklärten Operationen auf den Bereich $\mathbb{N}$ mit den dort bereits gegebenen arithmetischen Operationen übereinstimmen. Denn man braucht in obigen Definitionen für den Fall $a, b \in \mathbb{N}$ ja nur $n = m = 0$ zu wählen. Die Operationen $+, \cdot$ auf $\mathbb{E}$ sind also Fortsetzungen derjenigen von $\mathbb{N}$. Deswegen dürfen sie auch mit den gleichen Symbolen bezeichnet werden.

**Bemerkung.** Es sei mit Nachdruck darauf hingewiesen, dass in den angegebenen Definitionen keine Willkür liegt. Für die Summe z.B. ergibt sich die Definition bis auf äquivalente Formulierung zwangsläufig aus ihrer Entsprechung zur Addition in Größenbereichen. Sie ist durch Dezimalteilung von Einheiten in Größenbereichen leicht zu veranschaulichen. Es genügt vorerst, sich dies anhand von anschaulichen Beispielen zu verdeutlichen, wie etwa in $3{,}14m + 2,7m = 314cm + 270cm = 584cm = 5{,}84m$. In **10.3** werden wir rigoros beweisen, dass obige Definition der Summe sozusagen eine zwingende ist.

Hat man sich einmal überzeugt, dass die Summe so und nicht anders als angegeben definiert werden kann, so lässt sich allein aufgrund der Rechengesetze schon beweisen, dass auch das Produkt bis auf äquivalente Formulierung so und nicht anders definiert werden kann als dies oben geschehen ist, Übung 4. Ein anderes und ausschließlich didaktisches Problem ist es, die Multiplikation sachgerecht zu *veranschaulichen*. Möglichkeiten hierzu ergeben sich nicht nur bei der Flächenberechung und in der elementaren Kombinatorik, sondern ebenso bei der Deutung der Dezimalzahlen als Maßzahlen (Gleichung (1°) in **10.3**), oder als Operatoren, bei welchen die Multiplikation als Abbildungsprodukt in Erscheinung tritt. Auf diese Interpretation gehen wir in **10.7** ein.

Wir werden nun den einfachen Nachweis erbringen, dass $\mathbb{E}$ mit den oben erklärten Operationen einen $\mathcal{E}$-Bereich bildet, also sämtliche in **2.1** angegebenen Axiome überprüfen. Danach darf man mit den endlichen Dezimalzahlen wie gewohnt rechnen. Natürlich ist damit nicht nur bloßes Ausrechnen gemeint, sondern das bewusste und zielorientierte Anwenden der bekannten Rechengesetze, zum Beispiel bei der Termumformung.

## 4.2 Nachweis der Rechengesetze

Die Eigenschaften $\mathsf{V}$, $\mathsf{F}$ und $\mathsf{G}$ sind offensichtlich. Es seien nun $a, b, c$ beliebige Elemente aus $\mathbb{E}$. Ferner sei $n$ gemeinsame Stellenzahl von $a, b, c$, so dass $a\xrightarrow{n}$, $b\xrightarrow{n}$, $c\xrightarrow{n}$ natürliche Zahlen sind. Dann ergeben sich die weiteren Rechengesetze in folgender Weise.

$\mathsf{N}^+$ : $\quad a + 0 = (a\xrightarrow{n} + 0\xrightarrow{n})\xleftarrow{n} = a\xrightarrow{n}\xleftarrow{n} = a.$

$\mathsf{N}^\times$ : $\quad a \cdot 1 = (a\xrightarrow{n} \cdot 1\xrightarrow{0})\xleftarrow{n} = a\xrightarrow{n}\xleftarrow{n} = a.$

$\mathsf{K}^+$ : $\quad a + b = (a\xrightarrow{n} + b\xrightarrow{n})\xleftarrow{n} = (b\xrightarrow{n} + a\xrightarrow{n})\xleftarrow{n} = b + a.$

Analog verläuft der Nachweis von Axiom $\mathsf{K}^\times$.

$\mathsf{A}^+$ :
$$\begin{aligned} a + (b + c) &= a + (b\xrightarrow{n} + c\xrightarrow{n})\xleftarrow{n} \\ &= (a\xrightarrow{n} + (b\xrightarrow{n} + c\xrightarrow{n})\xleftarrow{n}\xrightarrow{n})\xleftarrow{n} = (a\xrightarrow{n} + b\xrightarrow{n} + c\xrightarrow{n})\xleftarrow{n}. \end{aligned}$$

Derselbe Wert ergibt sich für $(a + b) + c$, und damit ist $\mathsf{A}^+$ bestätigt.

$\mathsf{A}^\times$ : $\quad a \cdot (b \cdot c) = (a\xrightarrow{n} \cdot (b\xrightarrow{n} \cdot c\xrightarrow{n})\xleftarrow{2n}\xrightarrow{2n})\xleftarrow{3n} = (a\xrightarrow{n} \cdot b\xrightarrow{n} \cdot c\xrightarrow{n})\xleftarrow{3n}.$

Derselbe Wert ergibt sich für $(a \cdot b) \cdot c$.

$\mathsf{D}$ :
$$\begin{aligned} (a + b) \cdot c &= (a\xrightarrow{n} + b\xrightarrow{n})\xleftarrow{n} \cdot c \\ &= \big((a\xrightarrow{n} + b\xrightarrow{n})\xleftarrow{n}\xrightarrow{n} \cdot c\xrightarrow{n}\big)\xleftarrow{2n} = \big((a\xrightarrow{n} + b\xrightarrow{n}) \cdot c\xrightarrow{n}\big)\xleftarrow{2n} \\ &= (a\xrightarrow{n} \cdot c\xrightarrow{n} + b\xrightarrow{n} \cdot c\xrightarrow{n})\xleftarrow{2n} && \text{(weil D in } \mathbb{N} \text{ gilt)} \\ &= (a\xrightarrow{n} \cdot c\xrightarrow{n})\xleftarrow{2n} + (b\xrightarrow{n} \cdot c\xrightarrow{n})\xleftarrow{2n} && \text{(Definition der Summe)} \\ &= a \cdot c + b \cdot c && \text{(Definition des Produkts).} \end{aligned}$$

Es verbleibt der Nachweis von Axiom $\mathsf{E}$. Für $c \in \mathbb{E}$ ist $c\xrightarrow{n} = c \cdot 10^n$. Also folgt mit $\mathsf{D}$

$$(1) \quad (a + b)\xrightarrow{n} = a\xrightarrow{n} + b\xrightarrow{n} \qquad (a, b \in \mathbb{E}).$$

Seien nun $a, b$ $n$-stellig und zuerst angenommen, $a + d = b$ mit $d \neq 0$. (1) ergibt $a\xrightarrow{n} + d\xrightarrow{n} = b\xrightarrow{n}$. Weil $a\xrightarrow{n}, b\xrightarrow{n} \in \mathbb{N}$, ist $d\xrightarrow{n} \in \mathbb{N}$, und weil mit $d \neq 0$ auch $d\xrightarrow{n} \neq 0$, ist $a\xrightarrow{n} < b\xrightarrow{n}$. Folglich ist $a < b$ gemäß $\mathsf{S}_3$. Sei nun das Letztere vorausgesetzt und $a\xrightarrow{n}, b\xrightarrow{n} \in \mathbb{N}$. Nach $\mathsf{S}_3$ ist $a\xrightarrow{n} < b\xrightarrow{n}$. Daher gibt es ein $c \in \mathbb{N}_+$ mit $a\xrightarrow{n} + c = b\xrightarrow{n}$. Für $d := c\xleftarrow{n}$ ergibt sich dann $a + d = (a\xrightarrow{n} + d\xrightarrow{n})\xleftarrow{n} = (a\xrightarrow{n} + c)\xleftarrow{n} = b\xrightarrow{n}\xleftarrow{n} = b$, womit auch die andere Richtung von $\mathsf{E}$ gezeigt ist.

Damit sind sämtliche Axiome nachgewiesen und $\mathbb{E}$ hat sich als ein Elementarbereich erwiesen, der im Gegensatz zu $\mathbb{N}$ dicht geordnet ist. Es gelten für $\mathbb{E}$ dann automatisch auch alle Folgerungen der Axiome, wie etwa die Monotoniegesetze. Eine einfache, kaum bemerkbare Anwendung von $\mathsf{M}^+$ liefert z.B. für jedes $n$ die folgende, im weiteren oft verwendete Ungleichung, deren linke Hälfte natürlich klar ist:

$$(2) \quad a_{.n} \leq a < a_{.n} + \varepsilon_n \qquad (a \in \mathbb{D}).$$

Denn sei $a = z_0{,}z_1z_2\cdots$ und $a_{.n} + \varepsilon_n = z'_0{,}z'_1\cdots z'_n$. Wegen $a_{.n} < a_{.n} + \varepsilon_n$ gibt es ein $k \leq n$ mit $z_k < z'_k$ und $z_i = z'_i$ für $i < k$. Also gilt $a < a_{.n} + \varepsilon_n$ nach Definition von $<$.

## 4.3 Division endlicher Dezimalzahlen

Die Division ist im Zahlbereich $\mathbb{E}$ i.a. nicht ausführbar wie sich an einfachen Beispielen verdeutlichen lässt. So ist zum Beispiel dort die Gleichung $3 \cdot x = 1$ unlösbar. Denn angenommen es gibt ein $z_0{,}z_1 \cdots z_n$ mit $3 \cdot z_0{,}z_1 \cdots z_n = (3 \cdot z_0 \cdots z_n)\overset{n}{\leftarrow} = 1$, oder gleichwertig $3 \cdot z_0 \cdots z_n = 10^n$, wobei offenbar $z_n \neq 0$ angenommen werden kann. Für keine von 0 verschiedene Ziffer $z_n$ endet aber $3 \cdot z_n$ auf 0 was notwendig der Fall sein müsste, wäre $3 \cdot z_0 \cdots z_n = 10^n$.

Falls die Gleichung $b \cdot x = a$ (oder gleichwertig $x \cdot b = a$) für $a, b \in \mathbb{E}$ mit $b \neq 0$ lösbar ist, so nur auf eine Weise; denn $bc_1 = bc_2 \;\Rightarrow\; c_1 = c_2$ nach der Kürzungsregel. Die Lösung wird mit $\frac{a}{b}$ bezeichnet und heißt der *Quotient von $a$ durch $b$*. In **6** wird bewiesen, dass für $a, b \in \mathbb{E}$ die Gleichung $b \cdot x = a$ mit $b \neq 0$ *in* $\mathbb{D}$ stets lösbar ist. Nur ist der Quotient $\frac{a}{b}$ dann in der Regel nicht abbrechend. Anders als in $\mathbb{N}$ ist in $\mathbb{E}$ für $a \in \mathbb{E}$ und jedes $n$ immerhin die Gleichung $10^n x = a$ lösbar, nämlich durch $x = a\overset{n}{\leftarrow}$; denn $10^n a\overset{n}{\leftarrow} = a\overset{n}{\rightarrow}\overset{n}{\leftarrow} = a$. Man kann also in $\mathbb{E}$ stets durch $10^n$ dividieren und das Resultat entsteht durch $n$-fache Kommaverschiebung nach links; kurzum, $\frac{a}{10^n} = a\overset{n}{\leftarrow} = a\varepsilon_n$. Ist $a$ speziell eine Dezimalziffer $z$, liefert dies $\frac{z}{10^n} = z\overset{n}{\leftarrow} = 0{,}\underbrace{0 \cdots 0z}_{n}$ T. Demnach gilt

$$z_0 + \tfrac{z_1}{10} + \ldots + \tfrac{z_n}{10^n} = z_0 + 0{,}z_1 + 0{,}0z_2 + \ldots + 0{,}0 \cdots 0z_n .$$

Die Ausrechnung des rechten Terms mit dem vertrauten Additionsverfahren für mehrere Summanden, das ja nur eine zweckmäßige Schreibweise unseres auf $n$ Summanden erweiterten Additionsalgorithmus für endliche Dezimalzahlen darstellt, ergibt

$$\begin{array}{l} \quad z_0 \\ + \; 0{,}z_1 \\ + \; 0{,}0z_2 \\ \quad \vdots \\ + \; 0{,}0 \cdots 0z_n \\ \hline = z_0{,}z_1 \cdots z_n . \end{array}$$

Damit haben wir die wichtige und wohlbekannte, in **3.1** schon erwähnte Gleichung

$$(3) \qquad z_0{,}z_1 \cdots z_n = z_0 + \tfrac{z_1}{10} + \ldots + \tfrac{z_n}{10^n}$$

bewiesen, die uns in Abschnitt **7** zu den unendlichen Reihen führen wird. Man möge sich vergegenwärtigen, dass eine Bruchrechnung für die Herleitung dieser Gleichung nicht erforderlich war, denn die Terme $\frac{z_n}{10^n}$ sind wohlbestimmte Dezimalzahlen [2)].

In $\mathbb{E}$ kann nicht nur durch 10, sondern auch durch 2, durch 5, und allgemein durch jede natürliche Zahl der Form $2^i \cdot 5^k$ dividiert werden, und i.a. nur durch diese, Übung 1. Z.B. ist $\frac{a}{2} = 5a\overset{1}{\leftarrow}$, weil $a = 2 \cdot 5 \cdot a\overset{1}{\leftarrow}$, und analog ist $\frac{a}{5} = 2a\overset{1}{\leftarrow}$. Die letzte Gleichung ist nützlich für das praktische Rechnen: um eine Zahl durch 5 zu teilen, muss diese nur verdoppelt und im Ergebnis das Komma um eine Stelle verschoben werden.

[2)] Wer besonders pingelig sein will, kann (3) auch induktiv über $n$ beweisen. Dafür benötigt man nur die leicht beweisbare Gleichung $z_0{,}z_1 \cdots z_n = z_0{,}z_1 \cdots z_{n-1} + z_n\overset{n}{\leftarrow} = z_0{,}z_1 \cdots z_{n-1} + \frac{z_n}{10^n}$ $(n > 0)$.

Ganz wie in $\mathbb{E}$ hat die Gleichung $b \cdot x = a$ für $b \neq 0$ wegen der Streichungsregel in jedem $\mathcal{E}$-Bereich nur höchstens eine Lösung, welche stets mit $\frac{a}{b}$ bezeichnet wird, sofern sie existiert. Weil zum Beispiel $(1-c)\cdot(1+c+\ldots+c^n) = 1-c^{n+1}$ für $0<c<1$ nach (3) in **2**, gilt in jedem $\mathcal{E}$-Bereich die wichtige Formel

$$(4) \quad \tfrac{1-c^{n+1}}{1-c} = 1 + c + c^2 + \ldots + c^n \quad (0 < c < 1).$$

Auch $2x = n(n-1)$, und allgemeiner $k!\cdot x = n(n-1)\ldots(n-k+1)$, ist für $k \leq n$ bereits in $\mathbb{N}$ lösbar, nämlich durch $x = \binom{n}{k}$ (Übung 5). Dabei heißt die durch durch $0! = 1$ und $(k+1)! = k! \cdot (k+1)$ definierte Funktion von $\mathbb{N}$ nach $\mathbb{N}$ die *Fakultät-Funktion*, kurz $k! = 1 \cdot 2 \cdot \ldots \cdot k$. Die oben behauptete Identität $k! \cdot \binom{n}{k} = n(n-1)\ldots(n-k+1)$ ist nach Definition des Quotienten gleichwertig mit der leichter merkbaren Formel

$$(5) \quad \binom{n}{k} = \tfrac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{1\cdot 2\cdot\ \ldots\ \cdot k}.$$

Der Zähler rechts in (5) ist genau genommen erklärt als $\prod_{i<k}(n-i)$. Damit erkennt man leichter, dass dieser auch für $k=0$ wohldefiniert ist und den Wert 1 hat. Die Anfangsbedingungen der folgenden, von einer gegebenen Termfolge $t_0, t_1, \ldots$ ausgehenden rekursiven Definitionen sind nämlich in der Mathematik überall die gleichen:

$$\textstyle\sum_{i<0} t_i = 0\ ,\ \sum_{i<k+1} t_i = \sum_{i<k} t_i + t_k \quad ; \quad \prod_{i<0} t_i = 1\ ,\ \prod_{i<k+1} t_i = \prod_{i<k} t_i \cdot t_k.$$

Im Sinne der Konvention $\prod_{i<0} t_i = 1$ ist übrigens dann auch $k! = \prod_{i=1}^{k} i = 1$ für $k = 0$.

## 4.4 Übungen

1. Man zeige, die Gleichung $n \cdot x = m$ ist für teilerfremde positive $m, n$ in $\mathbb{E}$ lösbar genau dann, wenn $n$ außer 2 und 5 keine weiteren Primteiler hat.

2. Seien $a, b \in \mathbb{E}$ $n$-stellig. Man beweise $a < b \ \Leftrightarrow\ a + \varepsilon_n \leq b$.

3. Sei $b \in \mathbb{D}$ und $c \in \mathbb{E}$ $n$-stellig. Man zeige: $c = b_{.n}$ genau dann, wenn $c \leq b < c+\varepsilon_n$. Für $n = 0$ bedeutet dies speziell $k = \operatorname{Int} b$ genau dann, wenn $k \leq b < k+1$.

4. Sei $a + b$ für $a, b \in \mathbb{E}$ wie im Text definiert. Man zeige unter alleiniger Benutzung der Rechengesetze für $\mathcal{E}$-Bereiche, dass notwendigerweise $a \cdot b = (a\xrightarrow{n} \cdot b\xrightarrow{m})\xleftarrow{n+m}$ für $n$-stelliges $a$ und $m$-stelliges $b$.

5. Man beweise (5). Demnach gilt speziell $1 + 2 + \ldots + n = \binom{n+1}{2} = \frac{n(n+1)}{2} \quad (n > 0)$.

6. Die $n$-te *Rundung* $a_{(n)}$ einer Dezimalzahl $a = z_0{,}z_1 z_2 \cdots \in \mathbb{D}$ ist wie folgt erklärt:
$$a_{(n)} = \begin{cases} a_{.n} & \text{falls } z_{n+1} \leq 4, \\ a_{.n} + \varepsilon_n & \text{falls } z_{n+1} \geq 5. \end{cases}$$
Man beweise $a_{(n)} = (a_{.n+1} + 5\varepsilon_{n+1})_{.n} \ \big(= (\operatorname{Int} a\xrightarrow{n+1} + 5)\xleftarrow{n+1}{}_{.n}\big)$.
Für die wohlbekannte Kreiszahl $\boldsymbol{\pi} = 3{,}141592653589793\cdots$ ergibt sich demnach
$$\boldsymbol{\pi}_{(2)} = (3141+5)\xleftarrow{3}{}_{.2} = 3146\xleftarrow{3}{}_{.2} = 3{,}14 \ \ ; \ \ \boldsymbol{\pi}_{(3)} = (31415+5)\xleftarrow{4}{}_{.3} = 3{,}142.$$

# Abschnitt 5

# Arithmetik der nichtnegativen reellen Zahlen

Angesichts der im letzten Abschnitt deutlich gewordenen Mängel des Zahlbereichs $\mathbb{E}$ hinsichtlich Division gehen wir nun dazu über, die arithmetischen Operationen auf $\mathbb{D}$ zu erweitern. Damit erreicht man nun mehr, als bei oberflächlicher Betrachtung zu erhoffen wagt. $\mathbb{D}$ wird nämlich zu einem $\mathcal{E}$-Bereich, in welchem nicht nur unbeschränkt dividiert werden kann – Division durch 0 natürlich ausgenommen – sondern darüber hinaus hat der Zahlbereich $\mathbb{D}$ wegen der Gültigkeit des Satzes von der oberen Grenze alle jene Eigenschaften, welche die reelle Analysis überhaupt erst ermöglichen. In diesem Sinne ist die im vorliegenden Abschnitt begründete reelle Arithmetik das Fundament, auf dem sich das Riesengebäude der Analysis erhebt. Zwar fehlen zunächst noch die negativen Zahlen, aber dies ist unwesentlich; mehr noch, es macht die einfacheren mit dem Grenzwert zusammenhängenden Betrachtungen übersichtlicher. Was für die Definitionen der arithmetischen Operationen benötigt wird, ist übrigens sehr bescheiden. Lediglich die Grenzwerte monotoner beschränkter Folgen sind zu betrachten, welche wir der Kürze halber *schlichte* Folgen nennen werden.

Anschaulich ist klar, dass die Summe $a + b$ reeller Dezimalzahlen $a, b$ durch die Summe $a_{.n} + b_{.n}$ um so besser approximiert wird, je größer $n$ ist. In genau dieser Weise – und entsprechende Bemerkungen beziehen sich auf das Produkt – wurde die Summe reeller Zahlen erstmals in der Praxis definiert, zum Beispiel bei der Erstellung der Logarithmentafeln zu Beginn des 17. Jahrhunderts. Das geschah in einem aus heutiger Sicht blindem Glauben an die Rechtmäßigkeit dieses Verfahrens. Tatsächlich sind ja diese Verhältnisse anschaulich so klar, dass die Frage berechtigt ist, ob es hier überhaupt einer Rechtfertigung bedarf. Diese Frage muss allerdings klar bejaht werden. Denn es geht hier nicht um die Ausführbarkeit irgendwelcher Operationen, sondern um den Nachweis gewisser, meistens stillschweigend benutzter Rechengesetze. Diese Gesetze wurden von den Rechenexperten des 17. Jahrhunderts bei der Berechnung von Logarithmen- und anderen Tafeln gewissermaßen nur experimentell überprüft.

## 5.1 Schlichte Folgen

Wir betrachten zuerst Folgen $\langle a_n\rangle$ aus einer beliebigen geordneten Menge $M$. Eine solche Folge $M$ heißt *(monoton) wachsend*, falls stets $a_n \leq a_{n+1}$, *(monoton) fallend*, falls stets $a_n \geq a_{n+1}$, und *beschränkt*, wenn $a_n \,|\, n \in \mathbb{N}$ in $M$ beschränkt ist.

**Definition.** Eine wachsende beschränkte Folge $\langle a_n\rangle$ heiße eine *schlichte* Folge. Falls $a = \{\sup a_n \,|\, n \in \mathbb{N}\}$ existiert, heiße $a$ der *Grenzwert* oder *Limes* von $\langle a_n\rangle$, symbolisch $a = \lim_{n\to\infty} a_n$, oder $a = \lim\langle a_n\rangle$. Fällt $\langle a_n\rangle$ monoton und existiert $b = \{\inf a_n \,|\, n \in \mathbb{N}\}$, so heiße $b$ der Grenzwert von $\langle a_n\rangle$. In beiden Fällen sagt man, $\langle a_n\rangle$ ist *konvergent* oder *konvergiert* oder *strebt* gegen ihren Grenzwert.

In diesem Abschnitt betrachten wir nur schlichte Folgen aus $\mathbb{D}$. Derartige Folgen sind nach Satz 3.2 stets konvergent, und nach Übung 3.4 ebenso beliebige fallende Folgen aus $\mathbb{D}$. Zum Beispiel ist für $a \in \mathbb{D}$ die Näherungsfolge $\langle a_{.n}\rangle$ schlicht, denn sie wächst und ist durch $a$ beschränkt. Daher ist definitionsgemäß $\lim\langle a_{.n}\rangle = a$. Falls $\langle a_n\rangle$ ab einer gewissen Stelle $k$ konstant ist, d.h. falls $a_n = a_k$ für alle $n \geq k$, so ist $\lim\langle a_n\rangle$ natürlich mit $a_k$ identisch. Dieser Fall liegt für die Folge $\langle a_{.n}\rangle$ genau dann vor, wenn $a \in \mathbb{E}$.

Konvergenz im Sinne der obigen Definition wird beweisbar, wenn man von der folgenden, etwas allgemeineren Definition ausgeht, welche keine Ordnung, dafür aber das Vorhandensein einer geeigneten Abstandsfunktion $|\ |: M \times M \to \mathbb{D}$ voraussetzt. Diese haben wir zwar in jedem $\mathcal{E}$-Bereich (Übung 3.2), aber wir wollen ja erst zeigen, dass $\mathbb{D}$ ein solcher ist! Deswegen benutzen wir vorerst nur die anfängliche Definition. Die später verwendete, statt auf einer Ordnung auf einer Abstandsfunktion beruhende Definition lautet: $\langle a_n\rangle$ *konvergiert* und es sei $\lim\langle a_n\rangle = a$, falls es zu jedem $\varepsilon \in \mathbb{D}_+$ ein $n$ gibt mit $|a - a_m| < \varepsilon$ für alle $m \geq n$.

Alle Begriffe übertragen sich sinngemäß auf Folgen $\langle a_n\rangle_{n\geq k}$. Mit $\langle a_n\rangle$ ist offenbar auch $\langle a_n\rangle_{n\geq k}$ schlicht, und weil $\sup\{a_n \,|\, n \in \mathbb{N}\} = \sup\{a_n \,|\, n \geq k\}$, gilt $\lim\langle a_n\rangle = \lim\langle a_n\rangle_{n\geq k}$ für jedes $k$. Kurzum, man darf von einer schlichten Folge ein beliebiges „Anfangsstück“ weglassen und erhält eine schlichte Folge mit demselben Grenzwert. Dass für schlichte Folgen aus $\mathbb{D}$ der Grenzwert mit dem Supremum übereinstimmt, ist der Grund, warum diese Folgen in der Regel einfacher zu beherrschen sind.

## 5.2 Erweiterung der Rechenoperationen

Im folgenden werden wir zunächst den Fall schlichter Folgen vor uns haben, deren sämtliche Glieder endliche Dezimalzahlen sind. Sind $\langle a_n\rangle$ und $\langle b_n\rangle$ zwei derartige Folgen, so können auch die Folgen $\langle a_n + b_n\rangle$ und $\langle a_n \cdot b_n\rangle$ betrachtet werden. Es ist wegen der Monotoniegesetze offensichtlich, dass es sich hierbei wieder um schlichte Folgen handelt, die in $\mathbb{D}$ demnach wohlbestimmte Grenzwerte haben.

**Definition.** Die Summe $a + b$ reeller Zahlen $a, b \in \mathbb{D}$ sei erklärt als $\lim\langle a_{.n} + b_{.n}\rangle$, das Produkt $a \cdot b$ als $\lim\langle a_{.n} \cdot b_{.n}\rangle$.

So ist $3 \cdot 0{,}33\cdots = \lim\langle 3 \cdot 0,\underbrace{3\cdots 3}_{n}\rangle = \lim\langle 0,\underbrace{9\cdots 9}_{n}\rangle = 1$. Also löst $x = 0{,}333\cdots$ die Gleichung $3 \cdot x = 1$. Nach der in **4.3** eingeführten Bezeichnung ist also $\frac{1}{3} = 0{,}333\cdots$. Aus der Definition ergeben sich ferner sehr einfach die Formeln $a\xrightarrow{n} = a \cdot 10^n$ und $a\xleftarrow{n} = \frac{a}{10^n}$ (Übung 1). Genau wie in $\mathbb{E}$ bedeutet also auch in $\mathbb{D}$ Kommaverschiebung dasselbe wie Multiplikation bzw. Division mit einer entsprechenden Potenz von 10.

Vor weiteren Betrachtungen haben wir uns zunächst erst einmal zu überzeugen, dass Summe und Produkt – eingeschränkt auf $\mathbb{E}$ – mit den dort bereits gegebenen Operationen übereinstimmen. Dies aber ist leicht erkennbar. Sei $k$ so groß, dass $a = a_{.n}$ und $b = b_{.n}$ für alle $n \geq k$. Dann ist die Folge $\langle a_{.n} + b_{.n}\rangle$ von der Stelle $k$ ab konstant; also $\lim\langle a_{.n} + b_{.n}\rangle = a_{.k} + b_{.k} = a + b$, wo die rechts stehende Summe diejenige im alten Sinne ist. Das zeigt die Übereinstimmung. Dieselbe Überlegung gilt auch für das Produkt.

Für den Nachweis einiger Rechenregeln verwenden wir vorübergehend folgendes nützliche Kriterium. Darin ist $\varepsilon_n$ die in **3.2** definierte $n$-stellige Dezimalzahl $0{,}0\cdots 01$.

**Kriterium.** *Sei $c \in \mathbb{E}_+$ und seien $\langle a_n\rangle$, $\langle b_n\rangle$ schlichte Folgen endlicher Dezimalzahlen mit $\lim\langle a_n\rangle = a$ und $\lim\langle b_n\rangle = b$. Dann sind äquivalent*

(i) $a \leq b$ ; (ii) *zu jedem $n$ gibt es ein $m$ mit $a_n \leq b_m + c\varepsilon_n$.*

**Beweis.** Sei (i) erfüllt und $n$ vorgegeben. Entweder ist $a_n \leq c\varepsilon_n$ oder aber $c\varepsilon_n < a_n$. Im ersten Falle gilt natürlich $a_n \leq b_m + c\varepsilon_n$, sogar für beliebiges $m$. Im zweiten Falle ist $a_n - c\varepsilon_n < a_n \leq a \leq b$. Daher ist $a_n - c\varepsilon_n \leq b_m$ für ein gewisses $m$, also $a_n \leq b_m + c\varepsilon_n$ wie verlangt. Umgekehrt sei (ii) erfüllt und angenommen, dass $b < a$, also $b < a_k$ für ein gewisses $k$. Sei $d \in \mathbb{E}$ so gewählt, dass $b < d < a_k$. Für ein hinreichend großes $n \geq k$ ist dann $c\varepsilon_n \leq a_k - d$, also $d + c\varepsilon_n \leq a_k \leq a_n$. Wegen $b_m < d$ ergibt dies $b_m + c\varepsilon_n < d + c\varepsilon_n \leq a_n$ für alle $m$, was (ii) widerspricht. ■

**Satz 5.1.** *Sind $\langle a_n\rangle, \langle b_n\rangle$ schlichte Folgen endlicher Dezimalzahlen, so gelten*

(1) $\lim\langle a_n + b_n\rangle = \lim\langle a_n\rangle + \lim\langle b_n\rangle$ ; (2) $\lim\langle a_n \cdot b_n\rangle = \lim\langle a_n\rangle \cdot \lim\langle b_n\rangle$.

**Beweis.** Sei $a := \lim\langle a_n\rangle$, $b := \lim\langle b_n\rangle$, also $\lim\langle a_n\rangle = \lim\langle a_{.n}\rangle$, $\lim\langle b_n\rangle = \lim\langle b_{.n}\rangle$. Es ist $\lim\langle a_n + b_n\rangle = \lim\langle a_{.n} + b_{.n}\rangle$ zu beweisen. Nach dem Kriterium (angewandt mit $c = 1$) gibt es zu jedem $n$ ein $m$ mit $a_n \leq a_{.m} + \varepsilon_n$ und $b_n \leq b_{.m} + \varepsilon_n$ (o.B.d.A. kann hier dasselbe $m$ für beide Folgen genommen werden). Daher ist $a_n + b_n \leq a_{.m} + b_{.m} + 2\varepsilon_n$ für alle $n$, und abermalige Anwendung des Kriteriums ergibt $\lim\langle a_n + b_n\rangle \leq \lim\langle a_{.n} + b_{.n}\rangle$. Ganz analog beweist man $\lim\langle a_{.n} + b_{.n}\rangle \leq \lim\langle a_n + b_n\rangle$ und (1) ist bewiesen. Zum Beweis von (2) beachte man, $a_n \leq a_{.m} + \varepsilon_n$ und $b_n \leq b_{.m} + \varepsilon_n$ ergeben

$$\begin{aligned} a_n b_n &\leq (a_{.m} + \varepsilon_n)(b_{.m} + \varepsilon_n) \\ &= a_{.m}b_{.m} + (a_{.m} + b_{.m} + \varepsilon_n)\varepsilon_n \leq a_{.m}b_{.m} + (r + s + 1)\varepsilon_n, \end{aligned}$$

wobei $r, s \in \mathbb{E}$ mit $a \leq r$, $b \leq s$ beliebig gewählt seien. Nach dem Kriterium – man setze dort $c = r + s + 1$ – ist mithin $\lim\langle a_n \cdot b_n\rangle \leq \lim\langle a_{.n} \cdot b_{.n}\rangle$. Völlig analog zeigt man $\lim\langle a_{.n} \cdot b_{.n}\rangle \leq \lim\langle a_n \cdot b_n\rangle$, womit auch (2) bewiesen ist. ■

## 5.3 Nachweis der Rechengesetze

Nunmehr werden wir die recht einfache Aufgabe erledigen, auch für $\mathbb{D}$ alle Axiome für $\mathcal{E}$-Bereiche aus **2.1** nachzuweisen. $\mathsf{V}$ und $\mathsf{G}$ sind trivial. Axiom $\mathsf{F}$ folgt daraus, dass für $a, b \neq 0$ sicher $a_{.n}, b_{.n} > 0$ für ein $n$, also $0 < a_{.n} \cdot b_{.n} \leq \lim\langle a_{.n} \cdot b_{.n}\rangle = a \cdot b$.

$\mathsf{N}^+$: Weil $0_{.n} = 0$ für alle $n$, ist $a + 0 = \lim\langle a_{.n} + 0\rangle = \lim\langle a_{.n}\rangle = a$. Völlig analog beweist man $\mathsf{N}^\times$. Auch sind die Kommutativgesetze unmittelbar klar.

$\mathsf{A}^+$: $a+(b+c) = \lim\langle a_{.n}\rangle + \lim\langle b_{.n} + c_{.n}\rangle = \lim\langle a_{.n} + b_{.n} + c_{.n}\rangle$ nach Satz 5.1. Derselbe Wert ergibt sich für $(a + b) + c$. Ganz analog beweist man auch $\mathsf{A}^\times$.

$\mathsf{D}$: $(a + b) \cdot c = \lim\langle a_{.n} + b_{.n}\rangle \cdot \lim\langle c_{.n}\rangle = \lim\langle (a_{.n} + b_{.n}) \cdot c_{.n}\rangle = \lim\langle a_{.n} \cdot c_{.n} + b_{.n} \cdot c_{.n}\rangle$. Derselbe Wert ergibt sich erwartungsgemäß für $a \cdot c + b \cdot c$.

$\mathsf{E}$: Sei $a + d = b$ mit $d \neq 0$ und $n$ so gewählt, dass $d_{.n} \neq 0$, also $\varepsilon_n \leq d_{.n}$. Dann ist

$$a < a_{.n} + \varepsilon_n \leq a_{.n} + d_{.n} \leq a + d = b.$$

Also $a < b$. Etwas subtiler ist die andere Richtung von $\mathsf{E}$, d.h. der Existenzbeweis von $b - a$ für $a < b$. Zwar konvergiert $\langle b_{.n} - a_{.n}\rangle$ tatsächlich gegen $b - a$, diese Folge ist aber i.a. nicht monoton. Hier hilft nun die Einbeziehung einer anderen Näherung für $a$, nämlich $a^{.n} := a_{.n} + \varepsilon_n$. Sicherlich ist $a_{.n+1} = a_{.n} + z^a_{n+1}\varepsilon_{n+1}$. Damit erhalten wir $a_{.n+1} + \varepsilon_{n+1} \leq a_{.n} + (z^a_{n+1} + 1)\varepsilon_{n+1} \leq a_{.n} + 10\varepsilon_{n+1} = a_{.n} + \varepsilon_n$. Kurzum, $\langle a^{.n}\rangle$ fällt monoton. Sei nun $a < b$, also $a_{.n} \leq b_{.n}$ für alle $n$. Wir betrachten die Folge $\langle d_n\rangle$ mit

$$d_n = \begin{cases} 0 & \text{solange} \quad a_{.n} = b_{.n}, \\ b_{.n} - a^{.n} & \text{falls} \quad a_{.n} < b_{.n}. \end{cases}$$

$d_n$ ist wohldefiniert, weil $a_{.n} < b_{.n} \;\Leftrightarrow\; a_{.n} + \varepsilon_n \leq b_{.n} \;\Leftrightarrow\; a^{.n} \leq b_{.n}$, Übung 4.2. Wir zeigen nun, $\langle d_n\rangle$ ist beschränkt, wächst monoton, und $a + d = b$, mit $d := \lim\langle d_n\rangle$.

Ersteres ist klar, weil $d_n \leq b_{.n} \leq b$. Auch ist $d_n \leq d_{n+1}$, denn wegen $a^{.n} \geq a^{.n+1}$ ist $b_{.n} - a^{.n} \leq b_{.n+1} - a^{.n} \leq b_{.n+1} - a^{.n+1}$, solange $a^{.n} \leq b_{.n}$ (d.h. $a_{.n} < b_{.n}$). Sei nun $k$ nun minimal mit $a_{.k} < b_{.k}$. Nach Satz 5.1 ist $a + d = \lim\langle a_{.n}\rangle + \lim\langle d_n\rangle = \lim\langle a_{.n} + d_n\rangle$. Daher genügt zu zeigen $\lim\langle a_{.n} + d_n\rangle = b$. Für $n \geq k$ ist offenbar

$$a_{.n} + d_n = a_{.n} + b_{.n} - a^{.n} = b_{.n} - \varepsilon_n \leq b_{.n},$$

aber $a_{.n} + d_n \leq b_{.n}$ gilt auch für $n < k$, weil dann $a_{.n} = b_{.n}$ und $d_n = 0$. Folglich ist $a + d = \lim\langle a_{.n} + d_n\rangle \leq \lim\langle b_{.n}\rangle = b$. Daher ist nur noch $\lim\langle b_{.n}\rangle \leq \lim\langle a_{.n} + d_n\rangle$ zu zeigen. Für $n \geq k$ ist $b_{.n} = a^{.n} + d_n = a_{.n} + d_n + \varepsilon_n$, erst recht $b_{.n} \leq (a_{.n} + d_n) + \varepsilon_n$, und dies gilt natürlich auch für $n < k$. Daher ist Bedingung (ii) des Kriteriums mit $m = n$ und $c = 1$ erfüllt. Folglich ist $b \leq a + d$, und alles ist bewiesen.

Unser wesentliches Ziel ist hiermit erreicht. $\mathbb{D}$ hat sich mit der oben erklärten Addition und Multiplikation als ein $\mathcal{E}$-Bereich erwiesen, welcher den Satz von der oberen Grenze erfüllt. Anders formuliert: $\mathbb{D}$ ist ein lückenloser (Satz 3.3) und damit auch archimedischer $\mathcal{E}$-Bereich (Übung 4 in **3.5**). In **10** wird zudem bewiesen, dass es bis auf Isomorphie nur einen solchen Bereich gibt. Damit haben wir volles Recht, die Elemente von $\mathbb{D}$ als reelle

Zahlen zu bezeichnen. Mit diesen darf man von nun an wie gewohnt rechnen, unter Beachtung der vorläufig noch gültigen Einschränkungen hinsichtlich Subtraktion. Zum Beispiel folgt $a - \mathrm{Int}\, a < 1$ aus $a < \mathrm{Int}\, a + 1$ nur deshalb, weil der linke Term wegen $a \geq \mathrm{Int}\, a$ wohldefiniert ist. Auf die Definiertheit solcher Terme wird künftig nur dann explizit hingewiesen, wenn dies (wie in Abschnitt **9**) wesentlich ist.

## 5.4 Rechnen mit Näherungen

Bekanntlich operieren große und kleine Rechner mit endlichen Dezimalzahlen (eigentlich Dualzahlen, aber das ist hier weniger wichtig). Die hieraus resultierenden Probleme der Resultatsverfälschung bei umfangreichen numerischen Rechnungen sind theoretisch nur schwer zu beherrschen und erfordern eine sorgfältige Planung und Fehleranalyse. Es ist aus diesem und anderen Gründen unerlässlich, eine genauere Vorstellung von den Beziehungen der arithmetischen Operationen im Bereich reeller Zahlen und ihrer näherungsweisen Realisierung durch entsprechende Operationen in $\mathbb{E}$ zu besitzen.

In der Regel wird mit Rundungen einer durch die Hardware des Rechners festgelegten Stellenzahl $n$ gerechnet. Wir erörtern dies nicht in Einzelheiten, sondern interessieren uns hier nur für die „sicheren“ Kommastellen des Ergebnisses bei der Ausführung der elementaren Rechenoperationen. Gewiss ist $a_{.n} \leq a < a_{.n} + \varepsilon_n$ nach Formel (2) in **4.2**. Also ist $a_{.n} + b_{.n} \leq a + b < a_{.n} + b_{.n} + 2\varepsilon_n$. Es gilt gemäß Übung 5 aber eine etwas bessere Ungleichung für die $n$-te Näherung der Summe, nämlich

$$(3) \quad a_{.n} + b_{.n} \leq (a+b)_{.n} \leq a_{.n} + b_{.n} + \varepsilon_n,$$

was für $n = 0$ die Ungleichung $\mathrm{Int}\, a + \mathrm{Int}\, b \leq \mathrm{Int}\,(a+b) < \mathrm{Int}\, a + \mathrm{Int}\, b + 1$ umfasst. $(a+b)_{.n}$ wird also durch die Summe der $n$-ten Näherungen von $a, b$ mit einem minimalen Fehler in der letzten Dezimalen angegeben. Für $a = 1{,}213\cdots$ und $b = 2{,}381\cdots$ z.B. hat $(a+b)_{.3}$ gemäß (3) den Wert 3,594 oder 3,595. Die ersten zwei Dezimalen sind hier absolut gesichert, und die letzte hat eine minimale Unsicherheit. Ungeachtet dessen kann es vorkommen, dass für $a = z_0{,}z_1z_2\cdots$ und $b = z'_0{,}z'_1z'_2\cdots$ nicht einmal $\mathrm{Int}\,(a+b)$ exakt bestimmt werden kann, obwohl $a$ und $b$ ziffernweise berechenbar sind. Denn nach Übung 4 müssen zur Bestimmung von $\mathrm{Int}\,(a+b)$ die $z_i$ und $z'_i$ im Falle $z_i + z'_i = 9$ für $i = 1, 2, \ldots$ solange berechnet werden, bis erstmals entweder $z_n + z'_n < 9$, oder aber $z_n + z'_n > 9$, oder es ist beständig $z_i + z'_i = 9$. Im ersten Fall ist $\mathrm{Int}\,(a+b) = \mathrm{Int}\, a + \mathrm{Int}\, b$, in den beiden letzten Fällen ist $\mathrm{Int}\,(a+b) = \mathrm{Int}\, a + \mathrm{Int}\, b + 1$. Offenbar ist die Entscheidung, welcher Fall vorliegt, unter Umständen nicht zu treffen.

Ähnlich lässt sich leicht bestätigen, dass $a - b$ für $a > b$ und $a_{.n} > b_{.n}$ im Intervall mit den Grenzen $a_{.n} - b_{.n} - \varepsilon_n$ und $a_{.n} - b_{.n} + \varepsilon_n$ liegt. Bezüglich des Produkts ergibt sich

$$a_{.n}b_{.n} \leq ab \leq a_{.n}b_{.n} + \varepsilon_n(a_{.n} + b_{.n} + \varepsilon_n) < a_{.n}b_{.n} + \varepsilon_n(a+b) + \varepsilon_n^2.$$

Hieraus ersieht man, dass der Fehler in der Approximation von $ab$ durch $a_{.n}b_{.n}$ sehr groß werden kann. Will man $ab$ auf $m$ Kommastellen genau, so muss $n \geq m$ so gewählt werden, dass $\varepsilon_n(a+b) < \varepsilon_m$, also $a+b < 10^{n-m}$. Dies ergibt $\lg(a+b) < n-m$, oder $n > m + \lg(a+b)$; dabei wurde das quadratische Glied $\varepsilon_n^2$ vernachlässigt, weil es sich in der Regel erst in der $2n$-ten Dezimalen auswirkt. Die hier ausnahmsweise im voraus benutzte logarithmische Funktion lg wird in **7.4** behandelt.

**Beispiel.** Sei $a = 33333{,}3333\,3333\cdots$, $b = 3{,}6996\,3369\cdots$, und es soll $a \cdot b$ auf 2 Kommastellen genau ausgerechnet werden. Dazu benötigt man sogar die 7-ten Näherungen. Es ist nämlich $2 + \lg 33337{,}03\cdots = 6{,}5\cdots$. Rechnet man etwa nur mit $a_{.2}$ und $b_{.2}$, so erhält man $a_{.2} \cdot b_{.2} = 122999{,}9877$, während tatsächlich $a \cdot b = 123321{,}12\cdots$. Der (absolute) Fehler $ab - a_{.2}b_{.2}$ ist also größer als 321.

Diese hier nur angedeuteten Probleme des näherungsweisen Rechnens werden auch nicht aus der Welt geschafft, wenn mit den $n$-ten Rundungen statt mit den $n$-ten kanonischen Näherungen gerechnet wird. Eine genaue Kontrolle gewinnt man z.B. durch die so genannte Intervallarithmetik, auf die wir hier aber nicht eingehen können.

Meistens betrachtet man nicht den *absoluten Fehler* $|a' - a|$ einer Näherung $a'$ für $a$, sondern den *relativen Fehler* $r = |\frac{a'-a}{a}|$, der häufig in Prozenten angegeben wird. In obigem Beispiel ist dieser mit etwa 0,26% zwar noch recht klein, aber bei Tausenden von Multiplikationen in der Ausführung eines Rechenprogramms können sich auch diese Fehler erheblich summieren. Siehe etwa [38] für diesen Themenkreis.

## 5.5 Übungen

1. Man beweise $a \cdot 10^n = a^{\xrightarrow{n}}$ und $a \cdot \varepsilon_n = a^{\xleftarrow{n}}$ für *alle* $a \in \mathbb{D}$ und schließe daraus $(a+b)^{\xrightarrow{n}} = a^{\xrightarrow{n}} + b^{\xrightarrow{n}}$. Demnach gilt $a + b = (a^{\xrightarrow{n}} + b^{\xrightarrow{n}})^{\xleftarrow{n}}$ für beliebige $a, b \in \mathbb{D}$.

2. Eine Folge $\langle x_n \rangle$ aus $\mathbb{D}$ heiße eine *Nullfolge*, genauer eine *fallende Nullfolge*, wenn $x_0 \geq x_1 \geq \ldots$ und wenn es zu jedem $k$ ein $m$ gibt mit $x_m < \varepsilon_k$ (womit wegen der fallenden Monotonie dann $x_n < \varepsilon_k$ für alle $n \geq m$). Seien nun $a_n, a \in \mathbb{D}$ und sei die Folge $\langle a_n \rangle$ monoton wachsend. Man beweise die Äquivalenz von
   (i) $\langle a_n \rangle$ ist konvergent und $\lim \langle a_n \rangle = a$,
   (ii) $a_n \leq a$ für alle n und $\langle a - a_n \rangle$ ist eine Nullfolge.

3. Man beweise: Satz 5.1 gilt für beliebige schlichte Folgen $\langle a_n \rangle, \langle b_n \rangle$ aus $\mathbb{D}$.

4. Sei $a = z_0{,}z_1z_2\cdots$ und $b = z_0'{,}z_1'z_2'\cdots$, also $\operatorname{Int} a = z_0$, $\operatorname{Int} b = z_0'$. Man zeige
$$\operatorname{Int}(a+b) = \begin{cases} \operatorname{Int} a + \operatorname{Int} b, & \text{falls es ein } k > 0 \text{ gibt mit } z_k + z_k' < 9 \text{ und} \\ & z_i + z_i' = 9 \text{ für alle } i \text{ mit } 0 < i < k, \\ \operatorname{Int} a + \operatorname{Int} b + 1 & \text{sonst.} \end{cases}$$
   Demnach gilt im allgemeinen Falle nur $\operatorname{Int}(a+b) \geq \operatorname{Int} a + \operatorname{Int} b$.

