Einführung in die Fachdidaktik Mathematik

Universität Tübingen
Sommersemester 2013

Frank Loose

23. Mai 2013

Vorwort

Weil ich in der Vorlesung „Einführung in die Fachdidaktik“ doch nur relativ wenig an die Tafel schreibe, schreibe ich auf diesen Seiten in etwa das auf, was ich den Studierenden hauptsächlich mündlich in der Vorlesung erzählt habe. Ich hoffe damit, den Studierenden eine Hilfe zu geben in dem Sinne, dass sie hier noch einmal nachlesen können, was denn der Gegenstand der letzten Vorlesung war. Andererseits möge dieses Skript aber nicht dazu führen, dass man sich den Besuch der Vorlesung ersparen kann. Wieviel einfacher und angenehmer ist es doch, dem Dozenten direkt folgen zu dürfen als nur sein Skript zu lesen, dass doch wahrscheinlich vielfach viel langweiliger daher kommt als die eigentliche lebendige Vorstellung. In diesem Sinne bitte ich das Skript zu benutzen und bin dankbar für alle Hinweise auf Fehler und mögliche sinnvolle Ergänzungen.

Tübingen, im Frühjahr 2013

Inhaltsverzeichnis

1 Die natürlichen Zahlen

(1.1.) Nach einem Ausspruch von L. KRONECKER sind die natürlichen Zahlen von Gott, der Rest sei Menschenwerk. (Eigentlich heißt es dort: die ganzen Zahlen.)

Dennoch möchten wir die natürlichen Zahlen in diesem ersten Paragraphen vom Standpunkt der Mengenlehre aus beleuchten und ihre grundlegenden Eigenschaften aus den Prinzipien der Mengenlehre herleiten und beweisen.

„Was ist das Wesentliche an den natürlichen Zahlen?“, möchte man als erstes fragen. Fast alle, die man das fragen wird, sind geneigt zu sagen, dass natürliche Zahlen viel mit „zählen“ zu tun haben, etwa in folgendem Sinn: Die natürlichen Zahlen sind eine *Menge*, die wir mit $\mathbb{N}$ bezeichnen. Auf dieser Menge gibt es eine *Abbildung* $S\colon \mathbb{N} \to \mathbb{N}$, die das „Weiterzählen“ modelliert und von der wir gewisse Eigenschaften fordern werden. Schließlich gibt es ein besonderes *Element* in $\mathbb{N}$, welches wir mit 0 (*Null*) bezeichnen, aus dem sich durch wiederholtes Anwenden der Abbildung S alle anderen Elemente von $\mathbb{N}$ ergeben. Wir bezeichnen dann etwa

$$1 := S(0),\ 2 := S(1),\ 3 := S(2) \text{ usw.}$$

Um das „usw.“ zu präzisieren, kann man für das Tripel $(\mathbb{N}, 0, S)$ nach PÉANO folgenden axiomatischen Versuch machen, die natürlichen Zahlen auf der Basis einer *naiven Mengenlehre* zu definieren.

Definition 1.1 (Péano). Wir nennen ein Tripel $(\mathbb{N}, 0, S)$, bestehend aus einer Menge $\mathbb{N}$, einem Element $0 \in \mathbb{N}$ und einer Selbstabbildung $S\colon \mathbb{N} \to \mathbb{N}$ *natürliche Zahlen*, wenn folgendes gilt:

(a) S ist injektiv (also: aus $n \neq m$ folgt $S(n) \neq S(m)$, $\forall n, m \in \mathbb{N}$);

(b) $0 \notin \operatorname{im}(S)$ (also: es gibt kein $n \in \mathbb{N}$ mit $S(n) = 0$);

(c) ist $A \subseteq \mathbb{N}$ eine Teilmenge mit folgender Eigenschaft

- $0 \in A$;
- gilt für $n \in \mathbb{N}$, dass $n \in A$ ist, so ist auch $S(n) \in A$;

so muss A schon gleich $\mathbb{N}$ sein, $A = \mathbb{N}$.

Bedingung (c) ist sicherlich am schwierigsten zu akzeptieren. Sie wird eine gewisse Minimalität von $\mathbb{N}$ sicherstellen, so dass in $\mathbb{N}$ wegen (c) nicht zu

viele Elemente liegen werden, eben nur die, die darin liegen müssen (anschaulich nur jene, die sich aus $0 \in \mathbb{N}$ und „rekursives Anwenden“ von S darauf ergeben). Wir nennen (c) auch die mengentheoretische Formulierung des *Beweisprinzips der vollständigen Induktion*:

Ist $\mathcal{E}$ eine Eigenschaft der natürlichen Zahlen, die für $0 \in \mathbb{N}$ gilt - wir schreiben dafür: $\mathcal{E}(0)$ - und die für $S(n) \in \mathbb{N}$ gilt, falls sie für $n \in \mathbb{N}$ gilt - also $\mathcal{E}(n) \Rightarrow \mathcal{E}(S(n))$, $\forall n \in \mathbb{N}$ - dann gilt $\mathcal{E}$ für alle natürliche Zahlen.

Man setze nämlich einfach $A = \{n \in \mathbb{N} : \mathcal{E}(n)\}$ und wende dann (1.1.b) an.

(1.2) Wir wollen nun die Frage betrachten, ob es überhaupt natürliche Zahlen im Sinne von Definition (1.1) gibt, wie man also ihre Existenz sicherstellen kann, und in wie weit sie eindeutig bestimmt sind. Von technisch großer Wichtigkeit ist dabei folgender Satz, mit dessen Hilfe häufig neue Begriffe über natürliche Zahlen, z.B. die arithmetischen Operationen $+$ und $\cdot$, eingeführt werden.

Satz 1.2 (Rekursionssatz von R. Dedekind). *Gegeben seien einerseits natürliche Zahlen $(\mathbb{N}, 0, S)$ und andererseits ein weiteres Tripel (M, a, F), bestehend wieder aus einer Menge M, einem Element $a \in M$ und einer Selbstabbildung $F\colon M \to M$. Dann gibt es genau eine Abbildung $\varphi\colon \mathbb{N} \to M$ mit den Eigenschaften*

(a) $\varphi(0) = a$,

(b) $\varphi \circ S = F \circ \varphi$.

Wir werden gleich sehen, dass als eine Anwendung dieses Satzes die Eindeutigkeit der natürlichen Zahlen (bis auf einen adequaten Isomorphiebegriff) folgen wird. In diesem Sinne trägt dann die Eigenschaft der natürlichen Zahlen, die in diesem Satz beschrieben ist, die volle Information der natürlichen Zahlen, d.h., wir charakterisieren das Tripel $(\mathbb{N}, 0, S)$ mit dieser Eigenschaft und könnten daher mit dieser Eigenschaft auch die natürlichen Zahlen definieren. Tatsächlich werden wir das später etwa bei den Zahlbereichserweiterungen oder auch bei der Definition von Polynomen (vgl. § 2 bzw. § 4) so machen. Wir nennen sie die *universelle Eigenschaft der natürlichen Zahlen* und deuten sie mit der Kommutativität des folgenden Diagramms an, was gerade die Eigenschaft (b) im Rekursionssatz bedeutet.

Anschaulich stellt dieser Satz sicher, dass man eine Abbildung $\varphi\colon \mathbb{N} \to M$ dadurch angeben kann, was sie auf 0 macht, $\varphi(0) = a$, und was sie auf dem Nachfolger $S(n)$ macht, wenn man weiß, was sie auf n macht und dadurch, „rekursiv“ vermöge F, das Ergebnis auf $S(n)$ bekommt,

$$\varphi(S(n)) = F(\varphi(n)),\ \forall n \in \mathbb{N}.$$

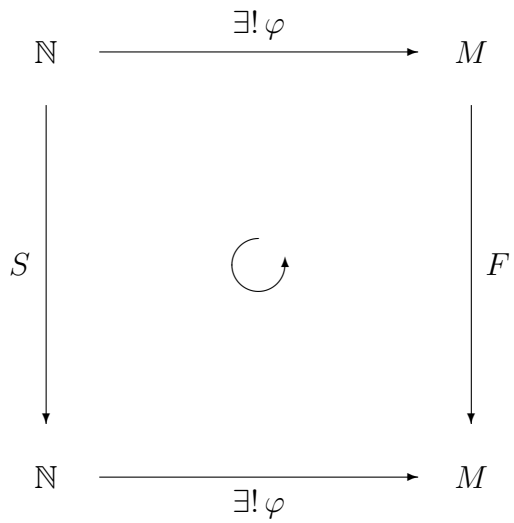

Abbildung 1: universelle Eigenschaft der natürlichen Zahlen

Wir wollen hier zumindest andeuten, wie sich die Existenz und Eindeutigkeit der Abbildung $\varphi\colon \mathbb{N} \to M$ nur mit Hilfe der Mengenlehre beweisen lässt:

Beweis von (1.2). (i) Eindeutigkeit: Nehmen wir also an, wir hätten zwei solche Abbildungen $\varphi_1, \varphi_2\colon \mathbb{N} \to M$ mit den Eigenschaften (a) und (b) aus (1.2). Wir betrachten dann die Teilmenge

$$A = \{n \in \mathbb{N} :\ \varphi_1(n) = \varphi_2(n)\}$$

der natürlichen Zahlen und stellen wegen (a) zunächst fest, dass $0 \in A$ ist, weil ja $\varphi_1(0) = a = \varphi_2(0)$ ist. Aber $A \subseteq \mathbb{N}$ ist wegen (b) auch *induktiv*, wie man sagt, d.h. mit $n \in A$ ist auch $S(n) \in A$. Ist nämlich $\varphi_1(n) = \varphi_2(n)$, also $n \in A$, so ist

$$\varphi_1(S(n)) = F(\varphi_1(n)) = F(\varphi_2(n)) = \varphi_2(S(n)),$$

und damit auch $S(n) \in A$. Nach Teil (c) von Definition (1.1) muss deshalb $A = \mathbb{N}$ sein und damit $\varphi_1 = \varphi_2$.

(ii) Für die Existenz von φ braucht man einige Axiome der unterliegenden Mengenlehre und ihre Folgerungen, die ich hier nicht vertiefen will, z.B. die Existenz des *cartesischen Produktes* $M \times N$ zweier Mengen M und N. Man betrachtet dann alle Teilmengen H des cartesischen Produktes $\mathbb{N} \times M$, die folgende Eigenschaft haben:

(α) $(0, a) \in H$;

(β) $(n, b) \in H \Rightarrow (S(n), F(b)) \in H$.

Z.B. hat $H = \mathbb{N} \times M$ sicher diese Eigenschaften. Dann definiert man $D \subseteq \mathbb{N} \times M$ als den Durchschnitt aller solchen Teilmengen (wobei die unterliegende Mengenlehre diese Durchschnittsbildung erlauben muss) und von diesem D beweist man dann (mit Hilfe vollständiger Induktion), dass es Graph von genau einer Abbildung $\varphi\colon \mathbb{N} \to M$ ist, d.h.: zu jedem $n \in \mathbb{N}$ gibt es genau ein $b \in M$, so dass $(n, b) \in D$ ist. Die Eigenschaften (α) und (β) von H, und damit von D, zeigen dann, dass φ die Eigenschaften (a) und (b) in (1.2) hat. (Für die Einzelheiten konsultiere man z.B. [3].) □

(1.3) Wir wollen nun als erste Anwendung des Rekursionssatzes illustrieren, wie man, bis auf geeignete Isomorphie, die Eindeutigkeit der natürlichen Zahlen bekommt. Genauer gilt folgendes.

Satz 1.3 *Seien* $(\mathbb{N}, 0, S)$ *und* $(\mathbb{N}', 0', S')$ *natürliche Zahlen. Dann gibt es eine eindeutig bestimmte Abbildung* $\varphi\colon \mathbb{N} \to \mathbb{N}'$*, so dass gilt:*

(i) $\varphi(0) = 0'$,

(ii) $\varphi \circ S = S' \circ \varphi$,

(iii) φ *ist bijektiv.*

(Wir sagen deshalb, dass $(\mathbb{N}, 0, S)$ *und* $(\mathbb{N}', 0', S')$ *kanonisch isomorph sind.)*

Beweis. Man wende zunächst den Rekursionssatz für $(\mathbb{N}, 0, S)$ und $(M, a, F) = (\mathbb{N}', 0', S')$ an, um $\varphi\colon \mathbb{N} \to \mathbb{N}'$ mit (i) und (ii) zu erhalten, wobei φ schon dadurch eindeutig bestimmt ist.

Es bleibt zu zeigen, dass φ bijektiv ist, also ein $\psi\colon \mathbb{N}' \to \mathbb{N}$ existiert, so dass

$$\psi \circ \varphi = \mathrm{id}_{\mathbb{N}} \text{ und } \varphi \circ \psi = \mathrm{id}_{\mathbb{N}'}$$

ist (und $\mathrm{id}_M\colon M \to M$ für eine Menge M die *Identität auf* M bezeichnet, $\mathrm{id}_M(x) = x$, $\forall x \in M$).

Dazu wenden wir den Rekursionssatz ein zweites Mal an, in dem wir nun die Rollen von $(\mathbb{N}, 0, S)$ und $(\mathbb{N}', 0', S')$ vertauschen, um ein $\psi\colon \mathbb{N}' \to \mathbb{N}$ zu erhalten mit $\psi(0') = 0$ und $\psi \circ S' = S \circ \psi$. Die Abbildungen $\psi \circ \varphi\colon \mathbb{N} \to \mathbb{N}$ bzw. $\varphi \circ \psi\colon \mathbb{N}' \to \mathbb{N}'$ erfüllen dann wieder den Rekursionssatz, dieses Mal für $(\mathbb{N}, 0, S)$ und $(\mathbb{N}, 0, S)$ bzw. für $(\mathbb{N}', 0', S')$ und $(\mathbb{N}', 0', S')$. Wegen der Eindeutigkeit im Rekursionssatz muss dann $\psi \circ \varphi = \mathrm{id}$ und $\varphi \circ \psi = \mathrm{id}$ sein, weil auch id den Rekursionssatz für diese Wahlen erfüllt. Damit ist φ auch bijektiv. □

(1.4) Wie kann man nun die Existenz eines solchen Tripels mit (1.1) sichern? Wiederum braucht man dazu eine gewisse Reichhaltigkeit des Axiomensystems unserer Mengenlehre (sagen wir das nach ZERMELO und FRAENKEL (ZF)), genauer das so genannte *Paarmengenaxiom* und das *Vereinigungsaxiom*, um aus zwei Mengen M und N die Menge $M \cup N$ bilden zu können, deren Elemente genau aus den Elementen von M und den Elementen von N besteht. Deshalb kann man dann für jede Menge M ihre so genannte *Nachfolgermenge*

$$M' := M \cup \{M\}$$

definieren.

Definition 1.4 Wir nennen eine Menge M *induktiv*, wenn mit jedem $x \in M$ auch $x' \in M$ ist.

(Elemente von Mengen sind wieder Mengen. In gewisser Weise ist „alles in der Mathematik eine Menge“.) Nun verlangen wir von unserer Mengenlehre, dass sie zunächst einmal das so genannte *Nullmengenaxiom* erfüllt, d.h., dass die *leere Menge* $\emptyset$ (also jene Menge, die überhaupt keine Elemente enthält) tatsächlich eine Menge ist. Das Axiomensystem einer Mengenlehre legt eben gerade fest, was Mengen sind, in dem es Operationen aufzeigt, wie man aus gegebenen Mengen neue bekommt. Damit es überhaupt Mengen in so einer Mengenlehre gibt, muss man irgendetwas irgendwo als Menge festlegen und das macht das Nullmengenaxiom mit der leeren Menge. Das folgende Axiom, welches wir nun auch als akzeptiert ansehen wollen, garantiert dann erst, dass es überhaupt Mengen gibt, die unendlich viele Elemente enthalten:

Induktionsaxiom 1.5 *Es gibt eine Menge M, die $\emptyset$ enthält und induktiv ist.*

Das sichert nun zunächst anschaulich die Existenz der folgenden Menge, die wir später als Definition der natürlichen Zahlen nach J. VON NEUMANN betrachten können, wobei die Pünktchen „...“ eben noch präzisiert werden müssen:

$$0 := \emptyset,\ 1 = 0' = \{\emptyset\},\ 2 := 1' = \{\emptyset, \{\emptyset\}\},\ 3 := 2' = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\},\ \ldots$$

Beachte, dass die Menge n dann genau aus n Elementen bestehen wird,

$$n = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \ldots\}.$$

Präzise erhält man nun aus dem *Aussonderungsaxiom* die Menge

$$\mathbb{N} := \bigcap \{x :\ x \text{ ist induktive Menge}\},$$

das Element $0 := \emptyset$ (da $\emptyset \in x$, für alle induktiven x) und die Abbildung $S\colon \mathbb{N} \to \mathbb{N}$, gegeben durch $S(x) = x'$. Das Tripel $(\mathbb{N}, 0, S)$ erfüllt dann die Eigenschaften (a)-(c) aus (1.1), wobei (b) unmittelbar klar ist, weil die Nachfolgermenge einer Menge stets Elemente enthält und auch (c) trivialerweise gilt, weil die dortige Teilmenge $A \subseteq \mathbb{N}$ selbst induktiv ist und daher $A \supseteq \mathbb{N}$ sein muss. Damit ist dann die Existenz der natürlichen Zahlen (und auch ihre Eindeutigkeit, wie wir gesehen haben) gesichert. (Für die Injektivität von $S = {}'$ siehe man z.B. [7] oder [8].)

(1.5) Nun kann man in den natürlichen Zahlen nicht nur „zählen“, sondern auch „rechnen“. Wir wollen deshalb nun hier auch die arithmetischen Operationen „$+$“ und „$\cdot$“ mit Hilfe unserer mengentheoretischen Hilfsmittel definieren und ihre elementaren Eigenschaften beweisen.

Bei der Addition „$m + n$“ lassen wir uns davon leiten, dass man „$m + n$“ erhält, in dem man bei m startet und n-mal weiterzählt (was etwas ganz anderes ist als „$n + m$“). Das „n-mal-Weiterzählen“ wird dabei durch eine Abbildung $\varphi_m\colon \mathbb{N} \to \mathbb{N}$ im Rekursionssatz (1.2) gesichert, in dem wir φ_m dadurch definieren, dass wir verlangen:

(i) $\varphi_m(0) = m$,

(ii) $\varphi_m(S(n)) = S(\varphi_m(n))$.

(Wir wenden also den Rekursionssatz auf das Tripel $(\mathbb{N}, m, S)$ an.) Dann setzen wir:

$$m + n := \varphi_m(n),$$

so dass z.B. gilt:

$$m + 1 = \varphi_m(1) = \varphi_m(S(0)) \overset{\text{(ii)}}{=} S(\varphi_m(0)) \overset{\text{(i)}}{=} S(m).$$

Mit dieser Bezeichnung wird dann (ii) zu der rekursiven Definition

$$m + (n + 1) := (m + n) + 1.$$

Nun hätte man Einiges zu tun, um die üblichen Rechenregeln für das Tripel $(\mathbb{N}, 0, +)$ zu beweisen (vgl. Übungen):

(i) $(\mathbb{N}, +)$ ist *assoziativ*, d.h.: $\forall k, m, n \in \mathbb{N}$ gilt:

$$(k + m) + n = k + (m + n),$$

(ii) 0 ist *neutral*, d.h.: $\forall n \in \mathbb{N}$ gilt:

$$n + 0 = 0 + n = n,$$

(iii) $+$ ist *kommutativ*, d.h.: $\forall m, n \in \mathbb{N}$ gilt:

$$m + n = n + m,$$

(iv) $(\mathbb{N}, +)$ erfüllt die *Kürzungsregel*: $\forall k, m, n \in \mathbb{N}$ gilt:

$$m + k = n + k \;\Rightarrow\; m = n.$$

(1.6) Ähnlich geht man bei der Multiplikation vor. Man führt sie auf die Addition und den Rekursionssatz zurück und lässt sich leiten von

$$(m + 1) \cdot n = m \cdot n + n.$$

(Es ist also $m \cdot n$ das Zusammenfassen von m Päckchen, wo jedes Päckchen n Elemente enthält. Das ist etwas ganz anderes als $m \cdot n$.) Betrachte also die Abbildung $F_n \colon \mathbb{N} \to \mathbb{N}$, die auf $k \in \mathbb{N}$ gerade (das feste) n aufaddiert, $F_n(k) = k + n$. Weiter sei nun $\varphi_n \colon \mathbb{N} \to \mathbb{N}$ nach dem Rekursionssatz gegeben durch

$$\varphi_n(0) = 0 \quad \text{und} \quad \varphi_n \circ S = F_n \circ \varphi_n.$$

In der Bezeichnung $S(m) = m + 1$ wird letzteres zu

$$\varphi_n(m + 1) = F_n(\varphi_n(m)) = \varphi_n(m) + n. \tag{1}$$

Nun setze

$$m \cdot n := \varphi_n(m),$$

so dass dann (1) wirklich zu

$$(m + 1) \cdot n = m \cdot n + n$$

wird. Nun kann man – und sollte man vielleicht zumindest stichprobenweise – beweisen:

(i) *Assoziativität*:

$$k \cdot (m \cdot n) = (k \cdot m) \cdot n, \;\forall k, m, n \in \mathbb{N},$$

(ii) *Neutralität von* 1:

$$n \cdot 1 = 1 \cdot n = n, \;\forall n \in \mathbb{N},$$

(iii) *Kommutativität*:

$$m \cdot n = n \cdot m, \;\forall m, n \in \mathbb{N},$$

(iv) *Kürzungsregel*:

$$m \cdot k = n \cdot k \;\Rightarrow\; m = n, \quad \forall k, m, n \in \mathbb{N} \text{ mit } k \neq 0.$$

Und darüber hinaus das verbindende Gesetz:

(D) *Distributivität*:

$$k \cdot (m + n) = k \cdot m + k \cdot n, \; \forall k, m, n \in \mathbb{N}.$$

In dieser Zeile vereinbaren wir dann auch, dass „Punktrechnung vor Strichrechnung“ gehen möge, wobei wir den Punkt für die Multiplikation ab jetzt auch oft gänzlich weglassen.

