Mathematik und Logik
β-Version

Franz Binder

29. Februar 2008

Inhaltsverzeichnis

Kapitel 1

Logik

Die *mathematische Logik* verwendet mathematische Methoden, um das logische Denken formal zu beschreiben. Neben dem klassischen Zweck, die verschiedenen Möglichkeiten des logischen Schließens zu beschreiben, zu präzisieren und darüber zu reflektieren, dient die mathematische Logik in jüngerer Zeit in zunehmendem Maße vor allem dazu, Wissen aller Fachgebiete exakt darzustellen, abzufragen, zu überprüfen und allgemein einer automatischen Verarbeitung zugänglich zu machen.

1.1 Aussagen und Beweise

Zentrales Objekt der Betrachtung in der Logik sind die Aussagen, zusammen mit Kriterien um festzustellen, ob (und warum) diese wahr oder falsch sind. Es liegt somit nahe, zuerst den Begriff *Aussage* näher zu präzisieren.

Nicht jeder Satz einer natürlichen Sprache kommt als Aussage in Frage. Frage- und Befehlssätze kommen selbstverständlich nicht als Aussagen in Betracht; aber auch viele Sätze, die in einem grammatikalischem Sinne Aussagesätze sind, bezeichnen nicht wirklich Aussagen, etwa: zynische oder aus dem Zusammenhang gerissene Sätze, oder allgemein Sätze, deren Bedeutung nicht klar definiert ist.

Da die Gültigkeit von Aussagen ein zentrales Thema ist, wird als populäre Definition gerne verwendet, daß genau diejenigen Sätze als Aussagen zugelassen sind, die entweder wahr oder falsch sind. Dabei tritt das prinzipielle Problem auf, daß eigentlich zuerst definiert werden sollte, wann ein Satz wahr oder falsch ist (was gerade die Aufgabe der Logik wäre). Diese Definition wird daher oft modifiziert zu: Aussagen sind Sätze, von denen es einen *Sinn* macht zu fragen, ob sie wahr oder falsch sind. Damit ist das Problem auf die Sinnfrage reduziert, welche außerhalb von Mathematik und Linguistik behandelt wird, und wohl für immer ungelöst bleiben wird.

Der klassische Ansatz, um dieses Problem zu lösen ist, eine formale Sprache zu definieren, welche genau die mathematischen Aussagen enthält, und von diesen dann implizit festzulegen, welche wahr und welche falsch sein sollten. Dieser Ansatz war tatsächlich sehr erfolgreich, und der Großteil der Kenntnisse in der mathematischen Logik wurde auf diese Weise gewonnen. Ob dadurch tatsächlich das logische Denken beschrieben wird, ist freilich eine andere Frage. Immerhin

kann man die Ergebnisse auf deren praktische Verwertbarkeit überprüfen.

Es stellt sich heraus, daß es für die Praxis gar nicht so wichtig ist, festzulegen, was Aussagen tatsächlich sind, und auch nicht, ob sie wahr oder falsch sind. Ein klassisches Resultat der Logik besagt sogar, daß dies sowieso unmöglich ist. Aus praktischer Sicht viel wichtiger ist vielmehr, ob man die Gültigkeit einer Aussage feststellen (*beweisen*) kann, oder nicht. Es geht also primär darum, welche Beziehungen zwischen einer Aussage und möglichen Beweisen bestehen. Wir präzisieren die beiden Begriffe daher vorerst nur durch eine einfache, sich in der Folge aber als sehr wirkungsvoll herausstellende nicht-formale Annahme.

1.1.1 Definition (Aussage). Eine *Aussage*(*proposition*) ist eine Konstruktion, durch welche festgelegt wird, wie ihre *Beweise*(*proof*) zu konstruieren sind. Eine Aussage, für welche ein Beweis konstruiert werden kann, heißt *wahr*(*true*).

Diese Definition vermeidet bewußt jede unnötige nähere Festlegung von Begriffen wie *Aussage*, *Beweis*, *wahr*, *falsch Sinn*, *Bedeutung*. Auch, ob es, neben den beweisbaren Aussagen, noch weitere wahre Aussagen gibt, wird durch diese Definition bewußt nicht festgelegt. Jedenfalls gilt eine Aussage aber nicht unbedingt als falsch, wenn sie nicht bewiesen werden kann.

Der hier verwendete Begriff des Beweises ist freilich noch näher festzulegen. Die allgemeine Vorgangsweise dazu ist die, daß verschiedene logische Operatoren definiert werden, mit der mehrere Aussagen verknüpft werden, um eine neue Aussage zu erhalten. Von jedem Junktor muß dann nur noch festgelegt werden, wie die Beweise der betroffenen Aussagen mit denen der neuen Aussage in Beziehung stehen.

1.1.2 Beispiel.

1. Aussagen wie $5 > 3$ sind trivial bzw. leicht nachzurechnen. Wie deren Beweise aussehen, hängt hauptsächlich davon ab, was man unter natürlichen Zahlen und dem Größerbegriff genau versteht, und ist für die Praxis irrelevant. Wichtig ist hier nur, ob es einen Beweis gibt oder nicht.

2. Die Aussage $3 > 5$ hat keinen Beweis.

3. Jede natürliche Zahl ist eine Beweis, dass es eine natürliche Zahl gibt. Daher kann man die Mengen der natürlichen Zahlen mit der Gesamtheit aller Beweise von „es gibt eine natürliche Zahl“ identifizieren.

4. Die Aussage „Zu jeder natürlichen Zahl gibt es eine größere“ erfordert als Beweis unter anderem eine Konstruktion, welche imstande ist, zu jeder beliebigen natürlichen eine größere zu konstruieren. Die „Funktion“ $(+1)$, also $x \mapsto x+1$, kann das. Ebenso $(+2)$, $(+100)$, $(2*)$, $(2\hat{})$, und viele mehr.

1.2 Aussagenlogik

Implikation, ($\Rightarrow$)

Formation

1.2.1 Definition. Sind P und Q Aussagen, dann bezeichnet $P \Rightarrow Q$ ebenfalls eine Aussage, die *Implikation* von P und Q.

Introduktion

1.2.2 Definition. Um $P \Rightarrow Q$ zu beweisen, muß man Q beweisen, wobei man einen Beweis von P voraussetzen darf.

Diese Definition wird formal durch die folgende Regel ausgedrückt:

1.2.3 Axiom ($\Rightarrow$-Einführungsregel).

$$\frac{\begin{array}{|c|}\hline P \\ \vdots \\ Q \\ \hline\end{array}}{P \Rightarrow Q} \Rightarrow\mathcal{I}$$

Dabei bezeichnet der eingerahmte Teil einen *Beweis-Rahmen*(*proof box*). Dieser bedeutet, daß man, mit Hilfe der oberhalb des Rahmens bezeichneten Annahme (*Hypothese*) P, die Aussage Q am unteren Rand des Rahmens herleiten kann.

In der Praxis werden Beweise verbal wiedergegeben. Die obige Regel liest sich dann etwa folgendermaßen:

Wir möchten $P \Rightarrow Q$ beweisen. Dazu nehmen wir an, es gelte P. Daraus leiten wird nun Q her.

Elimination

Wurde $P \Rightarrow Q$ mittels der Introduktionsregel bewiesen, so gibt es einen Beweis von Q, welcher unter der Annahme P funktioniert. Steht nun zusätzlich ein Beweis von P zur Verfügung, so hat man einen Beweis von Q:

1.2.4 Axiom ($\Rightarrow$-Eliminationsregel, *Modus Ponens*).

$$\frac{P \Rightarrow Q \qquad P}{Q} \Rightarrow\mathcal{E}$$

In verbalen Beweisen spiegelt sich die Eliminationsregel etwa folgendermaßen wieder:

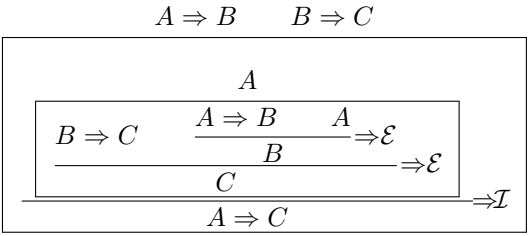
Da wir P bewiesen haben, gilt wegen $P \Rightarrow Q$ auch Q.

Transitivität

1.2.5 Satz. *Die Implikation ist transitiv, d.h. wir haben die hergeleitete Schlußregel*

$$\frac{A \Rightarrow B \qquad B \Rightarrow C}{A \Rightarrow C} \Rightarrow\text{-}\mathit{Transitivität}$$

Beweis.