5. Man beweise die Ungleichung $a_{.n} + b_{.n} \leq (a+b)_{.n} \leq a_{.n} + b_{.n} + \varepsilon_n$.

# Abschnitt 6

# Division und rationale Zahlen

Wir werden als erstes zeigen, dass die Division in $\mathbb{D}$ stets ausführbar ist, wenn nur der Divisor nicht Null ist. Genauer, die Gleichung $b \cdot x = a$ ist stets lösbar für $b \neq 0$, und wegen der Kürzungsregel dann auch eindeutig lösbar. Die Division ist, ähnlich wie die Subtraktion, eine nachträglich eingeführte partiell definierte Operation. Die für $b \neq 0$ eindeutig bestimmte Lösung von $b \cdot x = a$ heißt der *Quotient* von $a$ und $b$ und wird mit $\frac{a}{b}$ bezeichnet. Die Namen „Zähler" und „Nenner" für $a$ bzw. $b$ haben hier mit zählen und nennen gar nichts zu tun und sind insofern nur Relikte der Vergangenheit.

Aus der Definition des Quotienten gewinnt man leicht die Regeln der so genannten Bruchrechnung, die das Dividieren in den bisherigen Rechenkalkül einbeziehen. Es wäre verschwendete Mühe, diese für Quotienten natürlicher Zahlen extra zu formulieren, denn sie gelten für beliebige reelle Zähler und Nenner, und allgemeiner, für Elemente eines beliebigen $\mathcal{E}$-Bereichs mit ausführbarer Division, und sind für Quotienten aus natürlichen Zahlen auch nicht etwa einfacher zu beweisen. Den Algorithmus zur expliziten Berechnung von $\frac{a}{b}$ für reelle $a, b$ behandeln wir in **8.1**.

Die Quotienten $\frac{m}{n}$ für $m, n \in \mathbb{N}$ heißen *rationale* Zahlen. Deren Gesamtheit werde mit $\mathbb{Q}$ bezeichnet. In **9** ändern wir die Bezeichnung. Dann wird $\mathbb{Q}$ auch alle negativen rationalen Zahlen umfassen. Stets ist $\mathbb{Q}_+$ die Menge der positiven Elemente von $\mathbb{Q}$. Sowohl rationale als auch *irrationale* Zahlen (Elemente von $\mathbb{D}\backslash\mathbb{Q}$) sind im Sinne unserer Auffassung wohlbestimmte Dezimalzahlen. Auch ein elektronischer Rechner ignoriert die Einteilung in rationale und irrationale Zahlen, es sei denn, er werde speziell programmiert. Sonst erscheint in der Anzeige oder auf dem Bildschirm im Falle einer Division von $m$ durch $n$ immer nur das, was eine Bruchzahl gemäß unserer Auffassung ist, nämlich eine Dezimalzahl; genauer, es erscheint eine gewisse Rundung dieser Zahl.

Jede endliche Dezimalzahl $a$ ist rational; denn hat $a$ die Stellenzahl $n$, ist $10^n a = a\overset{n}{\rightarrow}$ eine natürliche Zahl, also $a = \frac{a\overset{n}{\rightarrow}}{10^n} \in \mathbb{Q}$. Aber es gibt sicher nichtabbrechende rationale Zahlen, z.B. $\frac{1}{3} = 0{,}333\cdots$, siehe **5.2**. Anderseits sind gewiss nicht alle Dezimalzahlen rational. Denn nach **6.2** ist jede rationale Dezimalzahl periodisch. Wir werfen schließlich einen kurzen Blick auf so genannte Stammbruchdarstellungen rationaler Zahlen, einer von vielen interessanten Anwendungen des Bruchrechnens in diesem Abschnitt.

## 6.1 Division und Bruchrechnung

Durch Anwendung des Satzes 3.2 von der obereren Grenze erhalten wir

**Satz 6.1.** *Für alle $a, b \in \mathbb{D}$ mit $b \neq 0$ gibt es genau ein $c \in \mathbb{D}$ mit $b \cdot c = a$.*

**Beweis.** Den Existenzbeweis kann man auf den Fall $a = 1$ beschränken. Denn ist ein $s \in \mathbb{D}$ mit $b \cdot s = 1$ konstruiert, so gilt für $c := sa$ dann $bc = bsa = 1a = a$. Sei also $b \neq 0$. Dann ist $X := \{x \in \mathbb{D} \mid bx \leq 1\}$ beschränkt. Denn ist $\varepsilon_m < b$ und $x \in X$, so folgt $\varepsilon_m x \leq bx \leq 1 = \varepsilon_m \cdot 10^m$, also $x \leq 10^m$. Sei $s := \sup X$. Wir beweisen $bs = 1$. Angenommen $bs < 1$. Wählt man $n$ so groß,dass $b\varepsilon_n = b\overset{n}{\leftarrow} \leq 1 - bs$, ist $b(s + \varepsilon_n) \leq 1$. Also wäre $s + \varepsilon_n \in X$, was unmöglich ist. Aber auch $bs > 1$ ist ausgeschlossen. Denn ist $b\varepsilon_n \leq bs - 1$ und $\varepsilon_n < s$, so folgt $bx \leq 1 \leq b(s - \varepsilon_n)$ für $x \in X$, also $x \leq s - \varepsilon_n$. Damit wäre bereits $s - \varepsilon_n$ obere Schranke für $X$, was nicht angeht. Also ist $bs = 1$. Die Eindeutigkeitsaussage folgt mit der Kürzungsregel aus $bc_1 = bc_2 \Rightarrow c_1 = c_2$. ∎

Diese Beweismethode ist sehr verallgemeinerungsfähig. So lässt sich ganz analog zeigen,dass die Gleichung $x^2 = a$ für gegebenes $a \in \mathbb{D}$ genau eine, mit $\sqrt{a}$ bezeichnete Lösung hat, **7.2**. Schon in der Antike wusste man im Prinzip, dass nicht nur $\sqrt{2} = 1{,}41\cdots$, sondern auch $\sqrt{n}$ – repräsentiert durch eine geeignete Streckenlänge – immer dann irrational ist, wenn $n$ einen Primfaktor in ungerader Potenz enthält.

Der nach Satz 6.1 für $b \neq 0$ existierende, in gewissen Zusammenhängen auch mit $b^{-1}$ bezeichnete Quotient $\frac{1}{b}$ heißt der *Kehrwert* von $b$, und man sagt, $b$ und $\frac{1}{b}$ seien *reziprok*. Es ist $\frac{1}{b} < 1$ für $b > 1$, und $\frac{1}{b} > 1$ für $0 < b < 1$ gemäß $\mathsf{M}^\times$. Besonders einfach lassen sich die Quotienten $\frac{1}{n}$ veranschaulichen, für $n > 1$ auch *Stammbrüche* genannt. Da

$$\underbrace{\tfrac{1}{n} + \ldots + \tfrac{1}{n}}_{n} = n \cdot \tfrac{1}{n} = 1,$$

ist $\frac{1}{n}$ nichts anderes als die $n$-Teilung der Zahl 1. So ist z.B. $3 \cdot \frac{1}{3} = 1$ und bereits in **5.2** wurde beispielhaft gezeigt, dass $\frac{1}{3} = 0{,}333\cdots$. Summiert man den Stammbruch $\frac{1}{n}$ $m$ mal, erhält man die rationale Zahl $\frac{m}{n}$. So sind rationale Zahlen historisch gesehen entstanden und sie behalten selbstverständlich auch ihre weitreichende innermathematische Bedeutung. Aber ihre Alltagsbedeutung habt sich angesichts der allgegenwärtigen automatischen Rechner erheblich vermindert und auch für die Begründung der reellen Arithmetik sind sie entbehrlich wie wir gesehen haben.

Sind $a, b$ irgendwelche Terme, so heiße die Schreibfigur $\frac{a}{b}$ ein *Bruchterm* oder kurz ein *Bruch*. Für $a, b \in \mathbb{D}$ und $b \neq 0$ bezeichnet $\frac{a}{b}$ die wohlbestimmte reelle Zahl $c$ mit $bc = a$. Unterschiedliche Bruchterme können, ebenso wie etwa die Differenzterme $3 - 2$ und $4 - 3$, dieselbe Zahl bezeichnen. Genaueres wird unten in $\mathsf{Q}_1$ formuliert.

Mitunter ist es von Vorteil, es bei der Darstellung eines Quotienten $c = \frac{a}{b}$ als Bruchterm zu belassen, vor allem dann, wenn $a$ und $b$ natürliche Zahlen sind. Oft ist die Kenntnis der dezimalen Darstellung von $c$ gar nicht erforderlich, z.B. weil sich Zähler oder Nenner im Verlaufe arithmetischer Umformungen gegen andere Faktoren wegheben. Im Sinne unserer Auffassung ist aber z.B. $\frac{1}{3}$ nur eine andere Schreibweise für $0{,}333\cdots$, die in besonders prägnanter Weise den Sachverhalt $3 \cdot 0{,}333\cdots = 1$ zum Ausdruck bringt.

Wir formulieren und beweisen jetzt die bereits erwähnten Regeln der Bruchrechnung. Hierin können $a$, $b$, $c$, $d$ beliebige natürliche Zahlen, aber ebenso beliebige Dezimalzahlen, ja Elemente eines völlig beliebigen $\mathcal{E}$-Bereichs bedeuten, in welchem die Division ausführbar ist. Denn weder die Formulierungen noch die Beweise dieser Regeln hängen von irgendeiner Zahldarstellung ab. Die einzige Einschränkung ist $b, d \neq 0$.

$\mathsf{Q}_0$ : $x = \frac{a}{b} \Leftrightarrow bx = a$, speziell $b\frac{a}{b} = a$,

$\mathsf{Q}_1$ : $\frac{a}{b} = \frac{c}{d} \Leftrightarrow ad = bc$, speziell $\frac{ae}{be} = \frac{a}{b}$ für $e \neq 0$ (Kürzungsregel),

$\mathsf{Q}_2$ : $\frac{a}{b} < \frac{c}{d} \Leftrightarrow ad < bc$,

$\mathsf{Q}_3$ : $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$, $\quad \frac{a}{b} - \frac{c}{d} = \frac{ad-bc}{bd}$ (falls $\frac{a}{b} \geq \frac{c}{d}$ und damit $ad - bc \geq 0$),

$\mathsf{Q}_4$ : $\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$, $\quad \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$ $\quad (b, c, d \neq 0)$,

$\mathsf{Q}_5$ : $(\frac{a}{b})^n = \frac{a^n}{b^n}$ $\quad (b \neq 0)$.

Es folgen die einfachen Beweise von $\mathsf{Q}_1$ - $\mathsf{Q}_5$ ($\mathsf{Q}_0$ gibt lediglich die Definition wieder).

$\mathsf{Q}_1$: $\frac{a}{b} = \frac{c}{d} \Leftrightarrow bd\frac{a}{b} = bd\frac{c}{d}$ (wegen der Kürzungregel in **2.2**)
$\Leftrightarrow ad = bc$ (weil $bd\frac{a}{b} = b\frac{a}{b}d = ad$ und $bd\frac{c}{d} = bc$ gemäß $\mathsf{Q}_0$).

$\mathsf{Q}_2$: $\frac{a}{b} < \frac{c}{d} \Leftrightarrow bd\frac{a}{b} < bd\frac{c}{d} \Leftrightarrow ad < bc$ (wegen $\mathsf{M}^\times$ und $\mathsf{Q}_0$).

$\mathsf{Q}_3$: $bd(\frac{a}{b} + \frac{c}{d}) = db\frac{a}{b} + bd\frac{c}{d} = ad + bc$. Daraus folgt die erste Gleichung mit $\mathsf{Q}_0$. Die zweite ergibt sich ebenso, weil $bd(\frac{a}{b} - \frac{c}{d}) = bd\frac{a}{b} - bd\frac{c}{d} = ad - bc$.

$\mathsf{Q}_4$: $bd \cdot \frac{a}{b}\frac{c}{d} = b\frac{a}{b} \cdot d\frac{c}{d} = ac$. Daraus folgt die erste Gleichung mit $\mathsf{Q}_0$. Entsprechend folgt die zweite, denn nach dem Bewiesenen und der Kürzungsregel ist $\frac{ad}{bc} \cdot \frac{c}{d} = \frac{adc}{bcd} = \frac{a}{b}$.

$\mathsf{Q}_5$: Nach $\mathsf{P}_\times$ aus **2.3** ist $b^n(\frac{a}{b})^n = (b\frac{a}{b})^n = a^n$, und die Behauptung folgt dann aus $\mathsf{Q}_0$.

Diese Regeln lassen sich auf elegante Art auch so beweisen: Man schreibe $b^{-1}a$ für $\frac{a}{b}$ (also $\frac{1}{b} = b^{-1}$) und bestätigte zuerst $b^{-1}d^{-1} = (bd)^{-1}$ und $(c^{-1})^{-1} = c$. Dann erhält man z.B. $\mathsf{Q}_3$ aus $b^{-1}a + d^{-1}c = b^{-1}d^{-1}da + b^{-1}bd^{-1}c = (bd)^{-1}(ad + cb)$.

Wegen $\mathsf{Q}_1$ hat jedes $r \in \mathbb{Q}_+$ eine *gekürzte* Darstellung $r = \frac{m}{n}$, womit gemeint ist, $m$ und $n$ seien teilerfremd. Aber was noch wichtiger ist: $\mathsf{Q}_3$ und $\mathsf{Q}_4$ zeigen, dass $\mathbb{Q}$ gegenüber Addition und Multiplikation abgeschlossen ist. Ebenso ist die Abgeschlossenheit gegenüber Subtraktion gewährleistet, falls diese (in $\mathbb{D}$) ausführbar ist. Dies genügt um festzustellen, dass $\mathbb{Q}$ als Teilbereich von $\mathbb{D}$ selbst einen $\mathcal{E}$-Bereich bildet; denn die Axiome in **2.1** sind einfach nachzuprüfen. So gilt $\mathsf{K}^+$ in $\mathbb{D}$, also speziell auch in $\mathbb{Q}$.

Von nun an verwenden wir die Rechenregeln in $\mathcal{E}$-Bereichen und die Bruchrechnung ohne besonderen Hinweis. Eine etwas anspruchsvollere Anwendung der letzteren ist der weiter unten geführte Beweis, dass $\langle e_n \rangle$ eine schlichte Folge ist, wobei $e_n := (1 + \frac{1}{n})^n$ für $n > 0$ und $e_0 := 1$. Deren Grenzwert wird allgemein mit $\mathbf{e}$ bezeichnet und auch die EULER*sche*

*Zahl* genannt. Also $\mathbf{e} = \lim\langle e_n\rangle$. In **7.1** wird ein bequemes Berechnungsverfahren für **e** angegeben, und in **8.3** wird nachgewiesen, dass **e** irrational ist. Hier beweisen wir nur $2 < \mathbf{e} < 3$. Nicht nur **e**, sondern auch die $e_n$, und allgemeiner die Zahlen $(1+\frac{x}{n})^n$ für reelle $x$ treten in vielen Anwendungen der Mathematik explizit auf.

Die Ungleichung (12) in **2.4** liefert offenbar $(1-\frac{1}{n^2})^n > 1-\frac{1}{n} = \frac{n-1}{n}$ für $n > 1$. Auch beachte man $1+\frac{1}{n-1} = \frac{n}{n-1}$, sowie $\frac{1+\frac{1}{n}}{1+\frac{1}{n-1}} = \frac{(n+1)(n-1)}{n^2} = 1-\frac{1}{n^2}$. Damit ergibt sich das (strikte) Wachstum von $\langle e_n\rangle$ aus der folgenden Ungleichung. Darin sei $n > 1$, weil $e_0 < e_1$ $(= (1+\frac{1}{1})^1 = 2)$ ohnehin klar ist.

$$\frac{e_n}{e_{n-1}} = \frac{(1+\frac{1}{n})^n\cdot(1+\frac{1}{n-1})}{(1+\frac{1}{n-1})^n} = \left(\frac{1+\frac{1}{n}}{1+\frac{1}{n-1}}\right)^n\cdot(1+\frac{1}{n-1}) = (1-\frac{1}{n^2})^n\cdot\frac{n}{n-1} > \frac{n-1}{n}\cdot\frac{n}{n-1} = 1.$$

Die Beschränktheit von $\langle e_n\rangle$ folgt wegen $\binom{n}{k} = \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{k!} \le \frac{n^k}{k!}$ ((5) in **4.3**) aus

$$\begin{aligned}(1+\tfrac{1}{n})^n &= \textstyle\sum_{k\le n}\binom{n}{k}\frac{1}{n^k} && \text{(Binomische Formel)}\\ &\le \textstyle\sum_{k\le n}\frac{n^k}{k!}\frac{1}{n^k} = \sum_{k\le n}\frac{1}{k!} \le 1+\sum_{k<n}(\frac{1}{2})^k\\ &= 1+\frac{1-(\frac{1}{2})^n}{1-\frac{1}{2}} < 1+\frac{1}{1-\frac{1}{2}} = 3 && \text{(Formel (4) in \textbf{4.3}).}\end{aligned}$$

## 6.2 Periodische Dezimalzahlen

Die reelle Zahl 1,2074 074 074 $\cdots$ ist Beispiel einer so genannten *periodischen* Dezimalzahl. Allgemein heiße $a = z_0{,}z_1z_2\cdots$ *periodisch*, wenn es Indizes $p \ge 1$ und $q \ge 0$ gibt mit $z_{p+q+i} = z_{q+i}$ für $i = 1,2,\ldots$ Sind $p,q$ minimal mit dieser Eigenschaft, so heiße $p$ die *Periodenlänge*, und $q$ auch der *Periodenindex* von $a$. Es ist demnach $a = z_0{,}z_1\cdots z_qz_{q+1}\cdots z_{q+p}z_{q+1}\cdots z_{q+p}z_{q+1}\cdots$ die allgemeine Form einer periodischen Dezimalzahl, und man schreibt meistens etwas kürzer $a = z_0{,}z_1\cdots z_q\overline{z_{q+1}\cdots z_{q+p}}$.

Die Folge $\langle z_{q+1},\ldots,z_{q+p}\rangle$ heiße die *Periode* und die natürliche Zahl $P = z_{q+1}\cdots z_{q+p}$ der *Periodenwert*, wobei die Nullen, mit denen $z_{q+1}\cdots z_{q+p}$ eventuell beginnt, zu unterdrücken sind. Ist $q = 0$, heißt $a$ *reinperiodisch*.

**Beispiele.** Für $a = 1{,}2\overline{074}$ ist $q = 1$ und $p = 3$. Der Periodenwert $P$ ist 074, also 74. Auch $a = 3{,}14 = 3{,}14000\cdots$ ist periodisch, mit $q = 2$, $p = 1$ und Periodenwert 0.

**Satz 6.2.** *Eine periodische Dezimalzahl $a = z_0{,}z_1\cdots z_q\overline{z_{q+1}\cdots z_{q+p}}$ ist rational. Genauer, ist $P = z_{q+1}\cdots z_{q+p}$ der Periodenwert, so gilt*

$$(1)\quad a = a_{.q} + \frac{P}{(10^p-1)\cdot10^q} = a_{.q} + \frac{P}{\underbrace{9\cdots9}_{p}\underbrace{0\cdots0}_{q}}.$$

**Beweis.** Sei (a) $R := 0{,}z_{q+1}\cdots z_{q+p}z_{q+1}\cdots = (a-a_{.q})\overset{q}{\to} = 10^q(a-a_{.q})$. Offenbar ist $10^pR = R\overset{p}{\to} = z_{q+1}\cdots z_{q+p}{,}z_{q+1}\cdots = P+R$, also (b) $R = \frac{P}{10^p-1}$. (a) und (b) ergeben $10^q(a-a_{.q}) = \frac{P}{10^p-1}$, also $a-a_{.q} = \frac{P}{(10^p-1)10^q}$, und damit (1). ■

Nach (1) ist z.B. $1{,}2\overline{074} = 1{,}2 + \frac{74}{9990} = \frac{163}{135}$. Irrationalzahlen wie $\sqrt{2}$, $\sqrt{10}$, $\mathbf{e}$ usw. sind also nicht periodisch. Obwohl das schon hier unschwer beweisbar wäre, wird sich in **8.1** ganz nebenbei ergeben, dass umgekehrt jede rationale Zahl auch periodisch ist. Die rationalen Dezimalzahlen sind demnach als die periodischen vollständig gezeichnet. Für eine reinperiodische Zahl $a$ ist $q = 0$, also $10^q = 1$; daher vereinfacht sich (1) zu

$$(2) \quad a = z_0 + \frac{P}{\underbrace{99\cdots9}_{p}}.$$

Nach dieser Formel ist z.B. $3{,}333\cdots = 3 + \frac{3}{9} = \frac{10}{3}$, aber $3{,}0303\cdots = 3 + \frac{3}{99} = \frac{100}{33}$. Ferner ist nach (2) z.B. $\frac{z}{9} = 0{,}zzz\cdots$, d.h. $\frac{1}{9} = 0{,}111\cdots$, $\frac{2}{9} = 0{,}222\cdots$ usw. Analog ist $\frac{z_1z_2}{99} = 0{,}z_1z_2z_1z_2\cdots$ und allgemein $\frac{z_1\cdots z_n}{10^n-1} = 0{,}\overline{z_1\cdots z_n}$. Auch ist $\frac{n}{11} = \frac{9n}{99} = 0{,}z_1z_2z_1z_2\cdots$ für $n \le 10$, mit $z_1z_2 = 9\cdot n$. Demnach ist $\frac{1}{11} = 0{,}\overline{09}$ $(= 0{,}0\overline{90})$, $\frac{2}{11} = 0{,}\overline{18}$ $(= 0{,}1\overline{81})$ usw. Taschenrechner liefern diese und weitere Einsichten sozusagen experimentell. Zum Beispiel sind die Topzahlen allen Aberglaubens, die Primzahlen 7 und 13, dadurch ausgezeichnet, dass gekürzte Bruchzahlen $\frac{m}{n} < 1$ mit Primnenner $n$ nur für $n = 7$ und $n = 13$ die Periodenlänge 6 haben. Nach Übung 8.2 ist nämlich $p$ die kleinste natürliche Zahl, so dass $10^p - 1$ durch $n$ geteilt wird. Für $n = 7$ ist dies die Zahl 6, denn 7 teilt $10^6 - 1 = 99999$, nicht aber $10^i - 1$ für $0 < i < 6$. Dasselbe gilt für $n = 13$. Weitere Möglichkeiten gibt es nicht. Denn es gilt $10^6 - 1 = 3^3 \cdot 7 \cdot 11 \cdot 13 \cdot 37$, und $\frac{1}{3} = 0{,}\overline{3}$, $\frac{1}{11} = 0{,}\overline{09}$, $\frac{1}{37} = 0{,}\overline{027}$ haben die Periodenlängen $p = 1$, $p = 2$, bzw. $p = 3$.

## 6.3 Stammbruchapproximation

Brüche wurden im alten Ägypten durch Stammbruchsummen dargestellt. Für genauere Rechnungen gab es Rechentabellen wie z.B. der Papyrus Rhind belegt. So wie wir heute reelle Zahlen gern durch 2-stellige Kommazahlen approximieren (etwa $\boldsymbol{\pi}$ durch 3,14), verwandte man damals oft 2-gliedrige Stammbruchsummen. Diese Methode ist für viele Zwecke ausreichend genau. Weil keine 2-gliedrige Stammbruchsumme echt zwischen $\frac{1}{2} + \frac{1}{3} = \frac{5}{6}$ und 1 fällt, liegt der Approximationsfehler für Werte $r$ mit $0 < r < 1$ durchgehend allerdings nur unterhalb $1 - \frac{5}{6} = \frac{1}{6}$. Vermutlich deshalb war $\frac{2}{3}$ der einzige häufiger anzutreffende Nicht-Stammbruch. Denn wegen $\frac{2}{3} + \frac{1}{4} = \frac{11}{12}$ ist dann der Fehler bei 2-gliedriger Approximation nur noch kleiner als $\frac{1}{12}$, wie man sich leicht überlegt.

Leonardo gibt in [22] u.a. folgendes Rezept an zur Verwandlung einer rationalen Zahl $0 < r < 1$ in eine Summe aus paarweise verschiedenen Stammbrüchen:

Man suche den größten Stammbruch $\frac{1}{g_0+1}$ mit $\frac{1}{g_0+1} \le r$, so dass $g_0$ eindeutig gekennzeichnet ist durch $\frac{1}{g_0+1} \le r < \frac{1}{g_0}$. Ist nicht schon $r = \frac{1}{g_0+1}$, suche man den größten Stammbruch $\frac{1}{g_1+1}$ mit $\frac{1}{g_1+1} \le r - \frac{1}{g_0+1}$ $(< \frac{1}{g_1})$, und so fort. Stets ist $g_i \ge 1$. Nach endlich vielen (sagen wir $k+1$) Schritten steht das Gleichheitszeichen und es ergibt sich

$$(3) \quad r = \frac{1}{g_0+1} + \ldots + \frac{1}{g_k+1} \quad (0 < r < 1 \text{ rational}).$$

So erhält man z.B. $\frac{5}{6} = \frac{1}{2} + \frac{1}{3}$. Wir zeigen, dass dieses Verfahren tatsächlich abbricht, und zwar induktiv über den Zähler $m$ von $\frac{m}{n}$ $(0 < m < n)$. Für $m = 1$ ist nichts zu zeigen. Sei angenommen, alle $r < 1$ $(r \in \mathbb{Q}_+)$ mit einem Zähler $< m$ hätten eine Darstellung (3). Sei $\frac{1}{g+1}$ größter Stammbruch mit $\frac{1}{g+1} \leq \frac{m}{n}$. Dann hat $r := \frac{m}{n} - \frac{1}{g+1} = \frac{m(g+1)-n}{n(g+1)}$ einen Zähler $< m$; denn $\frac{m}{n} < \frac{1}{g}$, also $mg < n$, und so $m(g+1) - n < m$. Daher hat das soeben gewählte $r$ eine Darstellung (3) gemäß Induktionsvoraussetzung, und wir erhalten somit $\frac{m}{n} = r + \frac{1}{g+1} = \frac{1}{g_0+1} + \ldots + \frac{1}{g_k+1} + \frac{1}{g+1}$. Damit wurde bestätigt, dass Leonardos Prozedur abbricht. Auch sind die $g_i$ paarweise verschieden. Denn nach Definition ist $\frac{1}{g_{i+1}+1} \leq (r - \frac{1}{g_0+1} - \ldots - \frac{1}{g_{i-1}+1}) - \frac{1}{g_i+1} < \frac{1}{g_i} - \frac{1}{g_i+1} = \frac{1}{g_i(g_i+1)}$. Also $g_{i+1} + 1 > g_i(g_i + 1)$. Daher genügen die Zahlen $g_i$ der Bedingung

(4) $\quad g_{i+1} \geq g_i(g_i + 1)$ für alle $i < k$.

Weil $g_i + 1 \geq 2$, gilt nach (4) sogar $g_{i+1} \geq 2g_i$. Stammbruchdarstellungen sind in der Regel nicht eindeutig. Zum Beispiel ist $\frac{7}{12} = \frac{1}{3} + \frac{1}{4} = \frac{1}{2} + \frac{1}{12}$. Siehe auch Übung 4. Eine Darstellung (3) mit der Eigenschaft (4) ist nach Übung 5 jedoch eindeutig.

Startet man Leonardos Verfahren mit irrationalem $r > 0$ (genauer, mit $r - \operatorname{Int} r$), so bricht das Verfahren nicht ab und man erhält eine Darstellung von $r$ als unendliche Stammbruchreihe, siehe **7.1**. Es ist z.B. $\boldsymbol{\pi} = 3 + \frac{1}{8} + \frac{1}{61} + \ldots$ Übrigens ist der relative Fehler der 2-ten Stammbruchapproximation $3 + \frac{1}{8} + \frac{1}{61}$ für $\boldsymbol{\pi}$ mit etwa 0,006% um fast eine Größenordnung besser als die dezimale Approximation 3,14 mit etwa 0,05%.

## 6.4 Übungen

1. Man zeige: $\langle a_n \rangle$ mit $a_n \neq 0$ für alle $n$ ist (fallende) Nullfolge genau dann, wenn $\langle \frac{1}{a_n} \rangle$ monoton und unbeschränkt wächst. Damit ist $\langle c^n \rangle$ für $c < 1$ eine Nullfolge, weil nach Übung 3.3 die Folge $\langle \frac{1}{c^n} \rangle = \langle (\frac{1}{c})^n \rangle$ unbeschränkt wächst.

2. Sei $e'_n = (1 - \frac{1}{n})^n$ mit $e'_0 := 0$. Man zeige: $\langle e'_n \rangle$ ist schlicht und $\lim \langle e'_n \rangle = \frac{1}{\mathbf{e}}$.

3. Sei $x > 0$ und $x_n = (1 + \frac{x}{n})^n$. Man zeige: $\langle x_n \rangle$ wächst strikt monoton und es ist $x_n \leq \sum_{k \leq n} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!}$. Ferner beweise man, $\langle \sum_{k \leq n} \frac{x^k}{k!} \rangle$ und damit auch $\langle x_n \rangle$ sind beschränkte und folglich schlichte Folgen.

4. Sei $\langle h_i \rangle$ eine Folge von Zahlen aus $\mathbb{N}_+$ mit $h_{i+1} = h_i(h_i + 1)$. Man zeige, für jedes $n$ gilt $\frac{1}{h_0} = \sum_{i<n} \frac{1}{h_i+1} + \frac{1}{h_n}$. Nach dieser Formel ist z.B. für $h_0 = 1$ und $n = 0, 1, 2, \ldots$
$$1 = 0 + 1 = \tfrac{1}{2} + \tfrac{1}{2} = \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{6} = \tfrac{1}{2} + \tfrac{1}{3} + \tfrac{1}{7} + \tfrac{1}{42} = \ldots$$
Dem entnimmt man leicht verschiedene Stammbruchdarstellungen von $\frac{1}{2}$, die aber Bedingung (4) verletzen. Zum Beispiel ist $41 \geq 6 \cdot 7$ nicht erfüllt.

5. Man zeige, eine Darstellung (3) mit der Eigenschaft (4) *ist* die Leonardosche und folglich eindeutig bestimmt.

# Abschnitt 7

# Beginnende Analysis – Unendliche Reihen, Potenzen, Logarithmen

Obwohl die negativen Zahlen noch nicht zur Verfügung stehen, lassen sich gewisse Elemente der analytischen, d.h. mit dem Grenzwert zusammenhängenden Betrachtungsweise bereits jetzt bequem entwickeln. Mehr noch, das Nichtvorhandensein negativer Zahlen macht den Kern mancher Konstruktionen durch den vorläufigen Wegfall von Vorzeichenbetrachtungen deutlicher sichtbar. Die Erweiterung der Begriffe unter Einschluss negativer reeller Zahlen ist später nur ein kleiner Schritt. Man benötigt lediglich Grenzwerte oder Suprema schlichter Folgen, um unendliche Reihen mit positiven Gliedern, Potenzen und Logarithmen einfach und gründlich behandeln zu können. Speziell ist der Nachweis von $z_0{,}z_1z_2\cdots = \sum_{i=0}^{\infty}\frac{z_i}{10^i}$ zu Beginn von **7.1** geradezu banal.

Ein Hauptziel dieses Abschnitts ist die Definition der Exponentialfunktion durch den lückenlosen Nachweis aller Potenzgesetze, auf denen z.B. die logarithmischen Formeln beruhen. Die Einführung der $n$-ten Wurzeln, beliebiger Potenzen und Logarithmen beruht meistens auf dem so genannten Zwischenwertsatz für stetige Funktionen. Nun haben die elementaren Funktionen eine die Gültigkeit dieses Satzes garantierende Eigenschaft, welche etwas einfacher handhabbar ist als die Stetigkeit und welche vor allem für die numerische Analysis von Bedeutung ist. Sie sind Lipschitz-stetig in allen Intervallen ihres Definitionsbereichs. Auch genügt es vorübergehend, sich auf monotone Funktionen dieser Art zu beschränken. Das führt uns zu dem Begriff der schlichten Funktion, einer sinngemäßen Verallgemeinerung schlichter Folgen, die z.B. auch nützlich ist für den Konvergenzssatz 8.5, einer Variante des BANACHschen Fixpunktsatzes für Funktionen, die nicht notwendig Kontraktionen sind.

Natürlich wird bei dieser Gelegenheit auch der für die höhere Analysis wesentliche Begriff der stetigen Funktion eingeführt. Schlichte Funktionen sind stetig und die Sätze dieses Abschnitts bleiben richtig, wenn schlichte Funktion überall durch stetige Funktion ersetzt wird - nur sind die Beweise dann weniger einfach. Auch lassen sich die Exponential- und Logarithmenfunktion auf gänzlich andere Weise ohne explizite Zwischenwertargumente einführen, nämlich als spezielle Isomorphismen, siehe hierzu **10.5**.

## 7.1 Unendliche Reihen

Nach Formel (3) in **4.3** haben wir die Darstellung $a_{.n} = z_0 + \frac{z_1}{10^1} + \ldots + \frac{z_n}{10^n}$ für die $n$-te kanonische Näherung der Dezimalzahl $a = z_0{,}z_1z_2\cdots$. Diese Summe liegt anschaulich um so näher bei $a$, je größer $n$ ist. Daher stellt sich die Frage, ob man der Gleichung

(0) $\quad a = z_0 + \frac{z_1}{10} + \frac{z_2}{100} + \ldots$

einen vernünftigen Sinn zu geben imstande ist. Dies ist in der Tat möglich. Bei dem Term auf der rechten Seite von (0) handelt es sich nämlich um eine spezielle unendliche Reihe. Diese haben eine lange Tradition. Sie waren in der Mathematik schon vor Erfindung des Differentialkalküls aufgetreten und spielen auch heute eine wichtige Rolle, etwa bei der Reihenentwicklung von Funktionen. Hier einige bekannte Beispiele:

(1) $\quad 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots$

(2) $\quad 1 + \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \ldots$

(3) $\quad 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots$

(4) $\quad \frac{1}{2} + \frac{1}{3} + \frac{1}{7} + \ldots + \frac{1}{k+1} + \frac{1}{k(k+1)+1} + \ldots$

Mit diesen Reihen, die man meistens in der Form $a_0 + a_1 + \ldots$ oder $\sum_i a_i$ schreibt [1)], wurde lange Zeit recht unbefangen operiert. Aber erst der Satz von der oberen Grenze liefert das geeignete Werkzeug zur Beherrschung dieser Reihen, und zwar gemäß folgender Idee: Statt einer vermeintlich unendlichen Summe $a_0 + a_1 + \ldots$ betrachtet man die wohldefinierten Summen $s_n = a_0 + \ldots + a_n$, welche die *Partialsummen* der Reihe heißen. Die Folge $\langle s_n \rangle$ ist offenbar monoton. Wenn sie nun überdies beschränkt ist, konvergiert sie gegen einen wohlbestimmten Grenzwert $s$. In diesem Falle heißt auch die Reihe *konvergent* (andernfalls *divergent*), und man nennt $s$ die *Summe* dieser Reihe, symbolisch $s = a_0 + a_1 + a_2 + \ldots$, oder kurz $s = \sum_i a_i$.

Für die Reihe rechts in (0) ist $s_n = z_0 + \frac{z_1}{10^1} + \ldots + \frac{z_n}{10^n} = z_0{,}z_1\cdots z_n = a_{.n}$. Die Partialsummenfolge $\langle s_n \rangle$ ist also mit der Näherungsfolge $\langle a_{.n} \rangle$ identisch und konvergiert damit gegen $a$. Der Term rechts in (0) erhält so nicht nur einen wohldefinierten Sinn, sondern darüber hinaus trifft (0) auch zu. Damit entpuppt sich in unserem Aufbau eine nichtabbrechende Dezimalzahl gewissermaßen nachträglich als diejenige unendliche Reihe, durch welche sie traditionsgemäß definiert wird.

Durch die angegebene Präzisierung erübrigt sich auch die spekulative Frage nach einer Summation unendlich vieler Reihenglieder, so dass eine vermeintliche Beantwortung in diesem oder jenem Sinne das Geschehen in keiner Weise beeinflußt. An ursprüngliche Vorstellungen erinnert nur noch die traditionelle Notation $a_0 + a_1 + a_2 + \ldots$ Allerdings ist hierbei zu beachten, dass dieser Term, ebenso wie $\sum_i a_i$, in doppelter Bedeutung auftritt. Er bezeichnet nicht nur die Reihe selbst, genauer deren Partialsummenfolge,

[1)] auch $\sum_{i \geq 0} a_i$ oder $\sum_{i=0}^{\infty} a_i$. Dabei heißt $a_n$ das *$n$-te Glied* oder der *$n$-te Summand* der Reihe. Falls $i$ einen anderen Definitionsbereich hat, z.B. $\{i \in \mathbb{N} \mid i \geq k\}$, ist dies klar zu kennzeichnen, z.B. durch $\sum_{i \geq n} a_i$. Die Laufvariable sei stets die unmittelbar neben dem $\sum$-Zeichen stehende.

sondern im Falle der Konvergenz auch deren Grenzwert. Eine solche Doppeldeutigkeit hat auch ihre guten Seiten; beabsichtigt oder nicht, zwingt sie zu erhöhter Aufmerksamkeit. Von zahlreichen Konvergenzkriterien erwähnen wir die folgenden:

**1**. $\sum_i a_i$ konvergiert genau dann, wenn $\sum_{i>n} a_i = a_{n+1} + a_{n+2} + \dots$ für beliebiges $n$ konvergiert, und im Falle der Konvergenz ist $\sum_i a_i = \sum_{i\leq n} a_i + \sum_{i>n} a_i$. Der Beweis ist so einfach, dass er gänzlich dem Leser überlassen werden kann.

**2**. Mit $\sum_i a_i$ ist auch $\sum_i ca_i$ für beliebiges $c \in \mathbb{D}$ konvergent und es gilt $\sum_i ca_i = c\cdot\sum_i a_i$. Dies ergibt sich unmittelbar aus $\sum_{i\leq n} ca_i = c \cdot \sum_{i\leq n} a_i$ für alle $n$.

**3**. (*Majorantenkriterium*). Sind (a) $\sum_i a_i$ und (b) $\sum_i b_i$ unendliche Reihen mit $a_i \leq b_i$ für alle $i$ – es heißt (b) dann eine *Majorante* für (a) – ist mit (b) auch (a) konvergent. Denn sind $s_n, t_n$ die Partialsummen von (a) bzw. (b), so gilt $s_n \leq t_n$, so dass mit $\langle t_n\rangle$ auch $\langle s_n\rangle$ beschränkt ist. Danach sind z.B. mit (2) auch (3) und (4) konvergent. Bei (3) betrachte man die Reihe $\frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots$ für die (2) eine Majorante ist, so dass mit (2) nach Kriterium **1** auch (3) konvergiert. (2) ist tatsächlich konvergent; sie hat die Partialsumme $s_n = 2 - \frac{1}{2^n}$ wie man der Formel (4) in **6.2** leicht entnimmt. Daher konvergiert $\langle s_n\rangle$ gegen 2, d.h. $\sum_i \frac{1}{2^i} = 2$.

**4**. (*Nullfolgenkriterium*). Die Reihe $\sum_i a_i$ konvergiert genau dann mit der Summe $s$, wenn $s_n = a_0 + \dots + a_n \leq s$ für alle $n$ und wenn die Folge $\langle r_n\rangle$ der so genannten Restglieder $r_n := s - s_n$ eine Nullfolge ist. Dies folgt unmittelbar aus Übung 5.2.

Von den angegebenen Reihen ist nur die erste, die so genannte *harmonische* Reihe, divergent. Ihre Partialsummenfolge wächst unbeschränkt; ist nämlich $a > 0$ vorgegeben, so ist wegen $2^i - 2^{i-1} = 2^{i-1}$ für jedes $n > 2a$

$$\begin{aligned} s_{2^n} &= 1 + \tfrac{1}{2} + \tfrac{1}{2+1} + \tfrac{1}{2^2} + \dots + \tfrac{1}{2^{n-1}+1} + \dots + \tfrac{1}{2^n} \\ &> \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{2}{2^2} + \dots + \tfrac{2^{n-1}}{2^n} = \tfrac{1}{2} \cdot n > a. \end{aligned}$$

Die harmonische Reihe entgeht in gewissem Sinne sehr knapp der Konvergenz und divergiert außerordentlich langsam, wovon man sich mit Hilfe eines programmierbaren Taschenrechners leicht überzeugt. Z.B. ist immer noch $s_{10\,000} < 10$. Dieses Beispiel zeigt auch, dass es für die Konvergenz von $\sum_i a_i$ i.a. nicht hinreicht, dass die Glieder $a_i$ selbst eine Nullfolge bilden.

In Lichte der unendlichen Reihen wollen wir uns eine aus der Antike überlieferte Paradoxie näher ansehen, nämlich den Wettlauf des Achilles mit der Schildkröte. Diese erhält einen Vorsprung von 100`m`. Achilles hat die Aufgabe, die Schildkröte bei gleichzeitigem Start einzuholen. Die konstante Laufgeschwindigkeit des Achilles sei 100`m` in 10`sec`. Die Schildkröte hingegen bewege sich mit einer konstanten Geschwindigkeit von 10`m` in 10`sec`, also 1`m` in einer Sekunde. Jedermann wird sagen, dass Achilles die Schildkröte nach einem Durchlauf von wenig mehr als 110`m` eingeholt haben wird.

Nun wird aber gerade diese Banalität durch folgendes Gedankenexperiment problematisiert. Nach Durchlauf von 100`m` ist die Schildkröte offenbar 10`m` weit gekommen; nachdem Achilles diesen Standort erreicht hat, ist die Schildkröte 1`m` vorangekommen.

Sobald Achilles diesen Ort erreicht hat, hat die Schildkröte abermals ein Stückchen Boden gutgemacht, und so ad infinitum. Jedesmal, wenn Achilles die Schildkröte scheinbar erreicht hat, ist diese wieder ein (wenn auch allmählich kleiner werdendes) Stückchen vorangekommen. Daher ist die Schildkröte anscheinend uneinholbar.

Das ist jedoch ein Trugschluss. Denn nach der $(n+1)$-ten Momentaufnahme des Geschehens sind gerade $t_n = 10+1+\frac{1}{10}+\cdots+\frac{1}{10^n} = 11{,}\underbrace{111\cdots1}_{n}$ Sekunden vergangen. Alle diese Zeitpunkte liegen vor dem Zeitpunkt $11{,}111\cdots$, dem Grenzwert der Folge $\langle t_n\rangle$. Nun ist aber gerade dieser der Treffzeitpunkt $\tau$. Denn bezeichnet $s_A$ den von Achilles bis zur Treffzeit zurückgelegten Weg, sowie $s_B$ den Weg der Schildkröte einschließlich der Vorgabe, so erhält man $\tau$ auf bekannte Weise aus der Gleichung $10\tau = s_A = s_B = \tau + 100$ zu $\tau = \frac{100}{9} = 11{,}111\cdots$.

Das Paradoxe an dem geistreichen Gedankenexperiment der Griechen erklärt sich also allein daraus, dass eine Folge von Zeitpunkten ins Auge gefasst wird, die sämtlich von der tatsächlichen Treffzeit $\tau$ liegen, und die auch noch gegen diese konvergiert. Damit erhält man dann auch eine Zerlegung des gesamten Zeitintervalls von der Start- bis zur Treffzeit in die unendliche Reihe $10 + 1 + \frac{1}{10} + \frac{1}{100} + \ldots$

Die hier auftretende Reihe $1 + \frac{1}{10} + (\frac{1}{10})^2 + \ldots$ ist, ebenso wie die Reihe (2), Spezialfall der so genannten *geometrischen Reihe*

$$(5) \qquad \sum_i c^i = 1 + c + c^2 + \ldots \qquad (0 < c < 1).$$

Gleichung (4) in **4.3** besagt $s_n := 1+c+\ldots+c^n = \frac{1-c^{n+1}}{1-c}$. Wegen der hieraus unmittelbar folgenden Abschätzung $s_n < \frac{1}{1-c}$ konvergiert die geometrische Reihe. Ihr Grenzwert ist $\frac{1}{1-c}$. Um dies zu bestätigen, beachte man zunächst, dass $\langle c^n\rangle$ nach Übung 6.1 eine Nullfolge ist, und damit sicher auch die Restgliedfolge $\langle\frac{c^{n+1}}{1-c}\rangle = \langle\frac{1}{1-c} - s_n\rangle$. Nach dem Nullfolgenkriterium ist also in der Tat $\sum_i c^i = \frac{1}{1-c}$. Ferner ist für gegebene Werte $b$ und $c < 1$ auch $\sum_i bc^i$ konvergent und es ist $\sum_i bc^i = \frac{b}{1-c}$. Auch die Reihe $\sum_i a_i c^i$ ist für $c < 1$ konvergent, wenn nur die $a_i$ alle durch ein $b \in \mathbb{D}$ beschränkt sind. Es ist dann $\sum_i bc^i$ nämlich eine Majorante. Speziell ist jede der unendlichen Reihen

$$(6) \quad z_0 + \tfrac{z_1}{g} + \tfrac{z_2}{g^2} + \ldots \qquad (2 \le g \in \mathbb{N}, \quad z_i \in \mathbb{N}, \quad z_i < g \text{ für alle } i > 0)$$

konvergent. Für $g = 10$ ist dies ohnehin klar gemäß (0), sofern $z_i < 9$ für unendlich viele $i$, d.h. falls $z_0{,}z_1\cdots z_n$ zulässig ist. Falls für alle $i > m$ aber $z_i = 9$ (oder $z_i = g-1$ im allgemeinen Falle), ist (6) eben auch konvergent. Falls $g = 10$ sieht man sehr leicht, dass dann gerade $z_0{,}z_1\cdots z_m + \varepsilon_m$ die Summe von (6) ist. Den Fall einer beliebigen Grundzahl $g \ge 2$ betrachten wir im nächsten Abschnitt.

Dieses Ergebnis wirft neues Licht auf unsere Ausgangsposition, nach der reelle Zahlen formal als Dezimalzahlen erklärt wurden; es verdeutlicht, dass wir von Anfang an statt $g = 10$ eine beliebige andere Grundzahl $g \ge 2$ hätten wählen können. Auch wird vollkommen klar, dass wir z.B. $0{,}999\cdots$ nicht unbedingt als unzulässig disqualifizieren mussten, sondern mit 1 von Anfang an hätten identifizieren können.

Für die Reihe (3) gilt mit $\mathbf{e} = \lim\langle(1+\frac{1}{n})^n\rangle$ (**6.1**), wie schon von EULER bemerkt,

$$(7) \quad 1 + \tfrac{1}{1!} + \tfrac{1}{2!} + \tfrac{1}{3!} + \ldots = \mathbf{e}.$$

Die gemäß (7) von EULER in [9] erstmals vorgenommene, auf 23 Dezimalen genaue Berechnung von $\mathbf{e}$ liefert $\mathbf{e} = 2{,}71828\,18284\,59045\,23536\,068\cdots$. Die Reihe (7) ist in der Tat sehr geeignet für die numerische Berechnung von $\mathbf{e}$ mit vorgeschriebener Stellenzahl. Denn sei $s := \sum_k \frac{1}{k!}$ und $s_n := \sum_{k\le n} \frac{1}{k!}$. Dann ergibt sich für die Summe der „Restreihe“ $r_n := s - s_n = \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \ldots$ für $n \ge 1$ die Abschätzung

$$\begin{aligned} r_n &= \tfrac{1}{(n+1)!}\cdot(1 + \tfrac{1}{n+2} + \tfrac{1}{(n+2)(n+3)} + \ldots) \\ &< \tfrac{1}{(n+1)!}\cdot(1 + \tfrac{1}{n+2} + \tfrac{1}{(n+2)^2} + \ldots) \le \tfrac{1}{(n+1)!}\cdot(1 + \tfrac{1}{3} + \tfrac{1}{3^2} + \ldots) = \tfrac{1{,}5}{(n+1)!}. \end{aligned}$$

Einfaches Ausrechnen ergibt z.B. $r_{13} < \frac{1{,}5}{14!} < 2\varepsilon_{11}$. Also gibt bereits die Partialsumme $s_{13}$ von (7) die EULERsche Zahl $\mathbf{e}$ auf mindestens 10 Kommastellen genau an.