2 Die ganzen Zahlen und die rationalen Zahlen

(2.1) Die natürlichen Zahlen $\mathbb{N}$ bilden zusammen mit ihrer Addition $+$ und dem Nullelement 0 eine *kommutative Halbgruppe* mit neutralem Element und Kürzungsregel (siehe §1). Allerdings kann man in ihnen nicht beliebig *subtrahieren*, z.B. hat die Gleichung

$$x + 2 = 1$$

in $\mathbb{N}$ keine Lösung. Äquivalent könnte man sagen, dass nicht jedes Element $a \in \mathbb{N}$ ein *Negatives* hat, d.i. ein Element $b \in \mathbb{N}$ mit $a+b = 0$. Dieses Element (welches in der Tat auch eindeutig wäre) würde man dann mit $-a$ bezeichnen und damit die Subtraktion in einer Gruppe durch

$$x - y := x + (-y) \tag{2}$$

ermöglichen. $(\mathbb{N}, 0, +)$ ist also keine (kommutative) *Gruppe*!

(2.2) Deshalb möchte man nun $(\mathbb{N}, 0, +)$ zu einer Gruppe $(\mathbb{Z}, 0, +)$ erweitern, d.h.: Man sucht eine Gruppe $(\mathbb{Z}, 0, +)$ zusammen mit einem Homomorphismus $\iota\colon \mathbb{N} \to \mathbb{Z}$ (also $\iota(0) = 0$ und $\iota(x) + \iota(y) = \iota(x) + \iota(y)$, für alle $x, y \in \mathbb{N}$), der sich als injektiv herausstellen wird. Man kann dann $(\mathbb{N}, 0, +)$ vermöge ι als eine *Unterhalbgruppe* von $(\mathbb{Z}, 0, +)$ auffassen. Schließlich ist die Forderung an $(\mathbb{Z}, \iota)$, dass dieses Paar in gewisser Weise wieder minimal ist, ähnlich wie wir es bei den natürlichen Zahlen gesehen haben (vgl. (1.2).

Definition 2.1 Ein Paar $(\mathbb{Z}, \iota)$, bestehend aus einer abelschen Gruppe $\mathbb{Z}$ und einem Homomorphismus $\iota\colon \mathbb{N} \to \mathbb{Z}$ heißt *ganze Zahlen*, wenn folgendes gilt: Ist (G, j) ein weiteres solches Paar, so gibt es genau einen Gruppenhomomorphismus $f\colon \mathbb{Z} \to G$ mit $f \circ \iota = j$.

Diese *universelle Eigenschaft* sichert, dass, wenn die ganzen Zahlen existieren, sie – bis auf kanonische Isomorphie – wieder eindeutig bestimmt sind (mit im Grunde dem gleichen Beweis wie in (1.3)).

(2.3) Für die Existenz geben wir wieder eine mengentheoretische Konstruktion an, die eine *Äquivalenzrelation* $\sim$ auf einer Menge M benutzt. Man erinnere sich daran, dass eine solche Relation die Menge M disjunkt in seine *Äquivalenzklassen*

$$[x] = \{y \in M : \ y \sim x\}$$

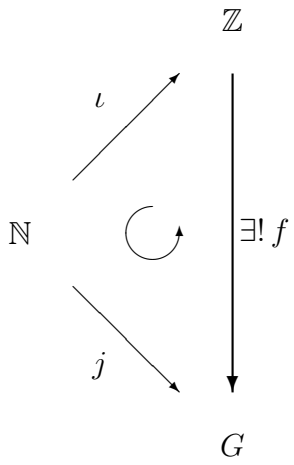

Abbildung 2: universelle Eigenschaft der ganzen Zahlen

zerlegt. Die Menge der Äquivalenzklassen

$$M/\sim = \{[x] \in \mathrm{Pot}(M) :\ x \in M\} \subseteq \mathrm{Pot}(M)$$

wird als *Quotient von M nach $\sim$* bezeichnet. ($\mathrm{Pot}(M)$ bezeichnet hier die *Potenzmenge einer Menge M*, d.i. die Menge aller Teilmengen von M.)

Die folgende Konstruktion, die brigens für jede (abelsche) Halbgruppe mit Kürzungsregel funktioniert, lässt sich davon leiten, dass jedes Element $x \in \mathbb{Z}$ (wenn man $\mathbb{Z}$ dann konstruiert hat) Differenz von zwei natürlichen Zahlen $a, b \in \mathbb{N}$ ist (wenn man $a \in \mathbb{N}$ mit $\iota(a) \in \mathbb{Z}$ identifiziert), $x = a - b$. Natürlich kann das auf mehrere Arten geschehen, $x = a - b = c - d$, aber das stimmt (nach den Rechenregeln in einer Gruppe) genau dann, wenn $a + d = b + c$ ist. (Beachte, dass man ja in $\mathbb{N}$ addieren kann, nicht jedoch subtrahieren, jedenfalls bis zu diesem Punkt.) Man setzt deshalb $M := \mathbb{N} \times \mathbb{N}$ und definiert darauf die Relation

$$(a, b) \sim (c, d) :\Leftrightarrow\ a + d = b + c.$$

Davon prüft man nun nach, dass $\sim$ tatsächlich eine Äquivalenzrelation ist, d.h., sie ist

- *reflexiv*, also $(a, b) \sim (a, b)$, $\forall (a, b) \in M$;
- *symmetrisch*, also $(a, b) \sim (c, d) \ \Rightarrow\ (c, d) \sim (a, b)$, $\forall (a, b), (c, d) \in M$;
- *transitiv*, also $(a, b) \sim (c, d)$, $(c, d) \sim (e, f) \Rightarrow\ (a, b) \sim (e, f)$, $\forall (a, b)$, $(c, d), (e, f) \in M$.

Man beachte, dass man, neben der Kommutativität und Assoziativität, für die Transitivität die Kürzungsregel braucht:

$$\begin{aligned}(a,b) \sim (c,d) &\Rightarrow a+d = b+c,\\ (c,d) \sim (e,f) &\Rightarrow c+f = d+e\end{aligned}$$

$$\begin{aligned}&\Rightarrow a+d+c+f = b+c+d+e\\ &\Rightarrow (a+f)+(c+d) = (b+e)+(c+d)\\ &\overset{KR}{\Rightarrow} a+f = b+e \Rightarrow (a,b) \sim (e,f).\end{aligned}$$

Wir setzen nun zunächst $\mathbb{Z} := M/\!\sim$ und $\iota\colon \mathbb{N} \to \mathbb{Z}$,

$$\iota(a) = [a,0].$$

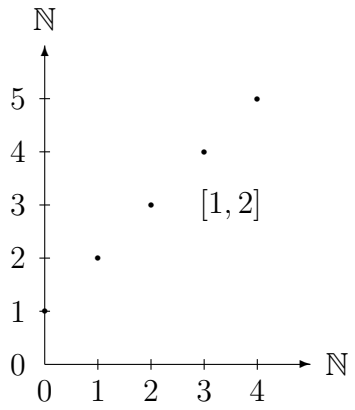

Abbildung 3: eine ganze Zahl als Äquivalenzklasse in $\mathbb{N} \times \mathbb{N}$

Dann betrachten wir auf $\mathbb{N}\times\mathbb{N}$ die natürliche Produkt-Halbgruppenstruktur, die gegeben ist durch die komponentenweise Addition

$$(a,b) + (c,d) := (a+b, c+d),$$

und diese kann man nun auf den Quotienten $M/\!\sim$ „herabdrücken“, weil die folgende Definition nicht von der Auswahl der Repräsentanten abhängt, d.h. die Setzung „*wohldefiniert*“ ist:

$$[a,b] + [c+d] := [a+c, b+d].$$

(Prüfe also dazu: $(a,b) \sim (a',b')$, $(c,d) \sim (c',d') \Rightarrow (a+c, b+d) \sim (a'+c', b'+d')$.) Es ist dann $\iota\colon \mathbb{N} \to \mathbb{Z}$ auch ein Homomorphismus, denn $\iota(0) = [0,0]$ (und letzteres ist das Nullelement in $\mathbb{Z}$) sowie

$$\iota(a+b) = [a+b,0] = [a,0] + [b,0] = \iota(a) + \iota(b),$$

für alle $a, b \in \mathbb{N}$.

Das Entscheidende ist aber nun, dass es in $\mathbb{Z}$ zu jedem Element $[a, b]$ tatsächlich ein Negatives gibt und $\mathbb{Z}$ dann mit dem so definierten Nullement und der so definierten Addition eine abelsche Gruppe ist. (Das Prüfen von Assoziativität, neutralem Element und Kommutativität wird als Übung überlassen.) Es ist nämlich $[b, a]$ Negatives zu $[a, b]$, weil

$$[a, b] + [b, a] = [a + b, a + b] = [0, 0]$$

ist. Insbesondere ist $[0, b] = -[b, 0] = -\iota(b)$ und man kann jedes Element $x = [a, b] \in \mathbb{Z}$ wie folgt schreiben, wenn wir $a \in \mathbb{N}$ mit $\iota(a)$ identifizieren:

$$[a, b] = [a, 0] + [0, b] = \iota(a) + (-\iota(b)) = \iota(a) - \iota(b) = a - b.$$

So hatten wir es vor.

Es bleibt damit nur noch die universelle Eigenschaft in Definition (2.1) zu prüfen. Ist also (G, j) ein weiteres solches Paar, so gibt es für einen Homomorphismus $f\colon \mathbb{Z} \to G$ mit $f \circ \iota = j$ nur einen Kandidaten. Wegen

$$f([a, b]) = f(\iota(a) - \iota(b)) = f(\iota(a)) - f(\iota(b)) = j(a) - j(b)$$

muss man es mit

$$f([a, b]) := j(a) - j(b)$$

versuchen. Es wird wieder dem Leser überlassen zu prüfen, dass dies wohldefiniert und tatsächlich ein Homomorphismus ist mit $f \circ \iota = j$. Die Existenz der ganzen Zahlen ist damit gesichert und damit die Erweiterung von $\mathbb{N}$ auf $\mathbb{Z}$ abgeschlossen.

(2.4) Man kann nun auch die Multiplikation von $\mathbb{N}$ auf $\mathbb{Z}$ ausdehnen, ohne dabei die Halgruppeneigenschaften (von $\mathbb{N}_+ := \mathbb{N} \setminus \{0\}$) mit Kürzungsregel und ihr Distributivgesetz zu verlieren. Wegen

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

denn das Distributivgesetz soll ja auch auf dem vergrößerten Zahlbereich der ganzen Zahlen gelten, hat man nur den Kandidaten

$$[a, b] \cdot [c, d] := [ac + bd, ad + bc]$$

und muss wieder Wohldefiniertheit nachprüfen. Dann stellt man fest, dass $\mathbb{Z}$ zusammen mit $+$ und $\cdot$ ein *Integritätsring* wird, d.h. ein kommutativer Ring mit Einselement, der *nullteilerfrei* ist, also

(a) $(\mathbb{Z}, 0, +)$ ist abelsche Gruppe,

(b) (i) $(x \cdot y) \cdot z = x \cdot (y \cdot z), \quad \forall x, y, z \in \mathbb{Z}$,

(ii) $x \cdot y = y \cdot x, \quad \forall x, y \in \mathbb{Z}$,

(iii) $\exists 1 \in \mathbb{Z}$ mit $1 \neq 0$, so dass $1 \cdot x = x$ ist, $\forall x \in \mathbb{Z}$,

(c) $x \cdot (y+z) = x \cdot y + x \cdot z, \forall x, y, z \in \mathbb{Z}$ (wobei wir wieder Punktrechnung vor Strichrechnung vereinbaren und den Punkt nun auch oft weglassen),

(d) $xy = 0 \Rightarrow x = 0$ oder $y = 0$ (Nullteilerfreiheit).

Man kann aber in $(\mathbb{Z}, +, \cdot)$ noch nicht *dividieren*, z.B. hat die Gleichung

$$2x = 3$$

in $\mathbb{Z}$ keine Lösung. Deshalb sucht man nun wieder, ähnlich wie bei dem Übergang von $\mathbb{N}$ nach $\mathbb{Z}$, nach einer Erweiterung $\mathbb{Q}$, d.h. zusammen mit einer *Einbettung* (d.i. hier ein injektiver *Ringhomomorphismus*) $\iota \colon \mathbb{Z} \to \mathbb{Q}$, wobei nun $(\mathbb{Q}, +, \cdot)$ ein *Körper* sein soll und $(\mathbb{Q}, \iota)$ wieder eine universelle Eigenschaft haben soll.

Definition 2.2 Ein Paar $(\mathbb{Q}, \iota)$, bestehend aus einem Körper $\mathbb{Q}$ und einem injektiven Ringhomomorphismus $\iota \colon \mathbb{Z} \to \mathbb{Q}$, heißt *rationale Zahlen*, wenn es zu jedem weiteren Paar (K, j) genau einen *Körperhomomorphismus* $f \colon \mathbb{Q} \to K$ gibt mit $f \circ \iota = j$.

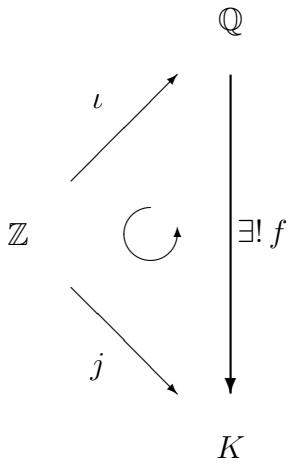

Abbildung 4: universelle Eigenschaft der rationalen Zahlen

Die universelle Eigenschaft stellt wiederum Eindeutigkeit von $(\mathbb{Q}, \iota)$ bis auf kanonische Isomorphie sicher (vgl. (1.3) und den Kommentar im Anschluss an Abbildung 2). Für die Existenz lässt man sich wieder von der Idee leiten, dass jedes Element $x \in \mathbb{Q}$ Quotient zweier Elemente $a, b \in \mathbb{Z}$ sein soll, wenn man denn in $\mathbb{Q}$ dividieren kann und $\mathbb{Z}$ vermöge ι wieder als *Unterring* von $\mathbb{Q}$ betrachtet. Natürlich muss hierbei $b \neq 0$ sein. Für zwei solche *Brüche*

$$\frac{a}{b} := ab^{-1} \quad \text{und} \quad \frac{c}{d},$$

die die gleiche rationale Zahl x repräsentieren, muss dann $ad = bc$ sein, so dass man auf $M := \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ nun folgende Äquivalenzrelation einführt:

$$(a, b) \sim (c, d) :\Leftrightarrow \ ad = bc.$$

Dann setzt man $\mathbb{Q} := M/\!\sim$ und notiert die Äquivalenzklasse nun direkt mit

$$\frac{a}{b} := [a, b],$$

weil sich später herausstellen wird, dass

$$\frac{a}{b} = \frac{\iota(a)}{\iota(b)} \tag{3}$$

sein wird, wobei der Bruchstrich rechts die Division im Körper $\mathbb{Q}$ meint.

Weiter setzt man natürlich $\iota\colon \mathbb{Z} \to \mathbb{Q}$,

$$\iota(a) = \frac{a}{1},$$

und weiter die Addition und Multiplikation durch

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

fest. Hier geht die Nullteilerfreiheit von $\mathbb{Z}$ ein, denn im Nenner muss $bd \neq 0$ sein, wenn $b \neq 0$ und $c \neq 0$ ist. Die Konstruktion funktioniert übrigens bei jedem Integritätsring, z.B. auch im Polynomring $K[X]$ für einen Körper K, von dem später noch die Rede sein wird (vgl. §4). Die Definitionen sind zwangsläufig, will man die üblichen Rechengesetze (das sind die Körperaxiome) in $\mathbb{Q}$ haben. Z.B. ist

$$\frac{a}{b} + \frac{c}{d} = \frac{ad}{bd} + \frac{bc}{bd} = \frac{1}{bd}(ad + bc) = \frac{ad + bc}{bd}.$$

Nun prüft man relativ leicht nach, dass $(\mathbb{Q}, +, \cdot)$ ein Körper ist, wobei Assoziativ-, Kommutativ- und Distributivgesetz letztlich von der Gültigkeit dieser

Gesetze in $\mathbb{Z}$ (und diese wiederum wegen ihrer Gültigkeit in $\mathbb{N}$) herrühren. Jedes Element $x \neq 0$ in $\mathbb{Q}$ hat nun auch tatsächlich ein *Inverses*, welches eindeutig bestimmt ist und wir mit x^{-1} bezeichnen. Die Division in $\mathbb{Q}$ wird dann vermöge

$$\frac{x}{y} := xy^{-1}$$

ermöglicht (vgl. die Konstruktion bei den ganzen Zahlen (2)). Ist nämlich $x = \frac{a}{b} \neq 0$, so muss (neben b) auch $a \neq 0$ sein, sonst wäre $x = 0$ (denn aus $0 \cdot 1 = 0 = 0 \cdot b$ in $\mathbb{Z}$ würde $\frac{0}{b} = \frac{0}{1}$ folgen und letzteres ist die Null in $\mathbb{Q}$). Aber dann ist $y = \frac{b}{a}$ tatsächlich invers zu x, denn

$$x \cdot y = \frac{a}{b} \cdot \frac{b}{a} = \frac{ab}{ab} = \frac{1}{1} = 1.$$

Beachte, dass hieraus auch klar wird, dass jedes Element $x = \frac{a}{b} \in \mathbb{Q}$ nun wie gewünscht auch Quotient von zwei ganzen Zahlen wird, denn

$$x = \frac{a}{b} = \frac{a}{1} \cdot \frac{1}{b} = \iota(a)\iota(b)^{-1} = \frac{\iota(a)}{\iota(b)}$$

(vgl. (3)), womit die zunächst doppelte Bedeutung des Bruchstrichs nun gerechtfertigt ist.

Man prüft nun wieder leicht nach, dass $\iota\colon \mathbb{Z} \to \mathbb{Q}$ ein injektiver Ringhomomorphismus ist und die universelle Eigenschaft aus (2.2) erfüllt, denn das einzige $f\colon \mathbb{Q} \to K$, welches in Frage kommt, ist durch

$$f(\frac{a}{b}) = j(a)j(b)^{-1}$$

gegeben und das tut es auch (Übung; beachte aber vielleicht, dass $j(b) \neq 0$ sein muss für $b \neq 0$, also j injektiv, was wir aber verlangt haben).

Die Erweiterung $\mathbb{N} \rightsquigarrow \mathbb{Z} \rightsquigarrow \mathbb{Q}$ ist nun von einem algebraischen Standpunkt aus befriedigend abgeschlossen. In $\mathbb{Q}$ sind alle Grundrechenarten möglich und $\mathbb{Q}$ ist im Sinne von (2.1) und (2.2) der kleinste *Oberkörper* von $\mathbb{N}$. Jeder injektive Homomorphismus $j\colon \mathbb{N} \to K$ von $\mathbb{N}$ in einen Körper K induziert nämlich einen (eindeutig bestimmten) Körperhomomorphismus $f\colon \mathbb{Q} \to K$, der dann notwendig auch injektiv ist, d.h: Enthält ein Körper K die natürlichen Zahlen $\mathbb{N}$, so enthält er auch die rationalen Zahlen $\mathbb{Q}$.

3 Die reellen Zahlen

(3.1) Einen Aspekt der natürlichen Zahlen, den wir bisher noch überhaupt nicht betrachtet haben, ist ihre *Anordnung*. Wir definieren nämlich dort für $a, b \in \mathbb{N}$

$$a < b :\Leftrightarrow \exists k \in \mathbb{N}_+ := \mathbb{N} \setminus \{0\} : \ b = a + k,$$

oder, wenn wir das von Neumannsche Modell einer kleinsten induktiven Menge nehmen: $a < b :\Leftrightarrow a \in b$. Wir stellen dann fest, dass $<$ eine so genannte *lineare Ordnung* auf $\mathbb{N}$ ist, d.h.: es tritt für jedes Paar $(a, b) \in \mathbb{N} \times \mathbb{N}$ genau einer der folgenden Fälle auf:

$$a < b, \quad , a = b, \quad a > b \text{ (d.h.: } b < a).$$

Diese lineare Ordnung erfüllt zudem folgende Verträglichkeit mit den algebraischen Strukturen $+$ und $\cdot$ auf $\mathbb{N}$:

(i) Ist $a < b$ und c beliebig $\Rightarrow a + c < b + c$,

(ii) ist $a < b$ und $c > 0$ (also $c \neq 0$ hier) $\Rightarrow ac < bc$

(siehe z.B. [8] oder [7]). Diese Anordnung ermöglicht also für je zwei natürliche Zahlen einen Größenvergleich, so dass wir uns die Elemente von $\mathbb{N}$ auf einem „Zahlenstrahl“ festgemacht denken, auf dem größere Elemente weiter rechts liegen.

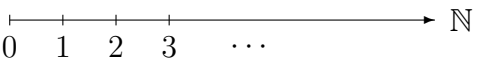

Abbildung 5: die natürlichen Zahlen auf dem Zahlenstrahl

Man beachte, dass man mit Hilfe dieser Anordnung das Beweisprinzip der vollständigen Induktion dadurch ausdrücken kann, dass jede nicht-leere Teilmenge $A \subseteq \mathbb{N}$ ein *Minimum* besitzt, d.h. ein Element $a \in A$, so dass $a \leq b$ (d.h. $a < b$ oder $a = b$) ist, für alle $b \in A$.