Oder verbal:

Annahmen: $A \Rightarrow B$ und $B \Rightarrow C$. Zu zeigen haben wir $A \Rightarrow C$. Gemäß $\Rightarrow\mathcal{I}$ nehmen die zusätzliche Hypothese A an und versuchen damit C herzuleiten. Zweimaliges Anwenden von $\Rightarrow\mathcal{E}$ liefert aus A und $A \Rightarrow B$ zuerst B, und daraus dann, zusammen mit $B \Rightarrow C$, die Behauptung C. □

Konjunktion (Logisches Und, $\wedge$, &)

Formation

1.2.6 Definition. Seien P und Q Aussagen. Dann bezeichnet $P \wedge Q$ ebenfalls eine Aussage, die *Konjunktion* von P und Q.

Introduktion

1.2.7 Definition. Um die Aussage $P \wedge Q$ zu beweisen, muß man sowohl P als auch Q beweisen.

Diese Definition läßt sich mit der folgenden *Einführungsregel*(*introduction rule*) formalisieren:

1.2.8 Axiom ($\wedge$-Einführungsregel).

$$\frac{P \qquad Q}{P \wedge Q} \wedge\mathcal{I} \tag{1.1}$$

Diese Regel ermöglicht es, eine Konjunktion aus anderen Aussagen herzuleiten, also in die Gesamtheit der bewiesenen Aussagen einzuführen. Bei dieser Notation steht die neu eingeführte Aussage unter dem Strich, und die Aussagen, welche vorher zu beweisen sind, stehen über dem Strich.

Durch die Regel $\wedge\mathcal{I}$ wird das umgangssprachliche Und, so wie es in der Mathematik gebraucht wird, vollständig beschrieben, d.h. es gibt keine weiteren Einführungsregeln für $\wedge$. Man beachte aber, daß das Wort „und“, vor allem außerhalb der Mathematik, oft auch in anderen Bedeutungen gebraucht wird, etwa „und daher“, „und trotzdem“, „und dann“, „aber“, usw.

Elimination

Die Bemerkung, daß es nur die eine Einführungsregel für die Konjunktion gibt, läßt sich nicht direkt in Form einer Regel anschreiben. Wir können diese Tatsache aber auch folgendermaßen verstehen: Wurde $P \wedge Q$ bewiesen, dann hat man auch einen Beweis von P und einen von Q, da ja ein Beweis von $P \wedge Q$

gerade aus diesen zwei Beweisen besteht. Auf diese Art läßt sich das Wissen um einen Beweis von $P \wedge Q$ verwerten. Für jede der Prämissen der Einführungsregel erhalten wir eine dazgehörige *Eliminationsregel* (*elimination rule*).

1.2.9 Axiom ($\wedge$-Eliminationsregeln).

$$\frac{P \wedge Q}{P} \wedge\mathcal{E}_0 \quad \frac{P \wedge Q}{Q} \wedge\mathcal{E}_1 \tag{1.2}$$

Die Eliminationsregeln drücken aus, daß wenn $P \wedge Q$ bewiesen worden ist, dass dann irgendwann einmal auch P und Q bewiesen worden sind, und daher auch deren Beweise zur Verfügung stehen müssen . Bei den Eliminationsregeln steht der betroffene Operator stets oberhalb des Striches.

Kommutativität

1.2.10 Satz. *Falls $A \wedge B$ bewiesen werden kann, dann auch $B \wedge A$.*

Beweis. Der formale Beweis

$$\dfrac{A \wedge B}{\boxed{\dfrac{\dfrac{A \wedge B}{B} \wedge\mathcal{E}_1 \qquad \dfrac{A \wedge B}{A} \wedge\mathcal{E}_0}{B \wedge A} \wedge\mathcal{I}}}$$

kann folgendermaßen gelesen werden (in (sehr) verbaler Ausführung): Die Annahme besagt $A \wedge B$. Zu zeigen ist $B \wedge A$. Aus der Annahme erhalten wir mit $\wedge\mathcal{E}_1$, daß B gilt, und ebenso mit $\wedge\mathcal{E}_0$, daß A gilt. Damit erhält man mit $\wedge\mathcal{I}$ schließlich $B \wedge A$, wie zu zeigen war. □

Analog gestaltet sich der Beweis für die *Assoziativität* der Konjunktion:

$$\frac{(A \wedge B) \wedge C}{A \wedge (B \wedge C)} \wedge\text{-associative}_0 \qquad \frac{A \wedge (B \wedge C)}{(A \wedge B) \wedge C} \wedge\text{-associative}_1$$

Logische Äquivalenz

1.2.11 Definition. Die logische *Äquivalenz* $P \Leftrightarrow Q$ ist definiert als Abkürzung für $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$.

1.2.12 Notation. Statt $P \Rightarrow Q$ kann man auch gleichbedeutend $Q \Leftarrow P$ schreiben.

1.2.13 Theorem. *Die logische Äquivalenz ist eine Äquivalenzrelation, d.h. es gelten*

Reflexivität *$A \Leftrightarrow A$;*

Symmetrie *Wenn $A \Leftrightarrow B$, dann auch $B \Leftrightarrow A$;*

Transitivität *Wenn $A \Leftrightarrow B$ und $B \Leftrightarrow C$, dann auch $A \Leftrightarrow C$.*

Disjunktion, Logisches Oder

Formation

1.2.14 Definition. Seien P und Q Aussagen. Dann bezeichnet $P \vee Q$ ebenfalls eine Aussage, die *Disjunktion* von P und Q.

Introduktion

1.2.15 Definition. Um die *Disjunktion* $A \vee B$ zu beweisen, ist P zu beweisen oder auch Q

Auch diese Definition läßt sich in Form von logischen Schlußregeln formulieren:

1.2.16 Axiom ($\vee$-Einführungsregeln).

$$\frac{P}{P \vee Q}\vee\mathcal{I}_0 \qquad \frac{Q}{P \vee Q}\vee\mathcal{I}_1$$

Hier erhalten wir zwei Einführungsregeln, weil es zwei Möglichkeiten gibt, eine Disjunktion zu beweisen.

Elimination

Jede der Einführungsregeln hat nur eine Prämisse, daher gibt es auch nur eine Eliminationsregel. Sie ist aber etwas komplizierter, weil zwei Fälle unterschieden werden müssen.

1.2.17 Axiom (Beweis durch Fallunterscheidung).

$$\frac{P \implies RQ \implies R}{P \vee Q \implies R}\vee\mathcal{E}$$

Distributivität

Wir zeigen eines der beiden Distributivgesetze:

1.2.18 Satz.

$$A \wedge (B \vee C) \Leftrightarrow (A \wedge B) \vee (A \wedge C).$$

Beweis. Diese Aussage ist eine Konjunktion (Äquivalenz ist ja nur die Abkürzung für die Konkjuntion der Implikationen in beide Richtungen). Wir haben also beide Richtungen zu zeigen.

$\implies$: Es gelte $A \wedge (B \vee C)$. Zu zeigen ist $(A \wedge B) \vee (A \wedge C)$. Mittels $\wedge\mathcal{E}$ erhalten wir $B \vee C$. Für jeden dieser Fälle zeigen wir $(A \wedge B) \vee (B \wedge C)$:

Fall: es gelte B. Wir erhalten A aus der Voraussetzung mittels $\wedge\mathcal{E}_0$, und daher mittels $\wedge\mathcal{I}$ auch $A \wedge B$, was mit $\vee\mathcal{I}_0$ zu $(A \wedge B) \vee (B \wedge C)$ wird.

Fall: es gelte C: Analog zum vorigen Fall, erhalten wir ebenfalls $(A \wedge B) \vee (B \wedge C)$.

$$\dfrac{\dfrac{A \wedge (B \vee C)}{B \vee C}\wedge\mathcal{E} \qquad \boxed{\dfrac{\dfrac{\dfrac{A \wedge (B \vee C)}{A}\wedge\mathcal{E} \qquad B}{A \wedge B}\wedge\mathcal{I}}{(A \wedge B) \vee (A \wedge C)}\vee\mathcal{I}} \qquad \boxed{\dfrac{\dfrac{\dfrac{A \wedge (B \vee C)}{A}\wedge\mathcal{E} \qquad C}{A \wedge C}\wedge\mathcal{I}}{(A \wedge B) \vee (A \wedge C)}\vee\mathcal{I}}}{(A \wedge B) \vee (A \wedge C)}\vee\mathcal{E}$$

$\Longleftarrow$: Wir nehmen wegen $(\Rightarrow\mathcal{I})$ an, daß $(A \wedge B) \vee (A \wedge C)$ und versuchen sowohl A als auch $B \vee C$ herzuleiten $(\wedge\mathcal{I})$. Gemäß $(\vee\mathcal{E})$ betrachten wir zuerst den Fall $(A \wedge B)$. Dann gilt sowohl A (mit $\wedge\mathcal{E}0$) als auch B (mit $\wedge\mathcal{E}1$), und damit $B \vee C$ (mit $\vee\mathcal{I}0$). Analog für den Fall $A \wedge C$. □

Die Umkehrung haben wir hier verbal, aber nicht streng formal gemacht; die andere Variante ist eine gute ***Übung!***

1.3 Konstruktionen/Beweise

Objekte unserer Betrachtung seien nun nicht mehr die Aussagen, sondern deren Beweise. Jeder Aussage entspricht eine gewisse Klasse von Beweisen, nämlich die Beweise dieser Aussage.