Es folgt nun der Beweis von (7). In **6.1** wurde für $e_n = (1+\frac{1}{n})^n$ bereits die Ungleichung $e_n \le s_n$ $(= \sum_{k\le n}\frac{1}{k!})$ gezeigt. Also $\mathbf{e} \le s = \sum_k \frac{1}{k!}$. Es genügt also nachzuweisen, dass $\langle s - e_n\rangle$ eine Nullfolge ist. Sei $\left[{n \atop k}\right] := \frac{n\cdot(n-1)\cdot\ldots\cdot(n-k+1)}{n^k}$. Dann gilt für $1 \le k \le n$ aufgrund von $1 \ge \left[{n \atop k}\right] \ge \frac{(n-k+1)^k}{n^k} = (1 - \frac{k-1}{n})^k$ und der Ungleichung (11) in **2.4** offenbar

$$1 - \left[{n \atop k}\right] \le 1 - (1 - \tfrac{k-1}{n})^k \le k\cdot\tfrac{k-1}{n} = \tfrac{1}{n}k(k-1).$$

Weil $\binom{n}{k} = \frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{k!}$ und $e_n = \sum_{k\le n}\binom{n}{k}\frac{1}{n^k} = 1 + \sum_{k=1}^n \left[{n \atop k}\right]\frac{1}{k!}$, folgt somit für $n \ge 2$

$$\begin{aligned} s - e_n = \textstyle\sum_{k=1}^n \frac{1}{k!}(1 - \left[{n \atop k}\right]) + \sum_{k>n}\frac{1}{k!} &\le \tfrac{1}{n}\textstyle\sum_{k=1}^n \frac{k(k-1)}{k!} + \frac{1}{n}\sum_{k>n}\frac{n}{k}\cdot\frac{1}{(k-1)!} \\ &< \tfrac{1}{n}(\textstyle\sum_{k<n}\frac{1}{k!} + \sum_{k\ge n}\frac{1}{k!}) = \frac{s}{n}. \end{aligned}$$

Damit ist $\langle s - e_n\rangle$ in der Tat Nullfolge und $\mathbf{e} = \sum_k \frac{1}{k!}$ ist bewiesen. Es gelten nun sogar

$$\text{(e1)}\quad \mathbf{e} - e_n < \tfrac{\mathbf{e}}{2n+1} \quad ; \quad \text{(e2)}\quad \mathbf{e} - e_n > \tfrac{\mathbf{e}}{2n+2},$$

Übung 3. Die Ungleichung (e2) verursacht ein überaus träges Konvergenzverhalten der strikt wachsenden Folge $\langle e_n\rangle$. Während z.B. bereits $\mathbf{e} - \sum_{i\le 8}\frac{1}{i!} < \varepsilon_5$, ist nach (e2) immer noch $\mathbf{e} - e_{100\,000} > \varepsilon_5$, also wird $\mathbf{e}$ von $e_{100\,000}$ bis höchstens zur 4. Dezimalen genau approximiert. Eine leichte Umformung von (e1),(e2) liefert die Ungleichungen

$$\text{(e3)}\quad \mathbf{e}\tfrac{2n}{2n+1} < e_n < \mathbf{e}\tfrac{2n+1}{2n+2}.$$

Dies ergibt unter Verwendung der (ausreichenden) Näherung 2,71828 für $\mathbf{e}$ das mit den angegebenen Ziffern genaue Ergebnis $e_{100\,000} = (1 + \frac{1}{100\,000})^{100\,000} = 2{,}71826\cdots$.

Wir haben hier die Irrationalzahl $\mathbf{e}$ benutzt, um eine Abschätzung über $e_{10^5}$ zu gewinnen. Tatsächlich ist es ohne theoretische Vorbereitung einfacher, $\mathbf{e}$ mit hoher Genauigkeit zu berechnen als die scheinbar harmlose, abbrechende Dezimalzahl $e_{10^5}$ mit ihren 500 000 Dezimalen (deren letzte eine 1 ist). Für $t_n := \frac{2n}{2n+1}\mathbf{e}$ folgt aus (e3)

$$\text{(e4)}\quad 0 < e_n - t_n < \tfrac{\mathbf{e}}{(2n+1)(2n+2)} < \tfrac{1}{n^2}.$$

Ein noch besserer Approximationsterm der Größenordnung $\frac{1}{n^3}$ wird nach [31] gegeben durch den Term $t_n := \frac{12n+5}{12n+11}\mathbf{e}$. Für diesen gilt sogar $0 < t_n - e_n < \frac{1}{10n^3}$, so dass $e_{10^5}$ durch $t_{10^5}$ auf mindestens 15 Dezimalen genau angegeben wird.

Reihe (4) hat die Summe 1. Denn nach Übung 6.4 ist $\frac{1}{h_0} = \sum_{i<n} \frac{1}{h_i+1} + \frac{1}{h_n}$. Das ergibt

$$(8) \quad \frac{1}{h_0} = \sum_i \frac{1}{h_i+1} \qquad \big(h_0 \in \mathbb{N}_+, \quad h_{i+1} = h_i(h_i+1)\big),$$

weil $h_n \geq 2^n$ und $\langle\frac{1}{h_n}\rangle$ daher Nullfolge ist. (4) ist der Sonderfall von (8) mit $h_0 = 1$ und hat deshalb die Summe 1. Setzt man LEONARDOs Algorithmus aus **6.3** auf ein $r < 1$ an, so erhält man nach (4) in **6.3** $r - \sum_{i<n} \frac{1}{g_i+1} < \frac{1}{g_n} \leq \frac{1}{2^n}$, was in $r = \sum_i \frac{1}{g_i+1}$ resultiert. Das ergibt für beliebiges $a \in \mathbb{D}_+$ wegen $r := a - \operatorname{Int} a < 1$ eine Darstellung

$$(9) \quad a = \operatorname{Int} a + \sum_i \frac{1}{g_i+1} \qquad \big(g_{i+1} \geq g_i(g_i+1)\big).$$

Dabei steht in der Nebenbedingung das =-Zeichen ab einer gewissen Stelle $n$ genau dann, wenn $r$, oder gleichwertig $a$, rational ist. Denn sei etwa $g_{i+1} = g_i(g_i+1)$ für alle $i \geq n$. Dann liefert (9) nach (8) mit $h_0 = g_n$ offenbar $\sum_{i \geq n} \frac{1}{g_i+1} = \frac{1}{g_n}$ und $a$ ist damit rational. Sei umgekehrt $a$ rational. Dann liefert die nach (8) mögliche Entwicklung des letzten Gliedes $\frac{1}{g_k+1}$ der LEONARDOschen Stammbruchdarstellung von $a - \operatorname{Int} a$ in eine unendliche Reihe auch eine die genannte Bedingung erfüllende Stammbruchreihe.

Zusammengefasst: jedes $a \in \mathbb{D}_+$ hat (genau) eine Darstellung (9); dabei steht in der Nebenbedingung ab einer gewissen Stelle das =-Zeichen genau dann, wenn $a$ rational ist. Bei irrationalem $r$ ist also $g_{i+1} > g_i(g_i+1)$ für unendlich viele $i$. Eindeutigkeit ergibt sich ähnlich wie in Übung 6.5. (9) heißt auch die SYLVESTER*sche Entwicklung* von $a$.

## 7.2 Der Zwischenwertsatz und erste Anwendungen

Die Lösbarkeit der Gleichung $x^n = a$ läuft auf die Frage hinaus, ob die Funktion $x \mapsto x^n$ den vorgegebenen Wert $a$ annimmt. Diese lässt sich z.B. mit dem Hinweis auf den Zwischenwertsatz für stetige Funktionen positiv beantworten. Nun kann man hier auch einen etwas einfacheren Begriff verwenden, der vor allem für die numerische Analysis bedeutsam ist, die intervallweise definierte Lipschitz-Stetigkeit, siehe auch **8.4**. Die elementaren Funktionen sind alle Lipschitz-stetig in geeignet gewählten Intervallen ihres Definitionsbereichs, ein bedeutender Vorteil für ihre numerische Berechnung. Außerdem genügt es für unsere Zwecke, sich auf monoton wachsende Funktionen dieser Art zu beschränken. Das führt zu folgender Definition, wobei für $u, v \in \mathbb{D}$ mit $u < v$ das Intervall $\{x \in \mathbb{D} \mid u \leq x \leq v\}$ wie üblich mit $[u, v]$ bezeichnet wird.

**Definition.** Sei $X \subseteq \mathbb{D}$. Eine Funktion $f : X \to \mathbb{D}$ heiße *schlicht* in $[u, v]$, wenn $f$ monoton wächst und ein $C \in \mathbb{D}$ existiert, so dass für alle $h$ mit $x, x+h \in X \cap [u, v]$

$$(10) \quad f(x+h) - f(x) \leq hC,$$

d.h. wenn $f$ dort Lipschitz-stetig ist. $f$ heiße schlicht schlechthin, wenn je zwei Punkte des Definitionsbereichs von $f$ einem Intervall angehören, in welchem $f$ schlicht ist.

Man beachte, dass $f$ nicht in ganz $[u,v]$ definiert sein muss, auch nicht an den Intervallgrenzen. In den Anwendungen wird meist $X = \mathbb{D}, \mathbb{E}, \mathbb{Q}$ oder ein bestimmtes Intervall hiervon sein. Um z.B. eine auf $\mathbb{E}$ überall definierte Funktion als schlicht (schlechthin) nachzuweisen, genügt es, die Schlichtheit in jedem Intervall $[0,n]$ zu bestätigen.

(10) ist offenbar gleichwertig mit $|fy - fx| \leq C|y-x|$ für alle $x, y \in X \cap [u,v]$. Setzt man $\Delta f := f(x+\Delta x) - f(x)$ mit $\Delta x := h$, so lässt sich (10) für $\Delta x \neq 0$ auch notieren als $\frac{\Delta f}{\Delta x} \leq C$. Schlichtheit umfasst also die Beschränktheit der *Differenzenquotienten* $\frac{\Delta f}{\Delta x}$ im gesamten Intervall $[u,v]$. Es ist anschaulich klar, dass diese für $f\colon x \mapsto x^2$ in jedem Intervall beschränkt sind. In der Tat, für $x, x+h \in [u,v]$ ist wegen $x + \frac{h}{2} < x + h \leq v$

$$f(x+h) - f(x) = (x+h)^2 - x^2 = 2h(x+\tfrac{h}{2}) < 2hv = Ch, \text{ mit } C = 2v.$$

Ferner ist für eine auf $X$ definierte, in $I = [u,v]$ schlichte Funktion $f$ mit jeder schlichten Folge $\langle x_n \rangle$ aus $X \cap I$ auch $\langle fx_n \rangle$ schlicht. Mehr noch, falls $\lim\langle x_n\rangle \in X$, gilt

$$(11) \qquad \lim\langle fx_n\rangle = f\xi \quad (\xi := \lim\langle x_n\rangle).$$

Denn mit $\langle \xi - x_n\rangle$ ist $\langle C(\xi - x_n)\rangle$ und daher sicher auch $\langle f\xi - fx_n\rangle$ eine Nullfolge. (11) besagt, dass $f$ im Punkte $\xi$ linksseitig stetig ist. $f$ ist in $\xi$ auch stetig schlechthin, und sogar gleichmäßig stetig in $I$, was hier nur erwähnt sei. Dabei heißt $f\colon X \to \mathbb{D}$ *stetig* im Punkte $\xi \in X$, wenn (11) für jede konvergente Folge $\langle x_n \rangle$ aus $X$ erfüllt ist.

Bedingung (10) muss nicht für alle $h$ geprüft werden. Es genügt, sich auf $h$-Werte $< 1$ zu beschränken. Denn jedes Intervall $I$ lässt sich in endlich viele Teilintervalle (mit gemeinsamen Randpunkten) zerlegen, die eine Länge $< 1$ haben, und die Schlichtheit in jedem dieser Teilintervalle impliziert offenbar die in $I$. Es genügt auch, (10) nur für die Werte $h = \varepsilon_n$ zu verifizieren, was häufig erheblich einfacher ist.

**Satz 7.1 (Zwischenwertsatz).** *Es sei $f$ eine im gesamten Intervall $[u,v] \subseteq \mathbb{D}$ definierte und dort schlichte reelle Funktion, sowie $fu < fv$. Dann nimmt $f$ jeden Wert $c$ mit $fu < c < fv$ an; kurz, die Gleichung $fx = c$ hat mindestens eine Lösung.*

**Beweis.** Sei $U$ die gewiss beschränkte nichtleere Menge $\{x \in [u,v] \mid fx \leq c\}$ und sei $s = \sup U$. Sicher ist $u \leq s \leq v$. Wir behaupten, $fs = c$. Gemäß Voraussetzung gibt es ein $C$, so dass $x, x+\varepsilon_n \in [u,v] \ \Rightarrow\ f(x+\varepsilon_n) - fx \leq C\varepsilon_n$. Angenommen $fs < c$. Dann ist notwendig $s < v$, und es gibt ein $n$ mit $s + \varepsilon_n \leq v$ und $C\varepsilon_n + fs \leq c$. Also ist $f(s+\varepsilon_n) \leq C\varepsilon_n + fs \leq c$, und damit $s + \varepsilon_n \in U$, im Widerspruch zu $s = \sup U$. Sei nun $fs > c$ angenommen, so dass sicher $u < s$, und sei $\varepsilon_n$ so gewählt, dass $u + \varepsilon_n \leq s$ und $C\varepsilon_n < fs - c$. Für $a = s - \varepsilon_n$ $(\geq u)$ ergibt sich dann

$$fs - f(s-\varepsilon_n) = f(a+\varepsilon_n) - fa \leq C\varepsilon_n < fs - c,$$

also $c < f(s-\varepsilon_n)$. Sei $x \in U$, mit $s - \varepsilon_n \leq x$. Dann ist $c < f(s-\varepsilon_n) \leq fx$, was $x \in U$ widerspricht. Also verbleibt in der Tat nur die Möglichkeit $fs = c$. ■

Es ist einfach nachzuweisen, dass Summe und Produkt schlichter Funktionen (mit demselben Definitionsbereich) wieder schlicht sind. Deswegen sind z.B. alle Funktionen $x \mapsto a_0 + a_1x + \ldots + a_nx^n$ schlicht. Satz 7.1 ergibt ferner: Ist $f\colon \mathbb{D} \to \mathbb{D}$ schlicht, strikt wachsend und unbeschränkt, so gibt es zu jedem $y > f0$ genau ein $x \in \mathbb{D}$ mit $fx = y$. Zum Beispiel erfüllt $x \mapsto ax$ für jedes $a > 0$ diese Bedingungen, womit nochmals die Existenz des Quotienten bewiesen wurde (dessen Existenz wird im Beweis von Satz 7.1 nicht benötigt!). Aber auch $x \mapsto x^n$ erfüllt für jedes $n \geq 1$ die genannten Bedingungen. Nach Satz 7.1 hat also $x^n = c$ für jedes reelle $c \geq 0$ genau eine Lösung. Es ist dies die mit $\sqrt[n]{c}$ bezeichnete *$n$-te Wurzel* aus $c$, für $n = 2$ auch *Quadratwurzel* aus $c$ genannt. Die Funktion $x \mapsto \sqrt[n]{x}$ ist demnach gerade die Umkehrfunktion von $x \mapsto x^n$.

Aus den in Abschnitt **2** genannten Regeln der Potenzrechnung mit Exponenten aus $\mathbb{N}$ ergeben sich leicht die drei wichtigsten Gesetze der Wurzelrechnung, nämlich

$$\mathsf{R}_1:\ \sqrt[n]{ab} = \sqrt[n]{a}\cdot\sqrt[n]{b} \quad ; \quad \mathsf{R}_2:\ (\sqrt[n]{a})^m = \sqrt[n]{a^m} \quad ; \quad \mathsf{R}_3:\ a < b \ \Leftrightarrow\ \sqrt[n]{a} < \sqrt[n]{b}.$$

Zum Beweis von $\mathsf{R}_1$ beachte man $(\sqrt[n]{a}\cdot\sqrt[n]{b})^n = (\sqrt[n]{a})^n\cdot(\sqrt[n]{b})^n = ab$. Also löst $\sqrt[n]{a}\cdot\sqrt[n]{b}$ die Gleichung $x^n = ab$, woraus wegen der Eindeutigkeit der Lösung $\mathsf{R}_1$ folgt. $\mathsf{R}_2$ ergibt sich aus $((\sqrt[n]{a})^m)^n = ((\sqrt[n]{a})^n)^m = a^m$ durch beidseitiges Ziehen der $n$-ten Wurzel. Schließlich ist $\mathsf{R}_3$ eine Folge des strikten Wachstums von $x \mapsto x^n$. Denn eine in ihrem Definitionsbereich $X$ strikt wachsende Funktion $f$ bildet $X$ bijektiv auf ihren Wertebereich $Y$ ab, und die Umkehrfunktion $f^{-1}: Y \to X$ ist stets wieder strikt wachsend.

Auch zeigt man leicht: Ist $f$ schlicht in $I$ und $\frac{\Delta f}{\Delta x} \geq D$ für ein $D > 0$ (d.h. werden die $\frac{\Delta f}{\Delta x}$ nicht zu klein), ist auch $f^{-1}$ im Bildintervall von $f$ schlicht, mit der Konstanten $\frac{1}{D}$. Für $f: x \mapsto x^2$ z.B. ist $\frac{\Delta f}{\Delta x} = 2x \geq 2u$ mit $x \in [u,v]$. Also ist $f^{-1}: x \mapsto \sqrt{x}$ schlicht in jedem Intervall $[u',v']$ $(= [fu, fv])$, mit der Konstanten $\frac{1}{2\sqrt{u'}} = \frac{1}{2u}$ $(u > 0)$.

Sei nun $f$ eine zunächst nur auf $\mathbb{E}$ definierte schlichte Funktion mit Werten aus $\mathbb{D}$. Ist $a \in \mathbb{D}$, so ist mit $\langle a_{.n}\rangle$ offenbar auch $\langle fa_{.n}\rangle$ wieder eine schlichte Folge, hat also einen wohlbestimmten Grenzwert $\lim\langle fa_{.n}\rangle$. Schon in Übung 3.5 wurde die durch

$$(12) \qquad \bar{f}a = \lim\langle fa_{.n}\rangle \quad (= \sup\{fa_{.n} \mid n \in \mathbb{N}\})$$

auf ganz $\mathbb{D}$ erklärte Funktion die *natürliche Fortsetzung* von $f$ genannt. Der folgende überaus nützliche Satz zeigt, dass durch natürliche Fortsetzung von $f$ auch dann keine anderen Funktionswerte entstehen, falls $f$ von vornherein schon einen größeren Definitionsbereich $X \supseteq \mathbb{E}$ besitzt (etwa $X = \mathbb{Q}$), solange $f$ dort monoton wächst.

**Satz 7.2.** *Sei $f: \mathbb{E} \to \mathbb{D}$ schlicht. Dann ist auch die natürliche Fortsetzung $\bar{f}$ schlicht auf ganz $\mathbb{D}$. Darüber hinaus ist $\bar{f}$ die einzige monotone Fortsetzung von $f$ auf ganz $\mathbb{D}$.*

**Beweis.** $\bar{f}$ ist wachsende Fortsetzung von $f$ nach Übung 3.5. Sei nun $\tilde{f}$ irgendeine wachsende Fortsetzung von $f$ auf $\mathbb{D}$. Wir zeigen zuerst, $\tilde{f}$ ist schlicht in $[0,n]$. Sei $a < b \leq n$. Man wähle $a', b' \in \mathbb{E} \cap [0,n]$ so, dass $a' \leq a$, $b \leq b'$ und $a - a', b' - b \leq \frac{b-a}{2}$. Dann ist $b' - a' = b' - b + b - a + a - a' \leq 2(b-a)$, und $fb' - fa' \leq C(b' - a')$ ergibt

$$\tilde{f}b - \tilde{f}a \leq \tilde{f}b' - \tilde{f}a' = fb' - fa' \leq C(b' - a') \leq 2C(b-a).$$

Daher ist $\tilde{f}$ schlicht in $[0,n]$, für alle $n$. Folglich gilt $\tilde{f}a = \lim\langle\tilde{f}a_{.n}\rangle$ gemäß (11). Also ist $\tilde{f}a = \lim\langle\tilde{f}a_{.n}\rangle = \lim\langle fa_{.n}\rangle = \bar{f}a$ nach (12), für alle $a \in \mathbb{D}$. Damit ist die Eindeutigkeit einer monotonen Fortsetzung von $f$ und zugleich deren Schlichtheit gezeigt. ∎

## 7.3 Potenzrechnung und Exponentialfunktion

Bisher kennen wir lediglich die Potenz $b^n$ ($b \in \mathbb{D}$) als $n$-fach wiederholte Multiplikation. Nun pflegt man $\sqrt[n]{b}$ auch in der Form $b^{\frac{1}{n}}$ zu schreiben, also z.B. $2^{\frac{1}{2}}$ für $\sqrt{2}$. Wie kommt man zu dieser Schreibweise? Die Antwort ist einfach: will man, dass auch nur einige der unten aufgelisteteten Potenzregeln für andere Exponenten als nur natürliche Zahlen gültig bleiben, dann wird die Gleichung $b^{\frac{1}{n}} = \sqrt[n]{b}$ schlichtweg erzwungen. Denn

$$b = b^1 = b^{\frac{n}{n}} = b^{\overbrace{\frac{1}{n}+\dots+\frac{1}{n}}^{n}} = \underbrace{b^{\frac{1}{n}} \cdot \ldots \cdot b^{\frac{1}{n}}}_{n} = (b^{\frac{1}{n}})^n,$$

also $b^{\frac{1}{n}} = \sqrt[n]{b}$. Dabei haben wir lediglich die Minimalforderungen $\mathsf{P}^1$ und $\mathsf{P}^+$ unten benutzt. Diese implizieren darüber hinaus $b^{\frac{m}{n}} = \underbrace{b^{\frac{1}{n}} \cdot \ldots \cdot b^{\frac{1}{n}}}_{m} = (\sqrt[n]{b})^m = \sqrt[n]{b^m}$.

Das bedeutet aber noch nicht, dass die Erklärung $b^{\frac{m}{n}} = \sqrt[n]{b^m}$ legal ist; denn dazu muss $b^{\frac{m}{n}} = b^{\frac{p}{q}}$ für $\frac{m}{n} = \frac{p}{q}$ gesichert sein. Bevor wir darauf eingehen, formulieren wir zuerst die Potenzgesetze, jetzt aber gleich für beliebige Exponenten $x, y \in \mathbb{D}$ mit einer beliebigen positiven reellen Basis $b$. Wir betrachten diese Gesetze oder wenigstens einige davon zunächst nur als Forderungen, die an eine sinnvolle Verallgemeinerung der Potenz zu stellen sind. Dass sich diese sogar alle erfüllen lassen, und zwar auf genau eine Weise, wird in Satz 7.4 bewiesen. Für Exponenten aus $\mathbb{N}$ gelten alle diese Regeln nach **2.3**.

**Regeln der allgemeinen Potenzrechnung**

$$\begin{array}{lll}
\mathsf{P}^1:\ b^1 = b, & \mathsf{P}^+:\ b^{x+y} = b^x \cdot b^y, & \mathsf{P}_{\geq 1}:\ b^x \geq 1 \text{ für } b \geq 1, \\
\mathsf{P}_1:\ 1^x = 1, & \mathsf{P}^\times:\ b^{x \cdot y} = (b^x)^y, & \mathsf{P}_{\leq 1}:\ b^x \leq 1 \text{ für } b \leq 1, \\
\mathsf{P}^0:\ b^0 = 1, & \mathsf{P}_\times:\ (a \cdot b)^x = a^x \cdot b^x, & \mathsf{P}^<:\ x < y \Rightarrow b^x < b^y \quad (b > 1).
\end{array}$$

In der linken Spalte stehen Randbedingungen, in der Mitte Gleichungen, und rechts Monotoniebedingungen. Stets ist $b^x \neq 0$, weil $b^x \cdot (\frac{1}{b})^x = 1$ nach $\mathsf{P}_\times$ und $\mathsf{P}_1$.

Zwischen obigen Regeln bestehen zahlreiche Abhängigkeiten. Zum Beispiel folgt $\mathsf{P}_1$ sofort aus $\mathsf{P}_{\geq 1}$ und $\mathsf{P}_{\leq 1}$. Diese aus besonderem Grunde hervorgehobenen Regeln sind unter Beachtung von $\mathsf{P}^1$, $\mathsf{P}^+$ und der aus $b^x(\frac{1}{b})^x = 1$ folgenden Gleichung $(\frac{1}{b})^x = \frac{1}{b^x}$ nur Spezialfälle von $\mathsf{P}^<$. Auch folgt $\mathsf{P}^0$ leicht aus $\mathsf{P}^1, \mathsf{P}^+$. Denn $b = b^{1+0} = b^1 \cdot b^0 = b \cdot b^0$, also $b^0 = 1$. Aus den angegebenen folgen leicht einige weitere Regeln, insbesondere

$$\mathsf{P}^-:\ b^{x-y} = \frac{b^x}{b^y} \quad (x \geq y), \qquad \mathsf{P}_\div:\ (\tfrac{a}{b})^x = \frac{a^x}{b^x}, \qquad \mathsf{P}_<:\ a < b \Rightarrow a^x < b^x \quad (x \neq 0),$$

$$\mathsf{P}_=:\ a^x = b^x \Rightarrow a = b \quad (x \neq 0), \qquad \mathsf{P}^=:\ b^x = b^y \Rightarrow x = y \quad (b \neq 1).$$

$\mathsf{P}^-$ folgt aus $b^{x-y}b^x = b^x$ gemäß $\mathsf{P}^+$, und ähnlich folgt $\mathsf{P}_\div$. Zum Beweis von $\mathsf{P}_<$ sei $0 < a < b$. Dann gibt es ein $c > 1$ mit $ac = b$. Also $a^x c^x = b^x$ und damit $a^x < b^x$, denn $c^x > 1$ für $x \neq 0$ gemäß $\mathsf{P}^<$ und $\mathsf{P}_1$. Für $b < 1$ ist $x \mapsto b^x$ strikt fallend, weil nämlich $\left(\frac{1}{b}\right)^x = \frac{1}{b^x}$, und weil $x \mapsto \left(\frac{1}{b}\right)^x$ nach $\mathsf{P}^<$ strikt wächst.

In Satz 7.4 wird die bemerkenswerte und keineswegs offensichtliche Tatsache bewiesen, dass alle erwähnten Potenzgesetze schon eine Folge sind aus $\mathsf{P}^1, \mathsf{P}^+, \mathsf{P}_{\geq 1}, \mathsf{P}_{\leq 1}$. Diese vier Gesetze seien daher die *Basisregeln* der Potenzrechnung genannt.

**Bemerkung 1.** Aus den Basisregeln folgt unmittelbar nur, dass $f : x \mapsto b^x$ für $b \geq 1$ monoton wächst: Sei $x < y$, also $x + t = y$ für ein gewisses $t$. Es ist $1 \leq b^t$ gemäß $\mathsf{P}_{\geq 1}$, also $b^x \leq b^x \cdot b^t = b^y$. Erst Satz 7.4 wird zeigen, dass die Basisregeln auch das strikte Wachstum von $f$ implizieren. Analoges gilt auch für $x \mapsto \frac{1}{b^x}$ im Falle $b < 1$, wegen $\mathsf{P}_{\leq 1}$.

**Bemerkung 2.** Aus obigen Ausführungen folgt, dass sich der Beweis aller erwähnten Potenzgesetze aus den Basisregeln auf den Nachweis von $\mathsf{P}^\times$, $\mathsf{P}_\times$ und $\mathsf{P}^<$ reduziert.

Wir hatten anfänglich schon erkannt, dass $a^r$ für $r = \frac{m}{n} \in \mathbb{Q}$ auf höchstens eine Weise sinnvoll erklärt werden kann. Nun steht auch einer Definition von $a^r$ als $\sqrt[n]{a^m}$ auch insofern nichts im Wege, als diese nicht von der Darstellung von $r$ abhängt. Denn für $\frac{m}{n} = \frac{p}{q}$ ist $mq = np$. Die Potenzgesetze in **2.3** und $\mathsf{R}_2$ ergeben dann

$$\sqrt[n]{a^m} = \sqrt[n]{\sqrt[q]{(a^m)^q}} = \sqrt[n]{\sqrt[q]{a^{m\cdot q}}} = \sqrt[n]{\sqrt[q]{a^{n\cdot p}}} = \sqrt[n]{\sqrt[q]{(a^p)^n}} = \sqrt[n]{(\sqrt[q]{a^p})^n} = \sqrt[q]{a^p}.$$

Erst diese Betrachtungen rechtfertigen die folgende

**Definition.** Für $a \in \mathbb{D}$ und $r = \frac{m}{n} \in \mathbb{Q}$ sei $a^r = \sqrt[n]{a^m}$.

Demnach ist z.B. $2^{0{,}2} = 2^{\frac{2}{10}} = 2^{\frac{1}{5}} = \sqrt[5]{2} = 1{,}148\cdots$. Mit ausreichender Geduld lassen sich mittels $\mathsf{R}_1 - \mathsf{R}_3$ alle anfänglich aufgelisteten Potenzgesetze für rationale Exponenten nachrechnen. Doch verwenden wir in Satz 7.3 eine ¨Überlegung, die uns – bis auf den Nachweis von $\mathsf{P}^+$ – jede Rechenarbeit erspart. In diesem Satz bezeichnen $r, s$ ausschließlich Elemente aus $\mathbb{Q}$. Wir benötigen dort außerdem

(13) $(1 + a)^{\varepsilon_n} \leq 1 + a\varepsilon_n$.

Nach der BERNOULLIschen Ungleichung ist nämlich $1 + a = 1 + 10^n a\varepsilon_n \leq (1 + a\varepsilon_n)^{10^n}$. Hieraus folgt (13) durch Ziehen der $10^n$-ten Wurzel auf beiden Seiten.

**Satz 7.3.** *Zu jeder positiven reellen Zahl $b$ gibt es genau eine Funktion $f : \mathbb{Q} \to \mathbb{D}$ mit den Eigenschaften*

$$\mathsf{P}^1: \quad f1 = b, \qquad \mathsf{P}^+: \quad f(r + s) = f(r) \cdot f(s),$$

*nämlich $f_b : r \mapsto b^r$ mit dem Definitionsbereich $\mathbb{Q}$. Darüber hinaus erfüllen die Funktionen $f_b$ alle weiteren Potenzgesetze. Schließlich ist $f_b$ für $b > 1$ schlicht und strikt wachsend; für $b < 1$ hat die Funktion $\frac{1}{f_b}$ diese Eigenschaften.*

**Beweis.** $f$ erfülle $\mathsf{P}^1, \mathsf{P}^+$. Dann ist, wie zu Beginn dieses Teilabschnitts schon festgestellt wurde, notwendigerweise $f\frac{m}{n} = \sqrt[n]{b^m}$. Denn man hätte dort doch nur überall $fr$ statt $b^r$ zu schreiben brauchen. Folglich ist $f = f_b$, und die Eindeutigkeit ist gezeigt. Gewiss gilt $\mathsf{P}^1$ für $f_b$. Sei nun $r = \frac{m}{n}$, $s = \frac{k}{n}$. Dann folgt $\mathsf{P}^+$ mit $\mathsf{R}_1, \mathsf{R}_2$ wie folgt:

$$b^{r+s} = b^{\frac{m}{n}+\frac{k}{n}} = b^{\frac{m+k}{n}} = \sqrt[n]{b^{m+k}} = \sqrt[n]{b^m \cdot b^k} = \sqrt[n]{b^m} \cdot \sqrt[n]{b^k} = b^r \cdot b^s.$$

Auch folgt $\mathsf{P}^<$ unmittelbar aus $\mathsf{R}_3$. Aufgrund von Bemerkung 2 verbleiben daher nur die Beweise von $\mathsf{P}_\times$ und $\mathsf{P}^\times$. Seien $a, b > 0$. Man betrachte $f : r \mapsto (ab)^r$ und $g : r \mapsto a^r \cdot b^r$. Beide Funktionen erfüllen offenbar $\mathsf{P}^1$, mit dem Wert $ab$ an der Stelle $r = 1$. Sie erfüllen auch $\mathsf{P}^+$. Für $f$ ist das klar, und nur für $g$ muss dies wirklich verifiziert werden:

$$g(r+s) = a^{r+s} \cdot b^{r+s} = a^r a^s b^r b^s = a^r b^r a^s b^s = g(r) \cdot g(s).$$

Nach der bereits bewiesenen Eindeutigkeitsaussage ist also notwendigerweise $f = g$ und $\mathsf{P}_\times$ ist bewiesen. Völlig analog zeigt man $\mathsf{P}^\times$ durch Vergleich von $f : s \mapsto (b^r)^s$ und $g : s \mapsto b^{rs}$ für festes $r \in \mathbb{Q}$, die bei $s = 1$ beide den Wert $b^r$ haben.

Schlichtheit folgt, wenn $f_b$ nur in jedem Intervall $[0, k]$ schlicht ist. Sei $b = 1 + a$ für ein $a \geq 0$, also $b^{\varepsilon_n} - 1 = (1 + a)^{\varepsilon_n} - 1 \leq a\varepsilon_n$ gemäß (13). Für $r \in \mathbb{Q} \cap [0, k]$ ist dann

$$\begin{aligned} b^{r+\varepsilon_n} - b^r \;=\; b^r b^{\varepsilon_n} - b^r &= b^r(b^{\varepsilon_n} - 1) \\ &\leq b^r a\varepsilon_n \\ &< b^k b\varepsilon_n \;=\; b^{k+1}\varepsilon_n \quad \text{(wegen } r \leq k \text{ und } a < b\text{)}. \end{aligned}$$

Folglich ist $f_b$ schlicht in $[0, k]$. Striktes Wachstum gilt wegen $\mathsf{P}^<$. Die Behauptung im Falle $b < 1$ folgt wegen $\frac{1}{b^r} = \left(\frac{1}{b}\right)^r$ in ganz entsprechender Weise. ■

Dieser Satz besagt zugleich, dass für rationale Exponenten alle Potenzregeln schon aus $\mathsf{P}^1$ und $\mathsf{P}^+$ folgen, d.h. werden $\mathsf{P}^1, \mathsf{P}^+$ von $f : \mathbb{Q} \to \mathbb{D}$ erfüllt, so erfüllt $f$ notwendigerweise auch alle übrigen Potenzgesetze.

Sei $b > 0$ eine beliebige reelle Zahl. Wir werden nun $b^x$ auch für irrationale $x$ erklären, und zwar einfach durch natürliche Fortsetzung. Nach der Eindeutigkeitsaussage von Satz 7.5 ist diese Art der Erklärung ohnehin die einzige, welche die Gültigkeit aller Potenzgesetze, ja nur der Basisregeln, für beliebige Exponenten zu sichern imstande wäre. Etwas genauer, wir erklären die Funktion $x \mapsto b^x$ für beliebige $x \geq 0$ durch natürliche Fortsetzung der entsprechenden, auf $\mathbb{E}$ eingeschränkten Funktion. Man benötigt dafür nur die Werte von $b^r$ für $r \in \mathbb{E}$, obwohl dieser Ausdruck für alle rationalen Exponenten eigentlich schon festgelegt wurde. Sie umfasst sozusagen noch einmal alle rationalen Exponenten $\frac{m}{n}$, die nicht zu $\mathbb{E}$ gehören. Nach einer Bemerkung vor Satz 7.2 folgt aber die Übereinstimmung von $a^{\frac{m}{n}}$ mit $\sqrt[n]{a^m}$ auch für $\frac{m}{n} \notin \mathbb{E}$. Man beachte, auch die automatischen Rechner bestimmen $b^x$ näherungsweise durch die Berechnung von $b^{x_{.n}}$ mit einer durch den Rechner gegebenen Stellenzahl $n$ solange sie numerisch rechnen, unabhängig davon ob $x$ rational ist oder nicht. Sie müssen $b^r$ nur für $r \in \mathbb{E}$ berechnen können.

**Definition.** Für $b \geq 1$ und $x \in \mathbb{D}$ sei $b^x = \lim\langle b^{x_{.n}}\rangle$. Für $0 < b < 1$ sei $b^x = \frac{1}{a^x}$, mit $a := \frac{1}{b}$. Ferner sei $0^x = 0$ für alle $x \neq 0$. Für $b > 0$ heißt $\exp_b : \mathbb{D} \to \mathbb{D}$ mit $\exp_b x = b^x$ die *Exponentialfunktion mit der Basis b*.

Die Definition liefert für den Fall $x \in \mathbb{E}$ (d.h. $x = x_{.n}$ für gewisses $n$) im Falle $b \geq 1$, und wegen $b^{x_{.n}} = \frac{1}{a^{x_{.n}}}$ nach Satz 7.3 auch im Falle $b < 1$ nichts neues. Daher gilt $b^x = \frac{1}{a^x}$ für alle $x \in \mathbb{D}$. Also steht uns bereits folgender Spezialfall von $\mathsf{P}_\times$ zur Verfügung:

(14) $\exp_{\frac{1}{b}} x = \frac{1}{\exp_b x}$ für alle $x \in \mathbb{D}$.

Weil ferner $g:\mathbb{E} \mapsto \mathbb{D}$ mit $gx = b^x$ nach Satz 7.3 für $b > 1$ schlicht und strikt monoton ist, erhalten wir nach Satz 7.2 dasselbe für deren natürliche Fortsetzung $\exp_b$, also das

**Korollar.** $\exp_b$ *ist für* $b > 1$ *schlicht und strikt monoton; ebenso* $\frac{1}{\exp_b}$ *für* $b < 1$.

Die Funktionenschar $\exp_b$ spielt in der Analysis eine fundamentale Rolle, wobei die Exponentialfunktion mit $b = \mathbf{e}$ aus mehreren Gründen besonders ausgezeichnet ist. Zum Beispiel ist sie die einzige durch den Punkt (0;1) verlaufene differenzierbare Funktion, die mit ihrer eigenen Ableitung übereinstimmt. Man schreibt meist nur exp für $\exp_{\mathbf{e}}$, so dass also $\exp(x) = \mathbf{e}^x$. Figur 2 gibt eine räumliche Darstellung dieser Funktionenschar für Werte der Basis zwischen $\frac{1}{\mathbf{e}} = 0{,}36\cdots$ und $\mathbf{e} = 2{,}71\cdots$ einschließlich, und dem Variablenbereich $0 \le x \le 2$. Es ist dies die auf der Fläche von links vorn nach rechts hinten gezeichnete Kurvenschar. Der rechte Rand des Flächenstücks zeigt gerade exp. Eine maßstabgetreue Abbildung zeigt auch Figur 7 in **9.3**, wo diese Funktion auf beliebige reelle Argumente unter Einschluss negativer Exponenten erweitert wird. $\exp_b$ wächst für $b > 1$ mit größer werdendem $x$ unbeschränkt und nimmt jeden Wert $\ge 1$ genau einmal an. Sie fällt anderseits für $b < 1$ sehr schnell auf nahezu 0 ab, wie dies die linke Randkurve des Flächenstücks verdeutlicht.

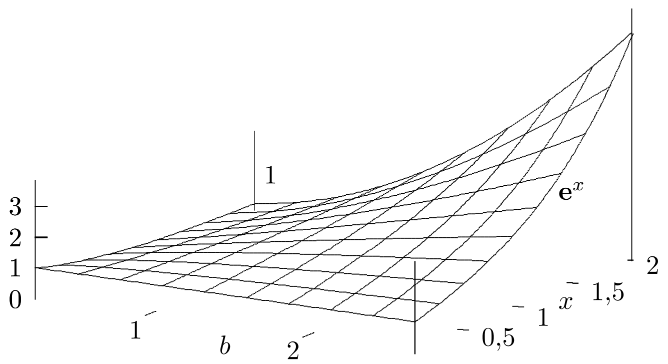



*Fig. 2 Die Schar der Exponentialfunktionen* $\exp_b$ *für* $\frac{1}{\mathbf{e}} \le b \le \mathbf{e}$

Es genügt, aus der Schar der Funktionen $\exp_b$ nur eine, etwa exp, zu kennen; die anderen kann man daraus leicht gewinnen. Denn offenbar ist

$$\exp(cx) = \mathbf{e}^{cx} = (\mathbf{e}^c)^x = b^x = \exp_b(x), \;\text{ mit } b = \mathbf{e}^c.$$

Für vorgegebenes $b > 1$ ist also $\exp_b$ für $b > 1$ durch exp gegeben; es ist nur das nach dem Zwischenwertsatz existierende, eindeutig bestimmte $c$ mit $b = \mathbf{e}^c$ zu berechnen, welches auch mit $\ln b$ bezeichnet wird, siehe **7.4**. Dieser enge Zusammenhang ist der Grund, warum einfach nur von *der* Exponentialfunktion gesprochen wird und womit in der Regel exp gemeint ist. Wegen (12) lässt sich auch die Exponentialfunktion mit einer Basis $< 1$ durch eine rationale Operation gewinnen. Nach Einführung der negativen Zahlen darf man auch schreiben $\exp_{\frac{1}{b}}(x) = \exp_b(-x)$, d.h. man wählt im Term $\exp(cx)$ einfach $c = -1$. Mit anderen Worten, $x \mapsto \mathbf{e}^{cx}$ mit $c \in \mathbb{R}$ ist die allgemeine Form einer Exponentialfunktion. Diese hat viele Darstellungsformen. Wir erwähnen vor allem

(15) $\mathbf{e}^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots,$

(16) $\mathbf{e}^x = \lim\langle(1+\frac{x}{n})^n\rangle,$

wobei die erste für die Analysis besonders wichtig ist. Die unendliche Reihe rechts in (15) ist Beispiel einer so genannten *Potenzreihe* und für jedes $x$ konvergent; denn deren Partialsummenfolge ist schlicht gemäß Übung 6.3. Dort wurde auch die Folge rechts in (16) als schlicht und damit konvergent erkannt.

Die Gleichung $\lim\langle(1+\frac{x}{n})^n\rangle = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots$ beweist man mittel der binomischen Formel fast genau so, wie dies in **7.1** für den Fall $x = 1$ geschah. Es genügt also, (16) auf elementarem Wege zu beweisen, um auch (15) ohne Differentialkalkül oder Potenzreihentheorie zu erhalten, siehe dazu Übung 4. Jede der beiden Gleichungen (15) und (16) eröffnet eine weitere Erklärungsmöglichkeit der Exponentialfunktion.

Mit einer bloßen Erklärung von $\exp_b$ oder $b^x$ ist allerdings noch nicht viel gewonnen. Worauf es vor allem ankommt ist der Nachweis der uneingeschränkten Gültigkeit aller Potenzgesetze. Wir werden diesen Nachweis nunmehr lückenlos erbringen und darüber hinaus beweisen, dass die Definition von $b^x$ durch natürliche Fortsetzung die *einzige* Erklärungsmöglichkeit darstellt, um auch nur einige der Potenzgesetze – nämlich die Basisregeln – zu erhalten. Letzteres ist nicht nur eine an sich interessante Tatsache, sondern sie wird uns helfen, ohne viel Rechnen auch die übrigen Potenzgesetze nachzuweisen. Dass diese dann eine notwendige Folge der Basisregeln sind, ergibt sich ganz nebenbei. Anders als in Satz 7.3 erscheinen in Satz 7.4 gewisse Ungleichungen. Wir erwähnen, dass diese oder äquivalente Bedingungen (z.B. die Monotonie oder Stetigkeit) für Satz 7.4 unverzichtbar sind. In Abschnitt **10** werden wir diesen Satz auf eine ganz andere Weise erhalten, nämlich als Nebenprodukt des Maßzahlensatzes.

**Satz 7.4.** *Zu jedem $b \in \mathbb{D}_+$ gibt es eine und nur eine Funktion $f : \mathbb{D} \to \mathbb{D}$ mit den Eigenschaften $\mathsf{P}_{\geq 1}$ : $fx \geq 1$ für $b \geq 1$ und $\mathsf{P}_{\leq 1}$ : $fx \leq 1$ für $b \leq 1$, sowie*

$$\mathsf{P}^1 : \ f1 = b, \qquad \mathsf{P}^+ : \ f(x+y) = f(x) \cdot f(y) \text{ für alle } x, y \in \mathbb{D},$$

*nämlich $f = \exp_b$. Darüber hinaus erfüllen die Funktionen $\exp_b$ alle Potenzgesetze.*

**Beweis.** Wir beweisen Eindeutigkeit zuerst und betrachten zunächst den Fall $b \geq 1$. $f$ erfülle die genannten Voraussetzungen. Die Einschränkung von $f$ auf $\mathbb{Q}$ genügt den Voraussetzungen von Satz 7.3. Also ist $f(\frac{m}{n}) = b^{\frac{m}{n}}$. Auch ist diese Einschränkung schlicht, und dasselbe gilt natürlich auch für die mit $g$ bezeichnete Einschränkung von $f$ auf $\mathbb{E}$. Weil die (gewöhnliche) Monotonie von $f$ nach Bemerkung 1 aus den Basisregeln direkt folgt – und nur hier benötigen wir die Eigenschaft $\mathsf{P}_{\geq 1}$ – ist $f$ monoton, und somit nach Satz 7.2 gerade die natürliche Fortsetzung von $g$; kurzum, $f = \exp_b$. Damit ist die Eindeutigkeit bewiesen und auch $\exp_b(\frac{m}{n}) = b^{\frac{m}{n}}$ ist gezeigt. Dasselbe gilt im Falle $b < 1$; denn $\frac{1}{\exp_b}$, und gemäß $\mathsf{P}_{\leq 1}$ auch $\frac{1}{f}$, sind monoton wachsende Fortsetzungen der nach Satz 7.3 schlichten Funktion $\frac{1}{g}$, und damit identisch.

Es verbleibt der Nachweis, dass die $\exp_b$ tatsächlich alle Potenzgesetze erfüllen. $\mathsf{P}^1$ ist klar. $\mathsf{P}^+$ ergibt sich im Falle einer Basis $b \geq 1$ wie folgt:

$$\begin{aligned} b^x \cdot b^y &= \lim\langle b^{x.n}\rangle \cdot \lim\langle b^{y.n}\rangle && \text{(Definition)}\\ &= \lim\langle b^{x.n} \cdot b^{y.n}\rangle && \text{(Spezialfall von Übung 5.3)}\\ &= \lim\langle b^{x.n+y.n}\rangle && \text{(Satz 7.3)}\\ &= b^{x+y} && \text{(wegen (11) und weil } x+y = \lim\langle x_{.n} + y_{.n}\rangle\text{)}. \end{aligned}$$

Der Fall $b < 1$ lässt sich auf den behandelten sofort zurückführen; denn ist $a = \frac{1}{b}$, so gilt $b^x = \frac{1}{a^x}$ nach (14), und weil $a > 1$, folgt $b^x \cdot b^y = \frac{1}{a^x} \cdot \frac{1}{a^y} = \frac{1}{a^{x+y}} = b^{x+y}$.