Satz 3.1 (vom kleinsten Element). *Jede nicht-leere Teilmenge $A \subseteq \mathbb{N}$ besitzt ein kleinstes Element. (Man sagt: $\mathbb{N}$ ist wohlgeordnet.)*

Für den Beweis davon siehe z.B. [8].

(3.2) Diese Anordnung kann man nun auf die Erweiterungen $\mathbb{Z}$ von $\mathbb{N}$ bzw. $\mathbb{Q}$ von $\mathbb{Z}$ in eindeutiger Weise so fortsetzen, dass man genau eine Ordnung auf $\mathbb{Q}$ erhält, die linear ist, die Verträglichkeitsbedingungen (i) und (ii) erfüllt

und die vorhandene Ordnung auf $\mathbb{N}$ fortsetzt. Dazu reicht es eine Menge $P \subseteq \mathbb{Q}$ auszuzeichnen, die man dann die Menge der *positiven Zahlen* nennt, die folgendes erfüllt:

$$\mathbb{Q} = (-P) \dot\cup \{0\} \dot\cup P, \tag{4}$$

d.h.: jede Zahl $x \in \mathbb{Q}$ ist entweder positiv, Null oder *negativ*, d.h.: $-x \in P$. Die Ordnung $<$ wird dann dadurch gegeben, dass man

$$x < y : \Leftrightarrow \; y - x \in P$$

setzt. Die obige Bedingung ist dann gleichbedeutend damit, dass $<$ eine lineare Ordnung auf $\mathbb{Q}$ ist. Die Verträglichkeitsbedingungen (i) und (ii) (für $\mathbb{Q}$) übersetzen sich dann für P in die Bedingungen

(i') $P + P \subseteq P$ (d.h.: $x, y \in P \;\Rightarrow\; x + y \in P$),

(ii') $P \cdot P \subseteq P$ (d.h.: $x, y \in P \;\Rightarrow\; xy \in P$).

Diese Bedingungen zwingen nun dazu, dass z.B. $1 \in P$ sein muss, denn da $1 \neq 0$ ist, muss also $1 \in P$ oder $-1 \in P$ sein. Aber

$$1 = 1 \cdot 1 = (-1) \cdot (-1) \in P.$$

Damit ist nach (i') dann auch $2 = 1 + 1 \in P$ und sukzesssive $n \in P$, für alle $n \in \mathbb{N}_+$. Das zeigt übrigens, dass die induzierte Ordnung auf $\mathbb{N}$ automatisch die vorher eingeführte ist, $P \cap \mathbb{N} = \mathbb{N}_+$, denn damit ist $m < n$, genau wenn $n - m \in P$ ist. Man bräuchte also nur (4) zusammen mit (i') und (ii') fordern.

Außerdem ist mit $x \in P$ auch $x^{-1} \in P$. Wenn nicht, so wäre ja dann $-x^{-1} \in P$, aber dann auch $x \cdot (-x^{-1}) = -1 \in P$, was falsch ist. Also müssen alle Brüche $x = \frac{a}{b}$ mit $a, b \in \mathbb{N}_+$ in P liegen und so definiert man $P \subseteq \mathbb{Q}$ dann auch. Mehr Brüche $\frac{a}{b}$ mit $a \in \mathbb{Z}$, $b \in \mathbb{Z} \setminus \{0\}$ können nicht in P liegen, sonst wäre die *Trichotomie* (4) zerstört. Es gibt also genau eine solche Anordnung P auf $\mathbb{Q}$, die (4) sowie (i') und (ii') erfüllt. Das ermöglicht also wieder den Größenvergleich von je zwei rationalen Zahlen und man kann diese sich wieder „linear“ auf einer „Zahlengeraden“ angeheftet denken.

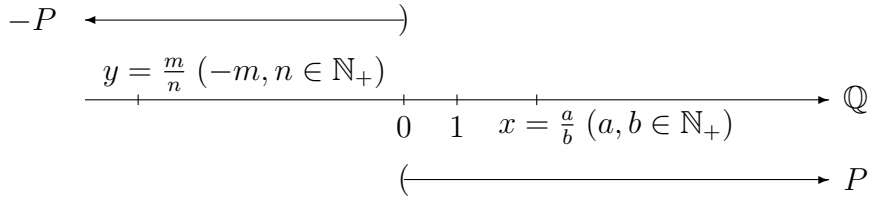

Abbildung 6: die rationalen Zahlen auf dem Zahlenstrahl

Das ermöglicht es schließlich, *Abstände* zwischen rationalen Zahlen mit Hilfe der *Betragsfunktion* $|\cdot|\colon \mathbb{Q} \to P \cup \{0\}$,

$$|x| := \begin{cases} x, & \text{falls } x \in P \\ 0, & \text{falls } x = 0 \\ -x, & \text{falls } x \in (-P) \end{cases}$$

zu definieren vermöge $d\colon \mathbb{Q} \times \mathbb{Q} \to P \cup \{0\}$,

$$d(x, y) := |y - x|.$$

Wir wollen zum Abschluss hier noch die *archimedische Eigenschaft* für diese Anordnung P auf $\mathbb{Q}$ erwähnen.

Satz 3.2 (Archimedisches Axiom). *Zu jedem $x \in \mathbb{Q}$ gibt es ein $n \in \mathbb{N}$ mit $n > x$.*

Beweis. Man kann annehmen, dass $x > 0$ ist, sonst wähle man z.B. $n = 1$. Ist dann $x = \frac{a}{b}$ mit $a, b \in \mathbb{N}_+$, so reicht es ein $n \in \mathbb{N}$ zu finden mit $n \cdot b > a$, weil dann $n > \frac{a}{b} = x$ ist. Aber dazu kann man z.B. einfach $n := a + 1$ wählen, denn dann ist

$$n \cdot b = (a + 1) \cdot b = a \cdot b + b > a \cdot b \geq a \cdot 1 = a.$$

□

(3.3) Da man die reellen Zahlen insbesondere auch zum Messen von Abständen, aber auch zum Messen von Flächeninhalten oder Volumina, benutzen wollte, hat man die Vorstellung entwickelt, dass ihre Elemente in eineindeutiger Beziehung zu allen Punkten auf einer (Zahlen-) Geraden stehen und dabei natürlich die algebraischen Rechenoperationen mit ihren üblichen Regeln erhalten bleiben sollen. Wir sagen heute, dass die Menge der reellen Zahlen ein *angeordneter Körper* sein soll, also eigentlich ein Paar $(\mathbb{R}, P)$, wobei $\mathbb{R}$ ein Körper ist und $P \subseteq \mathbb{R}$ eine Teilmenge, die die Axiome (4) und (i') und (ii') erfüllt. Wie soll man aber nun präzise ausdrücken, dass die Elemente von $\mathbb{R}$ in einer $1:1$-*Beziehung* zu den Punkten einer Geraden stehen soll?

Hier nun lässt man sich von der Idee leiten, dass, wenn schon nicht jeder Punkt auf der Zahlengeraden zu einer rationalen Zahl korrespondiert, so muss er aber doch durch die rationalen Punkte beliebig genau *approximierbar* sein. Dazu führt man zunächst das Konzept einer *konvergenten Folge* wie folgt ein.

Definition 3.3 Eine Folge $(x_n)_{n\in\mathbb{N}}$ rationaler Zahlen (d.h. eine Abbildung von $\mathbb{N}$ nach $\mathbb{Q}$, $n \mapsto x_n$), heißt *konvergent gegen eine rationale Zahl a*, falls es zu jedem (rationalem) $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ gibt, so dass für alle $n \in \mathbb{N}$ mit $n \geq n_0$ gilt:

$$|x_n - a| < \varepsilon.$$

Um nun die „Lücken“, die $\mathbb{Q}$ auf dem Zahlenstrahl noch hinterlässt, „aufzufüllen“, möchte man die Limiten von konvergenten rationalen Zahlenfolgen noch hinzunehmen, auch wenn diese gar nicht rational sind. Das ist die Heuristik.

Aber wie soll man ausdrücken, dass eine rationale Zahlenfolge konvergiert gegen eine Zahl, die man noch gar nicht hat? Hier hilft die geniale Idee einer *Cauchy-Folge*, die die Konvergenz einer Folge ausdrückt, ohne den Grenzwert dabei zu erwähnen. Man drückt lediglich aus, dass die Folgenglieder immer näher „zusammenrücken“.

Definition 3.4 Eine Folge rationaler Zahlen (x_n) heißt eine *Cauchy-Folge*, wenn es zu jedem (rationalen) $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ gibt, so dass für alle $m, n \in \mathbb{N}$ mit $m, n \geq n_0$ gilt:

$$|x_n - x_m| < \varepsilon.$$

Das ist nun der Malus, der beim Übergang von $\mathbb{Q}$ zu $\mathbb{R}$ noch zu beheben ist: dass es in $\mathbb{Q}$ Cauchy-Folgen gibt, die nicht konvergent (gegen $a \in \mathbb{Q}$) sind.

Beispiel 3.5 Die rekursiv definierte *(babylonische)* Zahlenfolge (x_n), rekursiv gegeben durch

$$x_0 = 1, \qquad x_{n+1} = \frac{1}{2}(x_n + \frac{2}{x_n})$$

(also $x_1 = \frac{3}{2}$, $x_2 = \frac{17}{12}$, $x_3 = \ldots$) ist eine Cauchy-Folge und wenn sie gegen eine Zahl $c \in \mathbb{Q}$ konvergiert, so muss gelten: $c^2 = 2$.

Beweis. Wir wollen zuerst bemerken, dass diese rekursiv definierte Folge tatsächlich wohldefiniert ist (also $x_n \neq 0$, $\forall n \in \mathbb{N}$), weil man den Rekursionssatz auf das Tripel $(M, a, F) = (\mathbb{R}_+, 1, F)$ mit

$$F(x) = \frac{1}{2}(x + \frac{2}{x})$$

anwenden kann. Etwas mehr ist richtig: Wir behaupten, dass alle Glieder der Folge im Intervall $[1, 2]$ liegen und damit der mögliche Grenzwert $c \in \mathbb{Q}$ auch, insbesondere ist dann auch $c \neq 0$. Für $n = 0$ ist das richtig und ist $x_n \in [1, 2]$, so gilt einerseits

$$x_{n+1} = \frac{1}{2}(x_n + \frac{2}{x_n}) \geq \frac{1}{2}(1 + \frac{2}{2}) = 1,$$

und andererseits

$$x_{n+1} = \frac{1}{2}(x_n + \frac{2}{x_n}) \leq \frac{1}{2}(2 + \frac{2}{1}) = 2.$$

Die Aussage folgt also mit vollständiger Induktion.

Dass nun (x_n) eine Cauchy-Folge ist, überlassen wir dem Leser als eine etwas schwierigere Übungsaufgabe (siehe z.B. [5]). (Es folgt z.B. daraus, dass $F|[1,2]$ eine Kontraktion ist, nämlich $|F(x)-F(y)| \leq \frac{1}{2}|x-y|$, $\forall x, y \in [1,2]$.) Wenn nun (x_n) gegen ein $c \in \mathbb{Q}$ konvergieren würde, also

$$\lim_{n\to\infty}(x_n) = c = \lim_{n\to\infty}(x_{n+1}),$$

rechnen wir mit den Rechenreglen für Grenzwerte aus Analysis I:

$$\begin{aligned} c &= \lim(x_{n+1}) = \lim \frac{1}{2}(x_n + \frac{2}{x_n}) \\ &= \frac{1}{2}(\lim(x_n) + \frac{2}{\lim(x_n)}) = \frac{1}{2}(c + \frac{2}{c}) \\ \Rightarrow \quad 2c &= c + \frac{2}{c} \quad \Rightarrow \quad c = \frac{2}{c} \quad \Rightarrow \quad c^2 = 2 \end{aligned}$$

(also c der Fixpunkt von $F|[1,2]$, vgl. auch den so genannten *Banachschen Fixpunktsatz*, z.B. in [**?**], für dieses Argument). □

Schon die Griechen aber wussten, und es kam einem geistigen Erdbeben bei ihnen gleich:

Satz 3.6 *Es gibt keine rationale Zahl c mit $c^2 = 2$.*

Beweis Angenommen doch. Dann sei o.E. $c > 0$ (sonst gehe zu $-c$ über). Also gibt es $m, n \in \mathbb{N}$ mit $(\frac{m}{n})^2 = 2$. Wir dürfen annehmen, dass m und n nicht beide gerade sind (sonst kürze hinreichend oft durch 2). Nun gilt aber:

$$\begin{aligned} &m^2 = 2n^2 \;\Rightarrow\; m^2 \text{ gerade } \Rightarrow\; m \text{ gerade, also } m = 2k \text{ (für ein } k \in \mathbb{N}) \\ &\Rightarrow\; 4k^2 = m^2 = 2n^2 \;\Rightarrow\; 2k^2 = n^2 \;\Rightarrow\; n^2 \text{ gerade } \Rightarrow\; n \text{ gerade,} \end{aligned}$$

Widerspruch! □

Von den reellen Zahlen will man ein solches Verhalten nicht. In ihnen soll jede Cauchy-Folge, auch solche, deren Glieder dann selbst reell sind, konvergent gegen eine reelle Zahl sein. (Man betrachte dazu die analogen Definitionen von konvergenten bzw. Cauchy-Folgen (3.3) und (3.4) für reelle Zahlenfolgen, wobei man auch für $\varepsilon > 0$ reelle Zahlen zulasse.) Man verlangt also die Gültigkeit des folgenden Axioms, welches die rationalen Zahlen nach (3.5) nicht erfüllen.

Vollständigkeitsaxiom 3.7 *Jede Cauchy-Folge konvergiert.*

Dieses Axiom, wenn es denn für $\mathbb{R}$ gültig ist, sichert dann z.B. die Existenz einer positiven Zahl c mit $c^2 = 2$, die wir $\sqrt{2}$ nennen. Wir verlangen also nun folgendes.

Definition 3.8 Ein angeordneter Körper $(\mathbb{R}, P)$ heißt *reelle Zahlen*, wenn die Anordnung archimedisch (siehe (3.2)) und vollständig (siehe (3.7)) ist.

Man beachte, dass ein angeordneter Körper (K, P) notwendig die natürlichen Zahlen enthält, denn die Abbildung $\iota\colon \mathbb{N} \to K$ mit

$$\iota(0) = 0, \quad \iota(n) = 1 + \cdots + 1 \quad (n\text{-mal})$$

(genauer mit dem Rekursionssatz $\iota(0) = 0$ und $\iota(n+1) = \iota(n) + 1$) injektiv ist (d.h.: die *Charakteristik von* K ist Null, $\operatorname{char}(K) = 0$), weil $1 \in P$ und damit auch $1 + \cdots + 1 \in P$ ist, also insbesondere $\neq 0$. Damit liegt dann auch $\mathbb{Q}$ in natürlicher Weise in K (vgl. den letzten Absatz in §2). In diesem Sinne kann man dann überhaupt das Archimedische Axiom (3.2) formulieren.

Beachte ferner, dass das Archimedische Axiom sicherstellt, dass $\mathbb{Q} \subseteq \mathbb{R}$ *dicht* liegen wird, wie man sagt, d.h.: zu je zwei reellen Zahlen $x, y \in \mathbb{R}$ mit $x < y$ gibt es eine rationale Zahl $r \in \mathbb{Q}$ mit $x < r < y$. Im Falle $0 < x < y$ etwa wähle man nämlich (nach Archimedes) ein $n \in \mathbb{N}$ mit $n > \frac{1}{y-x}$ und dann

$$m := \max\{k \in \mathbb{N} : \ \frac{m}{n} \leq x\}.$$

Es liegt dann $r = \frac{m+1}{n}$ echt zwischen x und y, wie man leicht sieht. Das stellt sicher, dass jede reelle Zahl Grenzwert einer rationalen Zahlenfolge (r_n) ist, $x = \lim(r_n)$. (Wähle dazu z.B. r_n so, dass $x - \frac{1}{n} < r_n < x$ ist, für alle $n \in \mathbb{N}$.) Es ist deshalb übrigens auch egal, ob man in den Definitionen (3.3) und (3.4) für rationale oder reelle Zahlenfolgen $\varepsilon > 0$ aus $\mathbb{Q}$ oder $\mathbb{R}$ zulässt.

Damit ist klar, dass $\mathbb{R}$ genau die Limiten von rationalen Cauchy-Folgen enthält und nicht mehr, so wie man es gewollt hat. Wir wollen aber hier schon erwähnen, dass in einem Sinn, den es noch zu präzisieren gilt, in $\mathbb{R}$ viel mehr Elemente liegen als nur die Lösungen von *algebraischen Gleichungen* der Form

$$x^n + r_1 x^{n-1} + \cdots + r_{n-1} x + r_n = 0 \tag{5}$$

mit $n \in \mathbb{N}_+$ und $r_1, \ldots, r_n \in \mathbb{Q}$, die man *algebraische Zahlen* nennt, z.B. $\sqrt{2}$ als Lösung von $x^2 - 2 = 0$. Auch die Länge des Kreises mit Radius 1 etwa, die wir später (siehe §**??**) mit 2π bezeichnen werden, wird eine reelle Zahl sein, und sie ist (nach einem Satz von F. Lindemann) nicht algebraisch. Aber dazu später mehr.

(3.4) Bevor wir uns über die Existenz eines solchen Körpers der reellen Zahlen klar werden, wollen wir wieder über seine Eindeutigkeit nachdenken. Tatsächlich wird das geforderte Axiomensystem, dass nämlich $(\mathbb{R}, P)$ ein *vollständig und archimedisch angeordneter Körper* ist, $\mathbb{R}$ bis auf kanonische Isomorphie festlegen. Das bedeutet aber dann für uns auch, dass alles Wesentliche über $\mathbb{R}$ in diesen Axiomen enthalten ist und dass wir, wenn wir uns der Existenz einmal versichert haben, um die Konstruktion keine Gedanken mehr machen brauchen. Mit den Elementen $x \in \mathbb{R}$ darf man wie im Körper $\mathbb{Q}$ rechnen, dazu kommt die Ordnung P mit ihrer Archimedischen Eigenschaft und ihrer Vollständigkeit. Das zu wissen reicht, und die Konstruktion darf man wieder vergessen.

Satz 3.9 *Seien (K_1, P_1) und (K_2, P_2) reelle Zahlen. Dann gibt es einen eindeutig bestimmten Körperisomorphismus $\varphi\colon K_1 \to K_2$, der die Anordnungen respektiert, d.h.: $\varphi(P_1) = P_2$.*

Beweis. Wie am Ende von §2 und unter (3.3) beschrieben, enthalten K_1 und K_2 in natürlicher Weise die rationalen Zahlen $\mathbb{Q}$. Natürlich setzt man $\varphi|Q = \mathrm{id}_{\mathbb{Q}}$, denn dazu wird man ja durch $\varphi(1) = 1$ und der Homomorphismuseigenschaft

$$\begin{aligned} \varphi(x_1 + x_2) &= \varphi(x_1) + \varphi(x_2), \\ \varphi(x_1 x_2) &= \varphi(x_1)\varphi(x_2), \end{aligned}$$

für alle $x_1, x_2 \in K_1$, gezwungen. Nun kann man aber wegen der Archimedischen Eigenschaft von K_1 jedes Element $x \in K_1$ als Grenzwert einer rationalen Folge (r_n) in K_1 schreiben. Weil die Einschränkungen der Ordnungen P_1 und P_2 auf $\mathbb{Q} \subseteq K_1$ bzw. $\mathbb{Q} \subseteq K_2$ die bekannte Ordnung auf $\mathbb{Q}$ induzieren (denn es gibt nur diese Ordnung auf $\mathbb{Q}$, wie wir wissen), ist (r_n) eine Cauchy-Folge auch in K_2. Da K_2 vollständig ist, muss diese einen Grenzwert in K_2 haben (und dieser ist auch eindeutig bestimmt). Man setzt deshalb

$$\varphi(\lim(r_n)) := \lim(r_n)$$

(d.h.: man setzt φ *stetig* fort (vgl. §**??**)). Hier hat man nun zu prüfen, dass φ wohldefiniert und ein Homomorphismus ist, was an den Rechenregeln für Grenzwerte liegt. Z.B. ist mit $x = \lim(r_n)$, $y = \lim(s_n)$:

$$\begin{aligned} \varphi(x + y) &= \varphi(\lim(r_n) + \lim(s_n)) = \varphi(\lim(r_n + s_n)) \\ &= \lim(r_n + s_n) = \lim(r_n) + \lim(s_n) = \varphi(x) + \varphi(y). \end{aligned}$$

Damit ist φ ein *Körperhomomorphismus*, nicht trivial und daher injektiv. Weil K_1 vollständig und K_2 archimedisch ist, ist φ auch surjektiv.

Für den Ordnungserhalt von φ beobachten wir, dass wegen der Vollständigkeit von K_1 jedes Element $a \in P_1$ eine *Wurzel* in P_1 hat. Man ersetze in (3.5) nämlich einfach die Zahl 2 (im Zähler des 2. Summanden in dem Rekursionsausdruck) durch ein beliebiges Element $a \in P_1$. P_1 ist also genau dadurch bestimmt, dass P_1 die Quadratzahlen (von Null verschieden) in K_1 sind, denn ein negatives Element kann nicht Quadratzahl sein, weil für $x \neq 0$ entweder $x \in P$ oder $-x \in P$ ist und damit $x^2 = x \cdot x = (-x) \cdot (-x) \in P$, wegen $P \cdot P \subseteq P$. Aber Quadratzahlen gehen unter Körperhomomorphismen auf Quadratzahlen, denn $\varphi(x^2) = \varphi(x)^2$. Also ist $\varphi(P_1) = P_2$ und damit φ wie gewünscht.