Bisher haben wir uns nur damit beschäftigt, ob eine Aussage einen Beweis hat oder nicht. Nun sehen wir uns die Beweise selbst etwas genauer an.

1.3.1 Notation. Ist x ein Beweis der Aussage A, dann schreiben wir $x\colon A$.

Umgekehrt kann man jede Zusammenfassung von Objekten mit einer Aussage identifizieren. Damit kann $x\colon A$ auch bedeuten, daß x den Datentyp A hat, was oft mit $x \in A$ bezeichnet wird.

Implikation

Konstruktor

Wir annotieren die Einführungsregel für die Implikation mit den konkreten Beweisen. Der Beweis von A ist rein hypothetisch und wird mit einer Variablen, die sonst noch keine Bedeutung hat, bezeichnet. Die Notation $t[x]$ bedeutet dann, daß t irgend eine Konstruktion ist, in der das x vorkommen kann.

1.3.2 Axiom (Funktionsdefinition).

$$\frac{\begin{array}{c} x\colon A \\ \boxed{\begin{array}{c}\vdots \\ t[x]\colon B\end{array}} \end{array}}{x \mapsto t[x]\colon A \Rightarrow B}\Rightarrow\mathcal{I}$$

Ein Beweis von $A \Rightarrow B$ ist also eine „Funktion“ von A nach B und die Einführungsregel entspricht der üblichen Art, wie man Funktionen definiert. Eine klassische Variante dieser Schreibweise ist $\lambda x\,.\,t[x]$ (daher λ-*Kalkül*). Statt einer Definition $f = x \mapsto x^2$ schreibt man üblicherweise $f(x) = x^2$. Dabei sind oft auch Muster erlaubt, etwa in

$$f(x, y) = (y, x).$$

Dies ist insbesondere im Zusammenhang mit Konstruktoren sinnvoll.

Selektor

1.3.3 Axiom (Funktionsanwendung).

$$\frac{f\colon A \Rightarrow B \qquad x\colon A}{fx\colon B}\Rightarrow\mathcal{E}$$

Einem Beweis mit der Eliminationsregel entspricht also Anwendung einer Funktion. Oft scheibt man statt fx auch $f(x)$. Für spezielle Funktionen sind auch andere Konventionen üblich, etwa die Formen xf oder x^f (vgl.: $5!$, $\sin'$).

Nun annotieren wir auch den Beweis der Transitivität:

$$\dfrac{\boxed{\dfrac{g\colon B \Rightarrow C \qquad \dfrac{f\colon A \Rightarrow B \qquad x\colon A}{fx\colon B}\Rightarrow\mathcal{E}}{g(fx)\colon C}\Rightarrow\mathcal{E}}^{A}}{x \mapsto g(fx)\colon A \Rightarrow C}\Rightarrow\mathcal{I}$$

Der Beweis der Transitivität der Implikation entspricht also der Hintereinanderausführung von Funktionen. Für die Konstruktion $x \mapsto g(fx)$ schreiben wir $g \circ f$.

Konjunktion

Konstruktor

Wir annotieren die Einführungsregel der Konkunktion mit den entsprechenden Beweisen:

1.3.4 Axiom (Paar-Konstruktor).

$$\frac{x\colon A \qquad y\colon B}{(x,y)\colon A \wedge B}\wedge\mathcal{I}$$

Die Konstruktion $(,)$ zum Paare bilden macht aus einem Beweis x von A und einem Beweis y von B den Beweis (x,y) von $A \wedge B$. Man spricht daher von einem *Konstruktor*.

Die Aussage $A \wedge B$ entspricht daher einem Datentyp, dessen Objekte aus zwei Komponenten bestehen, einem vom Typ A, und einem vom Typ B. Die meisten höheren Programmiersprachen bieten die Möglichkeiten, neue Datentypen auf diese Weise zu definieren. Dabei können natürlich auch mehr als zwei Objekte zu einem zusammengefaßt werden. Die Details, wie das gemacht wird, sind recht unterschiedlich, ebenso wie die Namen (Verbund-Datentypen, Records, Strukturen, Datensätze). In Java verwendet man Klassen. Einem Konstruktor entspricht in Java eine Konstruktor-Methode, mit der ein neues Objekt der Klasse angelegt wird.

Selektoren

Wir annotieren nun auch die beiden Eliminationsregeln mit den entsprechenden Beweisen.

$$\frac{(x,y)\colon A \wedge B}{x\colon A}\wedge\mathcal{E}_0 \qquad \frac{(x,y)\colon A \wedge B}{y\colon B}\wedge\mathcal{E}_1$$

Hier haben wir die Konstruktion des Beweises von $A \wedge B$ explizit angeschrieben (ein Beweis von $A \wedge B$ muß ja so aussehen). Die Konstruktion eines Beweises von A bzw. B besteht dann im Selektieren der passenden Komponente. Eine alternative Notation ist die folgende:

$$\frac{z\colon A \wedge B}{\mathrm{fst}\, z}\wedge\mathcal{E}_0 \qquad \frac{z\colon A \wedge B}{\mathrm{snd}\, z}\wedge\mathcal{E}_1$$

Hier war es nicht notwendig, die Konstruktion des Beweises der Prämisse anzuschreiben, dafür wurden die *Selektoren* explizit bezeichnet.

Auch in Java gibt es in einer Klasse üblicherweise Selektor-Methoden, mit denen auf die Komponenten der Objekte einer Klasse zugegriffen werden kann.

Kommutativität mit Beweisannotation:

$$\boxed{\dfrac{\dfrac{(x,y)\colon A\wedge B}{y\colon B}\wedge\mathcal{E}_1 \qquad \dfrac{(x,y)\colon A\wedge B}{x\colon A}\wedge\mathcal{E}_0}{(y,x)\colon B\wedge A}\wedge\mathcal{I}}$$

Damit haben wir die Regel für die *Kommutativität* der Konjunktion

$$\frac{A\wedge B}{B\wedge A}\wedge\text{-Kommutativität},$$

eine *hergeleitete Schlußregel* (*derived rule*) (kein Axiom!) hergeleitet; und ein Beweis davon ist die Konstruktion, welche die beiden Komponenten eines Paares vertauscht.

Disjunktion

Konstruktoren

1.3.5 Axiom.

$$\frac{x\colon A}{\mathrm{Left}\,x\colon A\vee B}\vee\mathcal{I}_0 \qquad \frac{y\colon B}{\mathrm{Right}\,y\colon A\vee B}\vee\mathcal{I}_1$$

Ein Beweis von $A\vee B$ ist also entweder ein Beweis von A oder ein Beweis von B; der verwendete Konstruktor gibt an, welcher dieser beiden Fälle eintritt. Es handelt sich also um die „Vereinigung“ der Beweise von A und der Beweise von B. Es liegt somit eine Vereinigung von zwei Datentypen vor.

Selektor

Wir erhalten damit die folgenden ergänzten Schlußregeln

1.3.6 Axiom (Fallunterscheidung, (if, switch, case)).

$$\frac{z\colon A\vee B \qquad \begin{array}{c}x\colon A\\ \boxed{\begin{array}{c}\vdots\\ a[x]\colon C\end{array}}\end{array} \qquad \begin{array}{c}y\colon B\\ \boxed{\begin{array}{c}\vdots\\ b[y]\colon C\end{array}}\end{array}}{\mathrm{switch}\,z\ \begin{array}{l}\mathrm{case\,Left}\,x\mapsto a[x]\\ \mathrm{case\,Right}\,y\mapsto b[y]\end{array}\ :C}\vee\mathcal{E}$$

Dabei sind wieder $a[x]$ und $b[y]$ irgendwelche Terme, in denen x bzw. y vorkommen können.

1.4 Prädikatenlogik

Um eine Aussage der Form $\bigwedge_{x\colon X} A[x]$ zu beweisen, muss man $A[x]$ für ein beliebiges $x\colon X$ zeigen. Die entsprechende Regeln lautet hiermit:

1.4.1 Axiom.

$$\dfrac{\begin{array}{c}x\colon X\\ \boxed{\begin{array}{c}\vdots\\ A[x]\end{array}}\end{array}}{\bigwedge\limits_{x\colon X} A[x]}\forall\mathcal{E}$$

Eine alternative Schreibweise ist $\forall(x\colon X).A[x]$.