Nach Satz 7.3 gilt für die Einschränkung $g$ von $\exp_b$ auf $\mathbb{E}$ sicher $gx > 1$ für $b > 1$, und $gx < 1$ für $b < 1$, wenn immer $x \in \mathbb{E}_+$. Dasselbe gilt dann offenbar auch für die natürliche Fortsetzung von $g$. Damit erhalten wir nicht nur $P_{\geq 1}, P_{\leq 1}$ sondern leicht auch $P^<$. Es verbleiben nach Bemerkung 2 also lediglich die Nachweise von $\mathsf{P}_\times$ und $\mathsf{P}^\times$. Hierfür verwenden wir ähnlich wie in Satz 7.3 die schon bewiesene Eindeutigkeitsaussage. Für $\mathsf{P}_\times$ betrachte man im Falle $a, b \geq 1$ oder $a, b < 1$ die Funktionen $f : x \mapsto (ab)^x$ und $g : x \mapsto a^x \cdot b^x$ zuerst für $a, b \geq 1$. Sowohl $f$ als auch $g$ erfüllen sicher $\mathsf{P}^1$ (mit $ab$ für $b$). Beide Funktionen erfüllen offenbar auch $\mathsf{P}^+$, sowie die Bedingungen $P_{\geq 1}$ und $P_{\leq 1}$. Damit folgt $f = g$, und $\mathsf{P}_\times$ ist für diesen Fall gezeigt. Die verbleibenden Fälle $a < 1 \leq ab \leq b$ und $a \leq ab < 1 \leq b$ lassen sich auf den behandelten leicht zurückführen. Zum Beispiel ist im ersten Fall nach dem Bewiesenen $b^x = (ab)^x(\frac{1}{a})^x = (ab)^x\frac{1}{a^x}$. In derselben Weise zeigt man $\mathsf{P}^\times$ durch Betrachtung von $f : x \mapsto (b^y)^x$ und $g : x \mapsto b^{y \cdot x}$. ■

## 7.4 Logarithmen

In der Vergangenheit war das logarithmische Rechnen, d.h. die Verwandlung von Multiplikationen und Divisionen in Additionen und Subtraktionen, eine der wesentlichen Anwendungen der Logarithmen. Heute ist diese Art des Rechnens kaum noch üblich. Aber das Vorhandensein der Potenzen und Logarithmen auf Taschenrechnern deutet darauf hin, dass sie unabhängig davon ein breites Anwendungsspektrum haben. Das rührt vor allem daher, dass in Wissenschaft und Technik häufig Formeln verwendet werden, die diese Funktionen explizit enthalten. Uns geht es hier weniger um eine Auflistung von Anwendungen, sondern um die Klärung des Begriffs selbst. Ein beliebtes Verfahren ist z.B. die Einführung der logarithmischen Funktion ln durch

$$\ln x = \int_1^x \frac{dt}{t} \qquad (x > 0).$$

Das setzt natürlich voraus, dass der Integralkalkül voll entwickelt ist, weil aus dieser Definition die Eigenschaften von ln herzuleiten sind. Die wohl eleganteste Einführung von ln ist ihre Definition als gewisser Isomorphismus von der multiplikativen Gruppe der positiven Zahlen in die additive Gruppe der reellen Zahlen, siehe hierzu **10.6**.

Nach den bisherigen Vorbereitungen können wir indes die Logarithmusfunktion direkt als Umkehrung der Exponentialfunktion erklären. Für $b > 1$ ist $\exp_b$ gemäß dem Korollar in **7.3** schlicht, strikt monoton wachsend und unbeschränkt. Nach dem Zwischenwertsatz existiert also zu jedem $y \geq 1$ genau ein $x \geq 0$ mit $b^x = y$. Ist aber $0 < b < 1$, so hat $x \mapsto \frac{1}{\exp_b x}$ diese Eigenschaften. Daraus folgt offensichtlich, dass $\exp_b$ jetzt jeden Wert $y \leq 1$ genau einmal annimmt.

**Definition.** Sei $b > 1$ und $x \geq 1$ oder aber $b < 1$ und $x \leq 1$. Dann sei $\log_b x$ der eindeutig bestimmte Wert $y$ derart, dass $b^y = x$. Die Funktion $x \mapsto \log_b x$ heißt die *Logarithmusfunktion zur Basis b*.

Für $\log_{10}$ schreibt man häufig lg (die BRIGG*schen Logarithmen*). Für $\log_2$ schreibt man meist ld (*logarithmus dualis*). Im Falle $b = \mathbf{e}$ spricht man vom *natürlichen Logarithmus* (*logarithmus naturalis*) und schreibt ln oder log für $\log_\mathbf{e}$. Man beachte, dass gemäß Definition stets $\log_b 1 = 0$, $\log_b b = 1$, und allgemeiner $\log_b b^n = n$. Konkrete Zahlenwerte sind z.B. $\ln 10 = 2{,}30258\cdots$ und $\log_8 10 = \frac{1}{\lg 8} = 1{,}1073\cdots$. Man beachte, dass stets $\log_b a \cdot \log_a b = \log_b(b) = 1$ gemäß $\mathsf{L}_3$ unten. Figur 3 am Schluss veranschaulicht die Funktion ld, das aus dem Bild von $\exp_2$ durch Spiegelung an der Achse $x = y$ entsteht. In **9** werden beide Funktionen erweitert. Dabei wird sich herausstellen, dass $\operatorname{ld} x$ für $0 < x < 1$ in natürlicher Weise nur als eine negative Zahl erklärt werden kann.

Aus der Definitionsidentität $b^{\log_b x} = x$ ergeben sich mit Hilfe der Potenzgesetze einige für das Rechnen mit Logarithmen maßgebliche Gleichungen, nämlich

$$\begin{aligned}
&\mathsf{L}_1: \quad \log_b xy = \log_b x + \log_b y,\\
&\mathsf{L}_2: \quad \log_b x^y = y \cdot \log_b x,\\
&\mathsf{L}_3: \quad \log_a x = \tfrac{\log_b x}{\log_b a} \quad \text{(insbesondere } \log_a b = \tfrac{1}{\log_b a}\text{)}.
\end{aligned}$$

Es folgt $\mathsf{L}_1$ aus $b^{\log_b xy} = xy = b^{\log_b x} {\cdot} b^{\log_b y} = b^{\log_b x + \log_b y}$ durch Vergleich der Exponenten, und $\mathsf{L}_2$ aus $b^{\log_b x^y} = x^y = (b^{\log_b x})^y = b^{y \cdot \log_b x}$. Schließlich ergibt sich $\mathsf{L}_3$ aus

$$b^{\log_b x} = x = a^{\log_a x} = (b^{\log_b a})^{\log_a x} = b^{\log_b a \ \cdot \ \log_a x}.$$

Gleichung $\mathsf{L}_1$ ist die Grundlage des logarithmischen Rechnens mittels einer Tafel. Um gegebene Werte $x, y$ zu multiplizieren, entnimmt man der Tafel die Werte $\log_b x$ und $\log_b y$, addiert diese und erhält $\log_b xy$. Sodann liest man aus der Tafel den Wert $x \cdot y$ ab. $\mathsf{L}_2$ sorgt dafür, dass Logarithmen nur in einem engeren Bereich tabuliert werden müssen, etwa für Werte $x$ mit $1 \leq x < 10$ bei $b = 10$; denn jedes $x \geq 10$ lässt sich für geeignetes $x'$ als $x = x' \cdot 10^n$ mit $1 \leq x' < 10$ schreiben, was auf bloße Kommaverschiebung in $x$ hinausläuft. $\mathsf{L}_3$ schließlich dient der Umrechnung der Logarithmen von einer Basis in die andere. Gemäß $\mathsf{L}_3$ ist z.B. $\lg x = \frac{1}{\ln 10} \cdot \ln x = \frac{1}{2{,}30\cdots} \cdot \ln x$ für alle $x \geq 1$.

Bei obigen Formeln und den Beweisen ist vorerst darauf zu achten, dass die Argumente jeweils im Definitionsbereich von $\log_b$ liegen, also z.B. $x, y \geq 1$ für eine Basis $b > 1$ in $\mathsf{L}_1$. Das bedeutet natürlich eine erhebliche Behinderung. Denn sobald ein Faktor $< 1$ ist, wäre man mit der Kunst des logarithmischen Multiplizierens am Ende, und die Wahl einer Basis $< 1$ schafft entsprechende Probleme mit Faktoren $> 1$. Hier also tritt die Notwendigkeit der Einbeziehung negativer Zahlen in den Kalkül mit aller Deutlichkeit hervor. Wären diese nicht schon da gewesen, hätte man sie nun erfinden müssen. Historisch gesehen waren es in der Tat die Logarithmen, durch die alle Vorbehalte gegen negative Zahlen endgültig schwanden. Allerdings definierte man bis ins 19. Jahrhundert hinein man negative Zahlen vorwiegend geometrisch und nur teilweise algebraisch, aber im Grunde nur intuitiv. Deswegen ist die Ablehnung negativer Zahlen durch VIETA (1540 – 1603), der die „Buchstabenrechnung" erfand, durchaus verzeihlich.

Taschenrechner ermitteln $\log_b x$ bekanntlich in Sekundenbruchteilen. Auch die Erfinder der ersten Logarithmentafel haben weniger gerechnet als man zunächst glaubt. Das beruht wesentlich auf einer geschickten Basiswahl. In allen Fällen lag und liegt diese auch in modernen Taschenrechnern in der Nähe von $\mathbf{e}$ oder dem Reziproken $\frac{1}{\mathbf{e}}$. Sieht man von Unerheblichkeiten ab, so hatte die erste Logarithmentafel von NAPIER im Jahre 1614 die Basis $(1-\frac{1}{10\,000\,000})^{10\,000\,000}$, welche näherungsweise $=\frac{1}{\mathbf{e}}$ ist, Übung 6.2. Für die Tafel von BÜRGI (1620) war sie $(1+\frac{1}{10\,000})^{10\,000}$ $(=2{,}71814\cdots$ gemäß (e4) in **7.1**). Diese stimmt immerhin in den ersten drei Kommastellen mit $\mathbf{e}$ überein.

Die Basis $b_n := e_{10^n} = (1+\frac{1}{10^n})^{10^n} = (1+\varepsilon_n)^{10^n}$ ist für die Berechnung einer $n$-stelligen Logarithmentafel u.a. deswegen sehr geeignet, weil $\vartriangle\log_{b_n} x$ nahezu gleich $\frac{\vartriangle x}{x}$ ist für kleine Schrittweiten $\vartriangle x$ von $x$, Übung 5. Um diesen Vorteil für $x=1$ etwas zu erläutern, setze man $x_k := (1+\varepsilon_n)^k$. Weil $b_n^{k\varepsilon_n} = (1+\varepsilon_n)^{10^n\cdot k\varepsilon_n} = x_k$, folgt $\log_{b_n} x_k = k\varepsilon_n$ und die Schrittweite dieser Werte für $k=0,1,2,\ldots$ ist gerade $\varepsilon_n$. Man erhält so den durch die Tabelle unten beschriebenen Anfang einer Logarithmentafel. Die Zahlenwerte in Klammern entsprechen der Wahl $n=4$, also denen einer 4-stelligen Tafel.

| $x$ | $(n=4)$ | $\log_{b_n} x$ | $(\log_{b_4} x)$ |
|---|---|---|---|
| 1 | (1,0000) | 0 | 0 |
| $1+\varepsilon_n$ | (1,0001) | $\varepsilon_n$ | (0,0001) |
| $(1+\varepsilon_n)^2$ | (1,0002) | $2\varepsilon_n$ | (0,0002) |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $(1+\varepsilon_n)^{100}$ | (1,0100) | $100\varepsilon_n$ | (0,0100) |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $(1+\varepsilon_n)^{141}$ | (1,0141) | $141\varepsilon_n$ | (0,0141) |
| $(1+\varepsilon_n)^{142}$ | (1,0143) | $142\varepsilon_n$ | (0,0142) |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |

*Logarithmentafel zur Basis* $e_{10^n}$

Nun ist, wie aus den Zahlenwerten der Tafel ersichtlich, für kleine $k$ die Schrittweite $x_{k+1}-x_k$ des Arguments (oder *Numerus*) $x$ auch nur $\varepsilon_n$. Denn $x_k = (1+\varepsilon_n)^k$ ist für nicht zu große $k$ nahezu gleich $1+k\varepsilon_n$. Für $k^2 \le 10^n$ ist $x_k$ gleich $1+k\varepsilon_n$ sogar bis auf $n$ Stellen genau. Denn für $\varepsilon := \varepsilon_n$ ist dann $k^{1+i}\varepsilon^i \le 1$ für alle $i \ge 1$, und wir erhalten

$$(1+\varepsilon)^k - (1+k\varepsilon) = \tbinom{k}{2}\varepsilon^2 + \tbinom{k}{3}\varepsilon^3 + \ldots + \tbinom{k}{n}\varepsilon^n \le \varepsilon\big(\tfrac{k^2\varepsilon}{2!} + \tfrac{k^3\varepsilon^2}{3!} + \ldots + \tfrac{k^n\varepsilon^{n-1}}{n!}\big)$$

$$\le \varepsilon\big(\tfrac{1}{2!} + \ldots + \tfrac{1}{n!}\big) < \varepsilon(\mathbf{e}-2) < \varepsilon.$$

So ist z.B. noch $k^2\varepsilon_4 \le 1$ für $k \le 100$, also ist $(1+\varepsilon_4)^{100} = 1+100\varepsilon_4$ in den Grenzen 4-stelliger Genauigkeit. Dies gilt sogar noch für $k=141$, denn $(1+\varepsilon_4)^{141} = 1{,}0141991\cdots$.

Erst ab $k = 142$ beginnt der Numerus beim Erstellen der Tafel von rechts nach links allmählich zu springen. Aufgrund dieser Verhältnisse ist es bequemer, nicht den Logarithmus auszurechnen, sondern man schreibt diesen schrittweise hin und berechnet den zugehörigen Numerus, wobei sich in der angegebenen Weise und mit anderen Tricks der Rechenaufwand ziemlich gering halten lässt. Auch moderne Rechner gehen etwa in dieser Weise vor; durch blitzschnelles Ausrechnen der Exponentialfunktion suchen sie den der Eingabe am nächsten gelegenen Numerus.

Obige Tafel entspricht derjenigen von BÜRGI. NAPIERs Tafel hingegen kann (mit umgekehrtem Vorzeichen) näherungsweise als eine solche zur Basis **e** angesehen werden, weil $\log_{\frac{1}{b}} x = -\log_b x$, **9.3**. Es ist deshalb unerheblich, ob $(1 + \varepsilon_n)^{10^n}$ oder $(1 - \varepsilon_n)^{10^n}$ die Berechnungsbasis einer $n$-stelligen Tafel darstellt. Beide Werte sind in den Grenzen n-stelliger Genauigkeit reziprok zueinander; denn wegen $\mathsf{P}_\times$ ist unter Beachtung der Ungleichung (12) in **3.4**

$$1 > (1 + \varepsilon_n)^{10^n}(1 - \varepsilon_n)^{10^n} = (1 - \varepsilon_n^2)^{10^n} > 1 - 10^n\varepsilon_n^2 = 1 - \varepsilon_n.$$

Hat man keine Tafel oder Rechner zur Hand, hilft oft eine ¨Überschlagsrechnung zur Ermittlung des Logarithmus einer Zahl etwa bis auf etwa eine Stelle nach dem Komma. Will man z.B. $\lg 2$ ermitteln, gehe man aus von $2^9 < 10^3 < 2^{10}$. Das ergibt offenbar $9 \lg 2 < 3 < 10 \lg 2$, also $\frac{9}{10} \log 2 < 0{,}3 < \lg 2$. Demzufolge ist

$$0{,}3 < \lg 2 < \tfrac{10}{9} \cdot 0{,}3 = \tfrac{1}{3} = 0{,}333\cdots.$$

Sei $k \in \mathbb{N}$. Wir wollen die Anzahl $\ell_2 k$ der Binärziffern von $k$ abschätzen, siehe **1.2**. Dies ist wichtig für die Planung des Umfangs von Speicherzellen von Rechnern, die mit vorgeschriebener, z.B. 10-stelliger Genauigkeit rechnen sollen. Setzt man $m := \operatorname{Int} \operatorname{ld} k$, so folgt $2^m \leq 2^{\operatorname{ld} k}$ $(= k) < 2^{m+1}$. Weil nun $2^m$ genau $m + 1$ Binärziffern hat und $2^{m+1}$ kleinste natürliche Zahl ist mit einer Binärziffer mehr, ergibt sich offenbar $\ell_2 k = m + 1$, also $\ell_2 k = \operatorname{Int} \operatorname{ld} k + 1$. Alles eben Gesagte gilt nun ganz analog für jede Grundzahl $g \geq 2$. Damit erhalten wir für die $g$-adische Länge von $k$ die Formel

$$(17) \quad \ell_g k = \operatorname{Int} \log_g k + 1 \qquad (g \geq 2,\ k > 0).$$

Gesetzt, wir verfügen im Moment über keine Rechenhilfsmittel und sind genötigt, die binäre Länge $\ell_2 10^{10}$ überschlagsmäßig zu ermitteln. Soviel Bitplätze benötigt man gerade, um eine beliebige der $10^{10}$ natürlichen Zahlen $< 10^{10}$, wie sie in der Anzeige eines Taschenrechners mit 10-stelligem Display erscheinen, intern zu speichern. Wegen $2^3 < 10 < 2^4$ ist $3 < \operatorname{ld} 10 < 4$. Das aber genügt nicht, um $\ell_2 10^{10}$ zu bestimmen, denn dazu benötigen wir nach obiger Formel $\operatorname{Int} \operatorname{ld} 10^{10} = \operatorname{Int}(10 \cdot \operatorname{ld} 10)$, also mindestens eine Kommastelle von $\operatorname{ld} 10$. Diese ermitteln wir durch *lineare Interpolation*: Der Verlauf von ld zwischen den Punkten (8;3) und (16;4) unterscheidet sich nur wenig von dem einer diese Punkte verbindenden Geraden $g$, der *Interpolationsgeraden*, Figur 3. Eine elementare ¨Überlegung zeigt, dass der Punkt (10;3,25) auf $g$ liegt. ld verläuft in geringem Abstand oberhalb von $g$, so dass $\operatorname{ld} 10 \approx 3{,}3$ angenommen werden kann. Es ist $\ell_2 10^{10} = \operatorname{Int}(10 \cdot \operatorname{ld} 10) + 1$ und dies ist gemäß Schätzung der Wert $10 \cdot 3{,}3 + 1 = 34$. Tatsächlich ist $\operatorname{ld} 10 = \frac{1}{\lg 2} = 3{,}32192\cdots$, so dass diese Schätzung nahezu exakt ist.

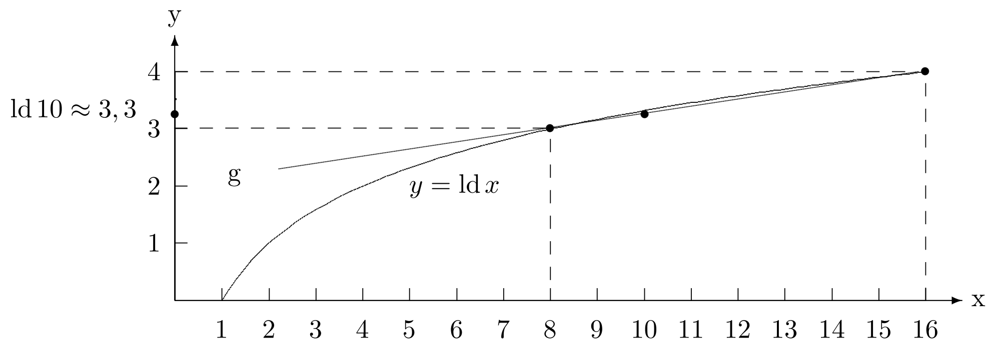



*Fig. 3 Die Funktion* ld *mit Interpolationsgerade durch* (8;3) *und* (16;4).

Wir haben hier stillschweigend benutzt, dass ld *konvex* ist, d.h. ist $1 \le a < x < b$ und $h$ die Interpolationsgerade durch die Punkte $(a; \mathrm{ld}\, a)$, $(b; \mathrm{ld}\, b)$, so ist $hx < \mathrm{ld}\, x$. Es genügt, dies für ln nachzuweisen (Übung 6); denn dasselbe gilt dann für alle Funktionen $\log_b$ mit $b > 1$, weil $\log_b x = c \cdot \ln x$ mit $c = \log_b \mathbf{e}$. Die Beobachtung, dass ld auf dem Bildschirm eines Rechners konvex erscheint, ist nur ein Hinweis, aber noch lange kein Beweis!

## 7.5 Übungen

1. Man zeige $\sqrt{2 \cdot \sqrt{2 \cdot \sqrt{2}}} \ldots = 2$, d.h. es ist zu zeigen, dass die (monotone) Folge $a_1 = \sqrt{2},\ a_2 = \sqrt{2\sqrt{2}},\ a_3 = \sqrt{2\sqrt{2\sqrt{2}}}, \ldots$ konvergiert und den Grenzwert 2 hat.

2. Sei $e_0 = 1,\ a_0 = 3,\ e_n = (1 + \frac{1}{n})^n,\ a_n = e_n(1 + \frac{1}{2n})\ (n > 0)$ und $b_n = e_n(1 + \frac{1}{2n+1})$. Man beweise: $\langle a_n \rangle$ konvergiert strikt fallend und $\langle b_n \rangle$ strikt wachsend gegen $\mathbf{e}$. Dies verdeutlicht, wie empfindlich das Monotonieverhalten gewisser Folgen ist.

3. Man beweise mit Hilfe von Übung 2 die Ungleichungen (e1) und (e2) in **7.1**.

4. Sei $f_n : x \mapsto (1 + \frac{x}{n})^n$ und $f : x \mapsto \lim \langle f_n x \rangle$. Man beweise auf elementarem Wege $fx = \mathbf{e}^x$, womit zugleich (15) und (16) in **7.3** bewiesen sind. (15) ergibt insbesondere $1 + x < \mathbf{e}^x$, oder gleichwertig, $\ln(1 + x) < x$, für alle $x \ge 0$.

5. Man beweise (a) $x - \frac{x^2}{2} < \ln(1 + x)$ für $x < 1$. Damit zeige man, für alle $x \ge 1$ und $0 < \vartriangle\!x < 1$ gilt (b) $0 < \frac{1}{x} - \frac{\Delta \ln x}{\vartriangle x} < \frac{\vartriangle x}{2x^2}$. Also ist $\frac{\Delta \ln x}{\vartriangle x} \approx \frac{1}{x}$, mit einem Fehler der Größenordnung $\vartriangle\!x$.

6. Sei $c > 0$. Man zeige (a) $f_c : x \mapsto (1 + \frac{c}{x})^x$ wächst monoton, oder gleichwertig, $x \mapsto (1 + cx)^{\frac{1}{x}}$ fällt monoton. (b) ln ist konvex. (c) Ist $\langle a_n \rangle$ eine (fallende) Nullfolge $(a_n \ne 0)$, so konvergiert $\langle \frac{\ln(1+a_n)}{a_n} \rangle$ wachsend gegen 1.

7. Sei $f$ die überall in $\mathbb{D}$ erklärte Funktion $x \mapsto \frac{2ax}{a+x^2}$ $(a > 0)$. Man zeige: $f$ ist schlicht im Intervall $[0, \sqrt{a}\,]$, und $x \le fx \le f(\sqrt{a}) = \sqrt{a}$ für alle $x \le \sqrt{a}$.

# Abschnitt 8

# Elementare Rechenverfahren

Dieser Abschnitt dient dem Verständnis praktisch wichtiger numerischer Algorithmen. Einige davon gehören in ihrer Grundstruktur zur Hardware automatischer Rechner, wie etwa der in **8.1** ausführlich behandelte Divisionsalgorithmus, eines der wichtigsten Rechenverfahren überhaupt. Bei seiner Ausführung muss nur addiert, subtrahiert und multipliziert werden. Mit demselben Verfahren bewerkstelligen die Computer auch die Umrechnung von einer Zahlendarstellung in die andere. Bekanntlich rechnen die meisten Computer binär. Um mit dem Nutzer kommunizieren zu können, müssen sie imstande sein, Zahldarstellungen in verschiedenen Systemen blitzschnell ineinander umzurechnen. Dies geschieht mittels des $g$-adischen Algorithmus, der nichts anderes ist als der Divisionsalgorithmus mit speziellen Parameterwerten, siehe **8.2**.

Eine elegante Verallgemeinerung des $g$-adischen Algorithmus ist der in **8.3** vorgestellte CANTORsche Algorithmus, bei dem die $g$-adischen Darstellungen als Sonderfälle eines allgemeineren Darstellungskonzepts erscheinen. Diese Algorithmen sind Beispiele so genannter *Iterationsverfahren*. Ein solches Verfahren ist grob dadurch gekennzeichnet, dass die Ausgangswerte einer oder mehrerer gegebener Operationen zu Eingangwerten derselben Operationen gemacht werden und die Schleife der wiederholten Anwendung derselben Operationen erst dann verlassen wird, wenn gewisse Bedingungen erfüllt sind, die bei jedem Durchlauf der Schleife vom Rechner zu testen sind. Auf diese Weise können z.B. gewisse Irrationalzahlen wie $\sqrt{2}$, $\boldsymbol{\pi}$ oder $\mathbf{e}$ heute problemlos und in kurzer Zeit auf Millionen von Dezimalstellen berechnet, Gleichungen mit einer riesigen Anzahl von Unbekannten gelöst und Differentialgleichungen näherungsweise integriert werden, deren praktische Lösung früher hoffnungslos war.

Mit den komplizierteren Rechenverfahren befasst sich die numerische Analysis, die durch die immer leistungsfähigeren Rechner zu einem faszinierenden Gebiet geworden ist. Die Begleitung der Theorie durch das Experiment am Computer kann unvermutete Sachverhalte offenbaren, auf die man sich allerdings erst dann wirklich verlassen kann, wenn sie mathematisch bewiesen sind. Diese Bemerkung betrifft natürlich auch andere Disziplinen, darunter nicht nur solche von mathematischem Charakter.

## 8.1 Der Divisionsalgorithmus

Es handelt sich hierbei um eines der grundlegenden Rechenverfahren, dessen Ausgangsobjekte in der einfachsten Version zwei natürliche Zahlen $a$ und $b$ sind, der *Dividend* und der *Divisor* und dessen Ergebnis eine abbrechende oder auch eine nichtabbrechende Dezimalzahl $u_0{,}u_1u_2\cdots$ ist, kurz notiert

$$a : b = u_0{,}u_1u_2\cdots.$$

Wir werden rigoros beweisen, und zwar ohne jeden Rückgriff auf unendliche Reihen, dass $u_0{,}u_1u_2\cdots$ der Quotient $\frac{a}{b}$ ist, also $\frac{a}{b} = u_0{,}u_1u_2\cdots$. Dies wird sogleich für beliebige Operanden $a, b \in \mathbb{D}$ nachgewiesen, mit der einzigen Einschränkung $b \neq 0$.

Für die Beschreibung des Algorithmus verwenden wir die Funktion Int. Mit ihrer Hilfe lässt sich für gegebene $a, b \in \mathbb{D}$ die größte natürliche Zahl $n$ bestimmen, so dass $nb \leq a$. Und zwar ist $n = \operatorname{Int}\frac{a}{b}$; denn $nb \leq a < (n+1)b$ ergibt nach Division durch $b$ die Ungleichung $n \leq \frac{a}{b} < n+1$, also $n = \operatorname{Int}\frac{a}{b}$. Im Hinblick auf eine im nächsten Teilabschnitt diskutierte Verallgemeinerung des Divisionsalgorithmus sei die Zahl 10 vorläufig mit $g$ bezeichnet. In **8.2** wird in diesem Algorithmus die Grundzahl $g = 10$ nämlich durch eine beliebige andere natürliche Zahl $\geq 2$ ersetzt werden.

Das Schema des schriftlichen Divisionsalgorithmus ist bekanntlich das folgende:

$$\begin{array}{l}
\ \ a \ : \ b \ = \ u_0{,}u_1u_2\cdots \\
-\underline{bu_0} \\
\quad r_0| \cdot g \\
\quad -\underline{bu_1} \\
\qquad\ \ r_1| \cdot g \\
\qquad\qquad \ddots
\end{array}$$

Das bedeutet im einzelnen: Zuerst wird die größte natürliche Zahl $u_0$ bestimmt, so dass gerade noch $bu_0 \leq a$. Mit anderen Worten, $u_0 = \operatorname{Int}\frac{a}{b}$. Weil $c - \operatorname{Int} c < 1$ für alle $c \in \mathbb{D}$, ist $r_0 := a - bu_0 = a - b\operatorname{Int}\frac{a}{b} = b(\frac{a}{b} - \operatorname{Int}\frac{a}{b}) < b$. Sodann wird die größte natürliche Zahl $u_1$ bestimmt mit $bu_1 \leq r_0 g$. Weil $\frac{r_0}{b} < 1$, ist $\frac{gr_0}{b} < g \ (= 10)$, also ist $u_1 = \operatorname{Int}\frac{gr_0}{b}$ eine Ziffer. Analog werden $r_1 := gr_0 - bu_1 \ (< b)$, $u_2 \ (< g)$, $r_2 \ (< b)$ bestimmt, usw. Formal lautet diese rekursive Definition der $u_n$ und $r_n$ wie folgt, wobei die Reste $r_n$ im Ergebnis nicht erscheinen und nur eine Hilfsrolle spielen:

$$(1) \qquad \begin{array}{rclrcll}
u_0 & = & \operatorname{Int}\frac{a}{b}, & r_0 & = & a - bu_0 & \big(= b(\frac{a}{b} - \operatorname{Int}\frac{a}{b})\big); \\
u_{n+1} & = & \operatorname{Int}\frac{gr_n}{b}, & r_{n+1} & = & r_n g - bu_{n+1} & \big(= b(\frac{gr_n}{b} - \operatorname{Int}\frac{gr_n}{b})\big).
\end{array}$$

Das nach (1) berechnete $u_{n+1}$ ist tatsächlich Dezimalziffer, also $u_{n+1} < g \ (= 10)$; denn $r_0 < b$ und ebenso $r_{n+1} = b(\frac{gr_n}{b} - \operatorname{Int}\frac{gr_n}{b}) < b$. Daher ist $\frac{r_n}{b} < 1$ für alle $n$, und folglich

$$u_{n+1} = \operatorname{Int}\frac{gr_n}{b} \leq \frac{gr_n}{b} = g\frac{r_n}{b} < g.$$

Offenbar ist eine schriftliche Division von Hand nichts anderes als ein ökonomisch geschriebenes Protokoll einer Berechnung gemäß den Rekursionsgleichungen (1).

Es ist leicht zu erkennen, dass die Ziffernfolge $\langle u_i \rangle$ des Ergebnisses effektiv bestimmt werden kann, wenn die Ziffern von $a$ schrittweise berechenbar sind und wenn der Divisor eine endliche Dezimalzahl ist. Hingegen ergeben sich bei einem beliebigen Divisor $b$ unter Umständen gewisse praktische Probleme, weil dann die Int-Operation in (1) möglicherweise nicht präzise genug ausgeführt werden kann. Schon um Int $\frac{a}{b}$ zu bestimmen, ist dasjenige $n$ mit $nb \le a < (n+1)b$ zu finden. Das gelingt im Falle $a > b$ durch schrittweise Subtraktion mit absoluter Sicherheit i.a. nur für abbrechende Operanden. Inwieweit man in der Praxis mit abbrechenden Näherungen von $b$ (oder auch $a$) rechnen darf, hängt natürlich von der Aufgabenstellung ab. Dies alles ist jedoch für die Theorie und den Beweis von Satz 8.1 unten völlig unerheblich.

Es seien nun $a$ und $b \neq 0$ gegeben und $\frac{a}{b} = z_0{,}z_1 z_2 \cdots$. Wir beweisen als erstes, dass dieser Algorithmus genau die Ziffern von $\frac{a}{b}$ liefert. Damit wird übrigens zugleich gezeigt, dass das Ergebnis des Algorithmus stets eine zulässige Dezimalzahl ist.

**Satz 8.1.** *Sei $a : b = u_0{,}u_1 u_2 \cdots$ das Ergebnis des Divisionsalgorithmus angewandt auf die Operanden $a \in \mathbb{D}$, $b \in \mathbb{D}_+$ und sei $\frac{a}{b} = z_0{,}z_1 z_2 \cdots$. Dann ist $u_n = z_n$ für alle $n$.*

**Beweis.** Wir verifizieren induktiv über $n$ die Gleichungen

$$(*) \qquad u_n = z_n \quad ; \quad r_n = b \cdot 0{,}z_{n+1} z_{n+2} \cdots$$

was die Behauptung des Satzes mit einschließt. Gewiss ist $u_0 = \operatorname{Int} \frac{a}{b} = z_0$ und daher

$$r_0 = b \cdot (\tfrac{a}{b} - u_0) = b \cdot (z_0{,}z_1 z_2 \cdots - z_0) = b \cdot 0{,}z_1 z_2 \cdots .$$

Sei $(*)$ für $n$ vorausgesetzt. Dann ist $\frac{g r_n}{b} = g \cdot 0{,}z_{n+1} z_{n+2} \cdots = z_{n+1}{,}z_{n+2} z_{n+3} \cdots$ und folglich $u_{n+1} = \operatorname{Int} \frac{g r_n}{b} = z_{n+1}$. Hieraus ergibt sich mit $r_n = b \cdot 0{,}z_{n+1} z_{n+2} \cdots$ dann auch

$$\begin{aligned} r_{n+1} &= g r_n - b \cdot u_{n+1} = b \cdot g \cdot 0{,}z_{n+1} z_{n+2} \cdots - b \cdot u_{n+1} \\ &= b(z_{n+1}{,}z_{n+2} z_{n+3} \cdots - u_{n+1}) \\ &= b \cdot 0{,}z_{n+2} z_{n+3} \cdots \qquad (\text{wegen } u_{n+1} = z_{n+1}). \quad \blacksquare \end{aligned}$$

Es ist klar, dass $\frac{a}{b} = \sup X$ mit $X = x \in \mathbb{D} \mid bx \le a$. Man bestätigt unschwer, dass der Divisionsalgorithmus im Grunde nur eine Spezialisierung des Berechnungsverfahrens der Ziffern von $\sup X$ gemäß dem Beweis von Satz 3.2 darstellt. Z.B. ist $u_0 = \operatorname{Int} \frac{a}{b}$ nichts anderes als die größte natürliche Zahl k, so dass gerade noch $k \le x$ für ein $x \in X$, ganz entsprechend der Definition von $z_0$ in Satz 3.2.

Der Divisionsalgorithmus kann auch in Gestalt eines so genannten *Flussdiagramms* veranschaulicht werden (Figur 4), einer nützlichen Vorstufe für dessen Programmierung auf einem Taschenrechner oder in irgendeiner Programmiersprache. Ein entsprechend dem Diagramm programmierter Rechner druckt die Ziffern von $\frac{a}{b} = u_0{,}u_1 u_2, \cdots$ schrittweise aus (PRINT-*Operation*). Er bricht die Rechnung ab, falls $\frac{a}{b} \in \mathbb{E}$, also nur endlich viele Ziffern zu bestimmen sind. Dies ist der Fall genau dann, wenn einer der Reste $r_n$ und damit alle nachfolgenden verschwinden, ebenso wie die Ziffern $u_{n+1}, u_{n+2}, \ldots$ Alle nicht durch INPUT belegten Variablen des Diagramms, nämlich $u, r, n$ haben anfänglich den Wert 0. Die Variable $g$ erhält zu Beginn den Wert 10. Sie wurde nur zwecks Nutzung des Algorithmus für die $g$-adische Entwicklung als Parameter eingeführt.

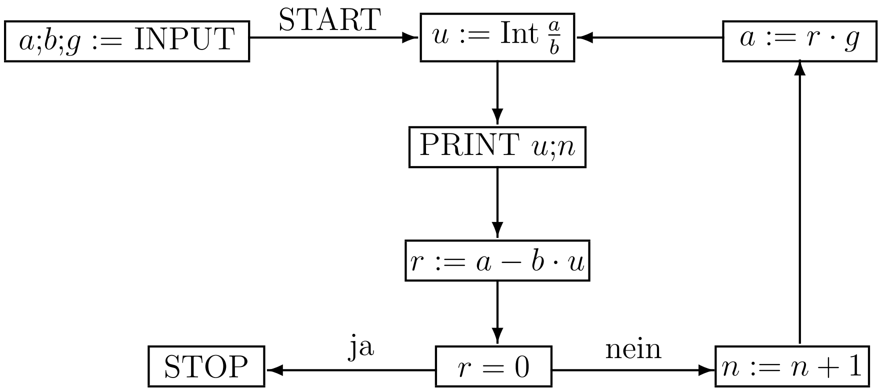



*Fig. 4 Flußdiagramm des Divisionsalgorithmus*

Rechtecke symbolisieren Operationsschritte. $\boxed{x := t}$ bedeutet, dass der Variablen $x$ der jeweilige Werte des Terms $t$ zugewiesen wird, wobei $x$ durchaus in $t$ vorkommen kann, wie etwa in $\boxed{n := n+1}$. Das Diagramm enthält die Testoperation $\boxed{r = 0\ ?}$, d.h. es ist die Frage ob $r$ im jeweiligen Rechenschritt den Wert 0 hat mit ja oder nein zu beantworten. Der Rechenfluss verzweigt sich entsprechend der Antwort. Im Verlaufe der Rechnung erhalten die Variablen i.a. neue Werte. Der Ablauf der Rechnung gemäß diesem Flussdiagramm ist einfach zu verfolgen und entspricht offenbar vollkommen dem Divisionsalgorithmus mit Abbruch der Rechnung, falls die Division „aufgeht“.

Man übersetzt das Diagramm sehr einfach in eine Programmiersprache (auch moderner Taschenrechner), z.B. durch Benutzung einer DO…UNTIL…END-Struktur und erhält so bei umsichtiger Formulierung des Programms die Dezimalen des Quotienten weit über die Stellenkapazität des Rechners hinaus. Das Diagramm nutzt zur Bestimmung von $\text{Int}\,\frac{a}{b}$ der Form nach die interne Rechnerdivision. Dies könnte wegen der Rundungsorganisation des Rechners zu Fehlern führen. Berücksichtigt man, dass $\text{Int}\,\frac{a}{b}$ nichts anderes ist als der Quotient bei der so genannten Division mit Rest, kann die besonders empfindliche Operation $\boxed{u := \text{Int}\,\frac{a}{b}}$ des Diagramms auch ohne Divisionen in einer Schleife fortlaufender Subtraktionen mit hoher Präzision ausgeführt werden.

Nun lässt sich der Divisionsalgorithmus natürlich auch so programmieren, dass das Verfahren bei überflüssigen Rechnungen überhaupt abbricht. Überflüssig wird weiteres Rechnen nicht nur wenn ein Rest $r$ gleich 0 wird, sondern auch dann, wenn einer der Reste einem früher berechneten gleicht. Sei nämlich $r_{q+p} = r_q$ für gewisse (minimal gewählte) $p, q$. Dann ist auch $r_{q+p+1} = r_{q+1}$, $r_{q+p+2} = r_{q+2}$ usw. Parallel dazu ergibt sich $u_{q+p+i} = u_{q+i}$ für $i = 1, 2, \ldots$ Kurzum, der Rechenprozess ist in eine Periode eingetreten und liefert die periodische Dezimalzahl $u_0{,}u_1 \cdots u_q\overline{u_{q+1} \cdots u_{q+p}}$ mit der Periodenlänge $p$. Es lässt sich durch eine auf der Hand liegende Erweiterung des obigen Diagramms problemlos erreichen, dass der Programmablauf bei Eintritt in eine Periode beendet wird und zwecks vollständiger Information zugleich die Periodenlänge ausgedruckt wird.

Periodizität tritt nun offenbar immer ein, wenn die Operanden $a, b$ natürliche Zahlen sind. Denn dann sind alle Reste $r_i$ natürliche Zahlen $< b$. Daher ist entweder einer der Reste $r_1, \dots, r_b$ (und damit alle nachfolgenden) gleich 0, oder aber alle diese Reste sind von 0 verschieden und mindestens zwei dieser Reste sind identisch. Damit haben wir nebenbei folgenden Satz bewiesen, der den Satz 6.2 gewissermaßen vervollständigt:

**Satz 8.2.** *Jede rationale Zahl* $\frac{m}{n}$ *ist periodisch. Dabei gilt* $p + q \leq n$ *und* $p < n$*, wobei* $p$ *die Periodenlänge,* $q$ *der Periodenindex ist.*

Den zweiten Teil des Satzes liefert ein einfaches Schubfach-Argument in der Betrachtung der Restfolge $r_0, r_1, \dots$ Übrigens folgt aus dem ersten Teil des Satzes mittels Formel (1) in Satz 6.2: Jede positive rationale Zahl $r$ hat mit geeignetem $m$ eine Darstellung

$$r = \frac{m}{(10^p-1)10^q} = \frac{m}{\underbrace{9\cdots9}_{p}\underbrace{0\cdots0}_{q}} \quad (p > 0).$$

Zum Beispiel sind $\frac{1}{13} = \frac{76923}{999999}$ und $\frac{1}{14} = \frac{714285}{9999990}$ solche Darstellungen, und zwar die kürzesten dieser Art. Es folgt hieraus offenbar gänzlich ohne Zahlentheorie, dass eine zu 10 teilerfremde natürliche Zahl Teiler ist von $10^p - 1$ für geeignetes $p$.

Wir illustrieren die Überlegenheit der programmierten Division an folgendem

**Beispiel.** Sei $a = 1{,}000\,009$ und $b = 17$. Division auf einem Rechner mit 13-stelligem Display liefert $1{,}000\,009 \div 17 = 0{,}058824\,058824\cdots$. Man könnte daher vermuten, dass die Periodenlänge des Quotienten 6 beträgt. Die genaue Berechnung nach einem Programm (oder eine Hobby-Rechnung von Hand) ergibt jedoch

$$\frac{1{,}000\,009}{17} = 0{,}058824\overline{0588235294117647}$$

mit der beim Nenner 17 maximalen Periodenlänge 16. Hierzu beachte man, dass die Periodenlängen von $r \in \mathbb{Q}$ und $r\overset{n}{\rightarrow}$ $(= 10^n r)$ übereinstimmen. Für $\frac{1000009}{17}$ ist diese aber $\leq 16$ gemäß Satz 8.2. Dass sie genau 16 ist, kann man mit etwas Zahlentheorie auch im voraus bestimmen: 1000009 und 17 sind teilerfremd und $p = 16$ ist das kleinste $p$, so dass $10^p - 1$ durch 17 teilbar ist; siehe hierzu Übung 2.

Eine Bemerkung zur Bestimmung des Quotienten $\frac{a}{b}$ für den Fall nichtabbrechender Operanden $a, b$. Dieser kann jedenfalls dann (näherungsweise) berechnet werden, wenn $a, b$ in dem Sinne berechenbar sind, dass durch ein Verfahren zu jedem $n$ Näherungen $a_n, b_n \in \mathbb{E}$ mit $|a - a_n|, |b - b_n| < \varepsilon_n$ effektiv bestimmt werden können, siehe auch **8.4**. Es lässt sich ähnlich wie für Summe und Produkt unschwer ausrechnen, wie groß $n$ sein muss, damit $|\frac{a}{b} - \frac{a_n}{b_n}| < \varepsilon_m$, wobei $\varepsilon_m$ die vorgeschriebene Genauigkeit sein soll. Allerdings handelt es sich hierbei nicht um die ziffernweise Berechenbarkeit. Denn ziffernweise Berechenbarkeit der Operanden überträgt sich i.a. auf keine der arithmetischen Operationen. Eine Zahl $a$ kann im erwähnten etwas allgemeineren Sinne berechenbar sein, ohne dass einem konkreten Berechnungsverfahren auch nur eine Information über den letztgültigen Wert von $\operatorname{Int} a$ entnommen werden kann, siehe **5.4**. Dieses Phänomen kann sehr störend sein bei Computerberechnungen mit empfindlichen Tests.

## 8.2 Der $g$-adische Algorithmus

Wie in **1.2** schon erwähnt wurde, hätte man sich auch auch auf ein $g$-adisches System mit einer Grundzahl $g \neq 10$ einigen können, und die Wahl $g = 8$ wäre in vieler Hinsicht sogar vorteilhafter gewesen. Unabhängig davon haben Computer-Experten bei komplexen Installationen und auch Nutzer (bei Änderungen der Konfiguration) häufig Veranlassung, Zahlen verschiedener $g$-adischer Systeme ineinander umzurechnen. Das betrifft vor allem die Grundzahlen $g = 2,\ g = 8,\ g = 10$ und $g = 16$. Im *Hexadezimalsystem* ($g = 16$) benutzt man neben den zehn Dezimalziffern die Buchstaben A, . . . ,F zur Bezeichnung der sechs fehlenden Ziffern. Diese Umrechnungen sind so von einer bloßen Spielerei zu einer praktisch wichtigen Angelegenheit geworden.

Jede positive natürliche Zahl $k$ hat im $g$-adischen System eine eindeutige Darstellung

$$k = z_0 g^n + z_1 g^{n-1} + \ldots + z_{n-1} g_0 + z_n = (z_0 z_1 \cdots z_n)_g \qquad (z_i < g,\ z_0 \neq 0),$$

die der dezimalen Darstellung vollkommen entspricht, siehe **11.6**. Wie man diese mit dem $g$-adischen Divisionsalgorithmus schnell errechnet, sehen wir weiter unten. Zuvor stellen wir die Frage, was entspricht den abbrechenden und nichtabbrechenden Dezimalzahlen im $g$-adischen System? Die Antwort ist im Prinzip einfach. Denn die in den Abschnitten **3 - 5** entwickelte Theorie der Dezimalzahlen kann ohne jede Änderung auf das $g$-adische System übertragen werden. Summe und Produkt lassen sich in derselben Weise definieren. Diese Bemerkung zeigt, dass nicht nur bezüglich der natürlichen Zahlen, sondern auch bezüglich der reellen Zahlen die Wahl von $g = 10$ nur die Wahl eines bestimmten Koordinatensystems ist. Alle wesentlichen Dinge bleiben bezüglich einer Transformation des Koordinatensystems invariant. Als Zugeständnis an die Tradition wollen wir jedoch an die Theorie der unendlichen Reihen anknüpfen und definieren $g$-adische reelle Zahlen durch unendliche Reihen der Gestalt $\sum_i \frac{z_i}{g^i} = z_0 + \frac{z_1}{g} + \frac{z_1}{g^2} + \ldots$ Dabei seien die $z_i$ für $i > 0$ Ziffern des $g$-adischen Systems, also natürliche Zahlen $< g$. Auch die natürliche Zahl $z_0$ sei $g$-adisch dargestellt. Obige Reihe ist nach den allgemeinen Ergebnissen in **7.1** konvergent und hat eine wohlbestimmte Summe $a$. Wir schreiben nun $z_0{,}z_1 z_2 \cdots_g$ für $z_0 + \frac{z_1}{g} + \frac{z_2}{g^2} + \ldots$ und nennen $z_0{,}z_1 z_2 \cdots_g$ eine *$g$-adische Darstellung* (oder Entwicklung) von $a$. Fragen, die deren Eindeutigkeit betreffen, erörtern wir am Schluss. In jedem Falle gilt $z_0{,}z_1 z_2 \cdots_g \xrightarrow{n} = z_0{,}z_1 z_2 \cdots_g \cdot g^n$, Übung 4.