Aus dieser Tatsache folgt übrigens auch, dass ein Körperisomorphismus stetig sein muss (Übung), woraus dann auch die Eindeutigkeit von φ folgt. □

(3.5) Es bleibt nun noch, die Existenz sicher zu stellen. Was soll man als Element definieren, dass als Grenzwert z.B. der babylonischen Folge (r_n) in $\mathbb{Q}$ auftreten soll? Die Antwort ist ganz einfach: *Die Folge selbst!* Allerdings muss man noch darauf achten, dass zwei Folgen mit dem gleichen Grenzwert auch identifiziert werden, d.h. genau dann, wenn ihre Differenz eine Nullfolge ist. Wir definieren also zunächst

$$M := \{(r_n) \in \mathbb{Q}^{\mathbb{N}} : \ (r_n) \text{ ist C-Folge}\}$$

und darin die Äquivalenzrelation

$$(r_n) \sim (s_n) :\Leftrightarrow \ (s_n - r_n) \text{ ist Nullfolge.}$$

Hierbei bezeichnen wir für zwei Mengen A und B mit A^B die Menge aller Abbildungen von B nach A, $A^B := \mathrm{Abb}(B, A)$, also mit $\mathbb{Q}^{\mathbb{N}}$ die Menge aller rationalen Folgen. Eine *C-Folge* ist eine Cauchy-Folge. Den Quotienten nehmen wir dann zunächst als *Menge der reellen Zahlen*, $\mathbb{R} := M/\sim$.

Auf M gibt es nun zunächst eine kommutative *Ringstruktur*, die durch die komponentenweise Addition und Multiplikation gegeben ist,

$$(r_n) + (s_n) := (r_n + s_n), \quad (r_n) \cdot (s_n) := (r_n s_n),$$

die zwar auch ein Einselement hat, aber nicht nullteilerfrei ist, geschweige denn eine Körperstruktur ist. Diese Strukturen drücken sich aber nun wieder auf $\mathbb{R} = M/\sim$ herunter, d.h.: man kann die Addition und Subtraktion für Äquivalenzklassen repräsentantenweise (wohl-) definieren (weil $I = \{(r_n) \in M : (r_n) \text{ ist Nullfolge}\}$ ein *Ideal* in M ist). Damit ist $\mathbb{R}$ zusammen mit den so definierten Verknüpfungen schon mal ein kommutativer

Ring mit Einselement, welches durch die Äquivalenzklasse der konstanten Folge $(1, 1, \ldots)$ gegeben ist,

$$1 = [1, 1, \ldots].$$

(Wir lassen die runde Klammer innerhalb der eckigen Klammer weg.) Die rationalen Zahlen liegen daher als die Äquivalenzklassen der konstanten Folgen in $\mathbb{R}$,

$$r = [r, r, \ldots]$$

(wenn wir auf der linken Seite den Buchstaben ι für die Einbettung von $\mathbb{Q}$ nach $\mathbb{R}$ unterdrücken). $\mathbb{R}$ ist aber nun, im Gegegnsatz zu M, sogar ein Körper (I ist ein *maximales Ideal*), weil für $x = [r_n] \neq 0$ *fast alle* Folgenglieder (im Sinne von: alle, bis auf endlich viele) von Null verschieden sind und es sogar ein $\varepsilon > 0$ gibt, so dass $|r_n| > \varepsilon$ ist, für fast alle $n \in \mathbb{N}$. Nach Wechsel des Repräsentanten darf man dies für alle $n \in \mathbb{N}$ annehmen, weil sich die Cauchy-Eigenschaft einer Folge nicht ändert, wenn man sie auf nur endlich vielen Stellen ändert und auch nicht ihre Äquivalenzklasse, weil die Differenz ja fast überall gleich Null und damit eine Nullfolge ist. Aber dann ist auch $(\frac{1}{r_n})$ eine C-Folge und ihre Klasse tatsächlich invers zu x, denn

$$[r_n] \cdot [\frac{1}{r_n}] = [r_n \frac{1}{r_n}] = [1] = 1.$$

Es ist also $\mathbb{R}$ ein Körper.

Für die Anordnung P auf $\mathbb{R}$ setzt man nun

$$[r_n] \in P :\Leftrightarrow \exists \varepsilon > 0 : \ r_n \geq \varepsilon, \text{ für fast alle } n \in \mathbb{N}$$

($\varepsilon \in \mathbb{Q}$) und prüft dann relativ leicht die Ordnungsaxiome nach. Diese Ordnung ist nun sicher archimedisch, denn zu $0 < x < y$ wähle man rationale $r, s > 0$ mit $0 < r < x$ und $y < s$. Nun gibt es ein $n \in \mathbb{N}$ mit $nr > s$, weil $\mathbb{Q}$ archimedisch ist, also auch

$$nx > nr > s > y.$$

Dass jedes Element $x = [r_n]$ Grenzwert einer rationalen Folge ist, ist klar, weil man r_n nun als Element in $\mathbb{R}$ betrachten kann und dann gilt (mit einer kleinen Überlegung): $x = \lim(r_n)$.

Schließlich ist $\mathbb{R}$ auch vollständig. Ist nämlich (x_n) eine C-Folge in $\mathbb{R}$, so wähle man zunächst $r_n \in \mathbb{Q} \subseteq \mathbb{R}$ mit $|x_n - r_n| < \frac{1}{n}$. Dann prüfe man, dass (r_n) auch eine C-Folge ist mit folgendem $\frac{\varepsilon}{3}$-*Argument*:

$$\begin{aligned} |r_n - r_m| &\leq |r_n - x_n| + |x_n - x_m| + |x_m - r_m| \\ &< \frac{1}{n} + \frac{\varepsilon}{3} + \frac{1}{m} < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon, \end{aligned}$$

falls $m, n \geq n_0$ und n_0 groß genug ist. Diese rationale C-Folge hat nun nach Konstruktion einen Grenzwert x in $\mathbb{R}$, nämlich die Klasse, die durch (r_n), repräsentiert wird, $x = \lim(r_n)$. Aber dieses x ist dann auch Grenzwert von (x_n), denn

$$|x_n - x| \leq |x_n - r_n| + |r_n - x| \leq \frac{1}{n} + \frac{\varepsilon}{2} < \varepsilon,$$

für n groß genug. Damit ist $\mathbb{R}$ ein archimedisch und vollständig angeordneter Körper und die Existenz solcher bewiesen.

(3.6) Wir wollen noch ein Argument geben, warum die reelle Zahlen x nicht alles *(verallgemeinerte) Wurzeln*, d.h. Nullstellen von Polynomen mit rationalen Koeffizienten, sein können,

$$x^n + r_1 x^{n-1} + \cdots r_{n-1} x + r_n = 0$$

(mit $n \in \mathbb{N}$, $r_1, \cdots, r_n \in \mathbb{Q}$). Bekanntlich sind ja die rationalen Zahlen $\mathbb{Q}$ *abzählbar*, d.h.: es gibt eine *Bijektion* $\mathbb{N} \to \mathbb{Q}$. Eine solche bekommt man z.B. dadurch, dass man $\mathbb{Q}_+ = \{\frac{m}{n} \in \mathbb{Q} : m, n \in \mathbb{N}_+\}$ so abzählt, wie in Diagramm 3.

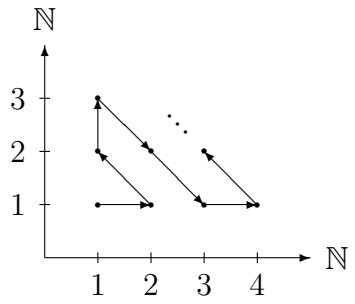

Abbildung 7: eine Abzählung der rationalen Zahlen

(Diese Abzählung ist nur surjektiv auf $\mathbb{Q}_+$, aber daraus kann man relativ leicht eine bijektive Abbildung nach $\mathbb{Q}$ bauen.)

Die reellen Zahlen sind dagegen nicht abzählbar, wie man an G. Cantors berühmten *Diagonalverfahren* sieht. Schon die Zahlen x im Intervall $[0, 1)$ sind es nicht. Um dies zu sehen, stellen wir die Zahlen im *Dezimalsystem* dar,

$$x = 0, a_1 a_2 \cdots, \tag{6}$$

was eine Abkürzung für die unendliche Reihe

$$x = \sum_{n=1}^{\infty} a_n 10^{-n}$$

ist. Man beachte, dass dies im Grunde auch eine rational C-Folge ist, deren einzelnen Glieder gerade die Partialsummen $\sum_{k=1}^{n} a_k$ sind. Man bekommt eine solche Dezimaldarstellung, in dem man im Grunde immer wieder eine Division mit Rest durchführt, von der in den nächsten beiden Paragraphen 4 und 5 ausführlich die Rede sein wird, nur dass hier der Prozess des Dividierens i.a. nicht aufhört. Man teilt nämlich zunächst $x \in [0,1)$ durch $\frac{1}{10}$ mit Rest, d.h.: man sucht die maximale Ziffer $a_1 \in \{0,1,\ldots,9\}$, so dass $a_1\frac{1}{10} \leq x$ ist. Der verbleibende Rest $r_1 = x - \frac{a_1}{10}$ muss dann kleiner als $\frac{1}{10}$ sein und diesen teilt man dann durch $\frac{1}{100}$ mit Rest. Man erhält dann a_2 derart, dass $r_2 := x - \frac{a_1}{10} - \frac{a_2}{100}$ kleiner als $\frac{1}{100}$ ist. Auf diese Weise fährt man fort und erhält die angegebene Dezimaldarstellung (6). Übrigens ist diese Darstellung dann auch eindeutig, wie man zeigen kann, bis auf *Neunerenden*. So beschreiben z.B. die Dezimalzahlen

$$0,1000\ldots \quad \text{und} \quad 0,0999\ldots \tag{7}$$

die gleiche reelle Zahl, nämlich $\frac{1}{10}$, weil ja die Differenz dieser beiden C-Folgen offenbar die Folge

$$(0, 0.1, 0.01, 0.001, \ldots)$$

ist, wobei die Kommas hier die Folgenglieder trennen und der Punkt das alte Komma in der Dezimaldarstellung markiert. Aber dies ist offenbar eine Nullfolge, d.h.: die beiden rationalen Folgen (7) sind äquivalent, d.h.: ihre Äquivalenzklassen sind die gleichen reellen Zahlen.

Nehmen wir nun mit Cantor an, dass wir doch eine Bijektion von $\mathbb{N}$ nach $[0,1)$ haben, so könnten wir also alle reelle Zahlen zwischen 0 und 1 aufzählen (eindeutig, wenn wir keine Neunerenden benutzen):

$$\begin{aligned} x_1 &= 0, a_{11}a_{12}a_{13}\ldots \\ x_2 &= 0, a_{21}a_{22}a_{23}\ldots \\ &\vdots \end{aligned}$$

Dann bildet man die reelle Zahl

$$x := 0, c_1c_2c_3\ldots,$$

wobei man (meinetwegen) $c_n := 5$ setzt, für alle $n \in \mathbb{N}_+$, außer wenn $a_{nn} = 5$ ist. In diesem Fall setzt man (z.B.) $c_n := 6$. Auf diese Weise ist nun sicher gestellt, dass $c_n \neq a_{nn}$ ist, für alle $n \in \mathbb{N}_+$. Und deshalb kann nun $x \in [0,1)$ doch nicht in obiger Liste auftauchen, denn würde es, sagen wir, an der n. Position sein, so müsste $c_n = a_{nn}$ sein. Genial!

Aber jetzt sieht man durch eine leichte Variation, dass auch die algebraischen Zahlen

$$A := \{x \in \mathbb{R} : \ \exists n \in \mathbb{N}_+, r_1, \ldots, r_n \in \mathbb{Q} : \ x^n + r_1 x^{n-1} + \cdots + r_n = 0\}$$

noch abzählbar sein müssen. Jedes solche Polynom hat nämlich, wie wir in §4 sehen werden, nur endlich viele Nullstellen, genauer: höchstens n. Und es gibt nur abzählbar viele solcher Polynome, weil auch $\mathbb{Q} \times \mathbb{Q}$, $\mathbb{Q}^3$ usw. und sogar deren Vereinigung

$$\mathbb{Q} \,\dot\cup\, \mathbb{Q}^2 \,\dot\cup\, \mathbb{Q}^3 \,\dot\cup \cdots$$

abzählbar sind. Damit gibt es nur abzählbar viele algebraischen Zahlen. Es gibt also (viel mehr) *transzendente Zahlen*, d.h. nicht-algebraische Zahlen. (Es ist allerdings überhaupt nicht einfach eine tranzendente Zahl konkret anzugeben. Wir haben schon erwähnt, dass beispielsweise die *Kreiszahl* π (siehe §**??**) transzendent ist. Auch die *Eulersche Zahl* e (siehe §**??**) ist transzendent (siehe z.B. [1] oder [9]).)

4 Polynome

(4.1) Sei K ein Körper, z.B. $K = \mathbb{Q}$ oder $K = \mathbb{R}$. Ein *Polynom mit Koeffizienten in K* ist ein formaler Ausdruck der Form

$$p = a_n X^n + a_{n-1} X^{n-1} + \cdots + a_1 X + a_0,$$

wobei $n \in \mathbb{N}$ ist und $a_0, \ldots, a_n \in K$. Ist $a_n \neq 0$, so heißt n der *Grad des Polynoms*, wir schreiben $n = \deg(p)$. Den Grad des *Nullpolynoms*, d.i., wenn alle *Koeffizienten* verschwinden, $a_0 = \ldots = a_n = 0$, setzen wir auf $-\infty$ fest. Der Buchstabe X hat zunächst keine Bedeutung. Er deutet lediglich an, wie man zwei Polynome zu addieren und zu multiplizieren hat. Man setzt nämlich

$$(a_n X^n + \cdots + a_0) + (b_m X^m + \cdots + b_0) := (a_n + b_n) X^n + \cdots + (a_0 + b_0),$$

wo hier $n \geq m$ angenommen wurde und $b_k = 0$ gesetzt sei für $k > m$. Für das Produkt setzt man

$$\begin{aligned}(a_n X^n + \cdots + a_0) \cdot (b_m X^m + \cdots + b_0) := \\ (a_n b_m) X^{n+m} + \cdots + (\sum_{i+j=k} a_i b_j) X^k + \cdots + (a_0 b_0).\end{aligned}$$

Man multipliziert sozusagen aus und sortiert dann nach Potenzen von X. Wem „formaler Ausdruck“ zu unspezifisch ist und auch die Rolle von X unheimlich ist, der mag ein Polynom p mit Koeffizienten in K als eine *abbrechende Folge* $(a_k)_{k \in \mathbb{N}}$ definieren, d.h.: es gibt ein $n \in \mathbb{N}$, so dass $a_k = 0$ ist, für alle $k > n$,

$$p = (a_0, a_1, \ldots, a_n, 0, 0, \ldots).$$

Die Addition von $p = (a_0, \ldots)$ und $q = (b_0, \ldots)$ geschehe dann komponentenweise und die Multiplikation wie oben beschrieben durch $p \cdot q = (c_k)$ mit $c_k = \sum_{i+j=k} a_i b_j$. Die Menge aller solchen Polynome werde mit $K[X]$ bezeichnet und die eingeführten Operationen $+$ und $\cdot$ machen dann $K[X]$ zu einem kommutativen Ring mit Eins. Die *Einheiten* in diesem Ring, das sind die Elemente, die ein Inverses besitzen, sind genau die *konstanten Polynome* ungleich Null, also die Polynome vom Grad 0. Ist nämlich $p, q \in K[X]$ mit $p \neq 0$ und $q \neq 0$, $n = \deg(p)$ und $m = \deg(q)$, so folgt unmittelbar aus dem Bildungsgesetz für die Multiplikation, dass auch $p \cdot q \neq 0$ ist und gilt (weil ja K nullteilerfrei ist):

$$\deg(pq) = \deg(p) + \deg(q). \tag{8}$$

Und diese Formel gilt auch, wenn $p = 0$ oder $q = 0$ ist, wenn wir noch $-\infty + d = -\infty$ vereinbaren, für alle $d \in \mathbb{N} \cup \{-\infty\}$. Damit ist $K[X]$ also

ein Integritätsring, d.h. nullteilerfrei (also $r_1 r_2 = 0 \Rightarrow r_1 = 0$ oder $r_2 = 0$) und weil die Eins in $K[X]$ durch das konstante Polynom 1 gegeben ist und damit Grad 0 hat, folgt mit (8), dass eine Einheit $p \in K[X]$, wo es also ein Element q mit $pq = 1$ gibt, Grad 0 haben muss,

$$\deg(p) + \deg(q) = 0 \quad \Rightarrow \quad \deg(p) = 0.$$

Die Elemente $p = a$, mit $a \in K^* := K \setminus \{0\}$, sind auch Einheiten, weil $q = a^{-1}$ invers ist. Überhaupt liegt K vermöge $\rho\colon K \to K[X]$, $a \mapsto a$, als Unterkörper in $K[X]$. Wir bezeichnen die Einheiten in einem (kommutativen) Ring R mit R^*. Das passt mit der vorigen Bezeichnung von $K^* = K \setminus \{0\}$ zusammen, weil in einem Körper alle Elemente ungleich 0 Einheiten sind.

Es hat aber $K[X]$ außerdem noch die Struktur eines *K-Vektorraumes*, wobei die Addition die gleiche ist wie bei der vorgestellten Ringstruktur. Die (äußere) Multiplikation $K \times K[X] \to K[X]$ wird dabei natürlich durch

$$\lambda \cdot (a_n X^n + \cdots + a_0) := (\lambda a_n) X^n + \cdots + (\lambda a_0)$$

gegeben. Diese äußere Multiplikation ist mit der inneren Multiplikation in dem Sinne verträglich, dass die natürliche Abbildung $\rho\colon K \to K[X]$ ein Ringhomomorphismus ist. Eine solche Struktur $(+, \cdot_a, \cdot_i)$ auf einer Menge R nennt man eine *K-Algebra-Struktur*. R zu sammen mit ihr ist dann eine *K-Algebra*. Sie ist gleichermaßen ein (kommutativer) Ring mit Eins und ein K-Vektorraum und beide Strukturen vertragen sich. Eine K-Algebra-Struktur auf einem Ring R wird durch einen Ringhomomorphismus $\rho\colon K \to R$ gegeben (und die skalare Multiplikation dann durch $\lambda \cdot r := \rho(\lambda) r$ definiert). Wir halten also fest, dass $K[X]$ mit den vorgestellten Strukturen eine (kommutative) K-Algebra ist.

(4.2) Nun wird das X in p oft als „Unbestimmte“ bezeichnet, für die „man etwas einsetzen kann“ kann. Das stimmt in folgendem präzisem Sinn, der übrigens als *universelle Eigenschaft* das Paar $(K[X], X)$ bis auf kanonische Isomorphie festlegt.

Satz 4.1 *Sei A eine beliebige K-Algebra und $a \in A$ ein beliebiges Element. Dann gibt es genau einen K-Algebra-Homomorphismus $\Phi_a\colon K[X] \to A$ mit $\Phi_a(X) = a$.*

Beweis. Ein Algebra-Homomorphismus $\Phi\colon K[X] \to A$ muss die Einsen ineinander überführen, $\Phi(1) = 1$, und damit wegen der K-Linearität $K \subseteq K[X]$ auf $K \subseteq A$, $\Phi(\lambda) = \lambda$, für alle $\lambda \in K$. (K liegt vermöge $\rho\colon K \to A$ auch in

A, da ρ injektiv sein muss.) Wegen der Multiplikativität von Φ kommt daher nur ein Kandidat in Frage, weil

$$\begin{aligned}\Phi(\lambda_n X^n + \cdots + \lambda_0) &= \Phi(\lambda_n)\Phi(X)^n + \cdots + \Phi(\lambda_0) \\ &= \lambda_n a^n + \cdots + \lambda_0\end{aligned}$$

ist. Man setzt also für X das Element $a \in A$ ein und rechnet in A, wo die vorkommenden Operationen $+, \cdot_i, \cdot_a$ erklärt sind. Nun rechnet man auch leicht nach, dass durch diese Setzung tatsächlich ein Algebra-Homomorphismus erklärt ist. □

Man nennt Φ_a auch den *Einsetzungshomomorphismus* zum Element $a \in A$ und notiert dann der Einfachheit halber auch für $p \in K[X]$ und $a \in A$

$$p(a) := \Phi_a(p) \in A.$$

Z.B. kann man den Satz für das Paar (A, a) mit $A = K[X]$ und $a = X$ anwenden, in welchem Fall $\Phi_X \colon K[X] \to K[X]$ natürlich die Identität wird. In diesem Sinne ist also $p = p(X)$ wovon wir gelegentlich in der Notation Gebrauch machen.