1.4.2 Bemerkung. Die Implikation ergibt sich als Spezialfall einer All-Aussage, in der $A[x]$ nicht von x abhängt.

Elimination

Wenn $A[x]$ für ein beliebiges x bewiesen wurde, dann kann man auch für die Variable x einen beliebigen Ausdruck t einsetzen. Die Notation $A[t]$ bezeichnet jene Aussage, die aus $A[x]$ entsteht, wenn alle Vorkommen der Variablen x durch t ersetzt werden.

1.4.3 Axiom.

$$\dfrac{\bigwedge\limits_{x\colon X} A[x] \qquad t\colon X}{A[t]}\forall\mathcal{E}$$

Konstruktor

Die Natur der Beweise einer All-Aussage ergibt sich wieder durch Annotation der Einführungsregel:

1.4.4 Axiom.

$$\dfrac{\begin{array}{c}x\colon X\\ \boxed{\begin{array}{c}\vdots\\ t[x]\colon A[x]\end{array}}\end{array}}{x\mapsto t[x]\ \bigwedge\limits_{x\colon X} A[x]}\forall\mathcal{E}$$

Damit ergibt sich prinzipiell derselbe Konstruktor wie bei der Implikation. Allerdings haben hier die Funktionswerte nicht alle denselben Datentyp, sondern den Typ $A[x]$, welcher von x abhängen kann.

Selektor

1.4.5 Axiom.

$$\frac{f\colon \bigwedge_{x\colon X} A[x] \qquad t\colon X}{ft\colon A[t]}\forall\mathcal{E}$$

Auch hier handelt es sich um eine Verallgemeinerung der üblichen Funktionsanwendung, bei der die Datentypen des Ergebnisses vom Input abhängen können.

Existenz-Quantor

Formation

$A[x]$ bezeichne eine Ausage, in er x vorkommen kann.

1.4.6 Definition. Ist $A[x]$ für jedes $x\colon X$ eine Aussage, dann ist auch $\bigvee_{x\colon X} A[x]$ eine *Existenz-Aussage*.

Eine alternative Schreibweise ist $\exists(x\colon X).A[x]$.

Introduktion

1.4.7 Definition. Um eine Aussage der Form $\bigvee_{x\colon X} A[x]$ zu beweisen, muss man $a\colon X$ konstruieren, sodaß man $A[a]$ beweisen kann.

1.4.8 Axiom ($\bigvee$-Einführungsregel).

$$\frac{a\colon X \qquad A[a]}{\bigvee_{x\colon X} A[x]}\exists\mathcal{I} \tag{1.3}$$

1.4.9 Bemerkung. Die Konjunktion ergibt sich als Spezialfall einer Existenz-Aussage, in der $A[x]$ nicht von x abhängt.

Elimination

1.4.10 Axiom.

$$\frac{\bigvee_{x\colon X} A[x] \qquad \boxed{\begin{array}{c} y\colon X \qquad A[y] \\ \vdots \\ C \end{array}}}{C}\exists\mathcal{E}$$

Diese Eliminationsregel entspricht der Einführung von „einem solchen x sodaß $A[x]$ gilt“. Genauer: es gelte $\bigvee_{x\colon X} A[x]$ und die Aussage C sei zu beweisen; dann darf man ein neues Symbol y einführen und annehmen, daß $A[y]$ gilt. Dieses neue Symbol darf in C nicht vorkommen (sonst wäre es nicht neu), es darf aber mitunter mit dem x übereinstimmen, da das x innerhalb der Existenzaussage nur innerhalb dieser eine Bedeutung hat (es ist durch den Existenzquantor gebunden).

Konstruktor

1.4.11 Axiom ($\bigvee$-Einführungsregel).

$$\frac{a\colon X \qquad t\colon A[a]}{(a,t)\colon \bigvee\limits_{x\colon X} A[x]}\exists\mathcal{I} \tag{1.4}$$

Der Beweis einer Existenzaussage ist somit ein Paar; allerdings darf der Datentyp der zweiten Komponente von der ersten Komponente abhängen.

Selektoren

1.4.12 Axiom.

$$\begin{array}{l} \\ (a,t)\colon \bigvee\limits_{x\colon X} A[x] \end{array} \quad \frac{\begin{array}{c} y\colon X \qquad u\colon A[y] \\ \boxed{\begin{array}{c} \vdots \\ f[y,a]\colon C \end{array}} \end{array}}{\{\,(y,u) := (a,t)C;\ f[y,u]\colon C\,\}}\exists\mathcal{E}$$

Der Selektor ist daher eine lokale Definition.

1.5 Boolsche Logik

Negation

Sind P und Q Aussagen, die bewiesen wurden, dann gilt jedenfalls $P \iff Q$, d.h. alle bewiesenen Aussagen sind äquivalent. Mit dem Symbol $\top$ (sprich: top) möchten wir irgendeine solche Aussage bezeichnen (andere Bezeichnungsweisen: T, W, true, wahre Aussage). Umgekehrt soll $\bot$ (sprich: bottom) einen Widerspruch (falsche Aussage) bezeichnen (andere Bezeichnungsweisen: F, false, falsum, $\lightning$). Wir können diese Begriffe wie gewohnt mit Schlußregeln definieren.

1.5.1 Axiom. *Sei P irgendeine Aussage. Dann gelten die Schlußregeln:*

$$\frac{P}{\top}\top\mathcal{I} \qquad \frac{\bot}{P}\bot\mathcal{E}$$

Es gibt aber keine $\top\mathcal{E}$- oder $\bot\mathcal{I}$-Regeln. Für $\top$ ist alles ein Beweis, dafür kann man aus $\top$ aber auch nichts schließen. Umgekehrt kann man aus $\bot$ alles herleiten, aber es gibt keine Möglichkeit, $\bot$ aus anderen Aussagen zu beweisen.

Der Sinn des Symbols $\bot$ besteht darin, die Negation von Aussagen zu definieren. Die Idee ist: wenn einer Aussage jede andere gefolgert werden kann, dann muß sie wohl falsch sein. Die Aussage „Jede Aussage ist wahr“ wäre eine derartige Aussage; aber auch aus etwas konkretem, wie etwa $0 = 1$, kann man jede Aussage herleiten. Ganz allgemein nennt man ein logisches System widersprüchlich, wenn darin jede Aussage herleitbar ist.

1.5.2 Definition. Die Formel $\neg A$ (nicht A) ist eine Abkürzung für $A \Rightarrow \bot$ und bedeutet, daß A falsch ist.

1.5.3 Satz. *Die aussagenlogische Formel*

$$A \wedge \neg A \implies \bot$$

ist allgemeingültig.

Beweis. Direkt aus der Definition durch einfache Anwendung der Implikations-Elimination. □

1.5.4 Satz. *Die Aussage*

$$A \implies \neg\neg A$$

ist allgemeingültig.

Beweis. Dies ist ein Spezialfall von $A \implies (A \implies C) \implies C$. □

1.5.5 Bemerkung. Bemerkenswerterweise läßt sich die Allgemeingültigkeit von $\neg\neg A \implies A$ mit den bisher besprochenen Regeln nicht herleiten.

1.5.6 Satz. *Die Aussage*

$$\neg\neg\neg A \implies \neg A$$

ist allgemeingültig.

Beweis. Dies ist ein Spezialfall von $(((A \implies C) \implies C) \implies C) \implies A \implies C$. □

1.5.7 Bemerkung. Dieser Satz besagt insbesondere, daß die Umkehrung von 1.5.4 zumindest für alle Aussagen der Form $\neg A$ gilt.

1.5.8 Satz. *Die aussagenlogische Formel*

$$\neg(A \vee B) \iff \neg A \wedge \neg B$$

ist allgemeingültig.

Beweis. Dies ist ein Spezialfall von $(A \vee B \implies C) \iff (A \implies C) \wedge (B \implies C)$. □

1.5.9 Folgerung. *Die Aussage*

$$\neg\neg(A \vee \neg A)$$

ist allgemeingültig.

Beweis. In 1.5.8 setzen wir statt $\neg A$ für B ein. Den Rest erledigt 1.5.3. □

Entscheidbare Aussagen

Die doppelte Negeation in 1.5.9 ist etwas unbefriedigend. Sicherlich interessant sind Aussagen, für welche dies vermieden werden kann.

1.5.10 Definition. Eine Aussage A heißt *entscheidbar*, wenn $A \vee \neg A$ bewiesen werden kann.

1.5.11 Bemerkung. Man beachte, daß der Entscheidbarkeitsbegriff im Kontext der klassischen Logik anders definiert wird.

1.5.12 Satz. *Die Aussage*

$$A \vee \neg A \implies (\neg\neg A \implies A).$$

ist allgemeingültig.