Die erste sich hier erhebende Frage ist, ob nun auch jede positive reelle Zahl eine $g$-adische Darstellung besitzt. Angesichts der Vorbemerkung über das „$g$-adische Koordinatensystem“ ist dies zwar klar, doch wollen wir uns auf eine formale $g$-adische Darstellung ja nicht berufen. Daher muss diese Darstellbarkeit bewiesen werden.

Die dazu erforderlichen Hilfsmittel stehen aber zur Verfügung. Wir betrachten den Divisionsalgorithmus mit folgender Änderung: $g = 10$ wird durch die Basis $g$ des gewählten $g$-adischen Systems ersetzt. $a$ sei irgend eine zunächst in Dezimaldarstellung gegebene reelle Zahl und nunmehr sei von Beginn an $b = 1$. Sonst bleibt alles beim alten. Die Formeln (1) zur Berechnung der Folgen $u_n$ und $r_n$ spezialisieren sich dann zu

$$(2)\qquad \begin{array}{rclrcll} u_0 &=& \operatorname{Int} a, & r_0 &=& a - u_0 & (= a - \operatorname{Int} a);\\ u_{n+1} &=& \operatorname{Int}(gr_n), & r_{n+1} &=& gr_n - u_{n+1} & (= gr_n - \operatorname{Int} gr_n). \end{array}$$

Offenbar ist $r_n < 1$ für alle $n$. Also $gr_n < g$, so dass $u_{n+1} = \operatorname{Int} gr_n \le gr_n < g$ eine $g$-adische Ziffer ist. Damit ist klar, dass der so genutzte Algorithmus, angewandt auf $a$, nach Umrechnung der natürlichen Zahl $u_0 = \operatorname{Int} a$ in das $g$-adische System eine wohlbestimmte reelle Zahl $u_0{,}u_1u_2\cdots_g = \sum_i \frac{u_i}{g^i}$ liefert. Wir werden nun gleich sehen, dass dieses Ergebnis mit $a$ identisch ist, also gerade eine $g$-adische Darstellung von $a$ liefert. Deshalb spricht man nunmehr vom *$g$-adischen Algorithmus*.

**Satz 8.3.** *Sei $u_0{,}u_1u_2\cdots_g$ das Ergebnis des $g$-adischen Algorithmus, angewandt auf eine reelle Zahl $a \ge 0$. Dann ist $a = u_0{,}u_1u_2\cdots_g$ $(= \sum_i \frac{u_i}{g^i})$.*

**Beweis.** Zuerst beweisen wir durch Induktion über $n$ die Formel

$$(*)\qquad a = u_0 + \tfrac{u_1}{g^1} + \ldots + \tfrac{u_n}{g^n} + \tfrac{r_n}{g^n}.$$

Für $n = 0$ heißt dies $a = u_0 + r_0$, und dies gilt gemäß der ersten Zeile von (2). Sei $(*)$ für $n$ vorausgesetzt. Gemäß (2) ist $gr_n = u_{n+1} + r_{n+1}$, also $\frac{r_n}{g^n} = \frac{u_{n+1}}{g^{n+1}} + \frac{r_{n+1}}{g^{n+1}}$. Daher gilt $(*)$ auch für $n+1$, und folglich für alle $n$. Um nun die Richtigkeit von $a = \sum_i \frac{u_i}{g^i}$ einzusehen, hat man sich nur davon zu überzeugen, dass $\langle \frac{r_n}{g^n} \rangle$ eine Nullfolge ist. Das aber ist klar; denn wie oben schon festgestellt wurde, ist $r_n < 1$, also $\frac{r_n}{g^n} < \frac{1}{g^n}$. ■

**Beispiel.** Wir stellen $\boldsymbol{\pi} = 3{,}14159\,26535\,89\cdots$ [1] mit dem 8-adischen Algorithmus als Oktalzahl dar. Wegen der Pünktchen bei $\boldsymbol{\pi}$ kennen wir nur die ausgeschriebenen Dezimalen von $8r_0, 8r_1, \ldots$ genau; diese bleiben unverändert, gleichgültig ob die nächste Ziffer von $\boldsymbol{\pi}$ eine 0 oder eine 9 ist. Damit sind alle Vorkommaziffern von $8r_0$ bis $8r_{10}$ korrekt, und damit auch die angegebenen 11 Kommastellen des Ergebnisses.

$$\begin{array}{ll} \underline{3{,}141592653589\cdots} & = 3{,}11037552421\cdots_8\\ \underline{1{,}13274122871\cdots} & (= 8r_0)\\ \underline{1{,}061929829\cdots} & (= 8r_1)\\ \quad\vdots & \quad\vdots \end{array}$$

Die Umrechnung einer Dezimalzahl in das Oktalsystem ist sehr nützlich als Zwischenstufe für ihre Umrechnung in das Binärsystem. Man braucht nämlich deren oktale Darstellung nur ziffernweise umzuschreiben, wobei den Ziffern $0, 1, 2, 3, \ldots, 7$ der Reihe nach die Tripel $000, 001, 010, 011, \ldots, 111$ von Binärziffern entsprechen, Übung 3. Für $\boldsymbol{\pi}$ erhält man deshalb ohne jede weitere Rechnung aus dem obigen Beispiel

$$\boldsymbol{\pi} = 3{,}11037552421\cdots_8 = 11{,}001\,001\,000\,011\,111\,101\,101\,010\,100\,010\,001\cdots_2\,.$$

Damit erhält man zugleich auch die Umrechnung einer Dezimalzahl in das Hexadezimalsystem. Jede Hexadezimalziffer $0, \ldots, 9, \mathrm{A}, \ldots, \mathrm{F}$ entspricht nämlich genau einem

[1] $\boldsymbol{\pi}$ wurde inzwischen (2005) auf über 200 Milliarden Dezimalziffern berechnet. Eines ihrer vielen Geheimnisse ist, ob die Folge dieser Ziffern, oder besser, der Ziffern ihrer Binärdarstellung (unten) den Kriterien einer Zufallsfolge genügt, und wenn nicht, welches Muster sich in der Ziffernfolge verbirgt.

Quadrupel von Binärziffern. Man erhält so insbesondere

$$\pi = 11{,}0010\,0100\,0011\,1111\,0110\,1010\,1000\,1000\,1\cdots_2 = 3,243\mathrm{F}6\mathrm{A}88\cdots_{16}\,.$$

Betrachten wir nun den Divisionsalgorithmus mit gegebener Grundzahl $g \geq 2$ und heben die oben gemachte Einschränkung $b = 1$ wieder auf. Man spricht dann vom *$g$-adischen Divisionsalgorithmus*, denn das Resultat $u_0{,}u_1u_2\cdots_g$ ist eine $g$-adische Entwicklung von $\frac{a}{b}$. Dies beweist man wörtlich wie Satz 1, ausgehend von einer durch Satz 8.3 garantierten Darstellung $\frac{a}{b} = z_0{,}z_1z_2\cdots_g$, welche die für den Beweis benötigten Eigenschaften (z.B. $\operatorname{Int}\frac{a}{b} = z_0$) tatsächlich auch besitzt. Eine $g$-adische Darstellung von $\frac{a}{b}$ bei dezimal gegebenen Operanden lässt sich also direkt ermitteln, ohne $\frac{a}{b}$ zuvor dezimal auszurechnen und dann mit dem $g$-adischen Algorithmus umzuwandeln. Noch wichtiger aber ist, dass der $g$-adische Divisionsalgorithmus es erlaubt, zugleich auch die natürliche Zahl $\operatorname{Int} a$ (speziell also jedes $a \in \mathbb{N}$) in das $g$-adische System umzurechnen. Denn sei $a < g^{n+1}$, also $\frac{a}{g^n} < g$, und sei $a : g^n = u_0{,}u_1u_2\cdots_g$ Ergebnis der Anwendung des $g$-adischen Divisionsalgorithmus, so dass $a = g^n \cdot u_0{,}u_1u_2\cdots_g = u_0{,}u_1u_2\cdots_g \xrightarrow{n}$. Wegen $\frac{a}{g^n} < g$ ist nun auch $u_0 = \operatorname{Int}\frac{a}{g^n}$ eine $g$-adische Ziffer und wir erhalten

$$a = u_0{,}u_1u_2\cdots_g \xrightarrow{n} = u_0\cdots u_n{,}u_{n+1}u_{n+2}\cdots_g\,.$$

Folglich ist $\operatorname{Int} a = (u_0\cdots u_n)_g$. Das alles gilt insbesondere für $a = \operatorname{Int} a$, d.h. $a \in \mathbb{N}$.

**Beispiel.** Es ist $100{,}6 < 8^3$, also $\frac{100{,}6}{8^2} < 8$. Der 8-adische Divisionsalgorithmus liefert

$$\begin{array}{lllll}
100{,}6 & : \; 64 & = & 1{,}44\,4631\,4631\cdots_8 & \\
\underline{-64} & (= -u_0\cdot 64) & & & \\
\;\;36{,}6 & \cdot\, 8 & = & \;\;292{,}8 & (= 8r_0) \\
 & & & \underline{-256} & (= -u_1\cdot 64) \\
 & & & \;\;\;\;36{,}8 & \cdot\, 8 = \underline{294{,}4}\;(= 8r_1) \\
 & & & & \ddots
\end{array}$$

Es ergibt sich also $100{,}6 = 1{,}44\overline{4631}_8 \xrightarrow{2} = 144{,}\overline{4631}_8 = 1\cdot 8^2 + 4\cdot 8 + 4 + \frac{4}{8} + \frac{6}{8^2} + \ldots$

Wie sieht es nun mit der Eindeutigkeit der $g$-adischen Darstellung aus? Für $g = 10$ ist z.B. $\frac{1}{10} + \frac{0}{10^2} + \frac{0}{10^3} + \ldots = \frac{1}{10} = \frac{0}{10} + \frac{9}{10^2} + \frac{9}{10^3} + \ldots$ Eindeutigkeit kann daher sicher nur unter zusätzlichen Bedingungen erwartet werden. Verlangt man $z_i \neq 9$ für unendlich viele $i$ in einer Darstellung $a = \sum_i \frac{z_i}{10^i}$, dann sind die Ziffern $z_i$ in der Tat eindeutig bestimmt. Denn die Zusatzbedingung besagt für $g = 10$ nichts anderes als $z_0{,}z_1z_2\cdots$ ist zulässig; und weil $a = z_0{,}z_1z_2\cdots$, sind die $z_i$ eindeutig bestimmt. Für beliebiges $g \geq 2$ gilt ein völlig analoger Sachverhalt: Sind $z_0$ und $z_0'$ natürliche Zahlen und sind $\langle z_n\rangle_{n>0}$ und $\langle z_n'\rangle_{n>0}$ Folgen $g$-adischer Ziffern, so dass $z_i \neq g-1$ für unendlich viele Indizes $i$ und gilt analoges für die $z_i'$, so impliziert $\sum_i \frac{z_i}{g^i} = \sum_i \frac{z_i'}{g^i}$ auch $z_i = z_i'$ für alle $i$. Den Beweis hierfür liefern wir in allgemeinerem Zusammenhang in **8.3**.

Man kann auch Eindeutigkeit erzwingen, indem man verlangt, dass in einer Darstellung $a = \sum_i \frac{z_i}{g^i}$ alle Ziffern $z_i$ ab einer gewissen Stelle identisch $g-1$ sind. Eine solche Darstellung existiert für eine beliebige Grundzahl $g \geq 2$. Z.B. ist $0{,}7_8 = 0{,}6777\cdots_8$.

## 8.3 Der CANTORsche Algorithmus

G. CANTOR hat einige wichtige Beiträge zur Theorie der unendlichen Reihen geliefert und unter anderem eine interessante Verallgemeinerung des $g$-adischen Algorithmus angegeben. Die folgenden Darlegungen hierüber verwenden nur einige Grundtatsachen über unendliche Reihen mit positiven Gliedern. Negative Zahlen treten hierbei nach wie vor nicht in Erscheinung.

Es sei $g_1, g_2, \ldots$ eine Folge natürlicher Zahlen $\geq 2$, die wir uns im folgenden beliebig, aber fest gewählt denken und eine CANTOR*sche Basis* nennen. Der CANTORsche Algorithmus verläuft wie der $g$-adische Algorithmus, nur wird im $n$-ten Schritt die Rolle der Zahl $g$ von $g_n$ wahrgenommen. Im Flussdiagramm der Figur 4 lässt sich dies durch Einfügung der Operation $\boxed{g := g_n}$ nach Ausführung von $\boxed{n := n+1}$ und Einfügung eines Unterprogramms zur Berechnung der $g_n$ problemlos berücksichtigen. Die Rekursionsformeln (2) zur Berechnung von $u_n$ und $r_n$ nehmen dann für den CANTORschen Algorithmus die folgende Gestalt an, wobei $a \in \mathbb{D}$ beliebig vorgegeben ist:

$$(3)\quad \begin{array}{rclrcll} u_0 &=& \operatorname{Int} a, & r_0 &=& a - u_0 & \big(= a - \operatorname{Int} a\big); \\ u_{n+1} &=& \operatorname{Int}(g_{n+1} r_n), & r_{n+1} &=& g_{n+1} r_n - u_{n+1} & \big(= g_{n+1} r_n - \operatorname{Int}(g_{n+1} r_n)\big). \end{array}$$

Diese produzieren eine Folge $\langle u_n \rangle$ natürlicher Zahlen mit $u_n < g_n$ für $n > 0$. Denn offenbar ist $r_n < 1$ für jedes $n$, und daraus folgt $u_{n+1} \leq g_{n+1} r_n < g_{n+1}$. Der $g$-adische Algorithmus ist demnach der Spezialfall des CANTORschen für $g_1 = g_2 = \ldots = g$. Natürlich kann im CANTOR'schen Algorithmus von den $u_n$ für $n > 0$ nur noch in sehr allgemeinem Sinne als den „Ziffern“ gesprochen werden. (3) impliziert offenbar $r_n = \frac{u_{n+1}}{g_{n+1}} + \frac{r_{n+1}}{g_{n+1}}$, also $a = u_0 + r_0 = u_0 + \frac{u_1}{g_1} + \frac{r_1}{g_1} = u_0 + \frac{u_1}{g_1} + \frac{u_2}{g_1 \cdot g_2} + \frac{r_2}{g_1 \cdot g_2}$ und allgemein

$$a = u_0 + \tfrac{u_1}{g_1} + \tfrac{u_2}{g_1 \cdot g_2} + \ldots + \tfrac{u_n}{g_1 \cdot \ldots \cdot g_n} + \tfrac{r_n}{g_1 \cdot \ldots \cdot g_n} \qquad (r_n < 1).$$

Setzt man zusätzlich $g_0 := 1$ und $G_n := g_0 \cdot \ldots \cdot g_n$, lässt sich dies auch schreiben als

$$(4)\qquad a = \textstyle\sum_{i \leq n} \frac{u_i}{G_i} + \frac{r_n}{G_n} \qquad (r_n < 1).$$

Daraus entnimmt man, dass die unendliche Reihe $\sum_i \frac{u_i}{G_i} = u_0 + \frac{u_1}{G_1} + \frac{u_2}{G_2} + \ldots$ konvergiert. Sie hat außerdem gerade den Wert $a$. Denn die jeweils letzten Summanden $\frac{r_n}{G_n}$ in (4) bilden eine Nullfolge, weil $G_n \geq 2^n$ für alle $n$, und damit $\frac{r_n}{G_n} < \frac{1}{G_n} \leq \frac{1}{2^n}$.

Die Darstellung $a = \sum_i \frac{u_i}{G_i}$ heiße die CANTOR*sche Entwicklung von a*. Wir wollen zuerst zeigen, dass diese noch die zusätzliche Eigenschaft $u_i \neq g_i - 1$ für unendlich viele Indizes $i$ besitzt, also im übertragenen Sinne zulässig ist. Denn angenommen, dies ist nicht der Fall, etwa $u_i = g_i - 1$ für alle $i > n$. Dann gilt $r_{i+1} = g_{i+1} \cdot r_i - (g_{i+1} - 1)$ für $i \geq n$ gemäß (3). Eine leichte Umformung hiervon ergibt $1 - r_i = \frac{1 - r_{i+1}}{g_{i+1}}$ und daher

$$1 - r_n = \tfrac{1 - r_{n+1}}{g_{n+1}} = \tfrac{1 - r_{n+2}}{g_{n+1} \cdot g_{n+2}} = \ldots = \tfrac{1 - r_{n+k}}{g_{n+1} \cdot \ldots \cdot g_{n+k}} \leq \tfrac{1}{g_{n+1} \cdot \ldots \cdot g_{n+k}} \leq \tfrac{1}{2^k}.$$

Diese Ungleichung gilt für beliebige $k$, woraus offenbar $1 - r_n = 0$ folgt, d.h. $r_n = 1$. Dies aber steht im Widerspruch zu $r_n < 1$. Es muss in der Tat also unendlich viele Indizes $i$ geben mit $u_i < g_i - 1$. Man beachte, dass damit auch eine Aussage über das Ergebnis

$\sum_i \frac{u_i}{g^i}$ des $g$-adischen Algorithmus gemacht wird, nämlich $u_i \neq g-1$ für unendlich viele Indizes $i$. Für $g = 10$ ist dies natürlich nach Satz 8.1 schon bekannt.

Unter einer CANTOR*schen Reihe* bezüglich einer gegebenen CANTORschen Basis $\langle g_n \rangle$ sei nun eine unendliche Reihe der Gestalt verstanden, wie sie sich beim CANTORschen Algorithmus ergibt, also eine unendliche Reihe folgender Gestalt, wobei die $z_i$ sämtlich natürliche Zahlen sind, und wie oben $G_n = g_0 \cdot \ldots \cdot g_n$ ($G_0 = g_0 = 1$) gesetzt sei:

$$(5) \quad z_0 + \tfrac{z_1}{G_1} + \tfrac{z_2}{G_2} + \ldots \quad (z_i < g_i \text{ für } i > 0;\ z_i \neq g_i - 1 \text{ für unendlich viele } i).$$

Bisher wurde nur bewiesen, dass eine derartige Reihe konvergiert, sofern sie Ergebnis des CANTORschen Algorithmus ist. Ferner zeigt dieser Algorithmus, dass jede reelle Zahl $a \geq 0$ auf wenigstens eine Weise als CANTORsche Reihe darstellbar ist. Nun ist aber jede CANTORsche Reihe konvergent; denn wegen $\frac{z_{i+1}}{g_{i+1}} < 1$ und $G_n \geq 2^n$ ist

$$\begin{aligned} z_0 + \tfrac{z_1}{G_1} + \tfrac{z_2}{G_2} + \ldots + \tfrac{z_n}{G_n} &= z_0 + \tfrac{z_1}{g_1} + \tfrac{1}{G_1} \cdot \tfrac{z_2}{g_2} + \tfrac{1}{G_2} \cdot \tfrac{z_3}{g_3} + \ldots + \tfrac{1}{G_{n-1}} \cdot \tfrac{z_n}{g_n} \\ &< z_0 + 1 + \tfrac{1}{G_1} + \tfrac{1}{G_2} + \ldots + \tfrac{1}{G_{n-1}} \\ &\leq z_0 + 1 + \tfrac{1}{2} + \tfrac{1}{2^2} + \ldots + \tfrac{1}{2^{n-1}} < z_0 + 2. \end{aligned}$$

Wir benötigen im folgenden eine schärfere Abschätzung. Dazu betrachten wir die unendlichen Reihen $\frac{g_{n+1}-1}{G_{n+1}} + \frac{g_{n+2}-1}{G_{n+2}} + \ldots$ und behaupten zuerst

$$(6) \quad \tfrac{g_{n+1}-1}{G_{n+1}} + \tfrac{g_{n+2}-1}{G_{n+2}} + \ldots = \tfrac{1}{G_n} \qquad (n = 0, 1, \ldots).$$

In der Tat, die Konvergenz der Reihe links in (6) folgt für beliebiges $n$ aus

$$\begin{aligned} \tfrac{g_{n+1}-1}{G_{n+1}} + \ldots + \tfrac{g_{n+i}-1}{G_{n+i}} &= \left(\tfrac{1}{G_n} - \tfrac{1}{G_{n+1}}\right) + \left(\tfrac{1}{G_{n+1}} - \tfrac{1}{G_{n+2}}\right) + \ldots + \left(\tfrac{1}{G_{n+i-1}} - \tfrac{1}{G_{n+i}}\right) \\ &= \tfrac{1}{G_n} - \tfrac{1}{G_{n+i}} = \tfrac{1}{G_n}\left(1 - \tfrac{1}{g_{n+1} \cdot \ldots \cdot g_{n+i}}\right) < \tfrac{1}{G_n}, \end{aligned}$$

und Gleichung (6) sodann aus $\frac{g_{n+1}-1}{G_{n+1}} + \frac{g_{n+2}-1}{G_{n+2}} + \ldots = \lim_{i\to\infty} \frac{1}{G_n}\left(1 - \frac{1}{g_{n+1} \cdot \ldots \cdot g_{n+i}}\right) = \frac{1}{G_n}$. (6) ergibt nun für die „Restsummen“ der Reihe $\sum_i \frac{z_i}{G_i}$ die Abschätzung

$$(7) \quad \textstyle\sum_{i>n} \frac{z_i}{G_i} = \frac{z_{n+1}}{G_{n+1}} + \frac{z_{n+2}}{G_{n+2}} + \ldots < \frac{g_{n+1}-1}{G_{n+1}} + \frac{g_{n+2}-1}{G_{n+2}} + \ldots = \frac{1}{G_n} \qquad (n \geq 0).$$

Das $<$-Zeichen in (7) ist richtig; denn obwohl i.a. nur $z_i \leq g_i - 1$, gilt für wenigstens ein $i \geq n$ tatsächlich $z_i < g_i - 1$, und dies genügt um die Ungleichung (7) zu sichern.

Wir beweisen nunmehr die folgende, alles weitere entscheidende Behauptung

(a) Das Ergebnis $\langle u_i \rangle$ der Anwendung des CANTORschen Algorithmus auf die Summe $a = \sum_i \frac{z_i}{G_i}$ einer CANTORsche Reihe ist gerade die Folge $\langle z_i \rangle$.

Mit dieser Behauptung hätten wir gleich zweierlei bewiesen, und zwar

(b) Jede Folge $\langle z_i \rangle$ natürlicher Zahlen mit $z_i < g_i$ für alle $i > 0$ und $z_i \neq g_i - 1$ für unendlich viele $i$ kommt als Ergebnis des CANTORschen Algorithmus wirklich vor, nämlich als Ergebnis der Anwendung auf die reelle Zahl $a = \sum_i \frac{z_i}{G_i}$,

(c) Die Darstellung einer reellen Zahl $a$ als CANTORsche Reihe ist eindeutig; denn ist $a = \sum_i \frac{z_i}{G_i} = \sum_i \frac{z'_i}{G_i}$, so ist $u_i = z_i$ und $u_i = z'_i$, also $z_i = z'_i$ für alle $i$. Dies betrifft insbesondere die $g$-adische Entwicklung, also den Fall $g_1 = g_2 = \ldots = g$.

Der Nachweis von (a) ist nun nach unseren Vorbereitungen nicht schwierig und erinnert an den Beweis von Satz 8.1. Sei $a = \sum_i \frac{z_i}{G_i} = z_0 + \sum_{i>0} \frac{z_i}{G_i}$ und $u_0, u_1, \ldots$ das Ergebnis des CANTORschen Algorithmus angewandt auf $a$. Wir behaupten, für jedes $n$ gilt

$$(*) \quad u_n = z_n \quad ; \quad r_n = G_n \cdot \sum_{i>n} \frac{z_i}{G_i} \; (= \frac{z_{n+1}}{g_{n+1}} + \frac{z_{n+2}}{g_{n+1} \cdot g_{n+2}} + \ldots).$$

In der Tat, es ist $u_0 = \operatorname{Int} a$ gleich $z_0$, weil $z_0 \leq a$ und $a - z_0 = \sum_{i>0} \frac{z_i}{G_i} < \frac{1}{G_0} = 1$ gemäß (7). Damit ist auch $r_0 = a - \operatorname{Int} a = \sum_{i>0} \frac{z_i}{G_i} = G_0 \cdot \sum_{i>0} \frac{z_i}{G_i}$, d.h. $(*)$ gilt für $n = 0$. Seien diese Formeln richtig für $n$. Dann ergibt sich wegen $g_{n+1} G_n = G_{n+1}$

$$g_{n+1} r_n = G_{n+1} \cdot \sum_{i>n} \frac{z_i}{G_i} = z_{n+1} + G_{n+1} \cdot \sum_{i>n+1} \frac{z_i}{G_i}.$$

Nach (7) ist $G_{n+1} \cdot \sum_{i>n+1} \frac{z_i}{G_i} < 1$. Also ist $\operatorname{Int}(g_{n+1} r_n) = z_{n+1}$, d.h. $u_{n+1} = z_{n+1}$, und folglich auch $r_{n+1} = g_{n+1} r_n - u_{n+1} = G_{n+1} \cdot \sum_{i>n+1} \frac{z_i}{G_i}$. Damit ist (a) bewiesen.

Wir betrachten nunmehr den Sonderfall CANTORscher Reihen für die Basis $\langle g_n \rangle$ mit $g_n = n + 1$, also $g_0 = 1, g_1 = 2$, $g_2 = 3$, usw. Dann erhalten wir eine der erstaunlichsten Unterscheidungen rationaler und irrationaler Zahlen, nämlich den folgenden

**Satz 8.4.** *Jede reelle Zahl $a \geq 0$ hat genau eine Darstellung der Gestalt*

$$(8) \qquad a = z_0 + \frac{z_1}{2!} + \frac{z_2}{3!} + \frac{z_3}{4!} + \ldots$$

*mit $z_i \in \mathbb{N}$ und $z_i \leq i$ für $i > 0$, sowie $z_i \neq i$ für unendlich viele $i$. Darüber hinaus ist $a$ rational genau dann, wenn ihre Darstellung* (8) *abbricht.*

**Beweis.** Der CANTORsche Algorithmus ergibt die Existenz einer Darstellung (8) mit $z_i < g_i = i + 1$, oder gleichwertig $z_i \leq i$, sowie $z_i \neq g_i - 1$ oder gleichwertig $z_i \neq i$ für unendlich viele $i$. Und (c) sichert deren Eindeutigkeit. Gewiss ist $a$ rational, wenn ihre Darstellung (8) abbricht. Sei anderseits $a = \frac{m}{n}$ gegeben. Dann ergibt (4) mit $z_i = u_i$

$$\frac{m}{n} = \frac{z_0}{1!} + \frac{z_1}{2!} + \ldots + \frac{z_{n-1}}{n!} + \frac{r_{n-1}}{n!} \qquad (0 \leq r_{n-1} < 1).$$

Multiplikation dieser Gleichung mit $n!$ ergibt

$$\frac{m \cdot n!}{n} = z_0 \cdot n! + \frac{z_1 \cdot n!}{2!} + \ldots + \frac{z_{n-1} \cdot n!}{n!} + r_{n-1}.$$

Offenbar lassen sich in allen Bruchtermen dieser Formel die Nenner wegkürzen, woraus sich schließen lässt, dass $r_{n-1}$ eine natürliche Zahl ist. Damit verbleibt – weil $r_{n-1} < 1$ – nur die Möglichkeit $r_{n-1} = 0$. Es sind dann gemäß (3) alle nachfolgenden Reste, und damit auch die natürlichen Zahlen $z_n, z_{n+1}, \ldots$ in (8) identisch gleich 0. ■

Für $\mathbf{e} = 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots = 2 + \frac{1}{2!} + \frac{1}{3!} + \ldots$ erhält man in einer Darstellung (8) gerade $z_0 = 2$, während die „Ziffern“ $z_i$ für $i > 0$ sämtlich gleich 1 sind. Also erhalten wir das

**Korollar.** *Die* EULER*sche Zahl* $\mathbf{e}$ *ist irrational.*

Man beachte, dass dieser Beweis der Irrationalität von $\mathbf{e}$ wesentlich auf der Eindeutigkeit der Darstellung (8) beruht. Ohne die Nebenbedingung $z_i \neq i$ für unendlich viele $i$ hat man gemäß (6) für $n = 0$ und mit $g_i = i + 1$ (also $G_i = (i + 1)!$ ) neben der trivialen Gleichung $1 = 1 + \frac{0}{2!} + \frac{0}{3!} + \dots$ auch die erwähnenswerte Gleichung

$$1 = \tfrac{1}{2!} + \tfrac{2}{3!} + \tfrac{3}{4!} + \dots$$

Es gibt noch manche andere Reihendarstellungen reeller Zahlen, siehe [27, 11]. Auch kann man unendliche Produkte und Quotienten, z.B. die numerisch interessanten Kettenbruchfolgen betrachten. Diese sind von der Gestalt $a_0 + \frac{1}{a_1}$, $a_0 + \frac{1}{a_1 + \frac{1}{a_2}}, \dots$ $(a_i > 0)$, und deren Limes bezeichnet man im Konvergenzfalle mit $a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}$.

## 8.4 Berechenbarkeit und Iteration

Wir geben im folgenden einen kurzen Einblick in einen Problemkreis, der von erheblichem Interesse für die Theorie und die Praxis mathematischer Anwendungen ist, den der expliziten Berechenbarkeit. Eine reelle Zahl $a$ $(\geq 0)$ ist im intuitiven Sinne berechenbar, wenn zu jedem $n$ eine $n$-stellige Näherung $a_n$ mit $|a - a_n| < \varepsilon_n$ explizit angegeben werden kann. Das ist i.a. etwas weniger als die explizite Berechenbarkeit ihrer Dezimalziffern, worauf in **5.4** schon hingewiesen wurde. Es lässt sich auch schreiben $a_n = N_a(n) \overset{n}{\leftarrow}$, mit $N_a(n) \in \mathbb{N}$ und einer gewissen berechenbaren oder allgemein-rekursiven Funktion $n \mapsto N_a(n)$. Auf diese Weise lässt sich reelle, und ähnlich auch komplexe Berechenbarkeit innerhalb der weit entwickelten Theorie der rekursiven (zahlentheoretischen) Funktionen präzisieren. Intuitiv ist klar und dies lässt sich streng beweisen, dass die rationalen Operationen $+, -, \times, \div$, aber auch nichtrationale Operationen wie Potenzierung und Logarithmierung, von in diesem Sinne berechenbaren Argumenten wieder zu berechenbaren Werten führen.

Ist $p(x) = c$ eine lösbare Polynomgleichung, so kann eine Lösung mit beliebig hoher Genauigkeit auch berechnet werden. Dies gilt allgemeiner für Gleichungen der Gestalt $f(x) = c$, wenn z.B. gewisse Differenzierbarkeitsforderungen an die Funktion $f$ gestellt werden, und legt die Frage nach der Formulierung möglichst allgemeiner Bedingungen über die Berechenbarkeit einer Lösung von $f(x) = c$ für berechenbare Zahlen $c$ und Funktionen $f$ nahe, falls eine Lösung überhaupt existiert.

Dazu muss offenbar gesagt werden, was die Berechenbarkeit einer reellen Funktion $f$ bedeuten soll. Hier bietet sich folgende Erklärung an. $f$ heiße *berechenbar*, wenn es ein effektives Verfahren[2)] gibt, mit dessen Hilfe sich für jedes berechenbare $x$ aus dem Definitionsbereich den Wert $fx$ beliebig genau berechnen lässt. Es sei zum Beispiel $f$ eine im Intervall $I = [0, v]$ schlichte und berechenbare Funktion, und $c$ eine berechenbare

[2)] Dieser Begriff wird in der Rekursionstheorie präzisiert. Mit den berechenbaren reellen Funktion befasst sich die so genannte *Rekursive Analysis*, wo Elemente der Analysis, der Numerik und der Rekursionstheorie zusammenfließen. Nach der Erörterung in 5.4 ist plausibel, dass z.B. die simple Funktion $a \mapsto \operatorname{Int} a$ nicht berechenbar ist. Es lässt sich zeigen, dass eine berechenbare Funktion immer stetig ist.

Zahl mit $f0 < c < fv$. Dann hat die Gleichung $fx = c$ nach Satz 7.1 die Lösung $\xi = \sup x \in I \mid fx \leq c$, deren Berechnung sich leicht programmieren lässt. Es kommt nur darauf an, ob der Test $\boxed{fx \leq c}$ mit $x = \xi_n + i\varepsilon_{n+1}$ für $i = 1, \ldots, 9$ effektiv ausführbar ist, wobei $\xi_n \leq \xi \leq \xi_n + \varepsilon_n$ ein schon berechneter Näherungswert ist. Dieser ist z.B. für $x \mapsto x^n$ präzise ausführbar, weil $x^n \in \mathbb{E}$ für $x \in \mathbb{E}$. Daher lässt sich $\sqrt[n]{c}$ sogar ziffernweise berechnen, falls die Information vorliegt ob $c \in \mathbb{E}$ oder nicht, und $c$ im letzteren Falle selbst auch ziffernweise berechenbar ist. Die Ausführbarkeit des obigen Tests beruht hier wesentlich darauf, dass mit $x \in \mathbb{E}$ auch $x^n \in \mathbb{E}$. Bei nichtrationalen berechenbaren Funktionen $f$ ist i.a. nur die übliche näherungsweise Berechenbarkeit einer Lösung von $fx = c$ garantiert. Das gilt auch dann noch, wenn $f$ in $I$ nur stetig und berechenbar ist, für jedes berechenbare $c$ zwischen $fu$ und $fv$.

Im Einzelfall werden den Gegebenheiten besonders angepasste Verfahren zur Berechnung von Zwischenwerten gewählt. So rechnen die Computer die $n$-te Wurzel meist logarithmisch und natürlich nur näherungsweise aus, es sei denn, man schreibt ein eigenes Programm. Die Möglichkeiten hierfür sind vielfältig. Am beliebtesten sind Iterationsverfahren, weil deren Programmierung besonders bequem ist. Beispiele behandeln wir unten. Sie beruhen auf folgendem Satz, einem von zahlreichen ähnlichlautenden Konvergenzsätzen der numerischen Analysis.

**Satz 8.5 (Fixpunktsatz).** *Sei $I \subseteq \mathbb{D}$ ein abgeschlossenes Intervall, sowie $f : I \to I$ eine in $I$ schlichte Funktion, und $\langle x_n \rangle$ eine beliebige Folge mit $x_0 \in I$* und

$$(9) \qquad x_{n+1} = fx_n \quad \text{für alle } n.$$

*Dann ist $\langle x_n \rangle$ konvergent, und $\lim\langle x_n \rangle$ ist Fixpunkt von $f$. Ist $f$ zudem schlicht in $I$ mit $C < 1$, hat $f$ in $I$ nur einen einzigen Fixpunkt.*

**Beweis.** $\langle x_n \rangle$ falle monoton und ist dann sicher konvergent; wegen $x_{n+1} = fx_n$ und $\xi := \lim\langle x_n \rangle = \lim\langle x_{n+1} \rangle$ ist auch $\xi = \lim\langle fx_n \rangle$. Weil $\lim\langle fx_n \rangle = f\xi$ nach (11) in **7.2**, ergibt sich $\xi = f\xi$. Sei nun $x_k < x_{k+1}$ $(= fx_k)$ für ein gewisses $k$. Ein simpler Induktionsschluss zeigt $x_n \leq x_{n+1}$ für alle $n \geq k$. Auch ist $\langle x_n \rangle_{n \geq k}$ beschränkt, weil diese Folge ganz in $I$ verbleibt. Also ist sie schlicht und damit konvergent, und es gilt $\xi := \lim\langle x_n \rangle_{n \geq k} = \lim\langle x_n \rangle$. Die Gleichung $f\xi = \xi$ erschließt man wie eben. Sei nun $fx - fy \leq C(x - y)$ mit $C < 1$ für alle $x, y \in I,\ x > y$. Ist $\xi_1 = f\xi_1,\ \xi_2 = f\xi_2$, und wäre etwa $\xi_1 < \xi_2$, so folgt $\xi_2 - \xi_1 \leq C(\xi_2 - \xi_1)$, also der Widerspruch $1 \leq C$. ∎

Dass eine den Voraussetzungen dieses Satzes genügende Funktion $f$ überhaupt einen Fixpunkt hat, ist keine Neuheit. Dazu genügt nach einem Beispiel in **3.3** bereits monotones Wachstum von $f$. Auch konvergiert dann immer noch die Folge $\langle x_n \rangle$, nur muss deren Limes nicht notwendig auch ein Fixpunkt von $f$ sein. Auch lässt sich die Voraussetzung $x_0 \in I$ offenbar noch abschwächen zu $x_i \in I$ für ein gewisses $i$, weil $\langle x_n \rangle$ und $\langle x_n \rangle_{n \geq i}$ dasselbe Konvergenzverhalten haben. Für Anwendungen von Satz 8.5 ist hauptsächlich der Fall interessant, dass $f$ genau einen Fixpunkt $\xi$ in $I$ hat. Jede gemäß (9) definierte Folge – eine so genannte *Iterationsfolge* – konvergiert dann automatisch gegen $\xi$. Oft weiß man aus anderen Gründen, dass $f$ in $I$ höchstens einen Fixpunkt hat, so dass die dafür hinreichende (aber nicht notwendige) Bedingung $C < 1$ gar keine Rolle spielt. Betrachten wir als Beispiel die bereits in Übung 7.7 im Intervall $[0, \sqrt{a}]$ als schlicht nachgewiesene Funktion $f : x \mapsto \frac{2ax}{a+x^2}$ für $a = 10$ nun in einem Intervall $I = [u, v]$ mit $0 < u < v := \sqrt{a}$, siehe hierzu Figur 5. $f$ bildet $I$ in sich ab und hat dort nur einen Fixpunkt $\xi$, nämlich am oberen Intervallende; denn die Gleichung $x = \frac{2ax}{a+x^2}$ hat nur die Lösungen $\xi = 0$ und $\xi = \sqrt{a}$. Die untere Intervallgrenze $u$ (= 0,5 in der Figur) ist nicht 0, um den Fixpunkt 0 aus der Betrachtung fernzuhalten und darf innerhalb des Spielraums $0 < u < \sqrt{a}$ frei gewählt werden. Nach Satz 8.5 konvergiert jede Iterationsfolge $\langle x_n \rangle$ mit $x_0 \in I$ gegen den einzigen Fixpunkt $\xi = \sqrt{a}$ von $f$ in $I$. Damit hat man ein bequemes iteratives Verfahren zur Berechnung von $\sqrt{a}$ mittels der rationalen Operationen. Man kann die Berechnung von $\sqrt{10}$ nach diesem Iterationsverfahren mit dem Eingangswert $x_0 = 1$ in der Figur längs der gestrichelten Linien deutlich verfolgen. Je größer $n$ ist, desto näher liegt $x_n$ bei $\sqrt{a}$. Gewisse Informationen über die Konvergenzgeschwindigkeit dieses Verfahrens, welche wesentlich durch die Anzahl der Durchläufe durch die Iterationschleife bestimmt wird, enthält die Bemerkung weiter unten.

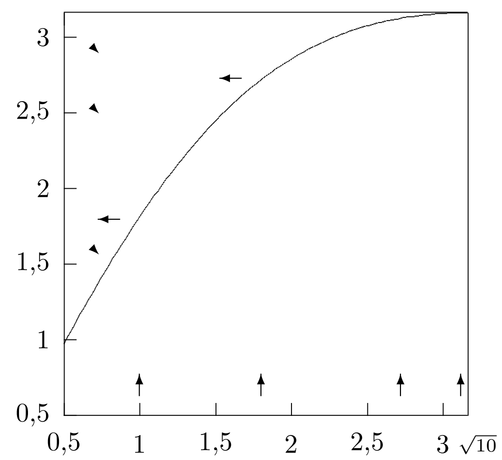
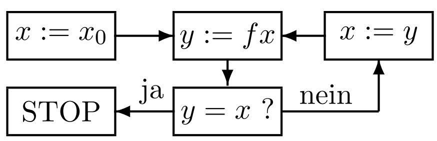



*Fig.¸5 Iterative Berechnung des Fixpunktes $\sqrt{a}$ von $x \mapsto \frac{2ax}{a+x^2}$ für $a = 10$.*




Die Programmierung des Verfahrens entsprechend nebenstehendem, leicht durchschaubaren Flussdiagramm verwendet den Test $\boxed{y = x\ ?}$, der angibt, ob die Berechnung nahe genug am Konvergenzpunkt liegt. Die Rechnung endet, wenn die Werte von $x$ und $y$ nach einer gewissen Anzahl von Schleifendurchläufen im Rahmen der Genauigkeit eines gegebenen Rechners ununterscheidbar sind, oder, wie man auch sagt, der Rechenprozess in den Bereich des „Rundungsrauschens“ gelangt. Auf dem programmierbaren Taschenrechner erkennt man dies bei einigen Iterationsverfahren nach Ersetzung des Abbruchtests durch $\boxed{\text{DISPLAY } x, y}$ an einem periodischen Wechsel der letzten Ziffer in der Anzeige. Hier die Berechnung von $\sqrt{10}$ nach dem obigen Verfahren mit der 0-ten Näherung $x_0 = 1$ auf einem Taschenrechner mit 12-stelligem Display:

$$x_1 = 1{,}81818181818, \qquad x_2 = 2{,}73291925465, \qquad x_3 = 3{,}12890616374,$$
$$x_4 = 3{,}16209970755, \qquad x_5 = 3{,}16227765516, \qquad x_6 = 3{,}16227766017.$$

Bereits nach dem 6. Iterationsschritt ändern sich die Folgenglieder nicht mehr und $\sqrt{10} = 3{,}16227\,76601\,6838\cdots$ ist so genau berechnet, wie es die Anzeige des Rechners zulässt. Für $x_0 = 3$ wären sogar nur 3 Iterationsschritte nötig. Übrigens muss man nicht auf $x_0 \leq \sqrt{a}$ bestehen. $x_0$ muss positiv, darf sonst aber beliebig sein. Falls $x_0 > \sqrt{a}$, liegt bereits $x_1 = fx_0$ wieder in $I$, wie man sich leicht überlegt.

**Weitere Beispiele.** (a) Man sieht leicht, dass $f : x \mapsto \sqrt{2x}$ in $[u, v]$ schlicht ist, wenn nur $u > 0$. Für $I := [1, 2]$ ist auch $f : I \to I$. Daher konvergiert die Iterationsfolge $\langle x_n \rangle$ mit $x_0 = 1$, also $x_1 = \sqrt{2}$, $x_2 = \sqrt{2\sqrt{2}}$, ... – siehe Übung 7.1 – gegen den einzigen Fixpunkt von $f$ in $I$, nämlich die Lösung $\xi = 2$ der Fixpunktgleichung $x = fx$, also von $x = \sqrt{2x}$. Damit erhalten wir einen leicht überschaubaren Beweis für $\lim\langle x_n \rangle = 2$.
(b) Sei $x_0 = 0$ und $x_{n+1} = \sqrt{2 + x_n}$, also $x_1 = \sqrt{2}$, $x_2 = \sqrt{2 + \sqrt{2}}$, ... Auch $\langle x_n \rangle$ ist konvergent mit $\lim\langle x_n \rangle = 2$. Denn $f : x \mapsto \sqrt{2 + x}$ ist schlicht in $I = [0, 2]$ mit $C = \frac{1}{\sqrt{8}}$ $< 1$; ferner gilt $f : I \to I$, und $\xi = 2$ ist nach Satz 8.5 einziger Fixpunkt von $f$ in $I$.
(c) Manche Computer vermeiden direkte Divisionen und ermitteln $\frac{1}{a}$ für $a \in \mathbb{D}_+$ mit dem folgendem Iterationsverfahren: Sie suchen schnellstmöglich einen Wert $x_0$ mit $0 < x_0 \cdot a \leq 1$ und berechnen dann iterativ $x_{n+1} = 2x_n - ax_n^2$ bis zum Eintritt des Rundungsrauschens. Auch hier sichert Satz 8.5 die Konvergenz von $\langle x_n \rangle$ gegen den Fixpunkt $\frac{1}{a}$ der in $[0, \frac{1}{a}]$ leicht als schlicht nachweisbaren Funktion $x \mapsto 2x - ax^2$.

**Bemerkung.** Eine konvergente Folge $\langle x_n \rangle$ mit $\xi = \lim\langle x_n \rangle$ konvergiert *linear*, *quadratisch* bzw. *kubisch*, wenn es eine Konstante $c$ gibt mit $d_{n+1} \leq c{\cdot}d_n^{\,k}$ für alle $n$, wobei $d_n = |\xi - x_n|$ und $k = 1, 2$ bzw. 3, sowie $c < 1$ für $k = 1$. Grob gesprochen, die Anzahl der genauen Dezimalen der Näherung $x_n$ von $x$ wächst linear, quadratisch bzw. kubisch mit $n$. So konvergiert die in den Beispielen oben zuletzt genannte Folge quadratisch, Übung 7. Auch die durch $x_{n+1} = \frac{2ax_n}{a+x_n^2}$ definierte Iterationsfolge konvergiert quadratisch, ebenso wie die des so genannten „Babylonischen“ Verfahrens in Übung 6, ein Spezialfall des NEWTONschen Näherungsverfahrens. Es geht noch besser: Die durch $x_{n+1} = \frac{x_n(x_n^2+3a)}{3x_n^2+a}$ definierte schlichte Folge $\langle x_n \rangle$ konvergiert sogar kubisch gegen $\sqrt{a}$. Beginnend mit $x_0 = 3$, hat man für $a = 10$ im ersten Iterationsschritt $\sqrt{10}$ bereits auf 3 Dezimalen, und im zweiten auf alle Dezimalen des Displays (in Wahrheit auf sogar 13 Dezimalen) genau. Hingegen konvergiert die Folge $\langle e_n \rangle$ aus **6.2** nicht einmal linear, wie mittels der Ungleichungen (e1),(e2) aus **7.1** unschwer erschlossen werden kann.