Welche K-Algebra A bietet sich nun als erstes an, deren Elemente $a \in A$ man in $p \in K[X]$ einsetzen kann? Als einfachste aller (nicht-trivialen) K-Algebren ist das doch wohl $A = K$ selbst. Hält man nun $p \in K[X]$ fest und lässt die Elemente $x \in K$ laufen, erhält man so eine Funktion $f \colon K \to K$, die wir *die zu $p \in K[X]$ gehörende Polynomfunktion* nennen und genauer mit f_p notieren. (Später wird f_p im Falle $K = \mathbb{R}$ häufig selbst wieder mit p notiert, was zu gewissen Verwirrungen führt. Wir kommen darauf zurück.) Die Funktion $f_p \colon K \to K$ wird also so definiert:

$$f_p(x) := p(x).$$

(Man setzt in p für X das Element $x \in K$ ein.) Aber man muss aufpassen. Es kann sein, dass die Funktion f_p nicht mehr die volle Kenntnis von p in sich trägt, d.h., dass der folgende Algebra-Homomorphismus nicht injektiv ist:

$$\Phi \colon K[X] \to \mathrm{Abb}(K, K), \quad \Phi(p) = f_p.$$

Betrachten wir etwa den Körper $K = \mathbb{F}_2$, der nur aus zwei Elementen $0, 1$ besteht (mit den offensichtlichen Verknüpfungen). Darüber betrachten wir (für jedes $n \in \mathbb{N}_+$) die *Monome* $p_n \in K[X]$, gegeben durch $p_n = X^n$, die sicher für $n \neq m$ verschieden sind. Die zugehörigen Polynomfunktionen $f_n := f_{p_n} \colon K \to K$ sind aber alle die gleichen. Sie bilden 0 auf 0 und 1 auf 1 ab,

$$f_n(0) = 0, \quad f_n(1) = 1,$$

also $f_n = \mathrm{id}$, für alle $n \in \mathbb{N}_+$. Man kann einer solchen Funktion also z.B. in keiner sinnvollen Weise *eine Vielfachheit* der Nullstelle 0 zuordnen (dem Polynom X^n natürlich sehr wohl, wie wir noch sehen werden). Es ist also Vorsicht damit geboten, ein Polynom $p \in K[X]$ mit ihrer zugehörigen Polynomfunktion $f_p \in \mathrm{Abb}(K, K)$ zu identifizieren.

Wir weren aber bald sehen, dass der Homomorphismus Φ sehr wohl injektiv ist, wenn der Körper unendlich viele Elemente besitzt. Z.B. im Falle $K = \mathbb{R}$ werden wir sehen (siehe §**??**), wie man durch Ableitungsprozesse an der Polynomfunktion f_p (im Nullpunkt) die Koeffizienten des Polynoms und damit das Polynom zurückermitteln kann. Für viele Prozesse, wie z.B. die folgende *Division mit Rest*, ist es aber konzeptuell besser, diese Prozesse in $K[X]$ zu denken als in dem Unterraum aller Funktionen $\mathrm{Abb}(K, K)$, der als Bild von Φ auftaucht.

(4.3) Wie bei den natürlichen Zahlen (vgl. §5) kann man zwischen zwei Polynomen dividieren, wenn man in Kauf nimmt, dass „ein Rest“ bleibt. Genauer gilt:

Satz 4.2 *Seien $a, b \in K[X]$ und $b \neq 0$. Dann gibt es genau ein Paar von Polynomen (q, r) (für Quotient und Rest), so dass gilt:*

$$a = qb + r \quad \textit{und} \quad \deg(r) < \deg(b).$$

Beweis. (i) Eindeutigkeit: Ist $q_1 b + r_1 = q_2 b + r_2$ und $\deg(r_1) < \deg(b)$, $\deg(r_2) < \deg(b)$, so ist wegen

$$\deg(f + g) \leq \max\{\deg(f), \deg(g)\},$$

für alle $f, g \in K[X]$, und

$$r_2 - r_1 = (q_1 - q_2)b \tag{9}$$

im Falle $q_1 - q_2 \neq 0$:

$$\begin{aligned} \deg(b) &\leq \deg(b) + \deg(q_1 - q_2) = \deg((q_1 - q_2)b) \\ &= \deg(r_2 - r_1) \leq \max\{\deg(r_1), \deg(r_2)\} < \deg(b), \end{aligned}$$

also doch $q_1 - q_2 = 0$, d.i. $q_1 = q_2$. Aber dann ist wegen (9) auch $r_1 = r_2$.

(ii) Existenz: Ist $\deg(a) < \deg(b)$, so setze $q = 0$ und $r = a$. Ist $\deg(a) \geq \deg(b)$, sagen wir

$$a = \lambda X^n + \cdots, \quad b = \mu X^m + \cdots$$

wo („$\cdots$“ Terme mit *niedrigerer Ordnung*, also kleinerer X-Potenz abkürzt), $\lambda, \mu \in K^*$ und $n \geq m$ ist, so setze zunächst

$$q_1 := \frac{\lambda}{\mu} X^{n-m}.$$

Dann ist $\deg(a - q_1 b) < \deg(a)$. Falls nun $\deg(a - q_1 b) < \deg(b)$ ist, setze $q := q_1$ und $r := a - q_1 b$. Ansonsten ist $a - q_1 b = \nu X^k + \cdots$ mit $k \geq m$ und $\nu \neq 0$ und wir setzen

$$q_2 := q_1 + \frac{\nu}{\mu} X^{k-m}.$$

Dann erhalten wir

$$\deg(a - q_2 b) = \deg(a - q_1 b - \nu X^k) < \deg(a - q_1 b).$$

Nach endlich vielen Schritten erhält man so ein $q \in K[X]$, so dass $\deg(a - qb) < \deg(b)$ ist und wir können $r := a - qb$ setzen. Es folgt dann wie gewünscht

$$a = qb + r \quad \text{und} \quad \deg(r) < \deg(b).$$

□

Dies hat verschiedene Konsequenzen, z.B. für die so genannten *Nullstellen* oder *Wurzeln* von $p \in K[X]$.

Definition 4.3 Sei $p \in K[X]$. Dann nennen wir $x \in K$ eine *Nullstelle von* p, wenn gilt: $p(x) = 0$.

Ist nämlich $a \in K$ eine Nullstelle von p, so kann man aus p den linearen Faktor $(X - a)$ *abklammern*, d.h.: es gibt ein $q \in K[X]$, so dass gilt:

$$p = (X - a)q.$$

Tatsächlich, wendet man Division mit Rest für p und $(X - a)$ an, so findet man zunächst Polynome $q, r \in K[X]$ mit $p = q(X - a) + r$ und $\deg(r) < \deg(X - a) = 1$, also $r = \text{const}$. Aber wegen

$$0 = p(a) = q(a)(a - a) + r(a) = r$$

muss $r = 0$ sein. Der Grad von q muss sich nun gegenüber dem von p um 1 erniedrigt haben. Deshalb gibt es zu $p \in K[X]$ vom Grad n höchstens n Elemente $a_1, \ldots, a_k \in K$ $(0 \leq k \leq n)$ und ein $q \in K[X]$ ohne Nullstellen, so dass gilt:

$$p = (X - a_1) \cdots (X - a_k)q, \quad q(x) \neq 0, \quad \forall x \in K,$$

wobei $a_1 \ldots, a_k$ nicht unbedingt paarweise verschieden sein müssen, sondern Nullstellen auch *mehrfach* vorkommen können (vgl. (4.5)).

Korollar 4.4 *Sei $p \in K[X]$ vom Grad $n \in \mathbb{N}$. Dann hat p höchstens n Nullstellen.*

Daraus folgt u.a., dass die Abbildung

$$\Phi\colon K[X] \to \mathrm{Abb}(K,K) \quad p \mapsto f_p,$$

injektiv ist, falls K unendlich viele Elemente hat. Da Φ K-linear ist, brauchen wir dazu nur zu prüfen, ob $\ker(\Phi) = (0)$ ist, und nicht größer. Ist aber $f_p \in \mathrm{Abb}(K,K)$ die Nullfunktion, so hat also p unendlich-viele Nullstellen. Das schafft aber nach (4.4) nur das Nullpolynom.

Eine weitere Konsequenz ist die, dass eine Polynom vom Grd n (bzw. ihre zugehörige Polynomfunktion) schon komplett festliegt, wenn man sie auf $n+1$ Stellen $a_1, \ldots, a_{n+1} \in K$ kennt (*Starrheit von Polynomfunktionen*). Sind nämlich $p_1, p_2 \in K[X]$ beide vom Grad n mit $p_1(a_k) = p_2(a_k)$, für $k = 1, \ldots, n+1$, so hat das Polynom $p := p_2 - p_1$ mindestens $n+1$ Nullstellen und $\deg(p) \leq n$. Das geht nur, wenn $p_2 - p_1 = 0$ ist, also $p_1 = p_2$.

Übrigens darf man auf $(n+1)$ paarweise verschiedenen Stellen $a_1, \ldots, a_{n+1} \in K$ andererseits beliebige Vorgaben $c_1, \ldots, c_{n+1} \in K$ machen und es existiert dann auch ein (dann notwendig eindeutiges) Polynom $p \in K[X]$ vom Grad kleiner oder gleich n mit $p(a_k) = c_k$ $(k = 1, \ldots, n+1)$. Die Bedingung an die $(n+1)$ Koeffizienten $\lambda_0, \ldots, \lambda_n \in K$ von p,

$$p = \lambda_0 + \lambda_1 X + \cdots + \lambda_n X^n,$$

ergibt dann nämlich ein *lineares Gleichungssystem*

$$A\lambda = c \tag{10}$$

mit $\lambda := (\lambda_0, \ldots, \lambda_n) \in K^{n+1}$, $c = (c_1, \ldots, c_{n+1}) \in K^{n+1}$ und $A \in \mathrm{Mat}_{n+1}(K)$ mit

$$A = \begin{pmatrix} 1 & a_1 & \cdots & a_1^n \\ \vdots & \vdots & & \vdots \\ 1 & a_{n+1} & \cdots & a_{n+1}^n \end{pmatrix}.$$

Da $\det(A) \neq 0$ ist (genauer gilt: $\det(A) = \prod_{i<j}(a_j - a_i)$, die so genannte *Vandermondesche Determinante*), existiert also tatsächlich eine Lösung $\lambda \in K^{n+1}$ von (10), d.i. ein $p \in K[X]$ mit $p(a_i) = c_i$ $(i = 1, \ldots, n+1)$. Das Argument gibt übrigens natürlich auch einen alternativen Beweis dafür, dass ein Polynom n-ten Grades höchstens n Nullstellen haben kann, denn für den Fall $c = 0$ in K^{n+1} ergibt sich ja wegen $\ker(A) = 0$ nur $\lambda = 0$ als Lösung von (10) und damit das Nullpolynom.

Schließlich kann man mit Hilfe der Division mit Rest *die Vielfachheit einer Nullstelle* wie folgt definieren.

Bemerkung und Definition 4.5 *Sei $p \in K[X] \setminus \{0\}$ und $a \in K$. Dann gibt es ein eindeutig bestimmtes $k \in \mathbb{N}$, so dass es ein $q \in K[X]$ gibt mit $q(a) \neq 0$ und es gilt:*

$$p = (X - a)^k q.$$

Es ist dann $0 \leq k \leq \deg(p)$ und wir nennen k die Vielfachheit der Nullstelle a in p.

Beweis Ist $p(a) \neq 0$, so setze $k = 0$ und $q = p$. Ansonsten klammere man so lange Faktoren $(X - a)$ ab, bis der Quotient q keine Nullstelle mehr in a hat, was spätestens nach $n = \deg(p)$ Schritten passiert, weil q dann konstant geworden sein muss.

Die Eindeutigkeit ist auch klar. Ist nämlich

$$(X - a)^{k_1} q_1 = (X - a)^{k_2} q_2$$

mit, sagen wir, $k_1 \leq k_2$, so folgt

$$(X - a)^{k_1}(q_1 - (X - a)^{k_2 - k_1} q_2) = 0,$$

also (wegen der Nullteilerfreiheit von $K[X]$)

$$q_1 = (X - a)^{k_2 - k_1} q_2$$

und damit $k_1 = k_2$, weil q_1 ja in a keine Nullstelle mehr hat. □

Man kommt also so auf eine Zerlegung eines beliebigen Polynoms $p \in K[X] \setminus \{0\}$ in ein Produkt der Form

$$p = (X - a_1)^{k_1} \cdots (X - a_r)^{k_r} q,$$

mit $q \in K[X]$ ohne Nullstellen in K, die Nullstellen von p genau die Elemente $a_1, \ldots, a_r \in K$ (nun paarweise verschieden) und ihre Vielfachheiten $k_1, \ldots, k_r \in \mathbb{N}_+$. Es gilt dabei die *Gradformel*

$$\deg(p) = k_1 + \cdots + k_r + \deg(q).$$

Ein Polynom vom Grad $n \in \mathbb{N}$ kann also, nun sogar mit Vielfacheiten gezählt, höchstens n Nullstellen haben, und es hat genau n Nullstellen (mit Vielfachheiten gezählt), wenn p, wie man sagt, *in Linearfaktoren zerfällt*, d.h.:

$$p = \lambda(X - a_1)^{k_1} \cdots (X - a_r)^{k_r},$$

mit $\lambda \in K^*$, $0 \leq r \leq n$, $a_1 \ldots, a_r \in K$ und $k_1, \ldots, k_r \in \mathbb{N}_+$.

(4.4) Etwas allgemeiner betrachtet, möchte man ein Polynom $p \neq \text{const}$ so lange in Faktoren zerlegen, bis es nicht mehr geht, d.h. die einzelnen Faktoren nicht mehr zerlegbar sind. Man definiert deshalb:

Definition 4.6 Ein Element $p \in K[X] \setminus K$ heißt *irreduzibel*, wenn folgendes gilt: Ist $p = ab$ für zwei Elemente $a, b \in K[X]$, so muss a oder b eine Einheit (also konstant ungleich Null) sein.

Z.B. sind (normierte) lineare Polynome $X - a$ irreduzibel, denn ist $X - a = fg$, so muss aus Gradgründen $\deg(f) = 0$ oder $\deg(g) = 0$ sein, also eine Einheit (da natürlich $f \neq 0$ und $g \neq 0$ ist). Bei (normierten) quadratischen Polynomen

$$p = X^2 + aX + b$$

hängt die Frage nach der Irreduzibilität schon sehr stark an dem unterliegenden Körper K. Um dies zu sehen, schreiben wir p mittels *quadratischer Ergänzung* als

$$\begin{aligned} p &= X^2 + aX + b = (X + \frac{a}{2})^2 + b - \frac{a^2}{4} \\ &= (X + \frac{a}{2})^2 - (\frac{a^2}{4} - b). \end{aligned}$$

Nehmen wir nun etwa $K = \mathbb{R}$ an. Dann gibt es zwei Fälle:

(i) Für die so genannte *Diskriminante von p*

$$\Delta := a^2 - 4b$$

gelte $\Delta \geq 0$. Dann kann man wegen der Vollständigkeit von $\mathbb{R}$ (genau) ein $c \geq 0$ finden mit

$$c^2 = \frac{1}{4}\Delta = \frac{a^2}{4} - b;$$

wir schreiben dafür wie gewohnt

$$c := \sqrt{\frac{a^2}{4} - b}.$$

Aber dann kann man nach der 3. *Binomischen Formel* p wie folgt zerlegen:

$$p = (X + \frac{a}{2})^2 - c^2 = (X + \frac{a}{2} + c)(X + \frac{a}{2} - c),$$

und p ist also nicht irreduzibel, sondern zerfällt in zwei Linearfaktoren. p hat dann im Falle $\Delta > 0$ zwei *einfache Nullstellen* (d.h. der Vielfachheit 1) und im Fall $\Delta = 0$ eine *doppelte Nullstelle* (d.h. der Vielfachheit 2), die durch

$$x_{1,2} = -\frac{a}{2} \pm \sqrt{\frac{a^2}{4} - b} \tag{11}$$

gegeben sind. Das nennt man wohl im Schwäbischen die „Mitternachtsformel“, weil angeblich der einstige Rektor des Tübinger Kepler-Gymnasiums und Mitautor des bekannten Schulbuches W. SCHWEIZER, seine Schüler um Mitternacht zusammenrief, um zu prüfen, dass sie Formel (11) um Mitternacht (sozusagen im Schlafe) fähig waren, aufzusagen.

(ii) $\Delta < 0$. Setzt man für $x \in \mathbb{R}$ dann $y := x + \frac{a}{2}$, so sieht man, dass eine eventuelle Nullstelle $x \in \mathbb{R}$ von p dann

$$0 = p(x) = (x + \frac{a}{2})^2 - \Delta, \quad \text{also} \quad y^2 = \Delta$$

erfüllt. Ein Quadrat in $\mathbb{R}$ kann aber, wie wir wegen $P \cdot P \subseteq P$ schon gesehen haben (vgl. den Beweis von (3.9)), nicht negativ sein. p hat also in diesem Fall keine Nullstelle und muss deshalb, wiederum aus Gradgründen, irreduzibel sein.

Es gibt also irreduzible Polynome vom Grad 2 in $\mathbb{R}[X]$. Dass es keine irreduziblen Polynome $p \in \mathbb{R}[X]$ höheren Grades gibt, ist ein berühmter Satz von C.F. GAUSS:

Fundamentalsatz der Algebra 4.7 *(reelle Version). Jedes normierte Polynom $p \in \mathbb{R}[X]$ vom Grad mindestens 1 zerfällt in ein Produkt von linearen und/oder quadratischen Faktoren,*

$$p = (X - a_1) \cdots (X - a_r) q_1 \cdots q_s,$$

mit $r, s \in \mathbb{N}$, $a_1, \ldots, a_r \in \mathbb{R}$ und $q_1, \ldots, q_s$ irreduzible quadratische Polynome.

Wenn wir die *komplexen Zahlen* kennengelernt haben, werden wir eine andere, bekanntere Version dieses berühmten Satzes vorstellen, die zu obiger Formulierung äquivalent ist.

Ist der Körper nicht $\mathbb{R}$, sondern z.B. $\mathbb{Q}$, so kann es sehr viele irreduzible Polynome geben. Z.B. ist das kubische Polynom $X^3 - 2 \in \mathbb{Q}[X]$ in $\mathbb{Q}[X]$ sicher irreduzibel, denn wenn es zerlegbar wäre, so hätte es aus Gradgründen auch einen Faktor vom Grad 1, d.h.: $X^3 - 2$ müsste eine rationale Nullstelle haben. Ähnlich wie bei $\sqrt{2}$ sieht man aber, dass auch $\sqrt[3]{2}$ nicht rational sein kann. $X^3 - 2$ hat also keine rationale Nullstelle.

Allgemein kann man aber festhalten, dass durch sukzessives Zerlegen jedes Polynom $p \neq \text{const}$ (jetzt wieder mit Koeffizienten in einem beliebigen Körper) eine Zerlegung

$$p = q_1 \cdots q_s$$

($s \in \mathbb{N}_+$) gestattet, wo alle Faktoren irreduzibel sind. Es gilt sogar, dass diese Zerlegung in dem Sinne eindeutig ist, als dass die Faktoren bis auf die Multiplikation mit Einheiten und die Reihenfolge eindeutig bestimmt sind (siehe z.B. [4]). Einen analogen Satz für die Zerlegung von ganzen Zahlen als ein Produkt von Primzahlen führen wir im nächsten Abschnitt vor (siehe §5).

(4.5) Hier wollen wir uns noch mit einer weiteren Zahlbereichserweiterung beschäftigen, die daraus resultiert, dass man sich nicht damit abfinden möchte, dass es Polynome gibt, die keine Nullstellen haben, z.B. das Polynom $X^2+1 \in \mathbb{R}[X]$. Wir führen hier eine berühmte Konstruktion von L. KRONECKER vor, die zeigt, wie man bei einem gegebenen Körper k und einem irreduziblen Polynom $p \in k[X]$, also eines, was insbesondere keine Nullstelle in k hat, einen *Oberkörper* $K \supseteq k$ mit einem Element $\alpha \in K$ bauen kann, welches Nullstelle von p ist. (Beachte, dass $K \supseteq k$ natürlich eine k-Algebra ist und man deshalb in p auch Elemente aus K einsetzen kann. Oder man betrachtet $p \in k[X]$ gleich als ein Polynom mit Koeffizienten in K, $p \in k[X] \subseteq K[X]$.) Was man dazu braucht, ist der berühmte *Agorithmus von Euklid*, bei dem man mit Hilfe der Division mit Rest aus zwei Polynomen $f, g \in k[X]$ einen gemeinsamen Teiler h (genauer sogar der *größten gemeinsamen Teiler von* f *und* g, vgl. §5) konstruiert, $h|f$ und $h|g$. (Wir schreiben $h|f$, wenn es ein a gibt mit $h \cdot a = f$.) Durch gegenseitige *Wechselwegnahme*, wie die Griechen das nannten, gibt es durch iteriertes Teilen mit Rest, bei – sagen wir – $\deg(f) \geq \deg(g)$, Polynome $q_1, \ldots, q_{r+1}$ und $a_1, \ldots, a_r$ ($r \in \mathbb{N}_+$), so dass

$$\deg(g) > \deg(a_1) > \cdots > \deg(a_r)$$

ist mit

$$\begin{aligned} f &= q_1 g + a_1 \\ g &= q_2 a_1 + a_2 \\ a_1 &= q_3 a_2 + a_3 \\ &\vdots \\ a_{r-2} &= q_r a_{r-1} + a_r \\ a_{r-1} &= q_{r+1} a_r, \end{aligned}$$

denn spätestens bei $\deg(g)$ Schritten muss der Rest a_{r+1} verschwinden, weil der Grad dann kleiner als Null sein muss (und einen solchen hat nur das Nullpolynom). Wir setzen dann $d := a_r$ und stellen fest, dass d ein Teiler sowohl von f als auch von g ist, denn zunächst ist $d|a_{r-1}$ wegen der letzten Zeile, dann wegen der vorletzten Zeile auch $d|a_{r-2}$, und dann hoch bis in die

ersten beiden Zeilen: $d|g$ und $d|f$. Außerdem stellen wir fest, dass wir d als „Linearkombination von f und g“ schreiben können, denn

$$\begin{aligned} d &= a_r = a_{r-2} - q_r a_{r-1} = a_{r-2} - q_r(a_{r-3} - q_{r-1}a_{r-2}) \\ &= (1 + q_{r-1}q_r)a_{r-2} + (-q_r)a_{r-3} \\ &= \dots \\ &= b \cdot g + c \cdot f, \end{aligned}$$

mit $b, c \in k[X]$. Wir bekommen so also:

Lemma 4.8 *(von* Bézout*). Seien $f, g \in k[X]$ teilerfremd (d.h.: ist $d|f$ und $d|g$, so ist $d \in k[X]^* = k^*$). Dann gibt es $a, b \in k[X]$ mit*

$$af + bg = 1.$$

Das benutzen wir, um den folgenden Satz von Kronecker zu beweisen.