Beweis. Gemäß zweimaliger Implikations-Introduktion nehmen wir die Aussagen $A \vee \neg A$ und $\neg\neg A$ an und haben A herzuleiten. Wir führen eine Fallunterscheidung ($\vee\mathcal{E}$) nach $A \vee \neg A$ durch. Im ersten Fall ist nichts mehr zu zeigen. Gilt aber $\neg A$, so können wir, zusammen mit $\neg\neg A$, mittels ($\Rightarrow\mathcal{E}$) den Widerspruch $\perp$ herleiten, woraus mit $\perp\mathcal{E}$ (unter anderem) A folgt. □

1.5.13 Bemerkung. Die Umkehrung von 1.5.4 gilt somit auch für alle entscheidbaren Aussagen.

1.5.14 Bemerkung. Entscheidbare Aussagen erfüllen somit in jedem Fall $A \vee \neg A$ und $\neg\neg A \iff A$, sowie, leicht anders formuliert:

$$(A \iff \top) \vee (A \iff \perp).$$

Das macht deutlich, daß die Menge der entscheidbaren Aussagen mit der logischen Äquivalenz als Gleichheitsbegriff nur zwei Elemente hat; man schreibt dafür

$$\mathbb{B} = \{\top, \perp\}.$$

Wahrheitstafeln

Die Gültigkeit von aus entscheidbaren Aussagen zusammengesetzten Formeln läßt sich stets durch Fallunterscheidungen überprüfen. Diese wird üblicherweise in Tabellenform vorgenommen, wie im folgenden Satz.

1.5.15 Satz. *Für alle entscheidbaren Aussagen A, B gilt*

$$\neg(A \wedge B) \Leftrightarrow (\neg A \vee \neg B).$$

Beweis. Wir untersuchen tabellarisch alle möglichen Wahrheitswerte für A und B:

A	B	$A \wedge B$	$\neg(A \wedge B)$	$\neg A$	$\neg B$	$\neg A \vee \neg B$	$\neg(A \wedge B) \Leftrightarrow (\neg A \vee \neg B)$
$\top$	$\top$	$\top$	$\bot$	$\bot$	$\bot$	$\bot$	$\top$
$\top$	$\bot$	$\bot$	$\top$	$\bot$	$\top$	$\top$	$\top$
$\bot$	$\top$	$\bot$	$\top$	$\top$	$\bot$	$\top$	$\top$
$\bot$	$\bot$	$\bot$	$\top$	$\top$	$\top$	$\top$	$\top$

Da in der letzten Spalte lauter $\top$ stehen, ist die Aussage $\neg(A \wedge B) \Leftrightarrow (\neg A \vee \neg B)$ somit für alle entscheidbaren Aussagen gültig. □

1.5.16 Bemerkung. Satz 1.5.15 ist dual zu Satz 1.5.8, der für allgemeine Aussagen bewiesen wurde. Ohne die Zusatzvoraussetzung gilt im Satz 1.5.15 nur die Implikation von rechts nach links.

1.5.17 Bemerkung. Man beachte, daß die Methode mit den Wahrheitstafeln ganz entscheidend davon abhängt, daß alle verwendeten Aussagen entscheidbar sind.

Gleichungen der Boolschen Algebra

1.5.18 Theorem. *Die Operationen $\wedge, \vee, \neg$ auf $\mathbb{B}$ erfüllen für entscheidbare Aussagen a, b, c die folgenden Gleichungen:*

$$(a \wedge b) \wedge c = a \wedge (b \wedge c) \qquad (a \vee b) \vee c = a \vee (b \vee c)$$
$$a \wedge b = b \wedge a \qquad a \vee b = b \vee a$$
$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c) \qquad a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$$
$$a \wedge \neg a = \bot \qquad a \vee \neg a = \top$$
$$\neg(a \wedge b) = \neg a \vee \neg b \qquad \neg(a \vee b) = \neg a \wedge \neg b$$

1.5.19 Definition. Jede Menge mit Operationen, welche diese Gleichungen erfüllen, heißt eine *Boolsche Algebra*.

Disjunktive Normalform

In diesem Abschnitt beschäftigen wir uns nur mit entscheidbaren Aussagen, bzw. gelte das Prinzip von ausgeschlossenen Dritten.

Eine Aussage wie $\neg(A \Rightarrow (B \wedge C))$ hat viele äquivalente Formen. Mitunter ist es sinnvoll eine *kanonische* Form zu berechnen. „Ausmultiplizieren“ ist eine

Möglichkeit. Dabei wird zuerst die jede Implikation der Form $A \Rightarrow B$ durch $\neg A \vee B$ ersetzt. Dann werden die Gesetze von De Morgan

$$\neg(A \wedge B) \Leftrightarrow \neg A \vee \neg B$$
$$\neg(A \vee B) \Leftrightarrow \neg A \wedge \neg B$$

immer von links nach rechts angewendet, sodaß alle Negationen immer weiter nach innen wandern, bis sie nur noch unmittelbar vor elementaren Aussagen vorkommen. Doppelte Negationen werden weggelassen. Ebenso werden die Distributivgesetze

$$A \wedge (B \vee C) \Leftrightarrow (A \wedge B) \vee (A \wedge C)$$
$$A \vee (B \wedge C) \Leftrightarrow (A \vee B) \wedge (A \vee C)$$

so angewendet, daß die Konjunktionen immer weiter nach innen wandern. Am Schluß ergibt sich damit eine Disjunktion von Konjunktionen von elementaren Aussagen oder deren Negation. Beispiel: $\neg(A \Rightarrow (B \wedge C)) = \neg(\neg A \vee (B \wedge C)) = \neg\neg A \wedge \neg(B \wedge C) = A \wedge (\neg B \vee \neg C) = (A \wedge \neg B) \vee (A \wedge \neg C)$. Dies ist allerdings noch nicht die disjunktive Normalform. Für diese ist zusätzlich gefordert, daß in jeder der Konjunktionen jede in der Gesamtaussage auftretende elementare Aussage vorkommt (negiert oder nicht). Dies erreicht man mit der allgemeingültigen Regel $A \Leftrightarrow (A \wedge B) \vee (A \wedge \neg B)$. Im Beispiel rechnen wir daher weiter: $= (A \wedge \neg B \wedge C) \vee (A \wedge \neg B \wedge \neg C) \vee (A \wedge B \wedge \neg C) \vee (A \wedge \neg B \wedge \neg C) = (A \wedge \neg B \wedge C) \vee (A \wedge \neg B \wedge \neg C) \vee (A \wedge B \wedge \neg C)$. Diese logische Formel ist jetzt eine disjunktive Normalform.

1.5.20 Definition. Eine logische Formel ist in *disjunktiver Normalform*, wenn sie eine Disjunktion von Konjunktionen aller darin vorkommenden elementaren Aussagen oder deren Negationen ist.

1.5.21 Satz. *Die disjunktive Normalform ist, bis auf Umordnungen, eindeutig bestimmt.*

Statt durch Ausmultiplizieren kann die disjunktive Normalform auch an der Wahrheitstafel abgelesen werden: Beispiel:

A	B	C	$B \wedge C$	$A \Rightarrow (B \wedge C)$	$\neg(A \Rightarrow (B \wedge C))$
$\top$	$\top$	$\top$	$\top$	$\top$	$\bot$
$\top$	$\top$	$\bot$	$\bot$	$\bot$	$\top$
$\top$	$\bot$	$\top$	$\bot$	$\bot$	$\top$
$\top$	$\bot$	$\bot$	$\bot$	$\bot$	$\top$
$\bot$	$\top$	$\top$	$\top$	$\top$	$\bot$
$\bot$	$\top$	$\bot$	$\bot$	$\top$	$\bot$
$\bot$	$\bot$	$\top$	$\bot$	$\top$	$\bot$
$\bot$	$\bot$	$\bot$	$\bot$	$\top$	$\bot$

Die drei Zeilen, in denen in der letzten Spalte ein $\top$ steht, entsprechen exakt der disjunktiven Normalform: $(A \wedge B \wedge \neg C) \vee (A \wedge \neg B \wedge C) \vee (A \wedge \neg B \wedge \neg C)$.

1.5.22 Satz. *In der Boolschen Logik sind die Aussagen $A \Rightarrow B$ und $\neg A \vee B$ gleichbedeutend.*

Beweis. **Übung!** □

Jeder Term, der sich aus Variablennamen und den Operationssymbolen $\wedge, \vee, \neg$ syntaktisch korrekt zusammensetzt, heißt ein *Boolscher Ausdruck*.