Eine in einem Intervall $I$ definierte Funktion $f$ heißt *Lipschitz-stetig in* $I$, wenn

(10) Es gibt eine Konstante $C$ mit $|fx - fy| \leq C|x - y|$ für alle $x, y \in I$

erfüllt ist. Eine in $I$ monoton wachsende Funktion ist dort Lipschitz-stetig genau dann, wenn sie schlicht ist. Falls in (10) $C < 1$ gewählt werden kann, heißt $f$ eine *Kontraktion.* Satz 8.5 ähnelt dem BANACH*schen Fixpunktsatz*, der behauptet, dass eine Kontraktion $f$ von $I$ genau einen Fixpunkt $\xi$ in $I$ hat und dass *jede* Iterationsfolge $\langle x_n \rangle$ mit $x_k \in I$ für ein $k$ gegen $\xi$ konvergiert. Eine von vielen Anwendungen dieses Satzes ist die folgende:

In [22] wird am Beispiel $a = 47$ ein iteratives Verfahren zur Berechnung von $\sqrt[3]{a}$ für $a \in \mathbb{N}_+$ suggeriert, welches das LEONARDO*sche Iterationsverfahren* genannt sei. Es ist vermutlich das erste, schriftlich fixierte numerische Iterationsverfahren im modernen Sinne und funktioniert für reelle $a \geq 1$ genauso wie in LEONARDOs Beispiel: Man errate zuerst ein $u \geq 1$ mit $u^3 \leq a < (u+1)^3$ – ein solches $u$ findet man stets auch in $\mathbb{N}$ – setze $x_0 := u$ und berechne sodann iterativ $x_{n+1} = x_n + \frac{a - x_n^3}{3(u+1)x_n}$. Tatsächlich konvergiert $\langle x_n \rangle$ nach dem BANACHschen Fixpunktsatz gegen die Lösung $\sqrt[3]{a}$ der Fixpunktgleichung $x = Fx := x + \frac{a - x^3}{3(u+1)x}$. Denn $F$ ist Kontraktion eines geeigneten Intervalls $[u, v]$ mit $v \geq u+1$, Übung 8. Das Verfahren konvergiert nicht sehr schnell. Beginnend mit $x_0 = 3$, benötigt man 11 Schritte, um z.B. $\sqrt[3]{47} = 3{,}608\cdots$ auf 10 Dezimalen zu berechnen. Aber für Handrechnungen ist es durchaus brauchbar. Zum Beispiel ist schon $x_1 = 3{,}555\cdots$. Hier noch eine weitere einfache Anwendung:

Die berühmte, durch $\varphi_0 = 1,\ \varphi_1 = 1$ und (a) $\varphi_{n+1} = \varphi_n + \varphi_{n-1}$ für $n > 0$ definierte Folge heißt die FIBONACCI–Folge. Also $\varphi_2 = 2,\ \varphi_3 = 3,\ \varphi_4 = 5$, aber z.B. ist bereits $\varphi_{58} = 956\,722\,026\,041$. Sei $q_n := \frac{\varphi_{n+1}}{\varphi_n}$. Wir beweisen, die Folge $\langle q_n\rangle$ konvergiert gegen $\frac{1+\sqrt{5}}{2} = 1{,}618\cdots$, der Verhältniszahl des Goldenen Schnitts. (a) ergibt $q_{n+1} = 1 + \frac{1}{q_n}$. Nun ist $f\colon x \mapsto 1 + \frac{1}{x}$ Kontraktion von $I := [\frac{3}{2}, 2]$; denn (10) ist mit $C = \frac{2}{3}$ erfüllt, wie man leicht sieht. Weil $q_2 \in I$, ist $\langle q_n\rangle$ nach dem BANACHschen Fixpunktsatz konvergent, und weil $\frac{1+\sqrt{5}}{2}$ die Fixpunktgleichung $x = 1 + \frac{1}{x}$ in $I$ löst, gilt $\lim\langle q_n\rangle = \frac{1+\sqrt{5}}{2}$. Auf $q_0$ (also auf $\varphi_0, \varphi_1$) kommt es demnach gar nicht an. Das erkennt man auch deshalb, weil $q_1 = 1 + \frac{1}{q_0},\ q_2 = 1 + \frac{1}{1+\frac{1}{q_0}}, \ldots$, und damit $\frac{1+\sqrt{5}}{2} = \lim\langle q_n\rangle = 1 + \cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{\ddots}}}$ .

## 8.5 Übungen

1. Man zeige: eine positive rationale Zahl $a$ ist im Dezimalsystem reinperiodisch genau dann, wenn es eine Darstellung $a = \frac{m}{n}$ gibt, so dass $n$ und 10 teilerfremd sind. Damit sind Summe und Differenz (im Falle ihrer Existenz), sowie das Produkt reinperiodischer Zahlen wieder reinperiodisch. Diese bilden also einen (in $\mathbb{D}$ dichten) $\mathcal{E}$-Bereich reeller Zahlen, der die Elemente aus $\mathbb{E}\backslash N$ sozusagen auslässt.

2. Seien $m, n$ teilerfremd und sei $a = \frac{m}{n}$ reinperiodisch mit Periodenlänge $p$. Man zeige (a) $n$ teilt $10^p - 1$, (b) $p$ ist die kleinste Zahl aus $\mathbb{N}$ derart, dass $n$ Teiler ist von $10^p - 1$. (Stets ist $p \leq$ Anzahl der zu $n$ teilerfremden $k \leq n$, EULER.)

3. Sei $a = z_0{,}z_1z_2\cdots{}_8$ eine Oktalzahl, also $0 \leq z_i < 8$ für $i > 0$, und auch $z_0$ sei im Oktalsystem geschrieben. Man zeige: die binäre Entwicklung von $a$ erhält man durch ziffernweise Umschrift der Oktalziffern von $a$ wie im Text geschildert.

4. Man betrachte die Kommaverschiebung für $g$-adische reelle Zahlen $a$ und beweise $a\overset{n}{\rightarrow} = a \cdot g^n$, sowie $a\overset{n}{\leftarrow} = a \cdot \frac{1}{g^n}$.

5. Man beweise, $a \in \mathbb{D}$ ist rational genau dann, wenn für jedes $g \geq 2$ die $g$-adische Entwicklung von $a$ periodisch ist. Diese Eigenschaft ist also basisinvariant.

6. Häufig wird folgendes Iterationsverfahren zur Berechnung von $\sqrt{a}$ für $a > 0$ betrachtet: $x_0 > 0$ sei beliebig, und $x_{n+1} = \frac{1}{2}(x_n + \frac{a}{x_n})$. Man beweise, dass $\langle x_n\rangle_{n>0}$ eine fallende, gegen $\sqrt{a}$ konvergierende Folge ist.

7. Sei $a > 0$ und $0 < ax_0 \leq 1$, sowie $x_{n+1} = 2x_n - ax_n^2$. Man zeige: $\langle x_n\rangle$ ist eine schlichte Folge und diese konvergiert quadratisch gegen ihren Grenzwert $\frac{1}{a}$.

8. Sei $u \geq 1$. Man zeige: $F\colon x \mapsto x + \frac{a-x^3}{3(u+1)x}$ kontrahiert $[u, v]$ für ein geeignetes $v > u + 1$. Dies sichert die Konvergenz des LEONARDOschen Verfahrens.

# Abschnitt 11

# Anhang – Ein Aufbau des Systems der natürlichen Zahlen

Beim Aufbau des Systems der reellen Zahlen in **3 – 5** sind wir davon ausgegangen, dass uns die natürlichen Zahlen uneingeschränkt zur Verfügung stehen. Mit einigen zusätzlichen (mengentheoretischen) Begriffen wie dem der Folge wurden sodann die reellen Zahlen konstruiert. Nun wird nachträglich dargelegt, wie sich das Konzept der natürlichen Zahl in den Rahmen klassischer Auffassung von Mathematik einordnet, nach der sich alle mathematischen Begriffe auf den Mengenbegriff reduzieren lassen. Diese Erkenntnis war das Fazit des gegen Ende des 19. Jahrhunderts von G. Frege (1848 - 1925) unternommenen und von B. Russell (1882 – 1970) u.a. weitergeführten Versuches, die Mathematik oder jedenfalls ihre wesentlichen Teile auf die Logik zu reduzieren (*Logizismus*). So weit zu gehen hat sich allerdings als unmöglich erwiesen, denn auch in der Mathematik kann man nicht erwarten, etwas für nichts zu erhalten.

Wir beginnen mit der Untersuchung der von R. Dedekind in [6] erstmals betrachteten Zählreihen, den Modellen des bekannten Peanoschen Axiomensystems. Eine Zählreihe dient vornehmlich dem Abzählen endlicher Mengen, und ihre Definition enthält genau die dazu erforderlichen Eigenschaften. Die Elemente einer Zählreihe nennen wir kurzerhand *natürliche Zahlen*. Diese Definition ist weniger vieldeutig als es den Anschein hat. Alle Zählreihen erweisen sich nämlich als isomorph. Daher kann nach Abschluss einer Reihe von Betrachtungen einfach von *der* Zählreihe oder *der* Menge der natürlichen Zahlen gesprochen werden.

In Zählreihen ist zunächst nur von einem Anfangselement und einer Nachfolgerfunktion die Rede. Sowohl die übliche Anordnung der natürlichen Zahlen als auch die arithmetischen Operationen müssen also definiert und ihre wesentlichen Eigenschaften nachgewiesen werden. Eine bekannte Definitionsmöglichkeit der arithmetischen Operationen ist ihre rekursive Einführung nach [6], siehe auch [19] oder [26]. Wir wählen indes einen Weg, der die axiomatische Behandlung der Zählreihen mit den Vorteilen einer mengentheoretischen Definition der arithmetischen Operationen verbindet. Der in **11.6** vorgenommene Nachweis aller einschlägigen Rechengesetze, genauer, der Axiome in Abschnitt **2**, ist dann überaus einfach. Ein weiterer Vorteil dieser Methode ist, dass sie durchweg den intuitiven Gegebenheiten folgt und nebenbei auch die Theorie endlicher Mengen mit begründet. Insbesondere wird durch sie der Begriff des Zählens, ähnlich wie der des Messens in Abschnitt **10**, auf eine solide Grundlage gestellt.

Die Existenz von Zählreihen wird mengentheoretisch gesichert. Diese Art des Vorgehens hat den Vorteil methodischer und konzeptioneller Klarheit. Man könnte prinzipiell jedoch auch anderen Konzeptionen folgen, z.B. einer solchen, welche die natürlichen Zahlen als von vornherein gegeben und als den Ursprung aller Mathematik ansieht. Diese Auffassungsweise wird besonders deutlich in dem berühmten Ausspruch L. Kroneckers (1823 – 1891): „Die natürlichen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk.“ Bei solcher Auffassung erübrigt sich ein Existenzbeweis für Zählreihen, weil es eine der Zahlentheorie übergeordnete Theorie gar nicht gibt. Allerdings käme man damit in der höheren Analysis nicht sehr weit, es sei dann, man verlegt die erforderlichen mengentheoretischen Hilfsmittel in eine Zahlentheorie 2. Stufe, zu deren Objekten auch Teilmengen natürlicher Zahlen gehören und in der reelle Zahlen durch gewisse Teilmengen natürlicher Zahlen kodiert werden.

Man möchte ohne Skrupel von der Menge aller messbaren Mengen reeller Zahlen, der Menge aller integrierbaren Funktionen usw. reden können und diese z.B. als Objekte eines metrischen Raumes, näher untersuchen. Hierbei verwendet man nichtkonstruktive und infinitistische Methoden durchweg. Um den Gebrauch dieser Methoden gegen jede Kritik zu schützen, entwickelte Hilbert das in wesentlichen Punkten jedoch gescheiterte Programm einer finiten Begründung der Mathematik. Dieses wurde revidiert, aber trotz des Fortschritts hat sich die Grundlagenproblematik ihrer vielschichtigen Aspekte wegen nicht erledigt. Immerhin befindet sich die Mathematik diesbezüglich in einer vergleichsweise besseren Situation als die meisten anderen Wissenschaften.

Viele Schlussweisen und Begriffsbildungen, die man etwas voreilig zu den naiven, der Logik angehörigen Hilfsmitteln zu zählen geneigt ist, haben bei genauerem Hinsehen kombinatorischen Charakter (sie lassen sich in einer formalisierten Zahlentheorie 1. Stufe, der so genannten Peano-Arithmetik begründen) oder sie gehen darüber hinaus und benutzen Hilfsmittel der transfiniten Mengenlehre. Wenn nun aber eine im Kern mengentheoretische Fundierung der Analysis geboten ist, macht es Sinn, hierin auch die natürlichen Zahlen einzubeziehen. Denn dazu benötigt man nur die allereinfachsten mengentheoretischen Werkzeuge. Ähnlich wie der Begriff des Dedekindschen Schnitts ist auch derjenige der Dezimalfolge der Form nach ein mengentheoretischer. Allerdings ließe sich dieser, wie auch die meisten Begriffe dieses Abschnitts, bereits im Rahmen einer erheblich eingeschränkten Mengenlehre behandeln[1)]. Nur der Bequemlichkeit halber stellen wir uns im folgenden auf den Boden der naiven Mengenlehre.

[1)]Mit Fragen, aus welchen Annahmen welche Sätze der Analysis und anderer mathematischer Disziplinen folgen und welche Tragweite diese Annahmen haben, befasst sich eine neuere Teildisziplin der Mathematischen Logik, die so genannte *Reverse Mathematics*, siehe etwa [36] für einen Überblick.

## 11.1 Zählreihen und der Rekursionssatz

Eine Zählung beginnt stets mit einem ersten Element und vollzieht sich schrittweise. Dieser schrittweise Charakter des Zählens lässt sich durch eine Funktion $n \mapsto n^+$ genauer beschreiben, welche einem Element $n$ einer Zählreihe das nachfolgende Element $n^+$ zuordnet. Deswegen gleicht das anschauliche Bild einer Zählreihe der einer mit einem Anfangselement beginnenden, sich beliebig fortsetzenden Perlenkette, siehe Figur 13. Die Präzisierung dieses anschaulichen Bildes führt so zu folgender Definition, einer modernen Ansprüchen genügende Formulierung des PEANOschen Axiomensystems [2)].

**Definition.** Ein Bereich, bestehend aus einer Menge $\mathbb{N}$, einer Funktion $^+ : \mathbb{N} \to \mathbb{N}$ genannt die *Nachfolgerfunktion* von $\mathbb{N}$, sowie einem besonderen, in der Regel mit 0 bezeichneten und zu $\mathbb{N}$ gehörenden Element, genannt das *Anfangselement*, heißt eine *Zählreihe*, wenn dieser Bereich die folgenden drei Axiome erfüllt:

(PI) $0 \neq n^+$ für alle $n \in \mathbb{N}$, (PII) $m^+ = n^+ \;\Rightarrow\; m = n$, für alle $m, n \in \mathbb{N}$,

(PIII) Ist $M \subseteq \mathbb{N}$ und erfüllt $M$ die beiden Bedingungen
(a) $0 \in M$, (b) $n \in M \;\Rightarrow\; n^+ \in M$, für alle $n$,
so ist $M = \mathbb{N}$.

Wir bezeichnen fortan eine Zählreihe und ihre Trägermenge mit dem gleichen Symbol $\mathbb{N}$. Missverständnisse werden daraus nicht erwachsen. $n^+$ heißt der *Nachfolger* von $n$, und $n$ der *Vorgänger* von $n^+$, wobei $n, m, i, k$ im folgenden ausschließlich Elemente von $\mathbb{N}$ bezeichnen. (PII) besagt, dass ein Element *höchstens* einen Vorgänger hat, und (PI) schließt die Existenz eines solchen für das Anfangselement definitiv aus. (PIII) heißt das *Induktionsaxiom*. Es sichert unter anderem, dass die Zählreihe keine überflüssigen, zum Zählen nicht benötigte Elemente enthält. Für die präzise Definition des in Figur 13 schematisierten Anfangs der Zählreihe sei auf **11.3** verwiesen.

$$\mathbb{N}:\ \underbrace{\overset{0}{\bullet} \text{——} \overset{0^+}{\bullet} \text{——} \bullet \text{——} \bullet \text{——} \overset{n}{\bullet}}_{A_n} \text{——} \overset{n^+}{\bullet} \text{——} \bullet \text{——} \bullet \;\bullet\; \bullet$$

*Fig. 13 Die Zählreihe und deren Anfang* $A_n = 0, \cdots, n$

0 hat gemäß (PI) also keinen Vorgänger. Doch lässt sich an den Axiomen nicht sofort erkennen, dass 0 das einzige Element ohne Vorgänger ist, und ebensowenig, dass der Vorgänger von $n$, wenn er existiert, von $n$ auch verschieden ist. Beide Tatsachen lassen sich jedoch durch *(vollständige) Induktion* beweisen, d.h. durch wesentliche Anwendung von (PIII). Solche und andere Beweise aus (PI) – (PIII) sind unabhängig von einer Existenzdiskussion; vielmehr betreffen sie ausschließlich die Struktur von Zählreihen.

---

[2)] Dieses stammt eigentlich von DEDEKIND, [6, §6], und das hier angegebene ist dem DEDEKINDschen ähnlicher. PEANO benutzte für die Darstellung seiner Axiome eine spezielle formale Sprache, wodurch gewisse Missverständnisse bei der Wiedergabe der Axiome in der mathematischen Literatur entstanden. Für einen sarkastischen Kommentar dieses Thema betreffend verweisen wir auf [10, Bd.**1**, S.144ff].

Die beiden im letzten Absatz erwähnten Behauptungen ließen sich getrennt durch Induktion beweisen. Man kann sie aber auch in die einzige Behauptung

$$(*)\quad \text{Für alle } n \in \mathbb{N}\text{: } n = 0 \text{ oder es gibt ein } m \neq n \text{ mit } m^+ = n$$

verpacken, deren Induktionsbeweis dann wie folgt aussieht:

Sei $M = n \in \mathbb{N} \mid n = 0$ oder es gibt ein $m \neq n$ mit $m^+ = n$. Offenbar ist $0 \in M$. Zum Nachweis von (b) in (PIII) sei $n \in M$ (*Induktionsannahme*). Im Falle $n = 0$ hat $n^+$ sicher einen Vorgänger $\neq n^+$, nämlich 0, denn nach (PII) ist $0 \neq 0^+$. Andernfalls (d.h. wenn $n \neq 0$) sei etwa $n = m^+$ mit $m \neq n$ gemäß Induktionsannahme. $n^+$ hat den Vorgänger $n = m^+$, und wegen $m \neq n$ ist nach (PII) auch $n = m^+ \neq n^+$. Daher hat auch $n^+$ einen Vorgänger $\neq n^+$, nämlich $n$, und gehört damit zu $M$. Nach (PIII) ist folglich $M = \mathbb{N}$, und $(*)$ ist bewiesen.

Obwohl alle Zählreihen isomorph sind (Satz 11.2), gibt es keine verbindliche Konvention hinsichtlich der Bezeichnung ihres Anfangselements, außer dass es regelmäßig entweder mit 0 oder mit 1 bezeichnet wird. Das liegt in der Natur der Sache. Denn für Abzählungen endlicher nichtleerer Mengen ist 0 eigentlich überflüssig. Da man eine $n$-elementige Menge gern mit den ersten $n$ natürlichen Zahlen abzählen möchte, müssten solche Zählungen mit 0 beginnen, wenn man 0 als natürliche Zahl ansieht. Anderseits beginnen Zählungen gewohnheitsgemäß mit 1 ($= 0^+$). Die 0 ist für Zählungen im ursprünglichen Wortsinne überflüssig. Hier liegt der eigentliche Grund für gefühlsmäßige Vorbehalte gegen die Bezeichnung von 0 als „natürliche“ Zahl. In der Mathematik beginnen Zählungen allerdings recht häufig auch mit 0. Man spricht etwa von den ersten $n$ Primzahlen $p_0, \ldots, p_{n-1}$. Ob 0 in einem Zählvorgang verwendet wird oder nicht, ist einzig und allein eine Frage der Bequemlichkeit. Auch an Alltagsbeispielen lässt sich der gelegentlich praktische Nutzen von Zählungen unter Einschluss von 0 demonstrieren. Es wäre z.B. sicher von allgemeinem Vorteil, für die Bezeichnung der Halteposition eines Fahrstuhls im Erdgeschoss international das Symbol 0 zu verwenden.

Da wir beabsichtigen, die natürlichen Zahlen zugleich als Anzahlmaße beliebiger endlicher Mengen einschließlich $\emptyset$ zu verwenden, sei das Anfangselement einer Zählreihe $\mathbb{N}$ mit 0 bezeichnet, solange nichts anderes gesagt wird. 0 wird als Anzahlmaß der leeren Menge dienen. Unabhängig von dieser Konvention behält man im übrigen die freie Entscheidung, ob man Abzählungen nichtleerer Mengen mit 0 oder 1 beginnen lässt.

Jedenfalls ist mit $\mathbb{N}$ offenbar auch $\mathbb{N}_+ := \mathbb{N}\{0\}$ eine Zählreihe, die *Zählreihe der positiven natürlichen Zahlen*. Denn $\mathbb{N}_+$ ist gegenüber der Nachfolgeroperation abgeschlossen (d.h. $n \in \mathbb{N}_+ \Rightarrow n^+ \in \mathbb{N}_+$) und hat das Anfangselement 1, weil dessen Vorgänger ja gerade entfernt wurde. Im PEANOschen Axiomensystem ist das Konstantensymbol 0 dann durch 1 zu interpretieren. $\mathbb{N}$ und $\mathbb{N}_+$ unterscheiden sich nach Satz 11.2 zwar nicht strukturell, weil aber $\mathbb{N}_+$ eine echte Teilmenge von $\mathbb{N}$ ist, lassen sich $\mathbb{N}$ und $\mathbb{N}_+$ nicht ohne weiteres identifizieren. Ganz allgemein ergibt sich aus den nachfolgenden Strukturuntersuchungen leicht, dass ein beliebiges Anfangsstück einer Zählreihe einfach weggelassen werden kann, ohne die Struktur der Zählreihe dadurch zu zerstören.

Wegen seiner herausragenden Bedeutung beweisen wir zuerst den berühmten Rekursionssatz, auch DEDEKIND*scher Rechtfertigungssatz* genannt. Dieser Satz wird nicht selten als unmittelbar einsichtig erachtet und bleibt oft unerwähnt. So wurde der Satz bereits in Abschnitt **2** bei der Definition der Potenz verwendet. Solange man mit natürlichen Zahlen und Funktionen naiv umgeht, erwächst hieraus kein Problem. Sobald man aber die in diesem Satz vorkommenden Begriffe im Rahmen einer Theorie präzisiert, erfordern Aussagen, welche diese Begriffe enthalten, auch entsprechende Beweise.

Der Rekursionssatz hat insofern mengentheoretischen Charakter als er über Funktionen redet, und diese sind spezielle Mengen geordneter Paare. Er führt kombinatorische Konstruktionen auf mengentheoretische zurück und besitzt wesentliche Verallgemeinerungen. Aus dieser Möglichkeit einer mengentheoretischen Reduktion sollte jedoch keine Religion gemacht werden. Man sollte sie angemessen bewerten; denn sie löst keine Grundlagenprobleme, sondern verlagert diese höchstens.

**Satz 11.1 (Rekursionssatz für $\mathbb{N}$).** *Es sei $\mathbb{N}$ eine Zählreihe, $A$ eine Menge, $a \in A$ und $F : \mathbb{N} \times A \to A$ eine beliebig vorgegebene Funktion. Dann existiert genau eine Funktion $f : \mathbb{N} \to A$ welche den folgenden Rekursionsgleichungen genügt:*

$$\text{(R)} \quad f0 = a \ \ ; \ \ fn^+ = F(n, fn), \text{ für alle } n \in \mathbb{N}.$$

**Beweis.** Die Eindeutigkeit ist gerade die Aussage von Satz 2.1. Also verbleibt der Existenzbeweis. Sei $f$ der Durchschnitt aller Mengen $P \subseteq \mathbb{N} \times A$ mit den Eigenschaften

$$\text{(i)} \ \ (0, a) \in P \quad ; \quad \text{(ii)} \ \ (n, y) \in P \ \Rightarrow \ (n^+, F(n, y)) \in P \quad \big(n \in \mathbb{N},\, y \in A\big).$$

Dann gelten (i) und (ii) offenbar auch für $P = f$. Wir beweisen nun nacheinander die Behauptungen (a) bis (c), von denen (c) die fragliche Existenzbehauptung sichert.

(a) $(0, y) \in f \ \Rightarrow \ y = a$, für alle $y \in A$.

Denn sei $(0, y) \in f$ und angenommen $y \neq a$, so dass $(0, a) \neq (0, y)$. Dann erfüllt auch $P := f\{(0, y)\}$ gewiss noch (i), und wegen $(0, y) \neq (n^+, F(n, y))$ auch (ii). Weil $f$ die kleinste (i) und (ii) erfüllende Menge ist, gilt $f \subseteq f\{(0, y)\}$. Dies bedeutet aber nichts anderes als $(0, y) \notin f$. Damit wurde die Annahme $y \neq a$ zum Widerspruch geführt.

(b) $(n^+, y') \in f \ \Rightarrow \ y' = F(n, y)$, sofern $(n, y) \in f$.

Denn sei $(n, y), (n^+, y') \in f$ und $y' \neq F(n, y)$. Dann gelten (i) und (ii) gewiss auch für $P := f\{(n^+, y')\}$. Also $f \subseteq f\{(n^+, y')\}$, d.h. $(n^+, y') \notin f$, was ein Widerspruch ist.

(c) $f$ ist eine (R) erfüllende Funktion von $\mathbb{N}$ nach $A$.

Dies heißt definitionsgemäß: zu jedem $n \in \mathbb{N}$ gibt es genau ein $y \in A$ mit $(n, y) \in f$. Das ist nach (i) und (a) sicher richtig für $n = 0$. Sei dies richtig für $n$ und sei $y$ dasjenige Element mit $(n, y) \in f$. Nach (ii) ist dann $(n^+, F(n, y)) \in f$; mehr noch, nach (b) gibt es auch nur genau ein $y'$ mit $(n^+, y') \in f$, nämlich $y' = F(n, y)$. Bezeichnet man nun dasjenige $y$ mit $(n, y) \in f$ wie üblich mit $fn$, so besagt (i) für $P = f$ nichts anderes als $f0 = a$, und (ii) besagt $y = fn \ \Rightarrow \ fn^+ = F(n, y)$. Dies aber ist äquivalent zu $fn^+ = F(n, fn)$. Damit wurde auch (c) vollständig bewiesen. ■

## 11.2 Eindeutigkeit und Existenz der Zählreihe

Wir zeigen in Satz 11.2 zuerst, dass es bis auf Isomorphie höchstens eine Zählreihe gibt. Zählreihen $\mathbb{N}$ und $\hat{\mathbb{N}}$ sind damit strukturell ununterscheidbar. Figur 14 veranschaulicht den Isomorphismus $f$ von $\mathbb{N}$ auf $\hat{\mathbb{N}}$. Dabei werden hier nur die Anfangselemente 0 von $\mathbb{N}$ bzw. $\hat{0}$ von $\mathbb{N}$ mit verschiedenen Symbolen bezeichnet. Aus der Gleichbezeichnung der beiden Nachfolgeroperationen erwachsen keine Missverständnisse. Die geordneten Paare, aus denen sich der mittels des Rekursionssatzes konstruierte Isomorphismus $f$ zusammensetzt, werden in der Figur durch senkrechte Pfeile dargestellt.

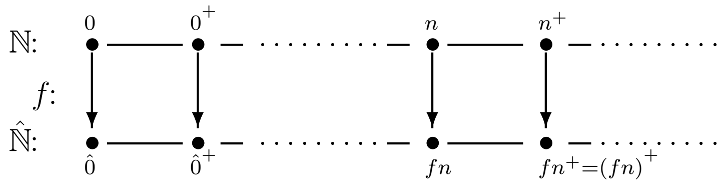



*Fig. 14 Der Isomorphismus zwischen Zählreihen*

**Satz 11.2 (Isomorphiesatz).** *Zählreihen* $\mathbb{N}$ *und* $\hat{\mathbb{N}}$ *mit den Anfangselementen* $0$ *bzw.* $\hat{0}$ *sind isomorph, d.h. es gibt eine Bijektion* $f:\mathbb{N}\to\hat{\mathbb{N}}$ *mit* $f0=\hat{0}$ *und* $fn^+=(fn)^+$.

**Beweis.** Wir definieren die gesuchte Abbildung $f$ von $\mathbb{N}$ nach $\hat{\mathbb{N}}$ mittels Satz 11.1 durch die Rekursionsgleichungen

$$f0=\hat{0} \quad ; \quad fn^+=(fn)^+.$$

Dies sind zugleich die Isomorphiebedingungen, so dass nur noch zu beweisen verbleibt, $f$ ist bijektiv. Sei $g:\hat{\mathbb{N}}\to\mathbb{N}$ analog wie $f$ durch $g\hat{0}=0$ und $gn^+=(gn)^+$ definiert. Dann gelten $g\cdot f=id_{\mathbb{N}}$ [3] und $f\cdot g=id_{\hat{\mathbb{N}}}$. In der Tat, es ist $g(f0)=g\hat{0}=0$, und falls $g(fn)=n$, ist auch $g(fn^+)=g((fn)^+)=(g(fn))^+=n^+$. Aus Symmetriegründen gilt auch $f\cdot g=id_{\hat{\mathbb{N}}}$. Ist nun $fn=fm$, so folgt $n=g(fn)=g(fm)=m$. Also ist $f$ injektiv. Ferner gilt $f(gm)=m$ für $m\in\hat{\mathbb{N}}$. Daher hat $m$ bezüglich der Abbildung $f$ das Urbild $gm$, und $f$ ist folglich auch surjektiv. ■

Nach den bisher gemachten Ausführungen ist eine Zählreihe strukturell zwar eindeutig bestimmt, doch fehlt offenbar ein rigoroser Existenzbeweis. Nun ist anschaulich zwar evident, dass z.B. eine Strichliste

$$| \quad || \quad ||| \quad |||| \qquad \ldots$$

eine Zählreihe darstellt, wobei die Nachfolgeroperation das jeweilige Hinzufügen eines Striches zu einem Listenglied bedeutet. Wenn man aber den Versuch unternimmt, den anschaulichen Begriff einer Strichliste exakt zu erfassen, treten allerlei Schwierigkeiten zutage. Unsere Anschauung über das Unendliche ist sehr unvollkommen. Genau dies hat ja zu den strengen Kriterien eines in sich geschlossenen Aufbaus der Mathematik geführt. Es hat sich bewährt, innerhalb der Mathematik nur das als existent anzusehen,

[3] $id_M$ ist die identische Funktion auf einer Menge $M$, also $id_M(x)=x$ für alle $x\in M$.

was aufgrund der mathematischen Konstruktionsprinzipien als existent nachgewiesen werden kann. Mit diesen Prinzipien lassen sich aller Erfahrung nach auch die kompliziertesten Erscheinungsbilder unserer Anschauung und der physikalischen Welt modellieren, ganz sicher also ein so banales Beispiel wie eine Strichliste.

Aber hat sich die Frage der Existenz einer Zählreihe mit dem Hinweis auf die Strichliste nicht schon erledigt? Hier verbirgt sich leider wie häufig das Problem im Detail. Selbst wenn man philosophischen Konzeptionen folgt, welche das Anschauliche vor die gedankliche Abstraktion stellen, ist nicht sichergestellt, ob über die unbegrenzte Strichliste als ganzes überhaupt geredet werden darf, z.B. weil diese grundsätzlich nicht vollständig herzustellen ist. Unsere Vorstellung überschreitet hier klar physikalische Grenzen. Es treten hierbei dieselben Probleme zutage, auf die in **10.1** im Zusammenhang mit der Betrachtung materieller Größenbereiche bereits aufmerksam gemacht wurde.

Die mengentheoretische Konstruktion einer Zählreihe ist hingegen recht einfach. Das liegt an der simplen Bauart des Begriffs, der nur die Grundeigenschaften des Zählens formalisiert und über Operationen wie Addition oder Multiplikation gar nicht redet. Eine raffinierte, heute zum Standard gewordene Konstruktion dieser Art wurde 1925 von J. v. Neumann (1903 – 1957) angegeben. Sie eignet sich gleichermaßen zur Konstruktion transfiniter Zahlen, siehe etwa [7].

Die v. Neumannsche Zählreihe $\omega$ sei definiert als Durchschnitt aller Mengen $\Omega$ mit den Eigenschaften $\emptyset \in \Omega$ und $x \in \Omega \Rightarrow x \cup x \in \Omega$. Dass es eine Menge mit diesen Eigenschaften überhaupt gibt, ist gerade die heute meist übliche Formulierung des Unendlichkeitsaxioms. $\emptyset$ ist Anfangselement von $\omega$, also $0 := \emptyset$, und $n^+ := n \cup n$ sei der Nachfolger von $n \in \omega$. Also $0^+ = 0 \cup 0 = 0, \quad 0^{++} = 0^+ \cup 0^+ = 0, 0^+$ usw.

Die Gültigkeit der Axiome (PI) und (PIII) für $\omega$ ist unmittelbar klar. Nur bei (PII) muss man rechnen. Zuerst zeigt man induktiv über $n \in \omega$, dass $n$ *transitiv* ist, was heißen soll $y \in x \in n \Rightarrow y \in n$, oder gleichwertig, jedes Element von $n$ ist auch Teilmenge von $n$. Dies ist trivial für $n = \emptyset$, und ist $n$ transitiv, so auch $n \cup n$. Denn ist $x \in n \cup n$, etwa $x \in n$, und daher auch $x \subseteq n$, ist auch $x \subseteq n \cup n$; für $x = n$ ist ebenfalls $x \subseteq n \cup n$. Hieraus ergibt sich leicht (PII). Denn sei $n \cup n = m \cup m$ und etwa $n \neq m$. Dann ist $n \in m$ und $m \in n$. Weil $n$ transitiv ist, folgt $n \in n$. Dies aber widerspricht dem induktiv beweisbaren Fakt $n \notin n$: Gewiss ist $0 \notin 0$. Sei $n \notin n$ gemäß Induktionsannahme und etwa $n^+ \in n^+$. Dann ist $n \cup n = n$ was $n \notin n$ widerspricht, oder $n \cup n \in n$, also $n \in n \cup n \in n$, was wegen der Transitivität $n \notin n$ gleichermaßen widerspricht.

Wir haben damit insgesamt gezeigt, dass es bis auf Isomorphie genau eine Zählreihe $\mathbb{N}$ gibt. Es ist unnötig, über die Natur der Elemente von $\mathbb{N}$ Festlegungen zu treffen, weil wir nur von den Struktureigenschaften von $\mathbb{N}$ Gebrauch machen werden. Man darf also die v. Neumannsche Definition im Prinzip wieder vergessen [4)]. Von nun an bezeichnen wir auch die Elemente von $\mathbb{N}$ in üblicher Weise mit $0, 1, 2, \ldots$, also $1 = 0^+$, $2 = 1^+$ usw.

[4)] Man könnte jedoch auch einen Schritt weitergehen und spezielle Eigenschaften von $\omega$ ausnutzen. Zum Beispiel lässt sich im v. Neumannschen Modell der Mengenterm $0, \ldots, n$ ohne Umschweife als die Menge $n^+$ definieren. Dies vereinfacht aber nicht die Betrachtungen im nächsten Teilabschnitt.

## 11.3 Die Anordnung der Zählreihe

Jedermann hat die lineare Anordnung der Zählreihe gemäß der sukzessiven Erzeugung ihrer Elemente sichtbar vor Augen. Doch ist das Vorhandensein einer derartigen Ordnung exakt zu bestätigen, und dies ist weniger einfach, als man auf den ersten Blick vermutet. Selbst DEDEKIND, der den Mathematikern diese und ähnliche Beweisnotwendigkeiten erst ins Bewusstsein rief, hat in [6] diese spezielle Aufgabe unterschätzt. Nicht nur die Auswahl von Eigenschaften der Zählreihe, die hier zum Erfolg führen, auch die Reihenfolge ihrer Beweise ist wichtig. Wir beginnen mit der Definition des Terms $0, \ldots, n$, den wir genau unter die Lupe nehmen werden.

**Definition.** Der von $n \in \mathbb{N}$ bestimmte *Anfang* $A_n \subseteq \mathbb{N}$ werde rekursiv erklärt durch $A_0 = 0$, $A_{n^+} = A_n \cup n^+$. Der Anfang $A_n$ wird auch mit $0, \ldots, n$ bezeichnet.

Stets ist $n \in A_n$. Offenbar ergibt sich $A_1 = 0, 1,\ A_2 = 0, 1, 2$ usw. Figur 13 in **11.1** vermittelt eine anschauliche Vorstellung von dem durch eine Zahl $n \in \mathbb{N}$ bestimmten Anfang $A_n = 0, \ldots, n$. Es ist plausibel, dass $A_n$ eine im naiven Sinne endliche Teilmenge von $\mathbb{N}$ ist. Aber auch dies kann erst dann wirklich bewiesen werden, wenn der Begriff *endlich* klar gefasst worden ist, was in **11.4** geschehen wird. Auch erst dann lässt sich nachweisen, dass $\mathbb{N}$ selbst nicht endlich ist.

Einige der im folgenden Lemma formulierten Eigenschaften sind trotz der Schreibweise $A_n = 0, \ldots, n$ keineswegs so einleuchtend wie sie aussehen. So lässt sich z.B. $n^+ \notin 0, \ldots, n$ sicher nicht für alle $n$ direkt verifizieren. Nur rigorose Beweise können zweifelsfrei bestätigen, dass präzise Definitionen das Gemeinte tatsächlich erfassen.

**Lemma.** *Für alle* $n, m \in \mathbb{N}$ *gelten*

(1) $0 \in A_n$, (2) $m^+ \in A_n \;\Rightarrow\; m \in A_n$,

(3) $n^+ \notin A_n$, (4) $m \neq n \in A_m \;\Rightarrow\; n^+ \in A_m$,

(5) $A_n = A_m \Rightarrow n = m$, (6) $A_n \subset A_m \Rightarrow A_{n^+} \subseteq A_m$.

**Beweis.** (1): Beweis durch Induktion über $n$. Dies ist klar für $n = 0$. Ist $0 \in A_n$ so gilt sicher auch $0 \in A_{n^+}$, einfach weil $A_n \subseteq A_{n^+}$.

(2): Beweis durch Induktion über $n$. Klar für $n = 0$, weil niemals $m^+ \in A_0 = 0$. Sei $m^+ \in A_{n^+}$. Im Falle $m^+ \in A_n$ ist $m \in A_n \subseteq A_{n^+}$ nach Induktionsannahme. Falls aber $m^+ = n^+$, ist $m = n$ und daher $m \in A_n \subseteq A_{n^+}$, also gleichwohl $m \in A_{n^+}$.

(3): Beweis durch Induktion über $n$. Es ist $0^+ \notin A_0 = 0$, weil $0^+ \neq 0$. Sei $n^+ \notin A_n$. Dann ist auch $n^{++} \notin A_n$ – sonst wäre $n^+ \in A_n$ nach (2), entgegen Annahme.Weil auch $n^{++} \neq n^+$, folgt die Induktionsbehauptung $n^{++} \notin A_{n+} = A_n \cup n^+$.

(4): Beweis durch Induktion über $m$. Die Behauptung ist trivial für $m = 0$, weil kein $n$ mit $0 \neq n \in A_0 = 0$ existiert. Induktionsschritt: $m^+ \neq n \in A_m \cup m^+$ ergibt offenbar $n \in A_m$. Daher ist entweder $m = n$ oder aber $m \neq n \in A_m$ und damit $n^+ \in A_m$ gemäß Induktionsannahme. In beiden Fällen ist dann aber $n^+ \in A_m \cup m^+ = A_{m^+}$.

(5): Sei $A_n = A_m$. Dann ist $n \in A_m$ und $n^+ \notin A_n = A_m$ nach (3). Daher $n = m$, denn wäre $m \neq n$ $(\in A_m)$, würde sich nach (4) der Widerspruch $n^+ \in A_m = A_n$ ergeben.

(6): Sei $A_n \subset A_m$. Um $A_{n^+} \subseteq A_m$ zu beweisen, genügt es $n^+ \in A_m$ zu zeigen. Wegen $A_n \subset A_m$ ist $m \neq n \in A_n \subseteq A_m$ gemäß (5), also $n^+ \in A_m$ nach (4). ■

Man könnte den Anfang $A_n$ auch explizit definieren, und zwar als die kleinste Teilmenge von $\mathbb{N}$, die das Element $n$ enthält und mit jedem Element dessen Vorgänger, sofern ein solcher existiert. Aber wir halten uns damit nicht auf und kommen sogleich zum Hauptresultat dieses Teilabschnitts.

**Satz 11.3 (Anordnungssatz).** *Sei $n < m$ in $\mathbb{N}$ definiert durch $n < m \Leftrightarrow A_n \subset A_m$. Dann ist die Relation $<$ eine Anordnung der Zählreihe $\mathbb{N}$.*

**Beweis.** $<$ ist offensichtlich transitiv. Zum Beweis der Vergleichbarkeit V (siehe **2.1**) beachte man zunächst, dass sich die Fälle $m < n$, $m = n$, $m > n$ gemäß Definition von $<$ sicher ausschließen. Wegen (5) im Lemma genügt es daher $n \leq m$ oder $m \leq n$, d.h.

$$(*) \quad \text{Für alle } n, m\text{: } A_n \subseteq A_m \textit{ oder } A_m \subseteq A_n$$

zu beweisen. Sei $M = n \in \mathbb{N} \mid A_n \subseteq A_m$ *oder* $A_n \subseteq A_m$, *für alle* $m \in \mathbb{N}$. Wir zeigen $M = \mathbb{N}$ induktiv über $n$, was offenbar nichts anderes besagt als $(*)$. Sicher ist $0 \in M$; denn $0 \in A_m$ für jedes $m$, d.h. $A_0 \subseteq A_m$, und damit erst recht $0 \in M$. Sei $n \in M$ gemäß Induktionsannahme, also $A_n \subset A_m$ oder $A_m \subseteq A_n$, für beliebiges $m$. Im Falle ist $A_n \subset A_m$ ist $A_{n^+} \subseteq A_m$ nach (6); im Falle $A_m \subseteq A_n$ ist offenbar $A_m \subseteq A_{n^+}$. Also $A_{n^+} \subseteq A_m$ oder $A_m \subseteq A_{n^+}$ in jedem Falle. Folglich ist auch $n^+ \in M$. Damit ist $M = \mathbb{N}$ und der Satz ist bewiesen. ■

Die so konstruierte Ordnung von $\mathbb{N}$ heißt die *natürliche*. Diese ist – unter unübersehbar vielen weiteren Anordnungsmöglichkeiten von $\mathbb{N}$ – eindeutig gekennzeichnet durch

(7) $\quad n < n^+$ für alle $n \in \mathbb{N}$,

Übung 1. Eigenschaft (7) gilt, weil $A_n \subseteq A_{n^+}$ und $n^+ \notin A_n$ nach (3), also $A_n \subset A_{n^+}$. Man vermerke, (6) lässt sich auch schreiben als $n < m \Rightarrow n^+ \leq m$, d.h. zwischen $n$ und $n^+$ liegen keine Elemente. Das ergibt $n < m \ \Rightarrow\ n^+ < m^+$, und wegen der Vergleichbarkeit gilt dann auch die Umkehrung hiervon. Wir erwähnen noch das nützliche

**Korollar.** $A_m = n \in \mathbb{N} \mid n \leq m$. *Insbesondere ist damit $m$ größtes Element in $A_m$.*

**Beweis.** Es ist zu zeigen $n \leq m \Leftrightarrow n \in A_m$, oder gleichwertig $A_n \subseteq A_m \Leftrightarrow n \in A_m$. Die Richtung $\Rightarrow$ ist wegen $n \in A_n$ trivial. Sei umgekehrt $n \in A_m$. Wäre $A_n \not\subseteq A_m$, so wäre $A_m \subset A_n$ nach Satz 11.3. Daher $m \neq n \in A_m$. und so $n^+ \in A_m \subseteq A_n$ nach (4), ein Widerspruch zu (3). ■

**Bemerkung.** Die obige Konstruktion der Anordnung natürlicher Zahlen ließe sich nicht unerheblich vereinfachen, wenn $n < m$ in $\omega$ durch $n \in m$ definiert wird. Die $\in$-Relation auf $\omega$ stellt nämlich schon eine Anordnung von $\omega$ dar. Dieses Vorgehen wäre jedoch ein Stilbruch insofern, als im axiomatischen Begriff der Zählreihe ja von Anordung keine Rede ist.

Die natürliche Anordnung von $\mathbb{N}$ hat darüber hinaus zwei weitere, für zahlreiche Anwendungen wichtige Besonderheiten, nämlich

(8) Jede nichtleere beschränkte Teilmenge von $\mathbb{N}$ hat ein größtes Element,

(9) Jede nichtleere Teilmenge $M \subseteq \mathbb{N}$ hat ein kleinstes Element.

Zum Beweis von (8) vergegenwärtige man sich zuerst, dass jede nichtleere Teilmenge $M$ von $A_n$ ein größtes Element hat. Dies ist klar für $A_0 = 0$. Ist $M \subseteq A_{n^+}$, so ist entweder $M \subseteq A_n$ und $M$ hat ein größtes Element gemäß Induktionsannahme – oder aber es ist $n^+ \in M$, in welchem Falle offenbar $n^+$ größtes Element von $M$ ist. Hieraus ergibt sich unmittelbar (8). Denn ist $M \subseteq \mathbb{N}$ durch $n$ beschränkt, so ist $M \subseteq A_n$ nach dem Korollar und hat damit ein größtes Element.

Von (9) spricht man als der *Wohlordnung* von $\mathbb{N}$. Allgemein heißt eine geordnete Menge *wohlgeordnet*, wenn jede ihrer nichtleeren Teilmengen ein kleinstes Element hat. Jedes $A_n$ ist wohlgeordnet. Das ist klar für $A_0$ und gilt dies für $A_n$ so auch für $A_{n^+}$. Denn sei $\emptyset \neq M \subseteq A_{n^+}$. Der interessante Fall ist $M \cap A_n \neq \emptyset$. Diese Menge hat ein kleinstes Element nach Induktionsannahme welches dann offenbar auch kleinstes Element in $M$ ist. Daraus folgt sofort (9). Denn ist $n \in M \subseteq \mathbb{N}$, so hat $A_n \cap M$ ein kleinstes Element und dieses ist nach dem Korollar zugleich das kleinste Element von $M$.

(8) und (9) implizieren direkt, dass $(\mathbb{N}, <)$ diskret geordnet ist, d.h. jeder Schnitt ist ein Sprung. Denn ist $(U, V)$ ein Schnitt von $\mathbb{N}$, so hat $U$ nach (8) ein größtes, und $V$ nach (9) ein kleinstes Element. Es gilt darüber hinaus, dass $(\mathbb{N}, <)$ bis auf Isomorphie die einzige diskret geordnete Menge mit kleinstem und ohne größtes Element ist, Übung 5. Dies ist eine bekannte ordnungstheoretische Charakterisierung von $\mathbb{N}$.

## 11.4 Abzählungen endlicher Mengen

Es ist wohlbekannt, dass sich Mengen ohne zu zählen hinsichtlich ihrer Anzahl (oder wie man auch sagt, ihrer *Mächtigkeit*) vergleichen lassen. Historisch gesehen war es jedoch ein bedeutender Fortschritt, Mengen zu *zählen*, statt sie durch direkte Zuordnung ihrer Elemente hinsichtlich ihrer Anzahl zu vergleichen. So kam z.B. ein Tauschgeschäft zwischen einem Viehzüchter und einem Ackerbauern der Urgesellschaft durch einen *Austausch von Informationen* (über die Anzahl der zu tauschenden Schafe und Scheffel Korn) zustande, nicht durch eine höchst beschwerliche direkte Gegenüberstellung der Tauschobjekte. Dieser Austausch von Informationen über Anzahlen ist eine entscheidende Voraussetzung für organisierte Arbeitsteilung, ohne die sich auch frühzeitliche Kulturen nicht hätten entwickeln können.

Mit der Zählreihe zählt man im Regelfalle die Elemente nichtleerer Mengen (gegen die leere Menge ist nichts einzutauschen). Das suggeriert, das erste Element der Zählreihe für die Zählung der Einermengen, den „kleinsten“ nichtleeren Mengen, zu verwenden. Es ist aber bequemer, das erste Element 0 der Zählreihe $\mathbb{N}$ als Anzahl der leeren Menge zu reservieren und Zählungen nichtleerer Mengen mit 1 zu beginnen. Das entspricht auch der traditionellen Zählweise. Wir bedienen uns hier also der Zählreihe $\mathbb{N}_+$ mit dem ersten Element 1. Deutlicher Unterscheidung wegen sei der durch $n$ bestimmte Anfang von $\mathbb{N}_+$ mit $A_n^+$ bezeichnet, also $A_n^+ = 1, \dots, n$. Außerdem sei $A_0^+ := \emptyset$.