Satz 4.9 *(Kronecker). Sei k ein Körper und $p \in k[X]$ ein irreduzibles Polynom. Dann gibt es eine endliche Körpererweiterung $K \supseteq k$ (d.h. $\dim_k K < \infty$, wenn man K als k-Vektorraum auffasst) und ein Element $\alpha \in K$ mit $p(\alpha) = 0$.*

Beweis. Betrache dazu den Polynomring $k[X]$ selbst und teile aus diesem das *Ideal*

$$(p) := \{f \cdot p : \ f \in k[X]\}$$

heraus,

$$K := k[X]/(p),$$

also $\bar{f} = \bar{g}$ in K, genau wenn $g - f \in (p)$, wenn also $g - f$ *Vielfaches von* p ist. K bleibt in natürlicher Weise ein Ring, weil man Addition und Multiplikation repräsentantenweise definieren kann und die Projektion

$$\pi\colon k[X] \to k[X]/(p) = K, \quad f \mapsto \bar{f},$$

wird ein Ringhomomorphismus.

Aber K ist (im Gegensatz zu $k[X]$) sogar ein Körper, weil nun jedes Element ungleich Null ein Inverses besitzt. Ist nämlich $\bar{q} \neq 0$, also $q \notin (p)$, so sind p und q teilerfremd, denn p ist irreduzibel. Nach (**??**) gibt es deshalb $a, b \in k[X]$ mit $ap + bq = 1$. Es folgt dann

$$\bar{b} \cdot \bar{q} = \bar{1} = 1_K,$$

denn $\bar{p} = 0$. Die Körpererweiterung $K \supseteq k$ ist auch endlich (genauer ist

$$[K : k] := \dim_k K = \deg(p)),$$

denn ist – sagen wir – p normiert vom Grad n,

$$p = X^n + a_1 X^{n-1} + \cdots + a_n,$$

so ist $(\bar{1}, \bar{X}, \ldots, \bar{X}^{n-1})$ ein Erzeugendensystem (sogar eine Basis) von K über k. Beachte dazu z.B., dass wegen $\bar{p} = 0$

$$\bar{X}^n = -a_1 \bar{X}^{n-1} - \cdots - a_n$$

ist.

Schließlich ist das Element $\alpha := \bar{X} \in K$ tatsächlich eine Wurzel von p in K, denn

$$\begin{aligned} p(\alpha) &= \bar{X}^n + a_1 \bar{X}^{n-1} + \cdots + a_n \\ &= \bar{X}^n + \bar{a_1} \bar{X}^{n-1} + \cdots \bar{a_n} \\ &= \overline{(X^n + a_1 X^{n-1} + \cdots + a_n)} \quad \text{(weil } \pi \text{ Homomorphismus ist)} \\ &= \bar{p} = 0. \end{aligned}$$

□

Wendet man diese Konstruktion z.B. auf $k = \mathbb{R}$ mit dem Polynom $X^2 + 1$ an, so erhalten wir

$$\mathbb{C} := \mathbb{R}[X]/(X^2 + 1),$$

die komplexen Zahlen. Nennen wir $i := \bar{X}$, so ist also $p(i) = 0$ für $p = X^2 + 1$, d.i.: $i^2 = -1$. Außerdem haben wir gesehen, dass $\dim_{\mathbb{R}} \mathbb{C} = 2$ ist und $\mathbb{C}$ als Vektorraum über $\mathbb{R}$ (nur mit Addition und skalarer Multiplikation) mit $\mathbb{R}^2$ identifiziert werden kann, wenn man z.B. $(1, i)$ als Basis nimmt und daher 1 mit $e_1 = (1, 0)$ und i mit $e_2 = (0, 1)$ identifiziert. Die neue Multiplikation ergibt sich dann aus

$$(x_1 + iy_1)(y_1 + iy_2) = (x_1 y_1 - x_2 y_2) + i(x_1 y_2 + x_2 y_1),$$

eine Formel, mit der man üblicherweise auf $\mathbb{R}^2$ eine Multiplikation einführt und dann zeigt, dass $\mathbb{R}^2$ damit zusammen einen Körper bildet (der natürlich zu unserem isomorph ist). In diesem Körper zerfällt dann das Plynom $p = X^2 + aX + b$ mit $a, b \in \mathbb{R}$, denn nun kann man $\pm\sqrt{\frac{a^2}{4} - b}$ im Falle $\Delta < 0$ als $\pm i\sqrt{b - \frac{a^2}{4}}$ lesen. Aber über diesem Erweiterungskörper $\mathbb{C} \supseteq \mathbb{R}$ zerfällt

sogar jedes *komplexe* quadratische Polynom $p = X^2 + aX + b$ mit $a, b \in \mathbb{C}$, denn wie man z.B. an der *Polardarstellung komplexer Zahlen* mit Hilfe der *komplexen Exponentialfunktion* (siehe §**??**) sieht, wonach man jede komplexe Zahl $z \in \mathbb{C}^*$ schreiben kann als

$$z = re^{i\varphi},$$

mit $r > 0$ und $\varphi \in \mathbb{R}$, gibt es für jedes $z \in \mathbb{C}$ ein $w \in \mathbb{C}$ mit $w^2 = z$, nämlich

$$w = \sqrt{r}e^{i\frac{\varphi}{2}},$$

weil nach dem *Exponentialgesetz*

$$e^z e^w = e^{z+w},$$

auch für alle $z, w \in \mathbb{C}$, gilt:

$$w^2 = (\sqrt{r} \cdot e^{i\frac{\varphi}{2}})^2 = (\sqrt{r})^2 \cdot e^{i\frac{\varphi}{2}} \cdot e^{i\frac{\varphi}{2}} = re^{i\varphi} = z.$$

Die andere Lösung ist dann

$$-w = re^{i(\varphi+\pi)},$$

so dass wir den Ausdruck $\pm\sqrt{z}$ als $\pm w$ erklären können. (In der so genannten *Funktionentheorie* (siehe auch §**??**) lernt man übrigens, dass man eine Auswahl dieser beiden Wurzeln auf ganz $\mathbb{C}^*$ nicht in *stetiger Weise* durchführen kann, es also keine *stetige Funktion*

$$\sqrt{}: \mathbb{C}^* \to \mathbb{C}^*$$

gibt mit $(\sqrt{z})^2 = z$, für alle $z \in \mathbb{C}^*$. Nur „$\pm\sqrt{}$“ macht in gewisser Weise Sinn.) In jedem Fall bekommt nun die Mitternachtsformel

$$z_{1,2} = -\frac{a}{2} \pm \sqrt{\frac{a^2}{4} - b}$$

für alle $a, b \in \mathbb{C}$ einen Sinn und $p = X^2 + aX + b$ zerfällt über $\mathbb{C}$ in Linearfaktoren,

$$p = (X - z_1)(X - z_2).$$

Tatsächlich besagt *der Fundamentalsatz der Algebra* in der komplexen Version, dass *jedes* Polynom $p \in \mathbb{C}[X]$ vom Grad mindestens 1 (und nicht nur solche vom Grad 2) in Linearfaktoren zerfällt,

$$p = \lambda(X - z_1) \cdots (X - z_n),$$

mit $\lambda \in \mathbb{C}^*$ und $z_1, \ldots, z_n \in \mathbb{C}$ (nicht notwendig verschieden natürlich). Oder, was äquivalent ist, weil man bei einer Nullstelle z stets den Linearfaktor $X - z$ abklammern kann:

Fundamentalsatz der Algebra 4.10 *(komplexe Version). Ist* $p \in \mathbb{C}[X]$ *mit* $\deg(p) \geq 1$*, so existiert ein* $z \in \mathbb{C}$ *mit* $p(z) = 0$*,*

Die reelle Version (4.7) bekommt man aus der komplexen Version so: Ist $p \in \mathbb{R}[X]$ und $z \in \mathbb{C}$ eine Nullstelle von p, so muss auch *das komplex Konjugierte* $\bar{z}$ (d.h. $\bar{z} = x - iy$, wenn $z = x + iy$ ist) eine Nullstelle sein, denn $\bar{}: \mathbb{C} \to \mathbb{C}$ ist ein Körperautomorphismus,

$$\overline{z_1 + z_2} = \bar{z_1} + \bar{z_2}, \quad \overline{z_1 z_2} = \bar{z_1}\bar{z_2},$$

für alle $z_1, z_2 \in \mathbb{C}$, und $\bar{}$ ist bijektiv, weil es offenbar *eine Involution* ist, d.h.: $\bar{} = \mathrm{id}$ ist). Es ist nämlich, wenn

$$p = X^n + a_1 X^{n-1} + \cdots + a_n$$

ohne Einschränkung (o.E.) *normiert* ist und $a_k \in \mathbb{R}$, also $\bar{a_k} = a_k$, $\forall k$,

$$\begin{aligned} p(\bar{z}) &= \bar{z}^n + a_1 \bar{z}^{n-1} + \cdots + a_n \\ &= \bar{z}^n + \bar{a_1}\bar{z}^{n-1} + \cdots + \bar{a_n} \\ &= \overline{z^n + a_1 z^{n-1} + \cdots + a_n} = \overline{p(z)} = \bar{0} = 0. \end{aligned}$$

Die komplexen Nullstellen von p sind deshalb entweder reell, oder sie treten, wenn $z \in \mathbb{C} \setminus \mathbb{R}$ eine Nullstelle ist, gepaart mit ihrem komplex Konjugierten $\bar{z}$ auf:

$$p = \lambda(X - c_1) \cdots (X - c_k) \cdot (X - z_1)(X - \bar{z_1}) \cdots (X - z_l)(X - \bar{z_l}),$$

mit $\lambda \in \mathbb{C}^*$, $c_1, \ldots, c_k \in \mathbb{R}$, $z_1, \ldots, z_l \in \mathbb{C} \setminus \mathbb{R}$ und $k, l \in \mathbb{N}$.

Das zeigt übrigens nebenher, dass ein reelles Polynom ungeraden Grades stets (mindestens) eine reelle Nullstelle haben muss, eine Tatsache die wir in §**??** erneut mit Hilfe des so genannten *Zwischenwertsatzes für stetige Funktionen* sehen werden.

Korollar 4.11 *Ist* $p \in \mathbb{R}[X]$ *mit* $\deg(p)$ *ungerade, so existiert ein* $a \in \mathbb{R}$ *mit* $p(a) = 0$*.*

Nun kann man nämlich die Faktoren $X - z_i$ und $X - \bar{z_i}$ für $i = 1, \ldots, l$ zusammenfassen und erhält für das ausmultiplizierte quadratische Polynom

$$X^2 - (z_i + \bar{z_i})X + z_i \bar{z_i},$$

dass die Koeffizienten

$$a := -(z_i + \bar{z_i}) = -2 \cdot \mathrm{Re}(z_i) \quad b = z_i \bar{z_i} = |z_i|^2 = |z_i|^2$$

reell sind. Hier bezeichne wir wie üblich für $z = x + iy \in \mathbb{C}$ $(x, y \in \mathbb{R})$ mit $x = \mathrm{Re}(z)$ den *Realteil von* z, mit $y = \mathrm{Im}(z)$ den *Imaginärteil von* z und mit

$$|z| = \sqrt{x^2 + y^2} = \sqrt{z\bar{z}} \tag{12}$$

den *Betrag von* z.

Die Aussage, dass bei einem quadratischen Polynom

$$p = X^2 + aX + b$$

mit Nullstellen c_1 und c_2,

$$p = (X - c_1)(X - c_2)$$

gerade

$$a = -(c_1 + c_2) \quad \text{und} \quad b = c_1 c_2$$

ist, wird auch als *Satz von* Vieta bezeichnet (und ergibt sich durch bloßes Ausmultiplizieren).

Bezeichnet man dieses quadratische Polynom $(X - z_i)(X - \bar{z}_i)$ mit $q_i \in \mathbb{R}[X]$ $(i = 1, \ldots, l)$, so sieht man, dass $p \in \mathbb{R}[X]$ die Zerlegung

$$p = \lambda(X - c_1) \cdots (X - c_k) q_1 \cdots q_l$$

zulässt, also gerade die Aussage von (4.7).

Wir wollen abschließend aber auch noch bemerken, dass der Übergang von $\mathbb{R}$ zu $\mathbb{C}$ auch etwas kostet. Der Körper der komplexen Zahlen kann nun nämlich nicht mehr, wie die reellen Zahlen, angeordnet werden, denn $-1 = i^2$ ist ja nun ein Quadrat, und müsste daher, genauso wie $1 = 1^2$, in den positiven Zahlen P liegen. Man beachte auch, dass die Betragsfunktion $|\cdot| \colon \mathbb{C} \to [0, \infty)$ aus (12) nicht von einer Ordnung stammt.

(4.6)[1] Man könnte nun auf die Idee kommen, auch für Polynome höheren Grades als 2, sagen wir mit komplexen Koeffizienten $a_1, \ldots, a_n \in \mathbb{C}$, eine Formel vom Typ der Mitternachtsformel zu suchen, die die n Nullstellen $z_1, \ldots, z_n \in \mathbb{C}$, mit Vielfachheiten gezählt,

$$X^n + a_1 X^{n-1} + \cdots + a_n = p = (X - z_1) \cdots (X - z_n),$$

durch die Koeffizienten $a_1, \ldots, a_n$ ausdrückt. Gedacht ist dabei daran, dass diese Formel nur die algebraischen Rechenoperationen beinhaltet, sowie iteriertes (aber freilich nur endlich oft mal iteriertes) k-tes Wurzelziehen in dem

[1] Dieser Abschnitt kann beim ersten Lesen übersprungen werden

Sinne, dass nur die Lösungen der Gleichung $X^k - b$, die man dann, wieder mehrwertig, mit $\sqrt[k]{b}$ bezeichnet, in der Formel vorkommen dürfen. Gibt es also eine solche Formel

$$z_{1,\dots,n} = z_{1,\dots,n}(a_1, \dots, a_n),$$

in der auf der rechten Seite nur *Radikale*, wie man sagt, vorkommen?

Übrigens: umgekehrt ist das Problem natürlich einfach. Hat man die Nullstellen $z_1, \dots, z_n$, so bekommt man die Koeffizienten $a_1, \dots, a_n$ in *Verallgemeinerung des Satzes von Vieta* als die so genannten *elementarsymmetrischen Funktionen* in $z_1, \dots, z_n$ (symmetrisch in dem Sinne, dass sich der Wert nicht ändert, wenn man die Reihenfolge von $z_1, \dots, z_n$ beliebig ändert),

$$a_k = (-1)^k \sigma_k(z_1, \dots, z_n),$$

mit

$$\sigma_k(z_1, \dots, z_n) := \sum_{1 \le i_1 < \dots < i_k \le n} z_{i_1} \cdots z_{i_k},$$

für $k = 1, \dots, n$. Insbesondere ist

$$\sigma_1(z_1, \dots, z_n) = z_1 + \cdots + z_n$$

und

$$\sigma_n(z_1, \dots, z_n) = z_1 \cdots z_n.$$

Ich möchte hier, zum Abschluss dieses Paragraphen, die Grundidee erklären, warum es eine solche Mitternachtsformel für den Grad $n = 5$ und höher nicht mehr geben kann. Dieses Problem hat die Algebra über lange Zeit hinweg sehr stark befruchtet und ist erst im 19. Jahrhundert, im Wesentlichen durch Galois, gelöst worden in dem Sinne, dass es für $n \ge 5$ nicht geht. Für $n = 3$ und $n = 4$ gibt es übrigens solche Formeln. Sie sind aber schon recht kompliziert und wurden von Cardano im Mittelalter gefunden.

Starten wir also nun mit einem normierten Polynom p vom Grad n mit Koeffizienten aus einem Grundkörper k,

$$p = X^n + a_1 X^{n-1} + \cdots + a_n,$$

$a_1, \dots, a_n \in k$. Man denke bei k durchaus an $\mathbb{Q}$. Sind $a_1, \dots, a_n$ komplexe Zahlen, so ist es gut, bei k an den kleinsten Unterkörper von $\mathbb{C}$ zu denken, der $a_1, \dots, a_n$ enthält, also alle rationalen Ausdrücke

$$\frac{f(a_1, \dots, a_n)}{g(a_1, \dots, a_n)},$$

die man mit Polynomen in n Veränderlichen (vgl. Abschnitt (4.7)) f und g (mit $g(a_1,\ldots,a_n)\neq 0$) in den Koeffizienten bilden kann. Überhaupt ist es eine Grundidee der folgenden Argumentation, die betrachteten Körper immer so klein wie möglich zu machen mit den Eigenschaften, die man von ihnen braucht. Wir notieren den beschriebenen Körper mit $\mathbb{Q}(a_1,\ldots,a_n)\subseteq\mathbb{C}$. Er wird also von $a_1,\ldots,a_n$ über $\mathbb{Q}$ in obigem Sinn *erzeugt*. Später werden wir übrigens für die Koeffizienten $a_1,\ldots,a_n$ auch Unbestimmte einsetzen, denn die gesuchte Mitternachtsformel für die Nullstellen von p soll ja für *alle* $a_1,\ldots,a_n$, und nicht nur für einen speziellen Satz von Koeffizienten, gelten.

Als nächtes betrachtet man den kleinsten Oberkörper $K\supseteq k$ von k, in dem das Polynom p zerfällt, also K alle Wurzeln von p enthält. Diesen kann man z.B. dadurch bekommen, dass man Kroneckers Konstruktion sukzessive so lange wiederholt, bis p über $K[X]$ zerfällt. Hat man eine Nullstelle von p an k *heranadjungiert*, wie man sagt, $K_1=k[X]/(p)$ (falls p irreduzibel ist, sonst nehme man einen irreduziblen Faktor von p), so kann man über K_1 aus p den Faktor $X-\alpha$ abklammern, wo $\alpha\in K_1$ eine Nullstelle von p ist, $p=(X-\alpha)p_1$ mit $p_1\in K_1[X]$. Mit p_1 wiederhole man den Vorgang. Man erhält spätestens nach n Schritten einen Oberkörper $K\supseteq k$, über dem p in Linearfaktoren zerfällt,

$$p=(X-c_1)\cdots(X-c_n),$$

und der von den Wurzeln $c_1,\ldots,c_n\in K$ über k erzeugt ist; wir schreiben wieder

$$K=k(c_1,\ldots,c_n).$$

Alternativ, wenn $k\subseteq\mathbb{C}$ ist, und man bereit ist, den Fundamentalsatz der Algebra einzuspeisen, kann man K als den Unterkörper von $\mathbb{C}$ nehmen, der von $c_1,\ldots,c_n$ über k erzeugt ist, $K=k(c_1,\ldots,c_n)\subseteq\mathbb{C}$. Ein solcher Körper $K\supseteq k$, wie immer man ihn konstruiert, ist bis auf Isomorphie eindeutig bestimmt. Er heißt der *Zerfällungskörper von p über k*. Eine brilliante Idee von Galois ist es nun, die so genannte *Galoisgruppe von p (über k)* zu betrachten, das sind alle Körperautomorphismen von K, die $k\subseteq K$ (punktweise) festlassen,

$$\mathrm{Gal}(p):=\{\varphi\in\mathrm{Aut}(K):\ \varphi|k=\mathrm{id}_k\}.$$

Ein solcher Automorphismus $\varphi\colon K\to K$ muss, ähnlich wie oben unter (4.5) die komplexe Konjugation $^-\colon\mathbb{C}\to\mathbb{C}$ die Nullstellen eines reellen Polynoms in sich überführte, die Wurzeln $c_1,\ldots,c_n$ von p in sich übertragen, denn

$$\begin{aligned}p(\varphi(c_i)) &= \varphi(c_i)^n+a_1\varphi(c_i)^{n-1}+\cdots+a_0\\ &= \varphi(c_i)^n+\varphi(a_1)\varphi(c_i)^{n-1}+\cdots+\varphi(a_0)\end{aligned}$$

$$\begin{aligned} &= \varphi(c_i^n + a_1 c_i^{n-1} + \cdots + a_0) \\ &= \varphi(p(c_i)) = \varphi(0) = 0. \end{aligned}$$

Weil aber $K \supseteq k$ von $c_1, \ldots, c_n$ erzeugt wird, liegt der ganze Automorphismus fest, wenn man ihn auf den Wurzeln kennt. Die Einschränkungsabbildung

$$\varphi \mapsto \varphi | \{c_1, \ldots, c_n\}$$

ist also injektiv und man kann deshalb $\mathrm{Gal}(p)$ als Untergruppe der *symmetrischen Gruppe* $\mathcal{S}_n$ aller Permutationen von n Buchstaben auffassen,

$$\mathrm{Gal} \hookrightarrow \mathcal{S}_n.$$

Ihre Ordnung stellt sich übrigens als der Körpergrad $[K : k]$ heraus. Diese Galoisgruppe wird es sein, die ein Hindernis gegen die Existenz einer solchen Mitternachtsformel für Polynome vom Grad n (für $n \geq 5$) darstellen wird.