1.5.23 Theorem. *Sei $f\colon \mathbb{B}^n \to \mathbb{B}$ eine Funktion in beliebig vielen Variablen mit Werten in $\mathbb{B}$. Dann gibt es einen boolschen Ausdruck $t[x_1, \ldots, x_n]$, sodaß $f(x_1, \ldots, x_n) = t[x_1, \ldots, x_n]$.*

Beweis. Ein passender Ausdruck (sogar in disjunktiver Normalform) kann unmittelbar aus der Wertetabelle abgelesen werden. □

1.5.24 Bemerkung. Dieser Satz hat die praktische Bedeutung, daß sich jede derartige Funktion (und damit eigentlich jede Funktion zwischen endlichen Mengen) aus einfachen Schaltkreisen für die logischen Operationen (UND-, ODER- und NICHT-Gatter) realisieren läßt. Freilich ist die Verwendung der disjunktiven Normalform für diesen Zweck nicht unbedingt sinnvoll; daher gibt es Verfahren, um die Anzahl bzw. Komplexität der verwendeten Bauelemente zu minimieren

Klassische Logik

Die *klassiche Logik* zeichnet sich durch Annahme eines der folgenden Prinzipien aus.

1.5.25 Axiom (Reductio ad absurdum). *Die aussagenlogische Formel $\neg\neg A \implies A$ ist allgemeingültig.*

1.5.26 Axiom (Tertium non datur). *Die aussagenlogische Formel $A \vee \neg A$ ist allgemeingültig.*

1.5.27 Theorem. *Die beiden Prinzipien* reductio ad absurdum *und* tertium non datur *sind äquivalent.*

Beweis. Satz 1.5.9 beweist die eine und Satz 1.5.12 die andere Richtung. □

1.5.28 Satz. *Bei Verwendung von klassischer Logik gilt stets*

$$\bigvee_{x\colon X} A(x) \iff \neg \bigwedge_{x\colon X} \neg A(x).$$

Beweis. Übung. □

1.5.29 Bemerkung. Die hier besprochenen Regeln gehen von der Idee aus, daß es eine absolute Wahrheit gibt, auch wenn sie uns Menschen verborgen bleibt. So ist es zwar möglich, daß eine Aussage weder beweisbar noch widerlegbar ist, aber im Prinzip muß sie doch wahr oder falsch sein (bzw.: wenn sie nicht falsch ist, so muß sie doch wahr sein).

Daneben hat die klassische Logik den Vorteil, daß sich einige Sätze der Mathematik einfacher formulieren oder beweisen lassen (z.B. durch Verzicht auf doppelte Negationen). Außerdem läßt sich die Gültigkeit aussagenlogische Formeln einfach mittels Wahrheitstabellen entscheiden.

Die überwältigende Mehrheit der mathematischen Literatur im 20. Jahrhundert wurde vom Standpunkt der klassischen Logik aus verfaßt, d.h. sie verwendet diese zusätzlichen Prinzipien ohne jeden Kommentar.

In neuerer Zeit gibt es aber immer mehr Anwendungen von nicht-klassischen Logik-Systemen. So lassen sich die in der Informatik so bedeutenden Begriffe der Entscheidbarkeit und der Berechenbarkeit in einer klassischen Logik nicht direkt ausdrücken, sodaß man gezwungen ist, diese Konzepte über die Hintertür (z.B. mittels Maschinenmodell) quasi in einer eigenen Welt künstlich einzuführen. In der Physik spielt die (dort so bedeutende) Frage der Meßbarkeit eine ähnliche Rolle.

Dies führt zu der Betrachtungsweise, daß ein Objekt, welches nicht berechnet (oder gemessen) werden kann, denselben Effekt hat wie eines, das gar nicht existiert.

In diesem Skriptum wird zumeist nicht mit der klassischen Logik, sondern mit der *konstruktiven* Logik (also ohne die soeben besprochenen zusätzlichen Prinzipien) gearbeitet, eben weil sich so einige wichtige Konzepte der Informatik treffender beschreiben lassen. Auf wesentliche Unterschiede zur klassischen Logik wird aber stets hingewiesen.

Kapitel 2

Mengen

2.1 Äquivalenzrelationen

Laut Cantor: *Eine Menge ist eine Zusammenfassung bestimmter wohlunterscheidbarer Objekte unserer Anschauung.*

Um zu bestimmen, welche Objekte in die Zusammenfassung aufgenommen werden sollen, ist festzulegen, wie diese zu konstruieren sind. Dies entspricht exakt unserer Definition des Begriffs *Aussage*.

Objekte, die zu einer Menge zusammengefaßt werden, sollten aber darüberhinaus auch *wohlunterscheidbar* sein. Das heißt, es muß ein vernünftiger Gleichheitsbegriff festgelegt werden. *Vernünftig* heißt hier, daß der Gleichheitsbegriff auf jeden Fall den Regeln für eine Äquivalenzrelation genügen muß.

2.1.1 Definition. Eine binäre Relation $(=)\colon X \to X \to \Omega$ heißt *Äquivalenzrelation*, wenn für alle $x, y, z\colon X$ gilt:

$$\begin{aligned} x = x & \qquad (\textit{reflexiv});\\ x = y \iff y = x & \qquad (\textit{symmetrisch});\\ x = y \wedge y = z \implies x = z & \qquad (\textit{transitiv}). \end{aligned}$$

2.1.2 Bemerkung. Um eine Menge vollständig zu definieren, ist also (im Prinzip) dreierlei zu tun:

- Festlegen, welche Objekte dazugehören;
- Eine Relation definieren;
- Beweisen, daß diese Relation tatsächlich die Eigenschaft einer Äquivalenzrelation erfüllt.

Diese Arbeit kann man sich natürlich sparen, wenn einfache Methoden zur Verfügung stehen, um aus bekannten Mengen neue zu konstruieren.

2.1.3 Bemerkung. Man beachte, daß je zwei Objekten lediglich eine Aussage zugeordnet werden muß; d.h. es muß klar sein, was bewiesen werden müßte, um zwei Objekte als gleich nachzuweisen. Es ist aber keinesfalls erforderlich, daß festgestellt werden kann, ob es so einen Beweis auch tatsächlich gibt.

2.1.4 Notation. Jede Menge kommt zusammen mit einem Gleichheitsbegriff. Mit $x = y : X$ ist gemeint, daß x und y gleiche Objekte von X bezeichnen. Oft ist aus dem Zusammenhang klar, daß x und y als Elemente von X zu betrachten sind, sodaß man stattdessen einfach $x = y$ schreibt. Die Notation $x\colon X$ kann auch als Abkürzung für $x = x : X$ aufgefaßt werden.

2.1.5 Definition. Eine Menge X heißt *diskret*, wenn für alle $x\colon X$ gilt:

$$x = x \vee \neg(x = x).$$

2.1.6 Bemerkung. Diese Bezeichnung macht keinen Sinn, wenn die Regel vom ausgeschlossenen Dritten (klassische Logik) angenommen wird. Für diesen Fall, muß *diskret* anders (etwa mit einem Maschinenmodell) definiert werden.

2.1.7 Beispiel. Die Gesamtheit aller logischen Aussagen, zusammen mit der logischen Äquivalenz, bildet eine Menge (es war einfach, nachzuweisen, daß die logische Äquivalenz tatsächlich ein Äquivalenzrelation ist). Wir bezeichnen sie mit Ω, und nennen sie manchmal auch die Menge aller Wahrheitswerte. In klassischer Logik (d.h. mit *reductio ad absurdum*) gilt natürlich $\Omega = \mathbb{B}$.

2.1.8 Beispiel. Die wahrscheinlich fundamentalste aller Mengen ist die Menge der natürlichen Zahlen $\mathbb{N}$. Diese ist diskret.

2.2 Konstruktionen für Mengen

Funktionen

Sind A und B Mengen, so ist auch die Gesamtheit aller *Funktionen* zwischen ihnen eine Menge. Dies ist als Verfeinerung des Begriffs der Implikation von Aussagen aufzufassen und wird daher mit den entsprechenden Introduktions- und Eliminationsregeln beschrieben:

2.2.1 Axiom (→-Introduktionsregel).

$$\dfrac{\begin{array}{c}x\colon A\\ \boxed{\begin{array}{c}\vdots\\ t[x]\colon B\end{array}}\end{array}}{x \mapsto t[x]\colon A \to B}\to\mathcal{I}$$

Genauer:

$$\dfrac{\begin{array}{c}x = y : A\\ \boxed{\begin{array}{c}\vdots\\ s[x] = t[y] : B\end{array}}\end{array}}{(x \mapsto s[x]) = (y \mapsto t[x]) : A \to B}\to\mathcal{I}$$

Dabei wird mit $t[x]$ der Term t bezeichnet und gleichzeitig ausgedrückt, daß darin an bestimmten Stellen ein x vorkommen kann.