Eine *Abzählung* von $M$ ($\neq \emptyset$) mittels $1, \dots, n$ sei nun nichts anderes als eine Bijektion $\alpha$ von $A_n^+$ auf $M$. Etwas genauer: $\alpha$ heiße eine *Abzählung von $M$ der Länge $n$*. Wir schreiben $\alpha_i$ für $\alpha(i)$ und bezeichnen $\alpha$ auch durch $\langle\alpha_i\rangle_n$, als Abkürzung für $\langle\alpha_i\rangle_{1\leq i\leq n}$. Ist $\alpha_i = a$, so heiße $i$ auch die *Nummer* von $a$ bei $\alpha$. Man kann sich eine Abzählung von $M$ veranschaulichen, indem man jedem $a \in M$ seine Nummer bildlich gesprochen „anheftet". Es ergibt sich so in natürlicher Weise folgende

**Definition.** Eine Menge $M$ heißt *endlich*, wenn $M = \emptyset$ oder wenn es ein $n \in \mathbb{N}_+$ und eine Abzählung von $M$ der Länge $n$ gibt. Andernfalls heißt $M$ *unendlich*.

Die Ausnahmerolle von $M = \emptyset$ in dieser Definition lässt sich vermeiden, indem man erklärt, $\langle\alpha_i\rangle_0 := \emptyset$ sei eine Abzählung von $M = \emptyset$ der Länge 0 (die „leere Abzählung"). Alle Einermengen $a$ sind endlich, denn $1 \mapsto a$ ist Bijektion von $A_1^+$ auf $a$ und damit Abzählung von $a$ der Länge 1. Zweiermengen $a, b$ (mit $a \neq b$) sind endlich, denn die Abbildung $\alpha$ mit $\alpha_1 = a$ und $\alpha_2 = b$ ist Abzählung von $a, b$ der Länge 2. Allgemein gilt: Ist $\alpha = \langle\alpha_i\rangle_n$ Abzählung von $M$ und $a \notin M$, so ist $\langle\alpha_i\rangle_{n^+}$, die „Verlängerung" von $\alpha$ mit $\alpha_{n^+} = a$, Abzählung von $M \cup a$ der Länge $n^+$. Kurzum, ist $M$ endlich, so auch die durch Hinzufügen eines beliebigen Elements zu $M$ entstehende Menge.

Nach dieser Definition sind endliche Mengen genau diejenigen, die Abzählungen einer gewissen Länge $n$ besitzen. Es gibt aber auch viele andere Definitionen. Die einfachste stammt von Dedekind; sie benutzt den Begriff der natürlichen Zahl überhaupt nicht: $M$ heißt endlich, wenn jede Injektion von $M$ in sich eine Bijektion ist.

Anschaulich ist klar, dass verschiedene Abzählungen ein und derselben endlichen Menge $M$ sämtlich von gleicher Länge sind. Dies besagt folgender Satz, ein Musterbeispiel dafür, dass eine physikalische Beobachtung eine rein mathematische Erklärung hat.

**Satz 11.4 (Zählsatz).** *Sind $\langle\alpha_i\rangle_n$ und $\langle\beta_i\rangle_m$ Abzählungen von $M$, so ist $n = m$.*

**Beweis** durch Induktion über $n$. Die Behauptung ist klar für $n = 0$, denn die leere Menge und nur diese hat die Abzählungslänge 0. Sei die Behauptung für $n \in \mathbb{N}$ als richtig angenommen und $\alpha = \langle\alpha_i\rangle_{n^+}$ Abzählung einer Menge $M$ der Länge $n^+$. Ist $\beta$ eine andere Abzählung von $M$, so hat $\beta$ jedenfalls nicht die Länge 0, also $\beta = \langle\beta_i\rangle_{m^+}$ für ein gewisses $m \in \mathbb{N}$. Es ist $m^+ = n^+$ zu beweisen. Betrachten wir zunächst den Sonderfall, dass $\alpha_{n^+} = \beta_{m^+}$ und sei $a := \alpha_{n^+}$ ($= \beta_{m^+}$). Dann sind $\langle\alpha_i\rangle_n$ und $\langle\beta_i\rangle_m$ Abzählungen von $M\backslash\{a\}$. Daher ist $m = n$ gemäß Induktionsannahme, also auch $m^+ = n^+$. Der allgemeine Fall wird nun auf den soeben behandelten Spezialfall zurückgeführt. Sei $a := \alpha_{n^+} = \beta_k$ und $b := \beta_{m^+}$. Es entstehe $\beta'$ aus $\beta$ durch Vertauschen der Nummern von $a$ und $b$, d.h. es sei $\beta'_{m^+} = a$, $\beta'_k = b$ und $\beta'_i = \beta_i$ sonst. $\beta'$ ist offenbar Abzählung von $M$ derselben Länge wie $\beta$ und hat nach dem Sonderfall dieselbe Länge wie $\alpha$. Damit haben auch $\alpha$ und $\beta$ dieselbe Länge. ∎

Unter der mit $|M|$ bezeichneten *Anzahl* oder *Kardinalzahl* einer endlichen Menge $M \neq \emptyset$ versteht man die nach letztem Satz eindeutig bestimmte Länge einer Abzählung von $M$. Ferner sei $|\emptyset| = 0$. Ist $|M| = n$, so heißt $M$ auch *$n$-zahlig*. Die Einermengen sind einzahlig, die Zweiermengen zweizahlig usw. Zu jedem noch so großen $n$ $(> 0)$ gibt es eine $n$-zahlige Menge; z.B. ist $|A_n^+| = n$, denn $id_{A_n^+}$ ist Abzählung von $A_n^+$ der Länge $n$.

$$\begin{tikzcd} M \arrow[dd, "\varphi"] & \\ & A_n^+ \arrow[ul, "\alpha^{-1}"'] \arrow[dl, "\beta"] \\ N & \end{tikzcd}$$

Mengen $M, N$ heißen *gleichmächtig*, symbolisch $M \sim N$, falls eine Bijektion von $M$ auf $N$ existiert. Für endliche $M, N$ ist dies gleichwertig zu $|M| = |N|$, also $|M| = |N| \;\Leftrightarrow\; M \sim N$. Denn sind $\alpha, \beta$ Abzählungen von $M$ bzw. N gleicher Länge, so ist $\varphi = \beta \cdot \alpha^{-1}$ Bijektion von $M$ auf $N$, wie sie die nebenstehende Figur verdeutlicht. Umgekehrt liefern eine Abzählung $\alpha$ von $M$ zusammen mit einer Bijektion $\varphi\colon M \to N$ die Abzählung $\beta = \varphi \cdot \alpha$ von $N$. Auch für unendliche Mengen erfordert die Gleichmächtigkeitsdefinition keinen Zahlbegriff. Dieser kommt erst ins Spiel, wenn ein *Maß* für den Umfang einer Menge benötigt wird.

Was endliche Mengen angeht, so sind neben dem Zählsatz die beiden folgenden Sätze für eine anschauungs-konforme Erklärung der arithmetischen Operationen wichtig.

**Satz 11.5.** *Mit $M$ ist auch jede echte Teilmenge $U \subset M$ endlich und es ist $|U| < |M|$.*

**Beweis** durch Induktion über $n = |M|$. Für $n = 0$ (also $M = \emptyset$) ist die Behauptung klar. Sei diese für alle $n$-zahligen Mengen vorausgesetzt, $\langle \alpha_i \rangle_{n^+}$ Abzählung von $M$ der Länge $n^+$ und $a = a_{n^+}$. Dann ist $\langle \alpha_i \rangle_n$ Abzählung von $M^- := M \backslash \{a\}$, also $|M^-| = n$. **Fall 1:** $U \subseteq M^-$. Dann ist gemäß Induktionsannahme $U$ endlich und darüber hinaus gilt $|U| \leq n < n^+$. **Fall 2:** $a \in U$. Dann ist $U^- := U \backslash \{a\} \subset M^-$. Also ist $U^-$ endlich und damit auch $U = U^- \cup a$. Ferner ist $|U^-| < n$, also $|U| = |U^-|^+ < n^+$. ∎

Dieser Satz ist charakteristisch für endliche Mengen $M$ und geht für unendliche Mengen verloren, wie das Beispiel von $\mathbb{N}$ zeigt. $^+ : \mathbb{N} \to \mathbb{N}^+$ ist Bijektion, obwohl $\mathbb{N}^+ \subset \mathbb{N}$. Entsprechendes ist für endliche Mengen $M$ ausgeschlossen; denn $M \sim U \subseteq M$ impliziert $|U| = |M|$, womit $U$ nach Satz 11.5 keine echte Teilmenge von $M$ sein kann. Zugleich wurde gezeigt, dass jede Injektion von $M$ in sich eine Bijektion ist. Die Dedekindsche Endlichkeitsdefinition ist also eine Folge der unseren. Erwartungsgemäß gilt auch

**Satz 11.6.** *Sind $M, N$ endlich, so auch $M \cup N$. Dasselbe gilt für $M \times N$.*

**Beweis** durch Induktion über $n = |N|$. Für $n = 0$ ist dies klar, weil $M \cup \emptyset = M$. Sei nun $\langle \alpha_i \rangle_{n^+}$ Abzählung von $N$ und $a := \alpha_{n^+}$, sowie $N^- = N \backslash \{a\}$. Dann ist $N^-$ $n$-zahlig und $M \cup N^-$ nach Induktionsannahme endlich. Dann gilt aber dasselbe auch für

$$M \cup N = M \cup (N^- \cup a) = (M \cup N^-) \cup a.$$

Die Endlichkeit von $M \times N$ folgt analog. Für $n = 0$ ist $M \times N$ leer. Sonst sei $a \in N$ wie oben, $N^- := N \backslash \{a\}$, so dass $M \times N^-$ gemäß Induktionsannahme endlich ist. Dasselbe gilt wegen $M \sim M \times a$ nach dem bereits Bewiesenen dann auch für $M \times N$, denn

$$M \times N = M \times (N^- \cup a) = (M \times N^-) \cup (M \times a). \quad ∎$$

## 11.5 Der kardinale und der ordinale Aspekt

Die in **11.4** ausführlich dargelegte Rolle der natürlichen Zahlen als Anzahlmaße endlicher Mengen sei als ihr *kardinaler Aspekt* bezeichnet, und zwar im Unterschied zu ihrem *ordinalen Aspekt*, der sich wie folgt beschreiben lässt:

Jede Abzählung $\alpha$ einer endlichen nichtleeren Menge $M$ aus $n$ Elementen induziert auf natürliche Weise eine mit $<_\alpha$ bezeichnete Anordnung von $M$, indem man erklärt

$$\alpha_i <_\alpha \alpha_j \;:\Leftrightarrow\; i < j \qquad (i, j \in 1, \ldots, n).$$

Mehr noch, *jede* Anordnung $<$ von $M$ $(\neq \emptyset)$ rührt her von einer Abzählung $\alpha$ von $M$, ist also mit $<_\alpha$ für geeignetes $\alpha$ identisch (Übung 3). Auch $M = \emptyset$ lässt sich hier einbeziehen, wenn man vereinbart, die leere Paarmenge eine „Anordnung“ von $\emptyset$ zu nennen und diese mit $<_\emptyset$ zu bezeichnen. Der Beweis von $< \,=\, <_\alpha$ für geeignetes $\alpha$ benutzt wesentlich, dass $M$ im nichtleeren Falle bezüglich einer beliebig vorgegebenen Anordnung ein größtes Element besitzt. Dies ist induktiv über die Anzahl der Elemente von $M$ leicht nachweisbar, beginnend mit dem Trivialfall $|M| = 1$.

Offenbar ist die Art der Abzählung ohne Bedeutung, wenn es nur um den Anzahlvergleich oder die bloße Feststellung der Anzahl geht. Andererseits können bestimmte Abzählungen in bestimmten Zusammenhängen relevant sein, etwa die Abzählung der Waggons eines Reisezuges, sagen wir von vorn nach hinten. Je nach dem Grad der Relevanz tritt der ordinale Aspekt der natürlichen Zahlen mehr oder weniger deutlich in den Vordergrund. Dieser Doppelaspekt der natürlichen Zahlen manifestiert sich in mehr oder weniger deutlich unterschiedenen Namen für natürliche Zahlen in ihrem kardinalen oder ordinalen Gebrauch. Im Deutschen gibt es eine wortstämmige Unterscheidung für 1, die *Eins*, das *Erste*; in den romanischen und slawischen Sprachen z.B. ist diese Unterscheidung auch noch für 2 klar ausgeprägt (etwa dwa=zwei und drugy=zweite im Polnischen). Die sprachliche Angleichung ab 3 deutet daraufhin, dass die Erkenntnis, wonach der kardinale und ordinale Gebrauch natürlicher Zahlen nur zwei Aspekte derselben Sache sind, sich frühzeitig im menschlichen Bewusstsein manifestiert hat.

Dass neben den natürlichen Zahlen weitere endliche Ordinalzahlen nicht erforderlich sind, liegt daran, dass eine endliche Menge $M$ im wesentlichen auf nur eine Weise geordnet werden kann. Genauer, $(M, <)$ und $(M, <')$ sind isomorph, wie immer die Anordnungen $<$ und $<'$ von $M$ beschaffen sind. Denn sei $< \,=\, <_\alpha$ und $<' \,=\, <_\beta$ (siehe oben). Nach dem Zählsatz sind $\alpha$ und $\beta$ gleichlang, und deshalb ist $\alpha_i \mapsto \beta_i$ ein Isomorphismus. Damit sind $(M, <)$ und $(N, <')$ auch dann isomorph, wenn nur $M \sim N$.

Lange vor CANTOR war den Mathematikern klar, dass sich die Verhältnisse bei dem Versuch, das scheinbare Chaos des Unendlichen zu durchdringen und auch unendliche Mengen sinnvoll zu zählen, vollkommen verändern. CANTOR hat, im wesentlichen nur unterstützt von DEDEKIND, diese Aufgabe konsequent in Angriff genommen. Von CANTOR stammen die Begriffe der transfiniten Kardinal- und Ordinalzahlen, die jeweils für sich sinnvolle Verallgemeinerungen der natürlichen Zahlen darstellen.

Beide Arten von Zahlen müssen aber deutlich voneinander unterschieden werden. Eine unendliche Menge $M$ hat, wie CANTOR bewies, überabzählbar viele paarweise nicht-isomorphe Wohlordnungen und jede dieser Wohlordnungen entspricht einer speziellen ordinalen Abzählung von $M$. Zum Beispiel kann man $\mathbb{N}$ auch wohlordnen, indem man erklärt $n < 0$ für alle $n \neq 0$, also willkürlich $1 < 2 < 3 < \ldots < 0$ setzt. Bezüglich dieser Wohlordnung wäre 1 das kleinste, 0 jedoch das größte Element. Die zur Standard-Wohlordnung von $\mathbb{N}$ gehörende Ordinalzahl pflegt man mit $\omega$ zu bezeichnen, und die Ordinalzahl der soeben künstlich konstruierten Wohlordnung mit $\omega + 1$.

Lediglich die Kardinalzahl $|M|$ von $M$ ist durch $M$ selbst eindeutig bestimmt. Man definiert diese so, dass $|M| = |N| \ \Leftrightarrow \ M \sim N$, und erhält so ein vielfach nützliches Maß auch für den Umfang unendlicher Mengen. Auch die transfiniten Kardinalzahlen lassen sich in natürlicher Weise ordnen und sogar wohlordnen. Deren kleinste ist die mit $\aleph_0$ bezeichnete Kardinalzahl der *abzählbaren* (unendlichen) Mengen, also $|\mathbb{N}| = \aleph_0$. Leider weiß man bis heute nicht, welchen Platz $|\mathbb{R}|$ in dieser Ordnung einnimmt. Denn die zu Beginn des 20. Jahrhunderts formulierten allgemein akzeptierten mengentheoretischen Axiome lassen nachweislich keine Entscheidung über diese Frage zu. Im wesentlichen weiß man nur, dass $\mathbb{R}$ *überabzählbar* ist, also $|\mathbb{R}| > \aleph_0$.

## 11.6 Arithmetik der natürlichen Zahlen

Wir definieren nun Summe und Produkt natürlicher Zahlen und erbringen den in unserem Aufbau des Zahlensystems noch fehlenden Nachweis, dass $\mathbb{N}$ einen $\mathcal{E}$-Bereich bildet.

**Definition.** Die *Summe* $m + n$ der Zahlen $m, n \in \mathbb{N}$ sei erklärt als $|M \cup N|$, wobei $M, N$ disjunkte Mengen sind mit $|M| = m$ und $|N| = n$.

Diese Erklärung ist sinnvoll. Denn wegen Satz 11.6 ist $M \cup N$ wieder endlich und nach Satz 11.4 ist die Erklärung unabhängig von der Wahl von $M$ und $N$. Man hat sich zudem davon zu überzeugen, dass für irgend zwei gegebene Zahlen $m, n \in \mathbb{N}$ disjunkte Mengen $M, N$ mit $|M| = m$ und $|N| = n$ auch wirklich existieren. Für $m = 0$ oder $n = 0$ ist dies klar und es ist $0 + n = n + 0 = 0$. Für $m, n \in \mathbb{N}_+$ wähle man $M = (i, 0) \mid i \in A_m^+$ und $N = (i, 1) \mid i \in A_n^+$. Dann ist $|M| = m$, $|N| = n$ und $M \cap N = \emptyset$ wie verlangt. Mit der Wahl $M = 1, \ldots n$ und $N = n^+$ erkennt man übrigens sofort $n^+ = n + 1$

Aus der Erklärung der Addition ergeben sich unmittelbar die Axiome $\mathsf{N}^+, \mathsf{K}^+, \mathsf{A}^+$ aus Abschnitt **2**. Zum Beispiel gilt $m + n = n + m$, weil $M \cup N = N \cup M$, für beliebige Mengen $M, N$. Auch verifiziert man leicht Axiom $\mathsf{E}$: Sei $m + k = n$ für $k \neq 0$, sowie $m = |M|$, $k = |K|$ und $M \cap K = \emptyset$. Dann ist $n = |M \cup K|$ und somit $m < n$ nach Satz 11.5, weil $M$ echte Teilmenge ist von $M \cup K$. Umgekehrt ergibt sich ein $k \neq 0$ mit $m + k = n$ für den Fall $m < n$ als $k := |A_n^+ \backslash A_m^+|$. Fast scheint es, als ließen sich die arithmetischen Gesetze rein logisch begründen; genau hier liegt die Wurzel des zu Beginn dieses Abschnitts erwähnten FREGEschen Logizismus.

Die Multiplikation führt man auf sehr natürliche Weise über das Kreuzprodukt ein.

**Definition.** Das Produkt $m \cdot n$ der Zahlen $m, n$ sei erklärt als $|M \times N|$, wobei $M, N$ (nicht notwendig disjunkte) Mengen sind mit $|M| = m$ und $|N| = n$.

Aus entsprechenden Eigenschaften des Kreuzproduktes ergeben sich unmittelbar die Eigenschaften $\mathsf{N}^\times$, $\mathsf{K}^\times$, $\mathsf{A}^\times$, $\mathsf{F}$ und auch $\mathsf{D}$. Die Beweise sind so einfach, dass wir sie übergehen können. Für $\mathsf{K}^\times$ zum Beispiel beachte man, dass eine natürliche Bijektion zwischen $|M \times N|$ und $|N \times M|$ besteht, indem man dem geordneten Paar $(a, b) \in M \times N$ das Paar $(b, a) \in N \times M$ zuordnet. Folglich sind beide Mengen gleichzahlig.

Damit sind wir jetzt dort angelangt, wo wir in Abschnitt **2** begonnen haben. Es ist der Faden geknüpft, der von den Mengen über die natürlichen zu den reellen Zahlen und damit zur Analysis führt. Wir beweisen zum Abschluss nur noch folgenden Satz, auf dem für den Fall $g = 10$ die elementaren Rechenalgorithmen mit natürlichen Zahlen beruhen, und der auch aus anderen Gründen ziemlich wichtig ist.

**Satz 11.7 (über die $g$-adische Darstellung natürlicher Zahlen).** *Sei $g \in \mathbb{N}$ und $g \geq 2$. Dann gibt es zu jedem positiven $k$ eindeutig bestimmte natürliche Zahlen $n$ und $z_0, \ldots, z_n < g$ mit $z_0 \neq 0$ derart, dass*

$$(10) \quad k = z_0 g^n + z_1 g^{n-1} + \ldots + z_{n-1} g + z_n \quad \left(= \textstyle\sum_{i \leq n} z_i g^{n-i}\right).$$

**Beweis.** Dies ist offenbar richtig für alle $k < g$, mit $n = 0$ und $z_0 = k$. Sei nun $k \geq g$ und die Behauptung für alle positiven Zahlen $< k$ vorausgesetzt [5]. Nach dem Satz von der Division mit Rest (Übung 6) ist (a) $k = q \cdot g + r$ für gewisse $q, r \in \mathbb{N}$ mit $r < g$. Wegen $k \geq g$ ist $q \neq 0$, also $q < q \cdot 2 \leq q \cdot g \leq k$. Also gibt es gemäß Induktionsannahme eine Darstellung (b) $q = \sum_{i \leq n} z_i g^{n-i}$. Dann ergeben (a) und (b) mit $z_{n+1} := r$ offenbar

$$k = (z_0 g^n + \ldots + z_n) g + r = z_0 g^{n+1} + \ldots + z_n g + z_{n+1},$$

also eine Darstellung der gewünschten Art. Auch ist diese Darstellung eindeutig. Denn sei $k = z'_0 g^m + \ldots + z'_{m-1} g + z'_m$ nebst (10) eine weitere Darstellung von $k$. Division mit Rest durch $g$ ergibt wegen deren Eindeutigkeit offenbar $z_n = z'_m$, womit nach Subtraktion von $z_n$ und anschließender Division durch $g$ sowohl die Behauptung $n = m$ als auch die Behauptung $z_i = z'_i$ für alle $i < n$ auf die Induktionsvoraussetzung einer eindeutigen Darstellung kleinerer Zahlen zurückgeführt wird. ∎

In der Darstellung (10) heißen $z_0, \ldots, z_n$ (genauer, deren Symbole) auch die *$g$-adischen Ziffern* von $k$ und man schreibt $k$ meistens in der Weise $(z_0 z_1 \cdots z_n)_g$. Der Beweis des Satzes liefert offenbar ein Verfahren zur schrittweisen Berechnung der $g$-adischen Ziffern mittels Division mit Rest (ein bequemeres Verfahren, der $g$-adische Divisionsalgorithmus, wird in **8.2** behandelt). So ergibt sich für $g = 8$ zum Beispiel

$$2000 = 250 \cdot 8 = (31 \cdot 8 + 2) \cdot 8 = ((3 \cdot 8 + 7) \cdot 8 + 2) \cdot 8 = 3 \cdot 8^3 + 7 \cdot 8^2 + 2 \cdot 8 + 0.$$

Das Jahr 2000 trüge im Oktalsystem also die klanglose Jahreszahl $3720_8$.

[5] Wir verwenden hier eine Form der vollständigen Induktion, die sich aufgrund der in **11.3** nachgewiesenen Wohlordnung von $\mathbb{N}$ leicht rechtfertigen lässt. Siehe Übung 2.

## 11.7 Übungen

1. Man zeige: die natürliche Anordnung von $\mathbb{N}$ ist die einzige Anordnung von $\mathbb{N}$ mit der Eigenschaft $n < n^+$ für alle $n \in \mathbb{N}$.

2. Man rechtfertige die folgende, z.B. im Beweis von Satz 11.7 verwendete Variante der vollständigen Induktion: *Sei $M$ eine Menge natürlicher Zahlen derart, dass für alle $m$: wenn $k \in M$ für alle $k < m$, ist auch $m \in M$. Dann ist $M = \mathbb{N}$* [6]).

3. Sei $M$ endlich und geordnet durch $<$. Man zeige, $< = <_\alpha$ für eine geeignete Abzählung $\alpha$ von $M$. Abzählungen und Ordnungen entsprechen einander.

4. Man beweise die Äquivalenz folgender Bedingungen für beliebige Mengen $M$, die eine weitere, vom Zählen unabhängige Definition endlicher Mengen liefert:
   (i) $M$ ist endlich,
   (ii) Es gibt eine Anordnung von $M$, in der jede nichtleere Teilmenge von $M$ ein kleinstes und ein größtes Element besitzt.

5. Man zeige: $(\mathbb{N}, <)$ ist bis auf Isomorphie die einzige diskret geordnete Menge mit kleinstem und ohne größtes Element. Ferner zeige man: $(\mathbb{Z}, <)$ ist bis auf Isomorphie die einzige diskret geordnete Menge ohne kleinstes und größtes Element.

6. Man beweise den **Satz von der Division mit Rest:** *Zu jedem $g \in \mathbb{N}_+$ und $k \in \mathbb{N}$ gibt es genau ein Paar natürlicher Zahlen $q, r$ mit $k = q \cdot g + r$ und $r < g$.*

7. Man beweise die folgende **modifizierte $g$-adische Darstellbarkeit natürlicher Zahlen** $\geq 1$, in der die Ziffer 0 nicht, dafür aber eine Ziffer für $g$ benötigt wird: *Jedes $k \geq 1$ hat eine eindeutige Darstellung $k = \sum_{i \leq n} z_i g^{n-i}$ mit $1 \leq z_i \leq g$.* Hier ist auch $g = 1$ zugelassen; man erhält so die „Strichlistendarstellung“ in **11.2**. Mit einer Ziffer für $g = 10$ (etwa Z) ergibt sich eine *quasidezimale* Darstellung. Sie ist ökonomischer als die dezimale. Die ersten 110 positiven Zahlen in dieser Darstellung sind $1, \ldots, 9, \mathrm{Z}, 11, \ldots, 19, 1\mathrm{Z}, 21, \ldots, 99, 9\mathrm{Z}, \mathrm{Z}1, \ldots, \mathrm{ZZ}$ $(= 110)$.

8. Sei $\mathbb{N}$ die Zählreihe, $m \in \mathbb{N}$ und $f_m : \mathbb{N} \to \mathbb{N}$ gemäß Rekursionssatz definiert durch
$$f_m 0 = m \quad ; \quad f_m n^+ = (f_m n)^+.$$
Man erkläre $m + n := f_m n$ und beweise auf der Grundlage dieser Definition das Assoziativgesetz $\mathsf{A}^+$ und das Kommutativgesetz $\mathsf{K}^+$. Das Analoge lässt sich für die Multiplikation beweisen, mit $m \cdot n := g_m n$; dabei ist für festes $m$
$$g_m 0 = 0 \quad ; \quad g_m n^+ = g_m n + m.$$
Dies sind die von Dedekind in [5] vorgeschlagenen Reduktionen der arithmetischen Grundoperationen auf die Nachfolgerfunktion.

[6])Man beachte, dass die Voraussetzung auch die Gültigkeit von $0 \in M$ sichert. Denn es gilt $k \in M$ für alle $m < 0$ trivialerweise, weil es keine natürlichen Zahlen $m < 0$ gibt. In der Praxis bedeutet dies jedoch meistens, dass $0 \in M$ explizit nachgewiesen werden muss.

# Lösungen der Übungen

## Abschnitt 2

**1**. $b \geq c$ impliziert $a+b \geq c$. Also ist $a+b-c$ wohlerklärt und (5) folgt mit $\mathsf{K}^+, \mathsf{A}^+$ aus $c+(a+(b-c)) = a+c+(b-c) = a+b$. Für (6) und (7) genügt es, die Richtung von rechts nach links zu zeigen, die sich gemäß $\mathsf{M}^+$ durch Addition von $b$ bzw. $c$ ergibt. (8) folgt mit (5) aus $b+c+(a-b-c) = b+(c+(a-b)-c) = b+a-b = a$. Ähnlich folgt (9); denn wegen $c+(b-c) = b$ ist $b-(b-c) = c$ für $b \geq c$, mit (5) ergibt sich also

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

**2**. Induktion über $k$ ergibt leicht $a^k > 1$ für $a > 1$ und alle $k > 0$. Sei nun $n < m$, etwa $n+k = m$, so daß $a^n \cdot a^k = a^m$ gemäß $\mathsf{P}^+$. Multiplikation von $1 < a^k$ mit $a^n$ liefert $a^n < a^n \cdot a^k = a^m$. Das beweist $\mathsf{P}^<$. Für $n = 1$ gilt sicher $\mathsf{P}_<$. Induktionsschritt: $a^n < b^n$ ergibt $a^{n+1} = a^n \cdot a < b^n a < b^n b = b^{n+1}$. Aus $\mathsf{P}^<$ folgt $\mathsf{P}^=$ durch Kontraposition: Ist $n \neq m$, z.B. $n < m$, ergibt sich $a^n < a^m$, also sicher $a^n \neq a^m$. Ebenso folgt $\mathsf{P}_<$ aus $\mathsf{P}_=$.

**3**. Sei $a \geq b$ und $c := a-b$, also $a = c+b$. Dann erhalten mit (5) in Übung 1

$$a^2+b^2 = c^2+2cb+b^2+b^2 = c^2+2b^2+2ab-2b^2 = c^2+2ab.$$

Also $a^2+b^2 = (a-b)^2+2ab$, und daher $a^2+b^2 \geq 2ab$ und $(a-b)^2 = a^2+b^2-2cb$.

**4**. (10): Klar für $n = 2$, denn $(1+a)^2 = 1+2a+a^2 > 1+2a$. Induktionsschritt:

$$\begin{array}{rcll}
(1+a)^{n+1} & = & (1+a)^n(1+a) & \text{(Definition)} \\
& > & (1+na)(1+a) & \text{(Induktionsannahme)} \\
& = & 1+(n+1)a+na^2 > 1+(n+1)a. &
\end{array}$$

(11): Aus Formel (3) folgt $1-c^n < (1-c)n$ für $c < 1 < n$. Setzt man $c = 1-a$, ist also $1-(1-a)^n < na$, oder gleichwertig $1+na < (1-a)^n$. Die rechte Ungleichung folgt für $n = 2$ aus Übung 3. Induktionsschritt: Wegen $(1-a)^{n+1} = (1-a)^n - (1-a)^n a$ ist

$$\begin{array}{rll}
(1-a)^{n+1}+(n+1)a & = (1-a)^n+na+(1-(1-a)^n)a & \\
& < (1-a)^n+na+na \cdot a & \text{(eben bewiesen)} \\
& \leq 1+\binom{n}{2}a^2+na^2 & \text{(Induktionsannahme)} \\
& = 1+\binom{n+1}{2}a^2 & \text{(weil } \binom{n}{2}+n = \binom{n+1}{2}\text{)}.
\end{array}$$

(12) ist nur eine Umschrift der linken Ungleichung von (11).

**5**. Sei $f : n \mapsto (a+b)^n$ und $g : n \mapsto \sum_{k \leq n} \binom{n}{k} a^{n-k} b^k$. Dann gilt $f0 = 1 = g0$ und $f(n+1) = (a+b) \cdot fn$. Nach Satz 2.1 genügt also der Nachweis von $(a+b) \cdot gn = g(n+1)$. Dies ergibt sich gemäß Definition der $\binom{n}{k}$ durch Addition in folgender Weise:

$$\begin{array}{rcccccccccc}(a+b)\cdot\sum_{k\le n}\binom{n}{k}a^{n-k}b^k & = & a^{n+1} & + & \binom{n}{1}a^nb & + & \dots & + & \binom{n}{n}ab^n & & \\ & & & + & \binom{n}{0}a^nb & + & \dots & + & \binom{n}{n-1}ab^n & + & b^{n+1} \\ & = & a^{n+1} & + & \binom{n+1}{1}a^nb & + & \dots & + & \binom{n+1}{n}ab^n & + & b^{n+1}.\end{array}$$

# Abschnitt 3

**1**. Der Nachweis der Rekursionsgleichung $f(n+1) = fn+n+1$ für $f : n \mapsto 1+2+\ldots+n$ und $f : n \mapsto \frac{n(n+1)}{2}$ ist fast trivial. Für $f : n \mapsto \binom{n+1}{2}$ folgt diese Gleichung durch Spezialisierung der Rekursionsgleichungen für die Binomialkoeffizienten in **2.3**.

**2**. Man kann in der Dreiecksungleichung $a, b$ vertauschen. Es genügt daher, den Fall $a \ge b$ zu betrachten. Für $a \ge b > c$ ist offenbar $a - b < a - c < (a - c) + (b - c)$. Für $a \ge c \ge b$ ist $a - b = (a - c) + (c - b)$, also sogar $|a - b| = |a - c| + |b - c|$. Im Falle $c > a \ge b$ schließlich ist $a - b \le c - b \le (c - a) + (c - b) = |a - c| + |b - c|$.

**3**. Sei $a = 1 + c$. Weil $a^n = (1 + c)^n \ge 1 + nc$ nach (10) in **2.4**, muß $n$ nur so gewählt werden, daß $1 + nc > b$, also $nc > b - 1$. Dann ist $a^n \ge 1 + nc > b$.

**4**. Sei $G$ nicht archimedisch, $a, b \in A_+$ und $na < b$ für alle $n$. Sei $U$ die Menge der $x \in A$ mit $x < na$ für ein gewisses $n$ und $V = A\backslash U$. Es ist $na \in A$ für alle $n$, und $b \in V$. Man zeigt dann unschwer $(U, V)$ ist eine Lücke.

**5**. (a): Ist $a \in \mathbb{E}$, so ist $a = a_{.m}$ für ein $m$. Also $\bar{f}a = \sup fa_{.n}|n \in \mathbb{N} = fa_{.m} = fa$. (b): Sei $a < b$, und sei $c \in \mathbb{E}$ so gewählt, daß $a < c < b$. Dann ist $fa_{.n} \le fc$ für alle $n$, also $\bar{f}a = \sup fa_{.n}|n \in \mathbb{N} \le fc$. Weil $c < b_{.m}$ für ein gewisses $m$, ergibt sich $fc \le fb_{.m} \le \bar{f}b$. Also $\bar{f}a \le fc \le \bar{f}b$, und daher $\bar{f}a \le \bar{f}b$. Sei $f$ nun strikt wachsend. Man wähle $c, d \in \mathbb{E}$ mit $a \le c < d \le b$. Dann ist $\bar{f}a \le fc < fd \le \bar{f}b$, also $\bar{f}a < \bar{f}b$.

# Abschnitt 4

**1**. Für $n := 2^i \cdot 5^k$ gilt $n\frac{2^{k-i}}{10^k} = 1$, falls $i \le k$; denn $2^i \cdot 5^k \cdot 2^{k-i} = 2^k \cdot 5^k = 10^k$. Im Falle $i > k$ ist analog $n\frac{5^{i-k}}{10^i} = 1$. Sei nun $n \cdot x = m$ für ein gewisses $x \in \mathbb{E}$ der Stellenzahl $k$, so daß $x\xrightarrow{k} \in \mathbb{N}$. Dann ist $nx\xrightarrow{k} = m\xrightarrow{k} = 10^k m$. Ist $p$ Primteiler von $n$, so teilt $p$ nicht $m$, teilt also $10^k$. Daher ist $p = 2$ oder $p = 5$.

**2**. Weil $a\xrightarrow{n}, b\xrightarrow{n}$ natürliche Zahlen sind, ergeben die Kommaverschiebungsregeln

$$a < b \;\Leftrightarrow\; a\xrightarrow{n} < b\xrightarrow{n} \;\Leftrightarrow\; a\xrightarrow{n} + 1 \le b\xrightarrow{n} \;\Leftrightarrow\; (a\xrightarrow{n} + 1)\xleftarrow{n} \le b \;\Leftrightarrow\; a + \varepsilon_n \le b.$$

**3**. Für $c = b_{.n}$ ist sicher $c \le b < c + \varepsilon_n$. Erfülle nun $c$ die genannten Bedingungen. Wäre $c < b_{.n}$, folgt $c + \varepsilon_n \le b_{.n} \le b$ nach Übung 2, im Widerspruch zu $b < c + \varepsilon_n$. Wäre $b_{.n} < c$, folgt $b < b_{.n} + \varepsilon_n \le c$, im Widerspruch zu $c \le b$. Also ist $c = b_n$.

**4**. Zuerst beweisen wir (a) $mc = (m \cdot c^{\xrightarrow{n}})^{\xleftarrow{n}}$ für beliebig gegebenes $n$-stelliges $c \in \mathbb{E}$ durch Induktion über $m$. Dies ist klar für $m = 0$. Induktionsschritt:

$$\begin{aligned}
(m+1)c &= mc + c = (m \cdot c^{\xrightarrow{n}})^{\xleftarrow{n}} + c \quad &&\text{(Induktionsannahme)}\\
&= (m \cdot c^{\xrightarrow{n}} + c^{\xrightarrow{n}})^{\xleftarrow{n}} &&\text{(Definition der Summe)}\\
&= ((m+1) \cdot c^{\xrightarrow{n}})^{\xleftarrow{n}}.
\end{aligned}$$

(a) ergibt für $m = 10^k$ wegen $\mathsf{S}_4$ $10^k c = (10^k c^{\xrightarrow{n}})^{\xleftarrow{n}} = c^{\xrightarrow{n+k}\xleftarrow{n}} = c^{\xrightarrow{k}}$. Also gilt die keineswegs offensichtliche Gleichung (b) $c^{\xrightarrow{k}} = 10^k c$ für alle $c \in \mathbb{E},\ k \in \mathbb{N}$. Mit (b), $\mathsf{A}^\times$, $\mathsf{K}^\times$ und wegen $10^{n+m} = 10^m \cdot 10^n$ erhalten wir für alle $a, b \in \mathbb{E}$

$$(ab)^{\xrightarrow{n+m}} = 10^{n+m}ab = 10^n a \cdot 10^m b = a^{\xrightarrow{n}} \cdot b^{\xrightarrow{m}}$$

und damit schließlich die eigentliche Behauptung $ab = (ab)^{\xrightarrow{n+m}\xleftarrow{n+m}} = (a^{\xrightarrow{n}} \cdot b^{\xrightarrow{m}})^{\xleftarrow{n+m}}$.

**5**. Wir zeigen durch Induktion über $n$: Für alle $k \le n$ ist $k!\binom{n}{k} = \prod_{i<k}(n-i)$. Für $n = 0$ ist dies klar. Sei dies richtig für alle $k \le n$. Induktionsschritt von $n$ zu $n+1$ (wir dürfen in der Induktionsbehauptung $k+1$ statt $k$ schreiben und $k+1 < n+1$ annehmen, weil die Induktionsbehauptung für $k = 0$ sicher gilt; ebenso gilt diese für $k = n+1$, denn es ist $(n+1)!\binom{n+1}{n+1} = (n+1)! = \prod_{i<n+1}(n+1-i)$ ):

$$\begin{aligned}
(k+1)!\tbinom{n+1}{k+1} &= (k+1)!\left(\tbinom{n}{k} + \tbinom{n}{k+1}\right) = (k+1)k!\tbinom{n}{k} + (k+1)!\tbinom{n}{k+1}\\
&= (k+1) \cdot \textstyle\prod_{i<k}(n-i) + \prod_{i\le k}(n-i) \quad \text{(Induktionsannahme)}\\
&= (k+1+n-k) \cdot \textstyle\prod_{i<k}(n-i) \quad \text{(Herausziehen des letzten Faktors)}\\
&= (n+1) \cdot \textstyle\prod_{i<k}(n-i) = \prod_{i<k}(n+1-i).
\end{aligned}$$

**6**. Die Behauptung ist $a_{(n)} = b_{.n}$ mit $b := a_{.n+1} + 5\varepsilon_{n+1}$. Für $z_{n+1} \le 4$ ist $a_{(n)} = a_{.n}$ und

$$a_{.n} \le a_{.n} + (z_{n+1} + 5)\varepsilon_{n+1}\ (= b) \le a_{.n} + 9\varepsilon_{n+1} < a_{.n} + \varepsilon_n.$$

Nach Übung 3 (mit $a_{.n}$ für $c$) ist $a_{.n} = b_{.n}$, also auch $a_{(n)} = b_{.n}$. Für $z_{n+1} \ge 5$ ist

$$a_{.n} + \varepsilon_n \le a_{.n} + (z_{n+1} + 5)\varepsilon_{n+1}\ (= b) < a_{.n} + 15\varepsilon_{n+1} < (a_{.n} + \varepsilon_n) + \varepsilon_n.$$

Wiederum nach Übung 3 ist $b_{.n} = a_{.n} + \varepsilon_n$; daher $b_{.n} = a_{(n)}$ auch in diesem Falle.

# Abschnitt 5

**1**. Wegen $a_{.n}{}^{\xrightarrow{m}} = a^{\xrightarrow{m}}{}_{.n-m}$ für $n \ge m$, und $\lim\langle c_n\rangle = \lim\langle c_n\rangle_{n\ge m}$ gilt für alle $m$

$$a \cdot 10^m = \lim\langle a_{.n} \cdot 10^m\rangle_{n\ge m} = \lim\langle a_{.n}{}^{\xrightarrow{m}}\rangle = \lim\langle a^{\xrightarrow{m}}{}_{.n-m}\rangle_{n\ge m} = \lim\langle a^{\xrightarrow{m}}{}_{.n}\rangle = a^{\xrightarrow{m}}.$$

Mit $a^{\xleftarrow{m}}$ für $a$ folgt $a^{\xleftarrow{m}}10^m = a^{\xleftarrow{m}\xrightarrow{m}} = a$, also $a^{\xleftarrow{m}} = a\frac{1}{10^m} = a\varepsilon_m$. Ebenso ergibt sich aufgrund von $\mathsf{D}$ offenbar $(a+b)^{\xrightarrow{n}} = (a+b)10^n = a10^n + b10^n = a^{\xrightarrow{n}} + b^{\xrightarrow{n}}$.

**2**. Gilt (i) und ist $\varepsilon > 0$, gibt es ein $n$ mit $a - \varepsilon \le a_n$, also $a - a_n \le \varepsilon$. Daher ist $\langle a - a_n\rangle$ Nullfolge. Gilt (ii), so ist $\langle a_n\rangle$ schlicht und man sieht leicht $\lim\langle a_n\rangle = a$.

**3**. Eine Inspektion zeigt, das Kriterium in **5.2** bleibt gültig für beliebige schlichte Folgen aus $\mathbb{D}$. Damit überträgt sich auch der Beweis von Satz 5.1 auf den allgemeinen Fall.

**4**. Falls die erste Bedingung gilt, ist $z_k + z'_k + 2 \le 10$ und wegen $10\varepsilon_n = \varepsilon_{n-1}$ folgt

$$\begin{aligned} z_0 + z'_0 \;\le\; a+b \;<\; a_{.k} + b_{.k} + 2\varepsilon_k &= z_0 + z'_0 + 0,\underbrace{9\ldots9}_{k-1} + (z_k + z'_k + 2)\varepsilon_k \\ &\le z_0 + z'_0 + 0,\underbrace{9\ldots9}_{k-1} + \varepsilon_{k-1} = z_0 + z'_0 + 1, \end{aligned}$$

daher $\operatorname{Int}(a+b) = z_0 + z'_0$ $(= \operatorname{Int} a + \operatorname{Int} b)$. Falls $z_n + z'_n = 9$ für alle $n > 0$, ergibt sich $a + b = \lim\langle z_0 + z'_0 + 0,\underbrace{9\ldots9}_{n}\rangle = z_0 + z'_0 + 1$, also $\operatorname{Int}(a+b) = z_0 + z'_0 + 1$. Dasselbe gilt, falls $z_i + z'_i = 9$ für alle positiven $i < k$ und $z_k + z'_k \ge 10$. Denn wir haben

$$\begin{aligned} z_0 + z_{0'} + 2 \;>\; a+b \;\ge\; a_{.k} + b_{.k} &= z_0 + z'_0 + 0,\underbrace{9\ldots9}_{k-1} + (z_k + z'_k)\varepsilon_k \\ &\ge z_0 + z'_0 + 0,\underbrace{9\ldots9}_{k-1} + \varepsilon_{k-1} = z_0 + z'_0 + 1. \end{aligned}$$

**5**.
$$\begin{aligned} (a_{.n} + b_{.n})\overset{n}{\rightarrow} &= a_{.n}\overset{n}{\rightarrow} + b_{.n}\overset{n}{\rightarrow} = a\overset{n}{\rightarrow}_{.0} + b\overset{n}{\rightarrow}_{.0} \quad (\text{wegen } c_{.n}\overset{n}{\rightarrow} = c\overset{n}{\rightarrow}_{.0}) \\ &\le (a\overset{n}{\rightarrow} + b\overset{n}{\rightarrow})_{.0} \quad (\text{Übung } 4) \\ &= (a+b)\overset{n}{\rightarrow}_{.0} \quad (\text{weil } a\overset{n}{\rightarrow} + b\overset{n}{\rightarrow} = (a+b)\overset{n}{\rightarrow},\ \text{Übung } 1) \\ &= (a+b)_{.n}\overset{n}{\rightarrow} \quad (\text{wegen } c_{.n}\overset{n}{\rightarrow} = c\overset{n}{\rightarrow}_{.0} \text{ für } c \in \mathbb{D}). \end{aligned}$$

Folglich $a_{.n} + b_{.n} \le (a+b)_{.n}$ Analog ergibt sich $(a+b)_{.n} \le a_{.n} + b_{.n} + \varepsilon_n$ aus

$$\begin{aligned} (a+b)_{.n}\overset{n}{\rightarrow} &= (a+b)\overset{n}{\rightarrow}_{.0} = (a\overset{n}{\rightarrow} + b\overset{n}{\rightarrow})_{.0} \\ &\le a\overset{n}{\rightarrow}_{.0} + b\overset{n}{\rightarrow}_{0} + 1 \quad (\text{Übung } 4) \\ &= a_{.n}\overset{n}{\rightarrow} + b_{.n}\overset{n}{\rightarrow} + \varepsilon_n\overset{n}{\rightarrow} = (a_{.n} + b_{.n} + \varepsilon_n)\overset{n}{\rightarrow}. \end{aligned}$$

# Abschnitt 6

**1**. Sei $\langle a_n\rangle$ (fallende) Nullfolge und $b > 0$ beliebig vorgegeben. Gemäß Voraussetzung ist $a_n < \frac{1}{b}$ für ein gewisses $n$, also $b < \frac{1}{a_n}$ nach den Regeln der Bruchrechnung. Daher wächst $\langle\frac{1}{a_n}\rangle$ unbeschränkt. Gelte nun das letztere und sei $\varepsilon_k$ beliebig vorgegeben. Man wähle ein $n$ mit $\frac{1}{a_n} > \frac{1}{\varepsilon_k}$. Dann ist $a_n < \varepsilon_k$. Somit ist $\langle a_n\rangle$ eine Nullfolge.