Nehmen wir nun an, wir hätten eine solche Formel, mit der sich $c_1, \ldots, c_n$ durch iteriertes Wurzelziehen aus $a_1, \ldots, a_n$ berechnen ließen. Wir könnten dann einen *Körperturm*

$$k =: L_0 \subseteq L_1 \subseteq L_2 \subseteq \cdots \subseteq L_r =: L$$

($r \in \mathbb{N}$) bauen, wobei wir in jedem Schritt den Zerfällungskörper eines Polynoms der Bauart $q_i = X^{k_i} - b_i$ betrachten, wobei $b_i \in L_i$ und $k_i \in \mathbb{N}_+$ ist, und L_{i+1} dann der Zerfällungskörper von q_i über L_i, also über L_i von den k_i-ten Wurzeln „ $\sqrt[k_i]{b_i}$“ erzeugt ist ($i = 0, \ldots, r-1$). Wenn sich alle $c_1, \ldots, c_n$ als so eine Radikalfunktion schreiben lassen, so muss dann schließlich K, der Zerfällungskörper von p, ein Unterkörper von L sein, $K \subseteq L$, denn K wäre der Unterkörper von L, der von $c_1, \ldots, c_n$ erzeugt wäre. Die *Galoisgruppe von* L,

$$\mathrm{Gal}_k(L) := \{\varphi \in \mathrm{Aut}(L) : \ \varphi | k = \mathrm{id}_k\},$$

kann man dann durch Einschränkung der Automorphismen surjektiv auf die Galoisgruppe von K, also $\mathrm{Gal}(p)$, abbilden, und der Kern besteht gerade aus den Automorphismen von L, die K punktweise festhalten,

$$\mathrm{Gal}(p) = \mathrm{Gal}_k(K) \cong \mathrm{Gal}_k(L)/\mathrm{Gal}_K(L). \tag{13}$$

Überhaupt gibt es nun den tiefer liegenden Satz von Galois (siehe z.B. [4] oder [2]), den so genannten *Hauptsatz der Galois-Theorie*, der eine eineindeutige Beziehung herstellt zwischen den Zwischenkörpern $k \subseteq M \subseteq L$ einerseits und den Untergruppen $(1) \subseteq H \subseteq G := \mathrm{Gal}_k(L)$ andererseits (vermöge $M \mapsto \mathrm{Gal}_M(L)$). Weil nun $L \supseteq k$ als Endprodukt einer Kette von

Körpererweiterungen zu Stande kam, korrespondiert nun dazu eine Kette von Untergruppen (inklusionsumkehrend)

$$G = H_0 \supseteq H_1 \supseteq H_2 \supseteq \cdots \supseteq H_r = (1), \tag{14}$$

wobei sogar $H_{i+1} \subseteq H_i$ als *Normalteiler* enthalten ist (d.h.: $\alpha H_{i+1}\alpha^{-1} \subseteq H_i$, $\forall \alpha \in H_i$), weil die Körpererweiterungen $L_{i+1} \supseteq L_i$ Zerfällungskörper waren. Die *Quotientengruppe* H_i/H_{i+1} ist dann die Galoisgruppe von L_{i+1} über L_i, und weil dies eine *Radikalerweiterung* war, d.i. der Zerfällungskörpers eines *reinen Polynoms* $X^k - b$, kann man ausrechnen, dass $\mathrm{Gal}_{L_i}(L_{i+1})$ *zyklisch* ist (falls L_i die k_i-ten Einheitswurzeln enthält). Insgesamt bekommt man so die Aussage, dass $G = \mathrm{Gal}_k(L)$ eine solche *Normalkette* von Untergruppen enthält (d.h.: H_i ist Normalteiler in H_{i+1}), derart, dass die Quotienten H_{i+1}/H_i zyklisch sind, also $H_{i+1}/H_i \cong \mathbb{Z}/l_i\mathbb{Z}$, für ein $l_i \in \mathbb{N}$ $(i = 0, \ldots, r-1)$. Solche Gruppen nennt man *auflösbar*, wobei der Name daher rührt, dass sie offenbar auftreten, wenn die Nullstellen eines Polynoms durch Radikale „auflösbar“ sind.

Die Auflösbarkeit vererbt sich dann auf Quotienten und damit auf unsere Galoisgruppe von p (vgl. (13)). Nun ist es aber so, dass die symmetrische Gruppe $\mathcal{S}_n$ für $n \geq 5$ *nicht* auflösbar ist. (Für $n < 5$ schon, da ist sie noch sehr klein.) Genauer ist die *alternierende Gruppe*

$$\mathcal{A}_5 := \{\sigma \in \mathcal{S}_5 : \ \mathrm{sgn}(\sigma) = +1\},$$

$\mathrm{sgn}\colon \mathcal{S}_n \to \{\pm 1\}$ die übliche *Vorzeichenfunktion*, eine Gruppe der Ordnung 60, sogar *einfach*, d.h.: sie hat, außer sich selbst und der trivialen Untergruppe, überhaupt keine Normalteiler mehr, geschweige denn so eine ganze Kette mit zyklischen Quotienten. Nun wissen wir an dieser Stelle aber noch nicht, ob die Galoisgruppe $\mathrm{Gal}(p) \subseteq \mathcal{S}_n$ die volle symmetrische Gruppe, oder wenigstens die alternierende Gruppe, sein kann. Die höheren symmetrischen bzw. alternierenden Gruppen $\mathcal{S}_n$ bzw. $\mathcal{A}_n$ sind dann automatisch auch nicht auflösbar, weil Untergruppen auflösbarer Gruppen, genauso wie deren Quotienten, wieder auflösbar sind. Da $\mathcal{S}_5 \subseteq \mathcal{S}_n$ ist, für $n \geq 5$, folgt dies aus der Nichtauflösbarkeit von $\mathcal{S}_5$.

(4.7)[2] Der letzte Schritt in der Argumentationskette wäre also nun, für jedes $n \geq 5$ ein (normiertes) Polynom p vom Grad n anzugeben (sagen wir mit Koeffizienten in $\mathbb{Q}$), so dass die Galoisgruppe von p die volle symmetrische Gruppe $\mathcal{S}_n$ (oder wenigstens die alternierende Gruppe $\mathcal{A}_n$) ist. Tatsächlich könnten wir das machen und das wird auch in einigen Büchern über Algebra

[2]Dieser Abschnitt ist eine Fortsetzung von (4.6) und kann daher beim ersten Lesen ebenso übersprungen werden

gemacht. Die Nullstellen von p müssen dabei in gewisser Weise „möglichst unsymmetrisch“ liegen. Sind die Nullstellen von p, sagen wir $z_1, \ldots, z_n \in \mathbb{C}$ (wenn wir jetzt mal $p \in \mathbb{Q}[X]$ annehmen), so darf z.B. nicht $z_2 = z_1^2$ sein, denn ein Automorphismus $\varphi\colon K \to K$ mit $K = \mathbb{Q}(z_1, \ldots, z_n)$, müsste ja dann automatisch auch $\varphi(z_2) = \varphi(z_1)^2$ erfüllen. Wenn aber $\mathrm{Gal}(p) = \mathcal{S}_n$ wäre, müsste man für $\varphi(z_1) \in \{z_1, \ldots, z_n\}$ die freie Auswahl haben, dann für $\varphi(z_2) \in \{z_1, \ldots, z_n\} \setminus \{\varphi(z_1)\}$ wieder die freie Auswahl usw. Z.B. ist die Galoisgruppe des Polynoms $X^n - 1$, deren Nullstellen $z_1, \ldots, z_n$ „äquidistant“ auf dem *Einheitskreis*

$$\mathbb{S}^1 = \{z \in \mathbb{C} : \ |z| = 1\}$$

liegen, „sehr weit weg davon“ die volle $\mathcal{S}_n$ zu sein ($n \geq 2$). Da alle n-ten *Einheitswurzeln* z_k, das sind die Nullstellen von $X^n - 1$, nämlich eine Potenz von $z_1 := e^{2\pi i/n}$ sind, genauer (bei geeigneter Nummerierung): $z_k = z_1^k$, für $k = 0, \ldots, n-1$, liegt ein Automorphismus $\varphi\colon K \to K$ schon durch seinen Wert auf z_1 fest. Darüber hinaus muss $\varphi(z_1)$ neben z_1 auch wieder ein Erzeuger der *Gruppe der n-ten Einheitswurzeln* (bzüglich der komplexen Multiplikation) $C_n = \mathbb{Z}/n\mathbb{Z}$ sein, seine so genannten *primitiven Einheitswurzeln*, und das sind genau die Einheiten in dem *Ring* C_n, also

$$\mathrm{Gal}(p) = (\mathbb{Z}/n\mathbb{Z})^*.$$

Das ist eine Gruppe, die nur

$$\varphi(n) = \sharp\{k \in \{1, \ldots, n\} : \ k \text{ und } n \text{ sind teilerfremd}\}$$

Elemente hat ($\sharp$ bezeichnet die Anzahl, $\varphi\colon \mathbb{N}_+ \to \mathbb{N}_+$ heißt *Eulersche Phi-Funktion*) und damit weit weg ist von $n! = \sharp\mathcal{S}_n$. So ein Polynom darf man eben gerade *nicht* nehmen. Es gibt unter seinen Wurzeln zu viel Symmetrie.

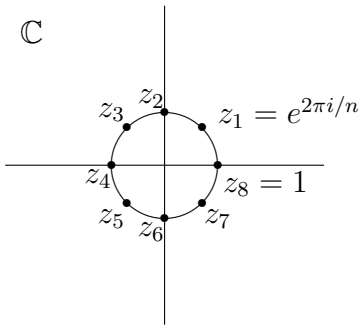

Abbildung 8: die 8. Einheitswurzeln

Das Polynom sollte also in einem gewissen Sinn „generisch“ sein. Wenn die Koeffizienten „zufällig“ gewählt werden, hat man vielleicht die größten Chancen, dass die Galoisgruppe optimal groß wird. In gewisser Weise machen wir das nun, in dem wir die Konstruktion vorstellen, die wir eingangs schon angedeutet haben: Wir lassen auch die Koeffizienten $a_1, \ldots, a_n$ *unbestimmt*. Beachte, dass wir damit dann aber „nur“ beweisen, dass es keine *allgemeine* Mitternachtsformel für die Wurzeln eines Polynoms vom Grad n ($n \geq 5$) geben kann, die sozusagen für alle Wahlen von Koeffizienten funktioniert. Es wird aber nicht ausgeschlossen, dass es für jedes spezielle Polynom, dessen Galoisgruppe auflösbar ist, eine spezielle Mitternachtsformel für dieses Polynom gibt, also sozusagen jedes *auflösbare Polynom* seine eigene Mitternachtsformel hat. Das ist weiterhin möglich.

Um nun die Koeffizienten $a_1, \ldots, a_n$ in dem normierten Polynom

$$p = X^n + a_1 X^{n-1} + \cdots + a_n$$

vom Grad n zu „Unbestimmten“ zu machen, betrachtet man (über einem festen Grundkörper k, sagen wir $k = \mathbb{Q}$) zunächst den *Polynomring in n Unbestimmten*, den man in ähnlicher Weise wie seinen Spezialfall $k[X]$ definiert,

$$R := k[Y_1, \ldots, Y_n] := \{ \sum_{0 \leq i_1 + \cdots + i_n \leq r} b_{i_1 \cdots i_n} Y_1^{i_1} \cdots Y_n^{i_n} : \ r \in \mathbb{N},\ b_{i_1 \cdots i_n} \in k \}.$$

Von diesem stellt man dann durch eine naheliegende Gradformel wie im Fall $n = 1$ wiederum fest, dass R ein *Integritätsring* ist, also keine Nullteiler hat. Deshalb kann man nun mit R die gleiche Erweiterungskonstruktion machen, wie wir sie mit $\mathbb{Z}$ gemacht haben, um einen Körper zu bekommen, in dem man (formale) Brüche $\frac{p}{q}$ mit $p \in R$ und $q \in R \setminus \{0\}$ betrachtet. Man erhält dann den so genannten *Funktionenkörper in n Veränderlichen über k*,

$$k(Y_1, \ldots, Y_n) := \mathrm{Quot}(k[Y_1, \ldots, Y_n]),$$

wobei wir mit Quot die Erweiterungskonstruktion meinen, die jedem Integritätsring seinen (dann so genannten) *Quotientenkörper* zuordnet. Durch Einsetzen könnte man daraus wieder „echte Funktionen“ (etwa auf $\mathbb{R}^n$ bei $k = \mathbb{R}$) bekommen (allerdings müsste man noch „Polstellen“ aus dem Definitionsbereich wegnehmen), was wir aber nicht machen wollen, sondern $\frac{p}{q} \in k(Y) := k(Y_1, \ldots, Y_n)$ wieder vielmehr als „formale Ausdrücke“ betrachten, in die noch nichts eingesetzt ist.

Jetzt sind wir fast fertig. Benennen wir die Unbestimmten $Y_1, \ldots, Y_n$ noch in $a_1, \ldots, a_n$ um, so können wir unser Polynom

$$p = X^n + a_1 X^{n-1} + \cdots + a_0$$

nun als Polynom mit Koeffizienten in dem Funktionenkörper $k(a_1, \ldots, a_n)$ betrachten: p ist also ein spezielles Polynom in $k(a_1, \ldots, a_n)[X]$. Und von diesem kann man nun die Galoisgruppe $G = \mathrm{Gal}_K(p)$ mit $K = k(a_1, \ldots, a_n)$ berechnen (siehe z.B. [2] oder [4]) und erhält dann tatsächlich für alle $n \in \mathbb{N}_+$: $G = \mathcal{S}_n$. Das *allgemeine Polynom vom Grad* $n \geq 5$ ist also in diesem Sinne nicht durch Radikale auflösbar.

5 Elementare Zahlentheorie

(5.1) Die elementare Zahlentheorie betrifft vor allem die Teilbarkeitstheorie in dem Integritätsring der ganzen Zahlen. In einem solchen Integritätsring, d.i. ein kommutativer Ring mit Einselement und ohne Nullteiler, sagt man, dass für zwei Elemente $x, y \in R$ *x ein Teiler von y ist*, oder *x teilt y*, notiert mit $x|y$, wenn es ein $a \in R$ gibt mit $x \cdot a = y$. Dabei sehen wir zunächst, dass jedes Element $x \in R$ die Null $y = 0$ teilt, denn für $a = 0$ ist $x \cdot a = x \cdot 0 = 0 = y$. Umgekehrt teilt die Null nur die Null. (Man beachte allerdings, dass, wenn man $a \neq 0$ fordern würde, kein Element $x \neq 0$ die Null teilt, weil R nullteilerfrei ist. Daher wohl die Bezeichnung.)

Eine weitere besondere Stellung nehmen die Einheiten des Ringes ein, die wir schon im Polynomring $R = K[X]$ über einem Körper K kennengelernt haben. Das sind jene Elemente $\varepsilon \in R$, die ein Inverses haben.

Definition 5.1 Sei R ein Integritätsring. Dann heißt ein Element $\varepsilon \in R$ eine *Einheit in R*, wenn es ein $a \in R$ gibt mit

$$\varepsilon \cdot a = 1.$$

Es ist ja dann $a = \varepsilon^{-1}$. Die Einheiten eines Rings R werden mit

$$R^* = \{\varepsilon \in R : \ \varepsilon \text{ ist Einheit}\}$$

notiert und bilden eine Gruppe bezüglich der Multiplikation in R, denn mit ε ud δ in R^* ist auch ε^{-1} und $\varepsilon\delta$ in R^* (mit $(\varepsilon\delta)^{-1} = \delta^{-1}\varepsilon^{-1}$). Ist R sogar ein Körper, so ist offenbar $R^* = R \setminus \{0\}$, ansonsten kann es schon sein, dass es nur sehr wenige Einheiten gibt. (Natürlich ist $\varepsilon = 1$ immer eine Einheit.) Für den Polynomring $R = K[X]$ (K ein Körper) hatten wir schon gesehen, dass R^* nur aus den konstanten Polynomen ungleich Null besteht, $R^* = K^*$ (wenn man $K \subseteq K[X]$ wie üblich als die konstanten Polynome betrachtet). Im Ring $R = \mathbb{Z}$ der ganzen Zahlen gibt es offenbar nur zwei Einheiten, nämlich ± 1, $\mathbb{Z}^* = \{-1, 1\}$.

Für die Teilbarkeitstheorie sind nun Einheiten in so fern von besonderer Bedeutung, weil sie, ähnlich wie das Einselement, einfach *alles* teilen. Denn ist $\varepsilon \in R^*$ und $y \in R$ beliebig, so kann man $a := \varepsilon^{-1}y$ wählen und erhält

$$\varepsilon \cdot a = \varepsilon \cdot \varepsilon^{-1}y = y, \quad \text{also} \quad \varepsilon|y.$$

In einem Körper (wie z.B. $R = \mathbb{Q}$ oder $R = \mathbb{R}$) gibt es deshalb keine (interessante) Teilbarkeitstheorie, weil „alles alles teilt“ (bis auf die Sonderstellung der Null, die nie eine Einheit ist (wir verlangen in einem Integritätsring $1 \neq 0$, sonst wird alles langweilig, da dann $R = \{0\}$ ist)).

Ähnlich verhält es sich mit den so genannten *assoziierten Elementen* eines Elementes $y \in R$.

Definition 5.2 Sei $y \in R$. Dann heißt ein Element $x \in R$ *assoziiert zu* y, wenn es eine Einheit $\varepsilon \in R$ gibt, so dass $x = \varepsilon y$ ist.

Im Falle $R = K[X]$ und $p \in R$ ist also $q \in R$ genau dann assoziiert zu p, wenn $q = \lambda p$ ist, für ein $\lambda \in K^*$. Man multipliziert das Polynom nur mit einer von Null verschiedenen Konstanten durch. Für $R = \mathbb{Z}$ gibt es zu $m \in \mathbb{Z}$, $m \neq 0$ (zum Nullelement ist natürlich nur die Null selbst assoziiert), genau zwei assoziierte Elemente, nämlich $\pm m = \pm 1 \cdot m$.

Für die Teilbarkeitstheorie beobachten wir nun, dass assoziierte Elemente zu $y \in R$ stets Teiler von y sind, denn für $x = \varepsilon y$ kann man einfach $a := \varepsilon^{-1}$ wählen:

$$x \cdot a = \varepsilon y \cdot \varepsilon^{-1} = y.$$

Außerdem haben assoziierte Elemente offenbar die gleichen Teiler, denn ist $x = \varepsilon y$ und $z|x$, also $za = x$ für ein a, so ist

$$za\varepsilon^{-1} = x\varepsilon^{-1} = y,$$

also auch $z|y$. Assoziiert zueinander sein ist aber symmetrisch, denn ist x assoziiert zu y, $x = \varepsilon y$ ($\varepsilon \in R^*$), so ist auch y assoziiert zu x, weil offenbar $y = \varepsilon^{-1}x$ ist. Daraus folgt, dass assoziierte Elemente tatsächlich die gleichen Teiler haben. Deshalb kann man sich im Polynomring $K[X]$ auf die normierten Polynome (vom Grad mindestens 1) und im Ring $\mathbb{Z}$ auf die natürlichen Zahlen (größer als 1) beschränken.

Jedes Element $y \in R$ hat also die *trivialen Teiler* $x = \varepsilon$, wo $\varepsilon \in R^*$ eine Einheit ist, oder $x = \varepsilon y$, wiederum mit $\varepsilon \in R^*$, also die Assoziierten zu y. Wenn es keine weiteren Teiler mehr gibt, so haben diese Elemente so etwas wie einen atomaren Charakter für die Teilbarkeitstheorie.

Definition 5.3 Ein Element $y \in R \setminus (R^* \cup \{0\})$ heißt *irreduzibel* (oder *unzerlegbar*), wenn gilt: Ist $x \in R$ ein Teiler von y, so ist x eine Einheit oder assoziiert zu y.

Einheiten und das Nullelement spielen für die Teilbarkeitstheorie wie beschrieben eine besondere Rolle. Wir wollen sie nicht irreduzibel nennen.

Die Definition stimmt mit der in §4 für $R = K[X]$ gegebenen offenbar überein, weil ein *echter Teiler* q von p (mit $\deg(p) \geq 1$, d.i. $p \in R\setminus(R^*\cup\{0\})$) einen Grad echt zwischen 0 und $\deg(p)$ hat. ($\deg(q) = 0$ würde gerade Einheit bedeuten und $\deg(q) = \deg(p)$ würde bedeuten, dass q assoziiert ist.)

Für $R = \mathbb{Z}$ stimmt diese Definition für positive Elemente $m \in \mathbb{Z}$, $m > 1$, offenbar mit der landläufigen Definition einer *Primzahl* überein: Ist $n > 0$ mit $n|m$, so ist $n = 1$ oder $n = m$.

Unglücklicherweise werden nun *Primelemente* in Integritätsringen üblicherweise anders definiert. Wir werden aber bald sehen, dass diese Definition für die Integritätsringe $R = K[X]$ und $R = \mathbb{Z}$ mit der Definition der Irreduzibilität zusammenfällt, so dass wir in diesen Ringen nicht zwischen irreduziblen Elementen und Primelementen unterscheiden müssen.

Definition 5.4 Sei $y \in R$, aber $y \notin R^* \cup \{0\}$. Dann heißt y ein *Primelement*, wenn gilt: Sind $a, b \in R$ und ist y ein Teiler von $a \cdot b$, so ist y ein Teiler von a oder ein Teiler von b,

$$y|ab \quad \Rightarrow \quad y|a \text{ oder } y|b.$$

Allgemein ist die Eigenschaft für ein Element $y \in R \setminus (R^* \cup \{0\})$ prim zu sein, stärker als irreduzibel, denn:

Bemerkung 5.5 *Sei $y \in R \setminus (R^* \cup \{0\})$. Dann gilt: Ist y prim, so ist y irreduzibel.*

Beweis. Sei also $x \in R$ ein Teiler von y und damit $x \cdot a = y$, für ein $a \in R$. Da y prim ist und $y|xa$ (denn $y \cdot 1 = ax$), folgt: $y|a$ oder $y|x$.