Für die Konstruktion $x \mapsto t[x]$ ist auch die Notation $\lambda x.t[x]$ gebräuchlich, und sie heißt daher auch λ-Abstraktion. In Java würde man für $f = (x \mapsto t[x]) : A \to B$ in etwa folgendes schreiben:

```
B f (A x){
   return (t[x]);
}
```

(Für `A` und `B` sind natürlich konkrete Java-Datentypen einzusetzen, für `t[x]` irgendein Ausdruck, in dem typischerweise das `x` vorkommt.

2.2.2 Axiom (→-Eliminationsregel).

$$\frac{f\colon A \to B \qquad x\colon A}{fx\colon B} \to \mathcal{E}$$

Genauer:

$$\frac{f = g : A \to B \qquad x = y : A}{fx = gy : B} \to \mathcal{E}$$

Der zur λ-Abstraktion gehörige Selektor ist damit die Funktionsanwendung (Funktionsapplikation). Statt fx ist die Notation $f(x)$ sehr gebräuchlich, wenn deutlich gemacht werden soll, daß es sich um keine Multiplikation handelt. Für viele spezielle Funktionen sind auch Postfix, Infix, oder andere spezielle Notationen gebräuchlich.

Der Zusammenhang zwischen der λ-Abstraktion und der Funktionsapplikation ergibt sich wieder aus den passenden β und η-Regeln.

2.2.3 Axiom (β-Regel). *Ist $t[x]\colon B$, für jedes $x\colon A$, und $a\colon A$, so gilt*

$$(x \mapsto t[x])a = t[a] : B.$$

Dabei bezeichnet $t[a]$ jenen Ausdruck, welcher entsteht, wenn in $t[x]$ jedes Vorkommen von x durch a ersetzt wird.

2.2.4 Axiom (η-Regel). *Für $f\colon A \to B$ gilt*

$$(x \mapsto fx) = f : A \to B.$$

2.2.5 Satz. *Zwei Funktionen $f, g\colon A \to B$ sind genau dann gleich, wenn $fx = gx : B$, für alle $x\colon A$.*

Dies bedeutet, daß es für die Gleichheit von Funktionen nicht entscheidend ist, wie diese konstruiert wurden, sondern ob sie bei jedem Input identische Ergebnisse liefern, also von *außen betrachtet* gleich erscheinen. (Auch eine eventuelle Laufzeit spielt für den Gleichheitsbegriff keine Rolle.)

Von der Implikation haben wir gezeigt, daß sie transitiv ist. Die entsprechende Konstruktion ist eine grundlegende Operation für Funktionen:

2.2.6 Definition. Seien $f\colon A \to B$, $g\colon B \to C$. Dann heißt

$$g \circ f = (x \mapsto g(fx)) : A \to C$$

die *Hintereinanderausführung* (*Komposition*) der Funktionen f und g.

2.2.7 Satz. *Die binäre Operation $(\circ)\colon (B \to C) \to (A \to B) \to (A \to C)$ ist assoziativ, d.h. für alle $f\colon A \to B$, $g\colon B \to C$, $h\colon C \to D$ gilt*

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

Insbesondere bildet daher die Menge der Funktionen $A \to A$, zusammen mit der Komposition, eine Halbgruppe; die identische Funktion $\mathrm{id}\, A = (x \mapsto x) : A \to A$ ist das neutrale Element dieser Halbgruppe.

Direktes Produkt

Sind A und B Mengen, so ist auch die Gesamtheit aller Paare, die man daraus bilden kann, eine Menge. Dies ist als Verfeinerung des Begriffs der Konjunktion von Aussagen aufzufassen und wird daher, so wie diese, mit Introduktions- und Eliminationsregeln beschrieben:

2.2.8 Axiom (×-Introduktionsregel).

$$\frac{x\colon A \qquad y\colon B}{(x,y)\colon A \times B}\times\mathcal{I} \tag{2.1}$$

Genauer:

$$\frac{x_1 = x_2 : A \qquad y_1 = y_2 : B}{(x_1,y_1) = (x_2,y_2) : A \times B}\times\mathcal{I} \tag{2.2}$$

2.2.9 Axiom (×-Eliminationsregel).

$$\frac{z\colon A \times B}{\operatorname{fst} z\colon A}\times\mathcal{E} \qquad\qquad \frac{z\colon A \times B}{\operatorname{snd} z\colon B}\times\mathcal{E} \tag{2.3}$$

Genauer:

$$\frac{z_1 = z_2 : A \times B}{\operatorname{fst} z_1 = \operatorname{fst} z_2 : A}\times\mathcal{E} \qquad\qquad \frac{z_1 = z_2 : A \times B}{\operatorname{snd} z_1 = \operatorname{snd} z_2 : B}\times\mathcal{E} \tag{2.4}$$

Die Notation (,) bezeichnet einen *Konstruktor*, während fst und snd *Selektoren* bezeichnen. Der Zusammenhang wird mit den folgenden Regeln beschrieben:

2.2.10 Axiom (β-Regel). *Für $x\colon A$ und $y\colon B$ gilt*

$$\operatorname{fst}(x,y) = x : A \qquad\qquad \operatorname{snd}(x,y) = y : B$$

2.2.11 Axiom (η-Regel). *Für $z\colon A \times B$ gilt*

$$(\operatorname{fst} z, \operatorname{snd} z) = z : A \times B$$

Damit ist garantiert, daß Paare genau dann gleich sind, wenn sie auf dieselbe Weise konstruiert wurden, d.h. wenn beide Komponenten übereinstimmen.

Direkte Summe

Sind A und B Mengen, so bildet auch die Gesamtheit derjenigen Objekte, welche zu einer dieser beiden Mengen gehören, ebenfalls eine Menge. Dies ist als eine Verfeinerung des Begriffs der Disjunktion aufzufassen und wird daher mit den entsprechenden Introduktions- und Eliminationsregeln beschrieben:

2.2.12 Axiom (+-Introduktionsregeln).

$$\frac{x\colon A}{\operatorname{Left} x\colon A + B}+\mathcal{I} \qquad\qquad \frac{x\colon B}{\operatorname{Right} x\colon A + B}+\mathcal{I}$$

Genauer:

$$\frac{x_1 = x_2 : A}{\operatorname{Left} x_1 = \operatorname{Left} x_2 : A + B}+\mathcal{I} \qquad\qquad \frac{y_1 = y_2 : B}{\operatorname{Right} y_1 = \operatorname{Right} y_2 : A + B}+\mathcal{I}$$

Statt $A + B$ ist auch die Bezeichnung $A \uplus B$ üblich.

Die Verwendung von zwei verschiedenen Konstruktoren stellt sicher, daß von jedem Element von $A + B$ entschieden werden kann, ob es aus A oder aus B stammt, insbesondere auch dann, wenn A und B dieselbe Menge bezeichnen.

2.2.13 Axiom (+-Eliminationsregel).

$$\frac{f\colon A \to C \qquad g\colon B \to C \qquad z\colon A + B}{\text{either } f\, g\, z\colon C}{+}\mathcal{E}$$

Genauer:

$$\frac{f_1 = f_2 : A \to C \qquad g_1 = g_2 : B \to C \qquad z_1 = z_2 : A + B}{\text{either } f_1\, g_1\, z_1 = \text{either } f_2\, g_2\, z_2 : C}{+}\mathcal{E}$$

So wie bei der Disjunktion, haben wir auch hier eine etwas kompliziertere Eliminationsregel. Sie ist aber etwas besser zu verstehen, wenn man folgendes beachtet: Für jedes $z\colon A+B$ gilt either $f\, g\, z\colon C$; daher ist either $f\, g\colon A+B \to C$. Der Selektor either entspricht damit einer Fallunterscheidung: either $f\, g$ verhält sich auf A so wie f, auf B aber so wie g. Dies verdeutlicht vor allem die folgende Regel:

2.2.14 Axiom (β-Regel). *Für $f\colon A \to C$, $g\colon B \to C$, $x\colon A$, $y\colon B$ gilt*

$$\text{either } f\, g\, (\text{Left}\, x) = fx : C \qquad\qquad \text{either } f\, g\, (\text{Right}\, y) = gy : C$$

2.2.15 Axiom (η-Regel). *Für $z\colon A + B$ gilt*

$$\text{either Left Right } z = z : A + B.$$

2.3 Potenzmenge

Teilmengen

Es sei G eine Menge, die im folgenden Grundmenge genannt wird. Für eine Eigenschaft $P\colon G \to \Omega$ sollte $\{\, g\colon G \mid P(g)\,\}$ eine Menge bezeichnen, die all jene Elemente von G umfaßt, welche die Eigenschaft P besitzen.