**2**. Die Folge $\langle e'_n\rangle$ ist durch 1 beschränkt. Auch ist $e'_0 < e'_1 < e'_2$. Für $n > 2$ ergibt sich

$$\tfrac{e'_n}{e'_{n-1}} = (1 + \tfrac{1}{n(n-2)})^n \cdot \tfrac{n-2}{n-1} > (1 + \tfrac{1}{n-2}) \cdot \tfrac{n-2}{n-1} = 1.$$

Also ist $\langle e'_n\rangle$ eine schlichte Folge, und daher auch $\langle c_n\rangle$ mit $c_n := e_n \cdot e'_n = (1 - \frac{1}{n^2})^n$. Wegen $(1 - \frac{1}{n^2})^n \ge 1 - \frac{1}{n}$ ist $\lim\langle c_n\rangle = 1$. Dann folgt die Behauptung aus

$$\mathbf{e} \cdot \lim\langle e'_n\rangle = \lim\langle e_n\rangle \cdot \lim\langle e'_n\rangle = \lim\langle e_n \cdot e'_n\rangle = \lim\langle c_n\rangle = 1.$$

**3**. Nach (12) in **2.4** ist $\frac{x_n}{x_{n-1}} = (1 - \frac{x}{n(x+n-1)})^n \cdot (1 + \frac{x}{n-1}) > \frac{n-1}{x+n-1} \cdot \frac{x+n-1}{n-1} = 1$ für $n > 1$. Also ist $\langle x_n \rangle$ strikt monoton. Für $\langle \sum_{k \le n} \frac{x^k}{k!} \rangle$ gilt dasselbe. Die Beschränktheit dieser Folge ergibt sich mit $m := \operatorname{Int} x$, $s := \sum_{k<m} \frac{x^k}{k!}$ und $t := \frac{x}{m+1} < 1$ aus

$$\begin{aligned} 1 + x + \tfrac{x^2}{2!} + \ldots + \tfrac{x^{m+n}}{(m+n)!} &= s + \tfrac{x^m}{m!}(1 + \tfrac{x}{m+1} + \tfrac{x^2}{(m+1)(m+2)} + \tfrac{x^n}{(m+1)\ldots(m+n)}) \\ &\le s + \tfrac{x^m}{m!}\,(1 + t + t^2 + \ldots + t^n) \\ &= s + \tfrac{x^m}{m!}\left(\tfrac{1-t^{n+1}}{1-t}\right) \qquad \text{(Formel (4) in } \mathbf{4}\text{)} \\ &< s + \tfrac{x^m}{m!} \cdot \tfrac{1}{1-t}. \end{aligned}$$

**4**. Induktion über $n$. Klar für $n = 0$. Induktionsschritt: Sei $\frac{1}{h_0} = \sum_{i<n} \frac{1}{h_i+1} + \frac{1}{h_n}$. Wegen $\frac{1}{h_n} = \frac{h_n+1}{h_{n+1}} = \frac{h_n}{h_{n+1}} + \frac{1}{h_{n+1}} = \frac{1}{h_n+1} + \frac{1}{h_{n+1}}$ ist auch $\frac{1}{h_0} = \sum_{i<n+1} \frac{1}{h_i+1} + \frac{1}{h_{n+1}}$.

**5**. Sei angenommen, $h_0 = g_0$, sowie $h_{i+1} = h_i(h_i + 1)$ für alle $i < k$. Dann ist $h_i \le g_i$ aufgrund von (4), und wegen $\frac{1}{g_i+1} < \frac{1}{h_i}$ für $i \le k$ folgt mit Hilfe von Übung 4

$$\frac{1}{g_0+1} \le r = \frac{1}{g_0+1} + \ldots + \frac{1}{g_k+1} < \frac{1}{h_0+1} + \ldots + \frac{1}{h_k} = \frac{1}{h_0} = \frac{1}{g_0}.$$

Fast genauso beweist man $\frac{1}{g_i+1} \le r - \frac{1}{g_{i-1}+1} < \frac{1}{g_i}$ für $i = 1, \ldots k$, wie es LEONARDOS Prozedur verlangt. Damit ist die vorgegebene Darstellung (3) die LEONARDOsche.

## Abschnitt 7

**1**. Es ist $a_n = 2^{\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^n}} = 2^{1-\frac{1}{2^n}} < 2$ und $\lim\langle 1 - \frac{1}{2^n}\rangle = 1$. Da $x \mapsto 2^x$ eine schlichte Funktion ist, folgt nach (11) in **7.2** offenbar $\lim\langle a_n \rangle = 2^{\lim\langle 1-\frac{1}{2^n}\rangle} = 2^1 = 2$.

**2**. Nach (11) in **2.4** ist $(1 - \frac{1}{n^2})^n \le 1 - \frac{n}{n^2} + \binom{n}{2}\frac{1}{n^4}$, $n > 0$. Weil $\binom{n}{2} = \frac{n(n-1)}{2}$, ist also

$$\frac{e_n}{e_{n-1}} = (1 - \tfrac{1}{n^2})^n \cdot \tfrac{n}{n-1} \le (1 - \tfrac{1}{n} + \tfrac{n-1}{2n^3})\tfrac{n}{n-1} = 1 + \tfrac{1}{2n^2}.$$

Daher ist $\frac{a_n}{a_{n-1}} = \frac{e_n}{e_{n-1}}(1 - \frac{1}{2n^2-n}) \le (1 + \frac{1}{2n^2})(1 - \frac{1}{2n^2}) < 1$. Also ist $\langle a_n \rangle$ strikt fallend. Induktion ergibt unschwer $(1-x)^n + nx + \binom{n}{3}x^3 \ge 1 + \binom{n}{2}x^2$ für $0 \le x \le 1$ und $n \ge 2$. Wegen $\binom{n}{3} = \frac{n(n-1)(n-2)}{3!}$ ergibt letztere Ungleichung für $x = \frac{1}{n^2}$ und $n \ge 2$

$$(1 - \tfrac{1}{n^2})^n \ge 1 - \tfrac{n}{n^2} + \tbinom{n}{2}\tfrac{1}{n^4} - \tbinom{n}{3}\tfrac{1}{n^6} = \tfrac{n-1}{n} + \tbinom{n}{2}\tfrac{1}{n^4} - \tbinom{n}{3}\tfrac{1}{n^6} = \tfrac{n-1}{n}(1 + \tfrac{1}{2n^2} - \tfrac{n-2}{6n^4}).$$

Hieraus folgt $\frac{e_n}{e_{n-1}} = (1 - \frac{1}{n^2})^n \cdot \frac{n}{n-1} \ge \frac{n-1}{n}(1 + \frac{1}{2n^2} - \frac{n-2}{6n^4})\frac{n}{n-1} = 1 + \frac{1}{2n^2} - \frac{1}{6n^3}$. Auch für $n = 1$ ist $\frac{e_n}{e_{n-1}} \ge 1 + \frac{1}{2n^2} - \frac{1}{6n^3}$ ersichtlich. Die strikte Monotonie von $\langle b_n \rangle$ ergibt sich dann aus $\frac{b_n}{b_{n-1}} = \frac{e_n}{e_{n-1}} \cdot \frac{2n^2+n-1}{2n^2+n} \ge (1 + \frac{1}{2n^2} - \frac{1}{6n^3})\frac{2n^2+n-1}{2n^2+n} \ge 1 + \frac{1}{6n^3} > 1$. Weil $b_n < a_n$, die Folge $\langle b_n \rangle$ monoton wächst, $\langle a_n \rangle$ aber fällt, erhalten wir $e_n < b_n < a_m$ *für alle* $n, m$. Dies ergibt über die Schlichtheit von $\langle b_n \rangle$ hinaus auch $(*)$ $\mathbf{e} \le \lim\langle b_n \rangle \le \lim\langle a_n \rangle$. Ferner ist $\lim\langle a_n \rangle \le a_n = \mathbf{e}(1 + \frac{1}{2n})$, also $\lim\langle a_n \rangle - \mathbf{e} \le \frac{\mathbf{e}}{2n}$ für alle $n$. Damit ist offenbar $\lim\langle a_n \rangle = \mathbf{e}$, und wegen $(*)$ dann auch $\lim\langle b_n \rangle = \mathbf{e}$.

**3**. Seien $a_n, b_n$ wie in Übung 2. Da die Folge $\langle a_n\rangle$ strikt fallend gegen $\mathbf{e}$ konvergiert, ist $\mathbf{e} < a_n = e_n(1+\frac{1}{2n})$. Damit ergibt sich (e1) aus der Äquivalenzkette

$$\mathbf{e} < e_n(1+\tfrac{1}{2n}) \;\Leftrightarrow\; \mathbf{e}(1-\tfrac{1}{2n+1}) < e_n \;\Leftrightarrow\; \mathbf{e} - e_n < \tfrac{\mathbf{e}}{2n+1}.$$

Analog ist (e2) äquivalent zu $b_n < \mathbf{e}$. Letzteres ist richtig, weil $\langle b_n\rangle$ nach Übung 2 strikt wachsend gegen $\mathbf{e}$ konvergiert.

**4**. Wir beweisen zuerst (a): $(1+\frac{p}{qn})^n < \mathbf{e}^{\frac{p}{q}}$, gleichwertig $(1+\frac{p}{qn})^{\frac{qn}{p}} < \mathbf{e}$, für alle $p,q,n \in \mathbb{N}^+$ (man beachte $a < b^x \;\Leftrightarrow\; a^{\frac{1}{x}} < b$). Für $n = pm$ folgt die Behauptung (a) aus $(1+\frac{p}{qn})^{\frac{nq}{p}} = (1+\frac{1}{qm})^{qm} < \mathbf{e}$. Wegen der wachsenden Monotonie von $\langle(1+\frac{x}{n})^n\rangle$ gilt dann (a) für alle $n$. Weil nun die Funktionen $f_n$ und exp schlicht sind, folgt $f_n x < \mathbf{e}^x$, für alle $x \geq 0$. Damit ist $fx\ (= \lim\langle f_n x\rangle) \leq \mathbf{e}^x$ bewiesen. Außerdem folgt aus der Monotonie der $f_n$ unmittelbar die der Grenzfunktion $f$. Wäre nun $fx < \mathbf{e}^x$ für ein $x$, so gilt auch $f\frac{p}{q} \leq fx < \mathbf{e}^{\frac{p}{q}} < \mathbf{e}^x$ für gewisses $\frac{p}{q} < x$. Also $(f\frac{p}{q})^{\frac{q}{p}} < \mathbf{e}$. Wählt man $n, m$ mit $n = pm$ derart, daß $\mathbf{e} - (1+\frac{1}{qm})^{qm} < \mathbf{e} - (f\frac{p}{q})^{\frac{q}{p}}$, ist $(1+\frac{1}{qm})^{qm} > (f\frac{p}{q})^{\frac{q}{p}}$; wegen $mp = n$ ist also $(1+\frac{1}{qm})^n = (1+\frac{1}{qm})^{qm\cdot\frac{p}{q}} > f\frac{p}{q}$, und weil $\frac{1}{qm} = \frac{p}{qn}$, erhalten wir den Widerspruch $f_n\frac{p}{q} = (1+\frac{p}{qn})^n > f\frac{p}{q}$. Es ist also tatsächlich $fx = \mathbf{e}^x$ für alle $x \geq 0$.

**5**. (a) ist gleichwertig mit $\mathbf{e}^x < (1+x)\mathbf{e}^{\frac{x^2}{2}}$; dies folgt wegen $1+y < \mathbf{e}^y$ gemäß (15) mit $\frac{x^2}{2}$ für $y$ aus $\mathbf{e}^x < (1+x)(1+\frac{x^2}{2}) = 1+x+\frac{x^2}{2}+\frac{x^3}{2}$. Das gilt für $0 < x < 1$, weil

$$\mathbf{e}^x - (1+x+\tfrac{x^2}{2}) = \tfrac{x^3}{3!} + \tfrac{x^4}{4!} + \ldots < x^3(\tfrac{1}{3!}+\tfrac{1}{4!}+\ldots) = x^3(\mathbf{e}-2{,}5) < \tfrac{x^3}{2}.$$

Nun sei $0 < \vartriangle x < 1 \leq x$, so daß $\vartriangle \ln x = \ln(x+\vartriangle x) - \ln x = \ln(1+\frac{\vartriangle x}{x}) < \frac{\vartriangle x}{x}$. Also gilt (c): $\frac{\vartriangle \ln x}{\vartriangle x} < \frac{1}{x}$. (a) ergibt $\frac{\vartriangle x}{x} - \frac{1}{2}(\frac{\vartriangle x}{x})^2 < \ln(1+\frac{\vartriangle x}{x}) = \vartriangle \ln x$, also (d): $\frac{1}{x} - \frac{\vartriangle \ln x}{\vartriangle x} < \frac{\vartriangle x}{2x^2}$. Aus (c) und (d) folgt unmittelbar (b).

**6**. (a): Zuerst beweist man unschwer, die Einschränkung $g_c$ von $f_c$ auf $\mathbb{Q}$ ist monoton, mit fast der gleichen Argumentation wie in der Lösung von Aufgabe 4. Ähnlich wie dort führt man auch die Annahme $x < y$ und $(1+\frac{c}{x})^x > (1+\frac{c}{y})^y$ zum Widerspruch. (b): Es genügt zu zeigen, $f : x \mapsto \ln(a+x) - \ln a$ ist konvex. Dies entspricht einer Verschiebung des Koordinatenursprungs (0;0), die diesen in den Punkt $(a;\ln a)$ verlegt. $h$ verläuft durch den neuen Nullpunkt und hat dann die Gleichung $y = \frac{\ln(1+\frac{b}{a})}{b}x$. Zu zeigen ist $hc = \frac{\ln(1+\frac{b}{a})}{b}c \leq fc = \ln(a+c) - \ln c = \ln(1+\frac{c}{a})$ für $a < c < b$, gleichwertig $\frac{\ln(1+\frac{b}{a})}{b} \leq \frac{\ln(1+\frac{c}{a})}{c}$, oder auch $(1+\frac{b}{a})^{\frac{1}{b}} \leq (1+\frac{c}{a})^{\frac{1}{c}}$. Die Behauptung reduziert sich so auf (a): $x \mapsto (1+\frac{x}{a})^{\frac{1}{x}}$ fällt monoton. (c): Sei $c_n = \frac{\ln(1+a_n)}{a_n}$. Daß $\langle c_n\rangle$ wächst, folgt aus der Konvexität von $x \mapsto \ln(1+x)$, weil $c_n$ die Steigung der Geraden durch (0;0) und $(a_n;c_n)$ angibt. $\lim\langle c_n\rangle = 1$ ergibt sich leicht aus Übung 5(b) für den Fall $x = 1$.

**7**. Monotones Wachstum $f(x) \leq f(x+h)$ erweist sich nach leichter Rechnung als äquivalent zu $x\cdot(x+h) \leq a$. Dies ist richtig für $x \leq x+h \leq \sqrt{a}$. Ferner ist $x \leq fx$ für $x \neq 0$ äquivalent zu $x^2 \leq a$ oder gleichwertig $x \leq \sqrt{a}$, so daß dann $x \leq fx \leq f(\sqrt{a})$.

# Abschnitt 8

**1**. Ein reinperiodisches $a$ ist gemäß (2) in **6.2** Bruch mit Nenner $10^p - 1$ der zu 10 offenbar teilerfremd ist. Sei umgekehrt $a = \frac{m}{n}$ und $r_i = r_k$ gemäß Divisionsalgorithmus, $0 < i < k$. Da $r_j = 10r_{j-1} - nu_j$ für alle $j$, folgt $10r_{i-1} - nu_i = 10r_{k-1} - nu_k$. Ist etwa $r_{i-1} \geq r_{k-1}$, so ergibt sich $10(r_{i-1} - r_{k-1}) = n(u_i - u_k)$. Sind $n$ und 10 teilerfremd und ist $u_i \neq u_k$, so ist $n$ Teiler von $r_{i-1} - r_{k-1} < n$; also $r_{i-1} = r_{k-1}$. Dasselbe gilt auch, falls $u_i = u_k$. Daher ist offenbar $r_0 = r_p$ für ein $p > 0$. Also $z_1 = z_{p+1},\ z_2 = z_{p+2}$, usw. Mit anderen Worten, $a$ ist reinperiodisch.

**2**. (a) Es ist $a = \frac{k}{10^p-1}$ für ein $k \in \mathbb{N}$ nach (2) in **6.2**. Also $m(10^p - 1) = nk$. Weil $n, m$ teilerfremd sind, ist $n$ Teiler von $10^p - 1$. (b) Sei $n$ Teiler von $10^q - 1$, $q \in \mathbb{N}_+$. Dann ist $a = \frac{k}{10^q-1}$ für ein gewisses $k$, also $a(10^q - 1) = k$, d.h. $a\overset{q}{\rightarrow} = a + k$. Ist etwa $a = z_0{,}z_1 \dots$, so heißt dies $z_0 \dots z_q{,}z_{q+1} \dots = z_0'{,}z_1 \dots$ mit $z_0' = z_0 + k$. Also $z_{q+i} = z_i$ für $i = 1, 2, \dots$ Folglich ist $q \geq p$ nach Definition der Periodenlänge $p$.

**3**. Es genügt, dies für $z_0 = 0$ zu zeigen, da sich der allgemeine Fall durch Kommaverschiebung nach links leicht auf diesen reduzieren läßt. Seien $u_i, v_i, w_i$ die drei zur Oktalziffer $z_i$ gehörenden Dualziffern, also $z_i = u_i \cdot 2^2 + v_i \cdot 2 + w_i$. Dann gilt sicher $\frac{z_1}{8^1} + \dots + \frac{z_n}{8^n} = \frac{z_1}{2^3} + \dots + \frac{z_n}{2^{3n}} = \frac{u_1}{2} + \frac{v_1}{2^2} + \frac{w_1}{2^3} + \dots + \frac{u_n}{2^{3n-2}} + \frac{v_n}{2^{3n-1}} + \frac{w_n}{2^{3n}}$. Dies ergibt

$$0{,}z_1 z_2 \dots {}_8 = \frac{z_1}{2^3} + \frac{z_2}{2^6} + \dots = \frac{u_1}{2} + \frac{v_1}{2^2} + \frac{w_1}{2^3} + \frac{u_2}{2^4} + \frac{v_2}{2^5} + \frac{w_2}{2^6} + \dots = 0{,}u_1 v_1 w_1 u_2 v_2 w_2 \dots {}_2.$$

**4**. Sei $a = u_0 u_1 \dots u_m{,}z_0 z_1 \dots$ mit $g$-adischen Ziffern $u_i$ und $z_i$. Dann ist nach **7.1**

$$\begin{aligned} ag^n &= (u_0 g^m + \dots + u_m g^0 + \tfrac{z_1}{g^1} + \tfrac{z_2}{g^2} + \dots)g^n \\ &= u_0 g^{m+n} + \dots + u_m g^n + z_1 g^{n-1} + \dots + z_n g^0 + \tfrac{z_{n+1}}{g} + \tfrac{z_{n+2}}{g^2} + \dots \\ &= u_0 \cdots u_m z_1 \cdots z_n{,}z_1 z_2 \cdots {}_g = a\overset{n}{\rightarrow}. \end{aligned}$$

Wegen $a\overset{n}{\leftarrow} g^n = a\overset{n}{\leftarrow}\overset{n}{\rightarrow} = a$ ist dann $a\overset{n}{\leftarrow} = \frac{a}{g^n}$. Für $n$-stellige $a, b \in \mathbb{E}^g$ sind $a\overset{n}{\rightarrow}, b\overset{n}{\rightarrow}$ natürliche Zahlen, also $(a+b)\overset{n}{\rightarrow} = (a+b)g^n = ag^n + bg^n = a\overset{n}{\rightarrow} + b\overset{n}{\rightarrow} \in \mathbb{N}$. Daher ist $a + b \in \mathbb{E}^g$. Für das Produkt zeigt man dies ganz analog. Hieraus folgt übrigens leicht, daß $\mathbb{E}^g$, genau wie $\mathbb{E} = \mathbb{E}^{10}$, einen $\mathcal{E}$-Bereich bildet.

**5**. Wegen Übung 4 übertragen sich die Beweise der Sätze 6.2 und 8.2 fast wörtlich.

**6**. Für $f : x \mapsto \frac{1}{2}(x + \frac{a}{x})$ und $g : x \mapsto \frac{2ax}{a+x^2}$ gilt $fx \cdot gx = a$. Im Falle $x_0 \leq \sqrt{a}$ ist $\frac{a}{x_1} = \frac{a}{fx_0} = gx_0 \leq \sqrt{a}$, siehe Text; daher $x_1 \geq \sqrt{a}$. Für $x \geq \sqrt{a}$ ist $x \geq fx \geq \sqrt{a}$. Also fällt $\langle x_n \rangle_{n>0}$ und ist damit konvergent. Wegen $\lim\langle x_n \rangle = \lim\langle x_{n+1} \rangle = \lim\langle fx_n \rangle$ genügt $b := \lim\langle x_n \rangle$ der Gleichung $x = \frac{1}{2}(x + \frac{a}{x})$, also $b = \sqrt{a}$.

**7**. Für $0 < x \leq \frac{1}{a}$ ist $0 \leq (1 - ax)^2 = 1 + a^2x^2 - 2ax$, was $f(x) := 2x - ax^2 \leq \frac{1}{a}$ ergibt. D.h. $f : I \to I$ mit $I := [0, \frac{1}{a}]$. Auch ist $f$ schlicht in $I$. Also konvergiert $\langle x_n \rangle$ gegen die Lösung $\xi = \frac{1}{a}$ von $x = 2x - ax^2$, und zwar quadratisch; denn wegen $0 < ax_n \leq 1$ ist

$$d_{n+1} = \tfrac{1}{a} - (2x_n - ax_n^2) = a(\tfrac{1}{a^2} + x_n^2 - 2\tfrac{x_n}{a}) = a(\tfrac{1}{a} - x_n)^2 = ad_n^2.$$

Damit ist die Bedingung $d_{n+1} \leq c \cdot d_n^2$ mit $c = a$ erfüllt.

**8**. Wir verwenden der Kürze halber den Differentialkalkül und negative Zahlen, sowie die Tatsache, daß für die Kontraktionseigenschaft einer in einem Intervall $I$ differenzierbaren Funktion $f$ der Nachweis von $(*)$ $|f'x| \leq C$ in $I$ für ein $C < 1$ hinreicht. $f'$ bezeichnet die Ableitung von $f$. Wir wählen $I = [u, u + \frac{13}{6}]$. Sei $b := \sqrt[3]{a}$ und $x \in I$. Falls $x \leq b$, also $x^3 \leq a < (u+1)^3$, gilt unter Beachtung von $u \geq 1$

$$u \leq x \leq Fx < u + 1 + \tfrac{(u+1)^3 - u^3}{3(u+1)u} = u + 1 + \tfrac{3u^2+3u+1}{3u^2+3u} = u + 2 + \tfrac{1}{3u^2+3u} \leq u + \tfrac{13}{6}.$$

Auch für $x > b$, also $x^3 - a > 0$, ist $Fx = x - \frac{x^3 - a}{3(u+1)x} < x \leq u + \frac{13}{6}$. In diesem Fall bestätigt man $u \leq Fx$ z.B. wie folgt: Es ist $Fb = b \geq u$, $F'b > 0$, $F$ ist konvex für $x \geq b$ – weil dort die 2. Ableitung $F \gg x < 0$ ist – und es ist noch $F(u + \frac{13}{6}) > u$. Daraus folgt offenbar $u \leq Fx$ im Intervall $[b, u + \frac{13}{6}]$. Nun beweisen wir $(*)$. Wegen $u \leq u^3 \leq a$ ist $F'(x) = 1 - \frac{a+2x^3}{3(u+1)x^2} \leq 1 - \frac{3u^3}{3(u+1)(u+\frac{13}{6})^2} < 1$; ferner, mit $c := \frac{2}{3}\frac{u+\frac{1}{3}}{u+1}$ ist $F'x \geq 1 - \frac{(u+1)^3}{3(u+1)u^2} - \frac{2}{3}\frac{u+\frac{1}{3}}{u+1} \geq 1 - \frac{4}{3} - c > -1$ für $u \leq x \leq u + \frac{1}{3}$; für $u + \frac{1}{3} < x \leq u + \frac{13}{6}$ hingegen ist $F'x \geq 1 - \frac{1}{3}\frac{(u+1)^2}{(u+\frac{1}{3})^2} - \frac{2}{3}\frac{u+\frac{13}{6}}{u+1} \geq 1 - \frac{1}{3}(1+\frac{2}{4})^2 - \frac{2}{3}(1+\frac{7}{12}) = 1 - \frac{3}{4} - \frac{19}{18} > -1$.

# Abschnitt 9

**1**. Für $x \in R$ haben $x$ und $x^n$ für ungerades $n$ nach (6) und (7) das gleiche Vorzeichen (d.h. beide sind entweder positiv, negativ oder beide sind 0). Daraus folgt leicht, daß $x \mapsto x^n$ in $R$ durchgehend strikt monoton wächst. Also hat $(*)$ dort höchstens eine Lösung. In $\mathbb{R}$ hat $x^n = |c|$ wegen des Zwischenwertsatzes gewiß eine nichtnegative Lösung $b$. Je nachdem ob $c \geq 0$ oder $c < 0$ ist dann offenbar $b$ oder $-b$ Lösung von $(*)$.

**2**. (i) $\Rightarrow$ (ii): Ist $-1 = \iota^2$ für $\iota \in K$, und $a = b^2$, so $-a = -1 \cdot b^2 = \iota^2 b^2 = (\iota b)^2$. Also ist auch $-a$ Quadrat. (ii) $\Rightarrow$ (iii): Wegen $1^2 = 1$ gibt es ein $\iota$ mit $\iota^2 = -1$. Für $a = 1$, $b = \iota$ ist dann $a^2 + b^2 = 0$. (iii) $\Rightarrow$ (i): Ist $a^2 + b^2 = 0$ und etwa $a \neq 0$, so ist $(\frac{b}{a})^2 = -1$.

**3**. Sei $\eta\colon K[\mathbf{i}] \to K'$ durch $\eta(a + \mathbf{i}a') = a + \iota a'$ erklärt. Offenbar ist $\eta(x+y) = \eta x + \eta y$ und $\eta(x \cdot y) = \eta x \cdot \eta y$ für alle $x, y \in K$. Die Abbildung $\eta$ ist injektiv; denn $a + \iota a' = b + \iota b'$ impliziert $a - b = \iota(b' - a')$, also $b' - a' = 0$ – sonst läge $\iota = \frac{a-b}{b'-a'}$ bereits in $K$. Daher ist $a' = b'$, $a = b$, und folglich ist $\eta$ eine Einbettung von $K[\mathbf{i}]$ in $K'$.

**4**. Sei $K' := K[\mathbf{i}]$ und $\iota := -\mathbf{i}$ $(\in K')$, sowie $\eta\colon K[\mathbf{i}] \to K'$ definiert wie in Punkt 3. Dann ist offenbar $\eta z = \overline{z}$ für alle $z \in K[\mathbf{i}]$. Weil $\eta$ eine Einbettung von $K[\mathbf{i}]$ in sich (sogar ein Automorphismus) ist, gelten die Gleichungen $\overline{x+y} = \overline{x} + \overline{y}$ und $\overline{x \cdot y} = \overline{x} \cdot \overline{y}$.

**5**. Seien $x = a + \mathbf{i}\alpha, y = b + \mathbf{i}\beta \in \mathbb{C}$, mit $a, b, \alpha, \beta \in \mathbb{R}$, also $x + y = a + b + \mathbf{i}(\alpha+\beta)$. Dann liefert die Definition leicht $\mathbf{e}^x \cdot \mathbf{e}^y = \mathbf{e}^a\mathbf{e}^{\mathbf{i}\alpha} \cdot \mathbf{e}^b\mathbf{e}^{\mathbf{i}\beta} = \mathbf{e}^{a+b} \cdot \mathbf{e}^{\mathbf{i}\alpha}\mathbf{e}^{\mathbf{i}\beta} = \mathbf{e}^{a+b} \cdot \mathbf{e}^{\mathbf{i}(\alpha+\beta)} = \mathbf{e}^{x+y}$.

**6**. Sei $c = r\mathbf{e}^{\mathbf{i}\alpha}$ mit $r = |c|$, $\alpha = \arg c$, $-\boldsymbol{\pi} \leq \alpha < \boldsymbol{\pi}$, sowie $s := \sqrt[n]{r}$, $\beta_k := \frac{\alpha + k\cdot 2\boldsymbol{\pi}}{n}$, und $x_k := s\mathbf{e}^{\mathbf{i}\beta_k}$ $(k = 1, \dots, n)$. Dann ist $x_k^n = s^n\mathbf{e}^{\mathbf{i}\beta_k n} = r\mathbf{e}^{\mathbf{i}(\alpha + k\cdot 2\boldsymbol{\pi})} = r\mathbf{e}^{\mathbf{i}\alpha} = c$. Also ist $x_k$ Lösung von $(*)$. Es genügt daher, $x_k \neq x_m$ für $k \neq m$ zu bestätigen. Das aber ist klar, denn $\arg x_k - \arg x_m = \beta_k - \beta_m = 2\boldsymbol{\pi}\frac{k-m}{n}$ ist kein ganzzahliges Vielfaches von $2\boldsymbol{\pi}$.

**7**. (a): Sei $b_n = \frac{\ln(1+a_n)}{a_n}$. Fällt $\langle a_n\rangle$, konvergiert $\langle b_n\rangle$ nach Übung 7.5 wachsend gegen 1. Analog konvergiert $\langle b_n\rangle$ fallend gegen 1, falls $\langle a_n\rangle$ wachsende Nullfolge aus negativen Gliedern ist. Sei nun $a_n := x^{\frac{1}{n}} - 1$ und $b_n$ wie oben. Also $1 + a_n = x^{\frac{1}{n}}$ und damit $n = \frac{\ln x}{\ln(1+a_n)}$. Es ist $\langle a_n\rangle$ wegen $\lim\langle x^{\frac{1}{n}}\rangle = \lim\langle \mathbf{e}^{\frac{\ln x}{n}}\rangle = \mathbf{e}^{\lim\langle\frac{\ln x}{n}\rangle} = 1$ für $x > 1$ fallende, und für $0 < x < 1$ wachsende Nullfolge; also konvergiert $\langle b_n\rangle$ wachsend bzw. fallend gegen 1. Da $\ln x > 0$ für $x > 1$ und $\ln x < 0$ für $0 < x < 1$, konvergiert $\langle\frac{\ln x}{b_n}\rangle$ in beiden Fällen fallend gegen $\ln x$. Es ist nun aber gerade $\frac{\ln x}{b_n} = \frac{\ln x}{\ln(1+a_n)}a_n = na_n = n(x^{\frac{1}{n}} - 1)$.

# Abschnitt 10

**1**: Sei $a = me$, $b = ne$. Dann ist $na = nme = mne = mb$; also wähle $i = m$, $k = n$. Sei nun $ia = kb$, wobei $i, k$ o.B.d.A. teilerfremd sind. Mittels des bekannten euklidischen Algorithmus zur Bestimmung des größten gemeinsamen Teilers ($= 1$ für $i, k$) erhält man natürliche Zahlen $p, q$ mit $pi + 1 = qk$ oder $pi = qk + 1$. Liege etwa der erste Fall vor. Dann ist $iqa = qia = qkb = (pi + 1)b > ipb$, also $qa > pb$. Es sind dann $a, b$ kommensurabel im definierten Sinne, mit $e := qa - pb$. Denn wegen $kq - pi = 1$ ist $ke = kqa - pkb = kqa - pia = (kq - pi)a = a$. Völlig analog zeigt man $ie = b$. Nach der soeben angegebenen Kennzeichnung ist die Kommensurabilität transitiv auf der Menge $G_+$. Reflexiv und symmetrisch ist sie trivialerweise. Schließlich sind die Größen $na + mb$ für teilerfremde $n, m$ paarweise inkommensurabel. Denn seien $c = na + mb$ und $d = pa + qb$ derartige Elemente und seien $c, d$ kommensurabel, etwa $ic = kd$. Die Inkommensurabilität von $a, b$ ergibt $in = kp$ und $im = kq$. Daraus folgt $mp = nq$. Damit teilen sich $n$ und $p$ gegenseitig; also $n = p$ und $m = q$. Daher ist $c = d$.

**2**. Sei $e \in G_+$. Dann ist $\rho : \mathbb{Q} \to G$ mit $\rho\frac{m}{n} = \frac{m}{n}e$ additiv und ordnungstreu. $\rho$ ist auch surjektiv, weil jedes $a \in G$ mit $e$ kommensurabel ist. Also ist $\rho$ ein Isomorphismus.

**3**. Sei $G'$ Menge aller Schreibfiguren $\frac{a}{n}$ ($a \in G$, $n \in \mathbb{N}_+$). Man setze $\frac{a}{n} = \frac{b}{m}$ falls $ma = nb$, und $\frac{a}{n} < \frac{b}{m}$ falls $ma < nb$, sowie $\frac{a}{n} + \frac{b}{m} = \frac{ma+nb}{nm}$. Diese Definitionen sind repräsentantenunabhängig, wie man leicht nachrechnet. Ferner überprüft man mühelos die Axiome für Größenbereiche. Man identifiziere $a \in G$ mit $\frac{a}{1}$; dies ist erlaubt, weil die Abbildung $a \mapsto \frac{a}{1}$ sicher eine Einbettung von $G$ in $G'$ darstellt. $G'$ ist dann ein teilbarer Größenbereich, wie man leicht sieht. Mit $G$ ist auch $G'$ archimedisch. Denn sei etwa $o < \frac{a}{m} < \frac{b}{k}$. Wählt man $n$ mit $na > mb$, folgt $n\frac{a}{m} = \frac{na}{m} > \frac{mb}{m} = \frac{b}{1} \geq \frac{b}{k}$.

**4**. Sei $G$ archimedisch, $(U, V)$ Schnitt von $G$ und $v \in V$. Es gibt ein $k$ mit $a := k\varepsilon \in U$, $b := (k + 1)\varepsilon \in V$ (sonst wäre $n\varepsilon < v$ für alle $n$), und gewiß ist $b - a \leq \varepsilon$. Seien nun Lücken beliebig klein und angenommen, $V := x \in G \,|\, x \geq na$ für alle $n$ sei nichtleer für ein $a \in G_+$, $U := G\mathtt{V}$, und $u \in U$, $v \in V$, $v - u < a$. Dann ist $u < na$ für ein $n$. Also $v - na < v - u < a$. Das ergibt $v < (n + 1)a$, ein Widerspruch zu $v \in V$.

**5**. Aus ($<$) folgt trivialerweise (i) und hieraus (ii). Sei letzteres angenommen. Dann ist $\varphi o = o$. Denn $o + o = a$ ergibt $\varphi o + \varphi o = \varphi o$, also $\varphi o = o$ wegen der Streichungsregel.

Sei nun $a < b$ für $a, b \in G$, also $a + c = b$ für ein $c \in G_+$. Dann ergibt sich $\varphi a + \varphi c = \varphi b$. Es ist $\varphi c \neq o$, denn $o \in G'$ ist als Bild von $o \in G$ schon vergeben. Daher ist $\varphi a < \varphi b$. Ferner: Nach dem Bewiesenen genügt es, (ii) zu beweisen. Da $\varphi$ nichttrivial ist, gibt es ein $a \in G$ mit $\varphi a > o$. Ixt $x \in G_+$, und etwa $nx \geq \varphi a$, so ist $n\varphi x = \varphi nx \geq \varphi a > o$, also notwendigerweise auch $\varphi x \neq o$.

**6**. Nach Bemerkung 1 besitzt $G$ eine Skalierung $\sigma : G \to \mathbb{D}$ und diese ist offenbar eine Einbettung von $G$ in den lückenlosen Größenbereich $\mathbb{D}$.

**7**. Man wähle ein $n$ mit $c < n(b - a)$, also (a) $na + c < nb$, und sodann ein $m$ mit $(m - 1)c \leq na < mc$, so daß (b) $na < mc \leq na + c$. Dann folgt $na < mc < nb$ unmittelbar aus (b) und (a).

**8**. Wir setzen $\beta a$ für $a$, $\alpha a$ für $b$, $\alpha x$ für $c$ in Übung 7. Dann ist $n\beta a < m\beta x < n\alpha a$ für gewisse $m, n$, und daher $\beta na = n\beta a < m\beta x = \beta mx$. Folglich $na < mx$. Mithin ist $m\beta x < n\alpha a = \alpha na < \alpha mx = m\alpha x$. Daher ist $\beta x < \alpha x$, d.h. $\alpha x > \beta x$.

**9**. Die ordnungstreuen Operatoren seien die positiven, die ordnungsinvertierenden die negativen Operatoren genannt. Zuerst zeigen wir: Ist $\gamma := \alpha + \beta$ nicht der Nulloperator, so ist $\gamma$ positiv oder negativ. Dies ist klar, falls $\alpha, \beta$ beide positiv oder negativ sind. Sei $\alpha$ positiv, $\beta$ negativ. Die Einschränkungen von $\alpha$ und $-\beta$ auf $G_o$ bilden $G_o$ in sich ab und sind dort offenbar Operatoren. Gewiß ist $\gamma a \neq o$ für ein $a > o$. Sei etwa $\gamma a > o$, also $\alpha a > -\beta a$. Nach Übung 8 ist $\alpha x \geq -\beta x$ für alle $x \in G_o$, also $\alpha x + \beta x \geq o$. Somit ist $\gamma = \alpha + \beta$ ordnungstreu zumindest auf $G_o$. Wegen $\gamma(-x) = -\gamma x$ wird $G_-$ durch $\gamma$ ordnungstreu nach $G_-$ abgebildet. Also ist $\gamma$ insgesamt ordnungstreu und damit positiv. Falls $\gamma a < o$, ist $\gamma$ negativ. Dies ergibt sich leicht durch Betrachtung von $\delta = -\gamma$ und Rückführung auf den diskutierten Fall. Damit ist $\Omega$ gegenüber $+$ abgeschlossen, ebenso wie gegenüber $\cdot$. Das Nachrechnen der Ringaxiome ist banal. Sei $\Pi = \alpha \in \Omega \mid \alpha$ *positiv.* Man definiere $\alpha < \beta \;\Leftrightarrow\; \beta - \alpha \in \Pi$. Die Relation $<$ ist irreflexiv und transitiv, weil $\gamma, \delta \in \Pi \;\Rightarrow\; \gamma + \delta \in \Pi$. Für $\alpha \neq 0$ ist entweder $\alpha \in \Pi$ oder $-\alpha \in \Pi$, also ist $\alpha < \beta$ oder $\beta < \alpha$ für $\alpha \neq \beta$. Auch überprüft man auf einfachste Weise die beiden Monotoniegesetze. Um schließlich nachzuweisen, daß $\Omega$ ein archimedischer Ring ist, sei $0 < \alpha < \beta$ und $a \in G_+$. Man wähle ein $n$ mit $n(\alpha a) > \beta a$. Bezeichnet $n\gamma$ die $n-$fache Summe eines Operators $\gamma$, folgt offenbar $(n\alpha)a = n(\alpha a) > \beta a$. Daher $(n\alpha - \beta)a > 0$, weswegen $n\alpha - \beta$ bereits positiv ist. Also $n\alpha > \beta$.

**10**. Nach dem Isomorphiesatz für lückenlose Größengruppen ist $G$ isomorph zur Größengruppe $\mathbb{R}$. Also ist der Operatorenring von $G$ isomorph zum Operatorenring $\Omega$ von $\mathbb{R}$. Daher genügt es, einen Isomorphismus des Körpers $\mathbb{R}$ auf $\Omega$ anzugeben, womit sich auch $\Omega$ als Körper erweist. Mit $(\mathbb{R}, +, 0, <)$ ist auch $(\mathbb{R}, +, 0, <^*)$ lückenlose Größengruppe, wobei $r <^* s \;\Leftrightarrow\; s < r$ für $r, s \in \mathbb{R}$. Nach dem Isomorphiesatz für lückenlose Größengruppen gibt es daher zu jedem $r \in \mathbb{R}$ genau ein $\alpha_r \in \Omega$ mit $\alpha_r 1 = r$. Auch ist $\Omega = \alpha_r \mid r \in \mathbb{R}$. Es ist $\alpha_{r+s} = \alpha_r + \alpha_s$, $\alpha_{r \cdot s} = \alpha_r \cdot \alpha_s$ (Beweis wie im Text für den Größenbereich $\mathbb{R}_0$). Also ist $r \mapsto \alpha_r$ Isomorphismus des Körpers $\mathbb{R}$ auf dem Ring $\Omega$.

# Abschnitt 11

**1**. Sei $<'$ eine beliebige Anordnung von $\mathbb{N}$ mit der Eigenschaft $k <' k^+$ für alle $k$. Es genügt, $(*)$ $n <' m \;\Rightarrow\; n < m$ zu beweisen; denn dann gilt hiervon auch die Umkehrung. $(*)$ folgt durch Induktion über $n$. Für $n = 0$ ist dies klar, denn $0 <' m$ impliziert $0 \neq m$ und daher $0 < m$. Es gelte nun $(*)$ für $n$. Sei $n^+ <' m$ angenommen. Wegen $n <' n^+$ folgt $n <' m$, und damit $n < m$ gemäß Induktionsannahme. Daher ist $n^+ \leq m$. Weil $n^+ <' m$, ist $n^+ = m$ ausgeschlossen. Also ist auch $n^+ < m$.

**2**. Es genügt zu zeigen $L := \mathbb{N}\mathtt{M} = \emptyset$. Andernfalls hätte $L$ nach (9) in **11.3** ein kleinstes Element $m$. Also $k \in M$ für alle $k < m$. Gemäß Voraussetzung ist dann aber auch $m \in M$, was $m \in L$ widerspricht.

**3**. Induktion über die Anzahl der Elemente von von $M$. Klar für $M = \emptyset$, mit der leeren Abzählung $<_\emptyset$. Sei nun $M = N \cup a$, $|M| = n^+$, $|N| = n$. Wir dürfen o.B.d.A. annehmen, $a$ sei größtes Element bezüglich der gegebenen Ordnung $<$ von $A$. Nach Induktionsannahme existiert eine Abzählung $\beta$ von $N$ derart, daß $<^N = <_\beta$, wobei $<^N$ die Einschränung der Ordnung $<$ auf $N$ bezeichne. Dann erfüllt die verlängerte Abzählung $\alpha = \beta \cup (n^+, a)$ von ganz $M$ offenbar die gewünschte Forderung $<_\alpha = <$. Denn ist $i < n^+$ so ist auch $a_i < a_{n^+} = a$, weil $a$ größtes Element von $A$ ist.

**4**. (i) $\Rightarrow$ (ii): Klar für $M = \emptyset$, denn (ii) ist mit der leeren Abzählung erfüllt. Sonst ist $< = <_\alpha$ für eine Abzählung $\alpha$ nach Übung 3. $A_n^+$ hat aber mit der Einschränkung der Ordnung von $n$ auf $A_n^+$ wegen (8) und (9) in **11.3** die in (ii) genannte Eigenschaft, die sich mittels des Isomorphismus $i \mapsto a_i$ $(i \in A_n^+)$ selbstredend auf $<$ überträgt.
(ii) $\Rightarrow$ (i): Klar für $M = \emptyset$. Habe die Ordnung von $M$ $(\neq \emptyset)$ die unter (ii) genannte Eigenschaft. Für $a \in M$ sei $a^+ = a$, falls $a$ größtes Element von $M$ ist; sonst sei $a^+$ das gewiß existierende kleinste Element von $x \in M \mid x > a$, der *unmittelbare Nachfolger* von $a$. Wir definieren nun $f : \mathbb{N} \to M$ rekursiv durch

$$f0 = \text{ kleinstes Element von } M \quad ; \quad fn^+ = (fn)^+.$$

Sei $U := fn \mid n \in \mathbb{N}$ das Bild von $f$. Durch Induktion über $n$ beweist man unschwer

$$\text{(a) } fn \neq fn^+ \;\Rightarrow\; fk < fm \;\; (k < m \leq n^+) \quad ; \quad \text{(b) } a \leq fn \;\Rightarrow\; a \in U.$$

Wir behaupten, $fn = fn^+$ für ein gewisses $n$. Denn andernfalls hat $U$ offenbar kein größtes Element, im Widerspruch zur Voraussetzung. Also $fn = fn^+$ für ein gleich minimal gewähltes $n$. Nach Definition ist $fn$ dann größtes Element von $M$, und (b) ergibt $U = M$. Also bildet $f$ den endlichen Anfang $A_n^+$ gemäß (a) ordnungstreu, insbesondere bijektiv auf $M$ ab. Folglich ist $M$ endlich.

**5**. Alle Definitionen und Bezeichnungen seien dieselben wie unter Punkt **4** oben. Nunmehr ist stets $fn \neq fn^+$; also hat $U$ kein größtes Element und ist deshalb nach Übung 3 in $M$ unbeschränkt. Ist daher $a \in M$, so ist $a \leq fn$ für ein $n$, und damit $a \in U$ gemäß (b). Also ist $U = M$, d.h. $f$ ist surjektiv. Nach (a) ist $f$ auch ordnungstreu, also ein

Isomorphismus von $(\mathbb{N},<)$ auf $(M,<)$. Das ergibt leicht die zweite Behauptung; denn nach dem Bewiesenen läßt sich $\mathbb{N}$ für $a \in M$ ordnungstreu auf $y \in M \mid y \geq a$ abbilden, und ebenso $x \in \mathbb{Z} \mid x < 0$ auf $y \in M \mid y < a$.

**6**. Beweis durch Induktion über $k$. Die Behauptung ist klar für $k = 0$. Ist $k = qg + r$, so ist für $r < g - 1$ offenbar $k^+ = qg + r^+$, und für $k = g - 1$ ist $k^+ = q^+g + 0$ eine gewünschte Darstellung. *Eindeutigkeit*: Wir behaupten $qg+r = q'g+r'$ impliziert $q = q'$ und $r = r'$. Denn sei etwa $q \geq q'$. Dann folgt $(q - q')g = r' - r$. Also ist $g$ ein Teiler von $r' - r < g$. Daher ist $r' - r = 0$, also $r' = r$, und folglich auch $q' = q$.

**7**. Für $1 \leq k \leq g$ ist dies klar. Ersetzt man im Divionssatz mit Rest den Fall $k = qg+0$ durch $k = (q-1)g+g$, ergibt sich ein *modifizierter Divisionssatz mit Rest*: für $k > g$ gibt es eindeutig bestimmte $q, r \in \mathbb{N}$ mit $k = gq + r$, wobei $1 \leq q < k$ und $1 \leq r \leq g$. Damit erhält man die Existenz und Eindeutigkeit der modifizierten $g$-adischen Darstellung von $k$ genau so wie im Beweis von Satz 11.7.

**8**. Der Beweis von $(k + m) + n = k + (m + n)$ erfolgt durch Induktion über $n$. Sicher ist $(k + m) + 0 = k + m = k + (m + 0)$. Sei $(k + m) + n = k + (m + n)$. Dann ist

$$\begin{aligned}(k+m)+n^+ &= ((k+m)+n)^+ \quad \text{(Definition)}\\ &= (k+(m+n))^+ \quad \text{(Induktionsannahme)}\\ &= k+(m+n)^+ = k+m+n^+ \quad \text{(Definition).}\end{aligned}$$

Wir beweisen zuvor induktiv über $m$ die Gleichung (a) $\quad n^+ + m = n + m^+$. Weil $n^+ + 0 = n^+ = (n + 0)^+ = n + 0^+$, ist (a) für $m = 0$ richtig. Induktionsschritt:

$$\begin{aligned}n^+ + m^+ &= (n^+ + m)^+ \quad \text{(Definition)}\\ &= (n + m^+)^+ \quad \text{(Induktionsannahme)}\\ &= n + m^{++} \quad \text{(Definition).}\end{aligned}$$

Damit beweisen wir nun $n + m = m + n$ durch Induktion über $n$, indem wir erst den Anfangsschritt $(n = 0)$ durch Induktion über $m$ verifizieren. $0 + m = m + 0$ ist klar für $m = 0$. Induktionsschritt:

$$\begin{aligned}0 + m^+ &= (0 + m)^+ \quad \text{(Definition)}\\ &= (m + 0)^+ \quad \text{(Induktionsannahme)}\\ &= m^+ = m^+ + 0 \quad \text{(Definition).}\end{aligned}$$

Der Induktionsschritt zum Nachweis von $\mathsf{K}^+$ sieht nunmehr wie folgt aus:

$$\begin{aligned}m + n^+ &= (m + n)^+ \quad \text{(Definition)}\\ &= (n + m)^+ \quad \text{(Induktionsannahme)}\\ &= n + m^+ \quad \text{(Definition)}\\ &= n^+ + m \quad \text{(gemäß (a)).}\end{aligned}$$

Auf ähnliche Weise, und wie man sieht nicht ganz mühelos, beweist man auch $\mathsf{A}^\times$, $\mathsf{K}^\times$ und weitere Rechengesetze.