(i) $y|a$.

$$\Rightarrow \exists b \in R: \ (ax)b = yb = a \ \Rightarrow \ a(xb-1) = 0 \ \Rightarrow \ xb - 1 = 0,$$

denn $a \neq 0$ und damit $x \in R^*$.

(ii) $y|x$. Es gibt also $\varepsilon \in R$ mit $y\varepsilon = x$. Andererseits gibt es auch ein $\delta \in R$ mit $x\delta = y$, nämlich $\delta = a$. Es folgt

$$y\varepsilon\delta = x\delta = y \ \Rightarrow \ y(\varepsilon\delta - 1) = 0 \ \Rightarrow \ \varepsilon\delta - 1 = 0,$$

denn $y \neq 0$. Es folgt $\varepsilon \in R^*$, d.h.: x ist assoziiert zu y.

y ist also irreduzibel. □

Wir nennen nun im Ring $R = \mathbb{Z}$ die positiven irreduziblen Elemente *Primzahlen* (und werden bald sehen, dass sie tatsächlich Primelemente sind). Nun ist zunächst klar, dass sich jede natürliche Zahl $n > 1$ als Produkt von Primzahlen schreiben ässt. Ist n nicht schon selber Primzahl, so gibt es $1 < a, b < n$ mit $a \cdot b = n$ und wenn a (oder b) nicht Primzahl sind, zerlegt man diese wieder. (Formal könnten wir das mit vollständiger Induktion machen, die Aussage für $2 \leq k \leq n-1$ annehmen und dann auf n schließen, weil wir auf a

und b die Induktionsvoraussetzung anwenden könnten. Der Induktionsanfang bei $n = 2$ ist einfach.) Es ist also

$$n = p_1 \cdots p_r$$

mit $r \in \mathbb{N}_+$ und $p_1, \ldots, p_r \in \mathbb{P}$, wobei wir

$$\mathbb{P} = \{n \in \mathbb{N} : \ n \text{ ist Primzahl}\}$$

setzen. Aber es gilt mehr als die bloße Existenz der Zerlegung in Primfaktoren. Diese Zerlegung ist auch, natürlich nur bis auf die Reihenfolge der Faktoren, eindeutig. Und es ist gar nicht so einfach, das einzusehen.

Hauptsatz der Arithmetik 5.6 *Zu jeder natürlichen Zahl $n > 1$ gibt es (bis auf die Reihenfolge) eindeutig bestimmte Primzahlen $p_1, \ldots, p_r$ ($r \in \mathbb{N}_+$), so dass gilt:*

$$n = p_1 \cdots p_r.$$

Beweis. Der Beweis erfolgt durch Induktion über n und die Verankerung für $n = 2$ ist klar.

$\{2, \ldots, n-1\} \to n$ (eine Variante des Induktionsprinzips): Sei also

$$n = p_1 \cdots p_r = q_1 \cdots q_s$$

mit $r, s \in \mathbb{N}_+$ und $p_1, \ldots, p_r, q_1, \ldots, q_s \in \mathbb{P}$.

(i) 1. Fall: $\{p_1, \ldots, p_r\} \cap \{q_1, \ldots, q_r\} \neq \emptyset$.
Sei dann (nach eventueller Vertauschung) $p = p_1 = q_1$. Auf

$$\frac{n}{p} = p_2 \cdots p_r = q_2 \cdots q_s$$

können wir dann die Induktionsvoraussetzung anwenden, da $\frac{n}{p} < n$ ist (und o.E. $\frac{n}{p} > 1$, sonst wäre $n = p$ und die Zerlegung sicher eindeutig). Dann folgt aber $r = s$ und (nach eventueller Vertauschung) $p_i = q_i$, für alle $i = 2, \ldots, r$ und für $i = 1$ sowieso.

(ii) 2. Fall: $\{p_1, \ldots, p_r\} \cap \{q_1, \ldots, q_s\} = \emptyset$. (Dieser Fall darf nicht auftauchen.)
Sei $p := p_1$, $q := q_1$, $a := p_2 \cdot p_r$, $b := q_2 \cdots q_s$, also

$$n = p \cdot a = q \cdot b. \tag{15}$$

Nach Induktionsvoraussetzung ist die Zerlegung $b = q_2 \cdots q_s$ eindeutig und damit p kein Teiler von b (sonst gäbe es auch eine Zerlegung mit einem Faktor p).

O.E. sei $p < q$. Es folgt dann $b < a$ aus (15) und wir setzen

$$\begin{aligned} m := n - p \cdot b &= qb - pb = (q-p)b \\ &= pa - pb = p(a-b). \end{aligned}$$

Auf alle 3 Elemente $m, q-p$ und b kann man die Induktionsvoraussetzung anwenden, weil sie kleiner als n sind. Da $p|p(a-b)$, also $p|(q-p)b$, und p nicht in der (eindeutigen) Zerlegung von b vorkommt, muss p in der eindeutigen Faktorzerlegung von $q-p$ vorkommen, weil m eine eindeutige Zerlegung hat, also: $p|(q-p)$. Also ist $p \cdot c = q - p$ und damit $p(c+1) = q$ für ein c, also $p|q$. Aber q war irreduzibel: Widerspruch!

□

Das ist nun der Grund, warum irreduzible Elemente in diesem Fall auch Primelemente sind: weil die Zerlegung eindeutig ist. (Tatsächlich gilt auch die Umkehrung: ist jedes irreduzible Element prim, so ist die Zerlegung eindeutig (siehe [4]).)

Korollar 5.7 *Jede Primzahl ist Primelement. (Sorry!)*

Beweis. Sei also jetzt $p \in \mathbb{N}$ Primzahl, d.h. irreduzibel und es teile p ein Produkt $a \cdot b$, mit $a, b \in \mathbb{N}_+$. Seien

$$a = p_1 \cdots p_r, \quad b = q_1 \cdots q_s$$

die Primfaktorzerlegungen von a und b. Dann ist

$$ab = p_1 \cdots p_r q_1 \cdots q_s$$

die Primfaktorzerlegung von ab und diese ist eindeutig. Deshalb muss p mit einem dieser Faktoren übereinstimmen, sonst gäbe es eine andere Zerlegung. Ist $p \in \{p_1, \ldots, p_r\}$, so gilt $p|a$, und sonst $p|b$. Das zeigt, dass p Primelement ist. (Ab jetzt machen wir keine Unterscheidung mehr zwischen „Primzahlen“ und „Primelementen“.) □

Wir wollen noch erwähnen, dass wir für den Ring $R = K[X]$ (K Körper) ähnlich argumentieren könnten und auch dort die Zerlegung in irreduzible Elemente eindeutig ist (und damit jedes irreduzble Element auch prim). Solche Integritätsringe werden *faktoriell* genannt (siehe z.B. [4]).

Nicht jeder Integritätsring hat diese Eigenschaft. Betrachtet man z.B. den Unterring der komplexen Zahlen

$$R := \mathbb{Z}[\sqrt{-5}] := \{x + iy\sqrt{5} \in \mathbb{C} : \ x, y \in \mathbb{Z}\}$$

(d.i. der kleinste Unterring von $\mathbb{C}$, der $\mathbb{Z}$ und $\pm\sqrt{-5}$ enthält, und den man auch mit Kronecker durch eine Adjunktion einer Wurzel von $X^2+5 \in \mathbb{Z}[X]$ an $\mathbb{Z}$ bekommen könnte), so gibt es für das Element $9 \in R$ die Zerlegungen in irreduzible Faktoren wie folgt (dass 3, $4 \pm i\sqrt{5}$ irreduzibel sind, sei zur Übung überlassen):

$$9 = 3 \cdot 3 = (4 + i\sqrt{5})(4 - i\sqrt{5}).$$

$\mathbb{Z}[\sqrt{5}]$ ist also nicht faktoriell.

Wir wollen hier noch nachtragen, dass es unendlich viele Primzahlen gibt, was schon EUKLID bekannt war:

Satz 5.8 *$\mathbb{P}$ ist unendlich.*

Beweis. Seien $p_1, \ldots, p_r \in \mathbb{P}$ (die ersten) r Primzahlen ($r \in \mathbb{N}_+$). Betrachte dann

$$n := p_1 \cdots p_r + 1.$$

Sei $p \in \mathbb{P}$ ein Primfaktor von n (den es ja gibt, weil man n zerlegen kann). Dann kann p nicht in $\{p_1, \ldots, p_r\}$ liegen, weil sonst $p|n$ und $p|p_1 \cdots p_r$ und damit auch $p|1$: Widerspruch! Also waren $\{p_1, \ldots, p_r\}$ nicht alle Primzahlen. □

(5.2) Wir wollen hier noch einen Aspekt der Teilertheorie innerhalb der ganzen bzw. natürlichen Zahlen beleuchten, der schon bei dem Studium der Polynome (über einem Körper) sehr hilfreich war: die Division mit Rest. Schon in der Grundschule lernt man dies beim „Schriftlichen Dividieren“:

Satz 5.9 (Division mit Rest). *Seien $a, b \in \mathbb{N}_+$. Dann gibt es genau ein Paar $(q, r) \in \mathbb{N} \times \mathbb{N}$, so dass gilt:*

(i) $b = q \cdot a + r$,

(ii) $r < a$.

Man sagt dann bekanntlich dazu, dass „b geteilt durch a gleich q Rest r“ ist, denn in $\mathbb{Q}$ würde ja dann (i) bedeuten, dass

$$\frac{b}{a} = q + \frac{r}{a}$$

ist.

Beweis. (a) Eindeutigkeit. Ist

$$q_1 a + r_1 = b = q_2 a + r_2,$$

mit $r_1 < a$ und $r_2 < a$, so ist also mit, sagen wir o.E., $r_1 \leq r_2$:

$$r_2 - r_1 = (q_1 - q_2)a, \tag{16}$$

woraus $q_1 - q_2 \geq 0$ folgen muss, da $r_2 - r_1 \geq 0$ ist. Aber mit $0 \leq r_1, r_2 < a$ ist auch $r_2 - r_1 < a$ und deshalb muss $q_1 - q_2 = 0$ sein, da sonst $(q_1 - q_2)a \geq a$ wäre. Also ist $q_1 = q_2$ und wegen (16) dann auch $r_1 = r_2$.

(b) Existenz. Nach der Archimedischen Eigenschaft von $\mathbb{Q}$ ist die Menge

$$A = \{m \in \mathbb{N} : \ m \cdot a \leq b\}$$

nach oben beschränkt und damit endlich. Es gibt also

$$q := \max\{A\}.$$

Wir setzen dann natürlich

$$r := b - qa.$$

Es folgt damit dann (i) und einerseits $r \geq 0$, weil $b \geq qa$ ist, und andererseits $r < a$, denn bei $r \geq a$ wäre auch $q + 1 \in A$:

$$(q+1)a = qa + a = b - r + a = b - (r - a) \leq b.$$

Das Paar $(q, r) \in \mathbb{N} \times \mathbb{N}$ erfüllt damit die Bedingungen (i) und (ii). □

Damit kann man für zwei Zahlen $a, b \in \mathbb{N}_+$ mit, sagen wir, $b \geq a$ den *Euklidischen Algorithmus* starten, der sukzessive b durch a mit Rest r_1 teilt, dann a durch r_1 mit Rest r_2 teilt, usw., so lange, bis zum ersten Mal kein Rest mehr bleibt, was nach endlich vielen Schritten passieren muss, da

$$a > r_1 > r_2 > \cdots \geq 0$$

ist. Es gibt also (eindeutig bestimmte) $k \in \mathbb{N}$, $q_1, \cdots, q_{k+1}, r_1, \cdots, r_k \in \mathbb{N}_+$ mit:

$$\begin{aligned} b &= q_1 a + r_1 \\ a &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{k-2} &= q_k r_{k-1} + r_k \\ r_{k-1} &= q_{k+1} r_k. \end{aligned}$$

Im Falle von $a|b$ ist $k = 0$ und wir setzen $r_0 := a$ in diesem Fall. Die Zahl

$$d := r_k,$$

d.i. also der letzte „echte Rest“, der beim Dividieren mit Rest auftritt, ist dann ein *gemeinsamer Teiler von a und b*, denn wegen der letzten Zeile ist $r_k | r_{k-1}$, wegen der vorletzen Zeile dann $r_k | r_{k-2}$, usw., wegen der 2. Zeile $r|a$ und wegen der 1. Zeile schließlich $r|b$. Wir hatten im Lemma von Bézout (siehe (4.8)) bei der Polynomdivision danach gesehen, was mit dem gleichen Argument auch hier gilt:

Korollar 5.10 *Seien $a, b \in \mathbb{N}_+$ und $d \in \mathbb{N}_+$ der Teiler von a und b, der sich mit dem Euklidischen Algorithmus aus a und b ergibt. Dann gibt es ganze Zahlen m, n, so dass gilt:*

$$m \cdot a + n \cdot b = d.$$

Beweis (noch einmal).

$$\begin{aligned} d = r_k &= (-q_k) r_{k-1} + 1 \cdot r_{k-2} = r_{k-2} - q_k(-q_{k-1} r_{k-2} + r_{k-3}) \\ &= (1 + q_{k-1} q_k) r_{k-2} + (-q_k) r_{k-3} \\ &\vdots \\ &= m \cdot a + n \cdot b \end{aligned}$$

mit geeigneten $m, n \in \mathbb{Z}$ (deren Existenz wir auch sauber durch vollständige Induktion über k zeigen könnten). □

Das zeigt nun etwas Erstaunliches. Ist nämlich d' ein beliebiger weiterer gemeinsamer Teiler von a und b, so gibt es also $k, l \in \mathbb{N}_+$ mit $d'k = a$ und $d'l = b$. Aber dann ist

$$d'(km + ln) = d'km + d'ln = am + bn = d.$$

Das bedeutet, dass d' nicht nur kleiner (wenn nicht gleich) d ist, sondern d sogar teilt. Wir machen diese Eigenschaft von d zu seiner Definition:

Definition 5.11 Seien $a, b \in \mathbb{N}_+$. Dann nennen wir $d \in \mathbb{N}_+$ den *größten gemeinsamen Teiler von a und b*, und schreiben $d = \mathrm{ggT}(a, b)$, wenn gilt:

(i) $d|a$ und $d|b$,

(ii) ist $d' \in \mathbb{N}_+$ mit $d'|a$ und $d'|b$, so gilt: $d'|d$.

Eine solche Zahl ist sicher eindeutig bestimmt (wenn sie existiert), denn für zwei solche $d_1, d_2 \in \mathbb{N}_+$ wäre ja $d_1|d_2$ und $d_2|d_1$, was nur bei $d_1 = d_2$ geht. Sie existiert auch, wie wir gerade gesehen haben, denn es ist die Zahl, die sich aus a und b mittels des Euklidischen Algorithmus' ergibt.

Um den größten gemeinsamen Teiler zweier natürlicher Zahlen mit Hilfe ihrer Primfaktorzerlegungen auszudrücken, notieren wir für $a \in \mathbb{N}_+$ und $p \in \mathbb{P}$ mit $\nu_p(a) \in \mathbb{N}$ die Anzahl der Faktoren p, die in der Primfaktorzerlegung von a vorkommen. Es ist also dann

$$a = \prod_{p \in \mathbb{P}} p^{\nu_p(a)},$$

wobei das Produkt auf der rechten Seite nur auf den ersten Blick wie ein unendliches Produkt aussieht, denn für *fast alle* (d.h. alle, bis auf endlich viele) $p \in \mathbb{P}$ ist ja $\nu_p(a) = 0$, nämlich all jene $p \in \mathbb{P}$, die gar nicht in der Zerlegung vorkommen. Der entsprechende Faktor in dem unendlichen Produkt ist dann $p^0 = 1$, und man könnte ihn also auch weglassen. Man beachte, dass wegen der Eindeutigkeit der Primfaktoren für zwei Zahlen $a, b \in \mathbb{N}_+$ gilt:

$$a = b \quad \Leftrightarrow \quad \nu_p(a) = \nu_p(b) \ \ \forall p \in \mathbb{P}.$$

Es gilt nun weiter:

Satz 5.12 *Seien $a, b \in \mathbb{N}_+$. Dann gilt:*

$$\mathrm{ggT}(a, b) = \prod_{p \in \mathbb{P}} p^{\min\{\nu_p(a), \nu_p(b)\}}.$$

Ein wichtiger Spezialfall ist der, wenn a und b *teilerfremd* sind, d.h.: $\mathrm{ggT}(a, b) = 1$. Das liegt also genau dann vor, wenn sie keinen gemeinsamen Primfaktor haben,

$$\min\{\nu_p(a), \nu_p(b)\} = 0, \ \ \forall p \in \mathbb{P}.$$

Beweis (von (5.12). Für ein $d \in \mathbb{N}_+$ ist d genau dann ein Teiler von $a \in \mathbb{N}_+$ wenn gilt:

$$\nu_p(d) \leq \nu_p(a), \ \ \forall p \in \mathbb{P}.$$

Das folgt aus den Primfaktorzerlegungen, weil $d \cdot e = a$, für ein $e \in \mathbb{N}_+$, gleichbedeutend ist mit

$$\nu_p(d) + \nu_p(e) = \nu_p(a), \ \ \forall p \in \mathbb{P}.$$

Es folgt, dass

$$d := \prod_{p \in \mathbb{P}} p^{\min\{\nu_p(a), \nu_p(b)\}}$$

Teiler von a und Teiler von b ist, und dass für jeden weiteren Teiler d' von a und b gelten muss: $d' | d$. Also ist $d = \mathrm{ggT}(a, b)$. □

Wir nennen auch $b \in \mathbb{N}_+$ ein *Vielfaches von* $a \in \mathbb{N}_+$, wenn $a | b$. In sehr ähnlicher Weise wie den $\mathrm{ggT}(a, b)$, für $a, b \in \mathbb{N}_+$, definiert man nun:

Definition 5.13 Seien $a, b \in \mathbb{N}_+$. Wir nennen eine Zahl $e \in \mathbb{N}_+$ das *kleinste gemeinsame Vielfache von a und b*, und schreiben $e = \mathrm{kgV}(a, b)$, wenn gilt:

(i) $a|e$ und $b|e$,

(ii) ist $e' \in \mathbb{N}_+$ mit $a|e'$ und $b|e'$, so gilt: $e|e'$.

Auch hier ist zunächst die Existenz einer solchen Zahl nicht unmittelbar klar, wohl aber ihre Eindeutigkeit (wenn sie existiert), denn für zwei solche Zahlen e_1 und e_2 wäre ja $e_1|e_2$ und $e_2|e_1$, was $e_1 = e_2$ bedeutet.

Die Existenz wird nun dadurch gesichert, dass wir, ähnlich wie in (5.12), $\mathrm{kgV}(a, b)$ für $a, b \in \mathbb{N}_+$ als Produkt von Primfaktoren angeben, wenn wir die Primfaktorzerlegungen von a und b kennen.

Satz 5.14 *Seien $a, b \in \mathbb{N}_+$. Dann existiert $\mathrm{kgV}(a, b) \in \mathbb{N}_+$, ist eindeutig bestimmt und ist gegeben durch*

$$\mathrm{kgV}(a, b) = \prod_{p \in \mathbb{P}} p^{\max\{\nu_p(a), \nu_p(b)\}}.$$

Beweis. Eine natürliche Zahl $e \in \mathbb{N}_+$ ist Vielfaches von $a \in \mathbb{N}_+$, genau wenn $\nu_p(e) \geq \nu_p(a)$, für alle $p \in \mathbb{P}$. Es ist $e \in \mathbb{N}_+$ also genau dann Vielfaches von a und b, wenn gilt:

$$\nu_p(e) \geq \max\{\nu_p(a), \nu_p(b)\}, \quad \forall p \in \mathbb{P}.$$

Es folgt, dass

$$e := \prod_{p \in \mathbb{P}} p^{\max\{\nu_p(a), \nu_p(b)\}}$$

tatsächlich kleinstes gemeinsames Vielfaches von a und b im Sinne von Definition (5.13) ist. □

Literatur

[1] M. Aigner und G. Ziegler: *Das BUCH der Beweise.* Springer-Verlag, Berlin, Heidelberg

[2] S. Bosch: *Algebra.* Springer-Verlag, Berlin, Heidelberg, New York

[3] H.-D. Ebbinghaus et al.: *Zahlen.* Springer-Verlag, Berlin, Heidelberg, New York, Tokio

[4] G. Fischer: *Lehrbuch der Algebra.* Friedr. Vieweg & Sohn Verlag, Wiesbaden

[5] O. Forster: *Analysis I.* Friedr. Vieweg Verlag, Braunschweig, Wiesbaden

[6] G. Frey: *Elementare Zahlentheorie.* Friedr. Vieweg & Sohn Verlag, Braunschweig, Wiesbaden

[7] U. Friedrichsdorf und A. Prestel: *Mengenlehre für den Mathematiker.* Friedr. Vieweg & Sohn Verlag, Braunschweig, Wiesbaden

[8] P. Halmos: *Naive Mengenlehre.* Vandenhoeck & Ruprecht Verlag, Göttingen

[9] M. Spivak: *Calculus.* Cambridge University Press, Cambridge

[10] U. Storch und H. Wiebe: *Lehrbuch der Mathematik, Band I: Analysis einer Veränderlichen.* Springer Verlag, Berlin, Heidelberg

Abbildungsverzeichnis

Index