2.3.1 Definition. Um ein Element der Menge $\{\, g\colon G \mid P(g)\,\}$ zu konstruieren, muß man ein Element $g\colon G$ konstruieren und dann $P(g)$ beweisen. Es ist also ein Beweis $(g, \varphi)\colon \bigvee_{g\colon G} P(g)$ zu konstruieren.

Als Gleichheitsbegriff legen wir fest

$$(g_1, \varphi_1) = (g_2, \varphi_2)\colon \{\, g\colon G \mid P(g)\,\} :\Longleftrightarrow\ g_1 = g_2\colon G,$$

was tatsächlich eine Äquivalenzrelation ist.

Die Menge $\{\, g\colon G \mid P(g)\,\}$ nennt man eine *Teilmenge* von G.

2.3.2 Bemerkung. Ist $(g, \varphi)\colon \{\, g\colon G \mid P(g)\,\}$, so ist wegen dem gewählten Gleichheitsbegriff der Beweis φ irrelevant; relevant ist nur, daß es einen solchen gibt. Dennoch ist es nicht ganz korrekt, wenn man $g\colon \{\, g\colon G \mid P(g)\,\}$ schreibt. Wir definieren deshalb ein geeigenetes 2-stelligen Prädikat.

2.3.3 Definition. Ist $(g, \varphi)\colon \{\, g\colon G \mid P(g)\,\}$, dann sagen wir, daß g in der Teilmenge $\{\, g\colon G \mid P(g)\,\}$ enthalten ist und schreiben für diesen Sachverhalt $g \in \{\, g\colon G \mid P(g)\,\}$. D.h.

$$g \in \{\, g\colon G \mid P(g)\,\} :\iff P(g).$$

Damit definieren wir eine Ordnungsrelation für Teilmengen.

2.3.4 Definition. Seien A und B Teilmengen von G. Dann definieren wir

$$A \subseteq B :\iff \bigwedge_{g \in G} (g \in A \implies g \in B),$$

welche reflexiv und transitiv ist, und gemäß Antisymmetrie den folgenden Gleichheitsbegriff

$$A = B \iff A \subseteq B \wedge B \subseteq A,$$

ergibt, welcher bedeutet, daß zwei Teilmengen genau dann gleich sind, wenn sie dieselben Elemente enthalten. Damit ist auch die Gleichheit von Teilmengen festgelegt; Die (geordnete) Menge aller Teilmengen von G heißt deren *Potenzmenge* und wird mit $\mathscr{P}G$ bezeichnet.

2.3.5 Bemerkung. Leicht anders formuliert haben wir

$$A = B : \mathscr{P}(G) \iff \bigwedge_{g\colon G} \big(g \in A \iff g \in B\big).$$

Dies bedeutet, daß zwei Teilmengen genau dann als gleich betrachtet werden, wenn sie dieselben Elemente haben. Man beachte aber, daß dies nur dann wirklich Sinn macht, wenn klar ist, auf welche weise die betroffenen Mengen in einer gemeinsamen Grundmenge enthalten sind.

Mengenalgebra

Sind A, B zwei Mengen, die nichts miteinander zu tun haben, und ist $a\colon A$, $b\colon B$, dann macht es im allgemeinen keinen Sinn zu fragen, ob a ein Element von B ist, oder ob $a = b$. Es muß zuerst festgelegt werden, wie die Elemente von A und B zu identifizieren sind. Dies ist insbesondere dann klar, wenn beide Teilmengen einer gemeinsamen Grundmenge sind. Dann lassen sich die üblichen Mengenoperationen definieren:

2.3.6 Notation. Statt $\neg(g \in G)$ schreibt man üblicherweise $g \notin G$.

2.3.7 Definition. Seien $A, B\colon \mathscr{P}(G)$. Mengentheoretische Vereinigung, Durchschnitt und Differenz werden definiert durch

$$\begin{aligned} A \cup B &= \{\, g\colon G \mid g \in A \vee g \in B\,\}, \\ A \cap B &= \{\, g\colon G \mid g \in A \wedge g \in B\,\}, \\ A \setminus B &= \{\, g\colon G \mid g \in A \wedge g \notin B\,\}. \end{aligned}$$

Eine spezielle Rolle spielen die *leere Menge* $\varnothing = \{\, g\colon G \mid \bot\,\}$ und die Gesamtmenge $\{\, g\colon G \mid \top\,\}$, welche üblicherweise mit der Grundmenge G selbst identifiziert wird. Ist die Grundmenge aus dem Zusammenhang klar, definiert man auch das mengentheoretischen Komplement

$$\complement A = G \setminus A \qquad\qquad (\textit{Komplement}).$$

Für das Komplement sind auch die Notationen A' und $\overline{A}$ gebräuchlich.

Ist $G \subseteq H$, so kann jede Teilmenge von G auch als Teilmenge von H aufgefaßt werden. Für die meisten der oben definierten Operationen spielt dies keine Rolle, wohl aber beim Komplement: denn wäre $T\colon \mathscr{P}H$, dann wäre $\complement T = H \setminus T$, und nicht $G \setminus T$.

2.3.8 Bemerkung. Identifiziert man G mit $\{g\colon G \mid \top\}$, so verschwindet formal der Unterschied zwischen $g\colon G$ und $g \in G$. Ersteres ist eine deklarative Variante, d.h. g wird als Element von G eingeführt. Letzteres dagegen ist eine Aussage, welche gegebenenfalls zu beweisen ist.

Die Rechenregeln für diese Operationen ergeben sich ziemlich direkt aus den entsprechenden Regeln für Aussagen. Im folgenden beschäftigen wir uns aber nur mit einer speziellen Klasse von Teilmengen.

Ohne *reductio ad absurdum* kann es natürlich passieren, daß weder $g \in B$ noch $g \notin B$ bewiesen werden kann.

2.3.9 Definition. Sei G eine Menge. Ein $A\colon \mathscr{P}(G)$ heißt *entscheidbar* wenn für alle $g\colon G$ gilt

$$g \in A \vee g \notin A.$$

Diese Definition macht natürlich nur Sinn, wenn *reductio ad absurdum* nicht generell angenommen wird. In klassischer Logik definiert man entscheidbare Teilmengen daher anders (nämlich, als mittels einer abstrakt definierten Maschine entscheidbar).

2.3.10 Satz. *Die entscheidbaren Teilmengen sind gerade diejenigen, welche sich als*

$$A = \{\, g\colon G \mid P(g) \,\},$$

darstellen lassen, mit $P\colon G \to \mathbb{B}$ (nicht nur mit $P\colon G \to \Omega$).

2.3.11 Bemerkung. Hier ist vor allem gemeint, daß das P tatsächlich berechenbar ist.

2.3.12 Satz. *Die entscheidbaren Teilmengen bilden mit Vereinigung, Durchschnitt und Komplement eine Boolsche Algebra.*

Beweis. Alle erforderlichen Gleichheiten ergeben sich unmittelbar aus der Definition und der entsprechenden Gleichheit in der Boolschen Logik. □

2.3.13 Bemerkung. Die Menge der entscheidbaren Teilmengen ist schon wesentlich besser handhabbar als die ganze Potenzmenge, aber für den praktischen Einsatz, etwa in Datenbanken, immer noch zu groß bzw. kompliziert. Daher beschränkt man sich bei derartigen Anwendungen üblicherweise auf die Menge aller endlichen Teilmengen. Allerdings bilden diese keine Boolsche Algebra mehr, da das Komplement einer endlichen Menge nicht mehr endlich sein muß (wenn die Grundmenge unendlich ist). Man kann aber stattdessen die Mengendifferenz verwenden (die Differenz endlicher Teilmengen einer diskreten Menge ist wieder endlich). Die Gesetze von De Morgan müssen ersetzt werden durch

$$A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C),$$
$$A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C);$$

und die Regeln für das Komplement durch

$$A \cap (B \setminus A) = \varnothing \qquad\qquad A \cup (B \setminus A) = B.$$

Eine solche algebraische Struktur heißt *Boolscher Ring*. Mit endlichen Teilmengen einer diskreten Menge können alle relevanten Operationen und Vergleiche effektiv durchgeführt werden (für bloß entscheidbare Teilmengen hat man dagegen z.B. kein allgemeines Verfahren, um festzustellen, ob eine solche leer ist. In der Theorie der formalen Sprachen werden diverse Zwischenstufen diskutiert. So bilden etwa die regulären Sprachen eine Boolsche Algebra, die auch unendliche Mengen umfaßt und in der alles entscheidbar und berechenbar bleibt.

Beispiele

Sei $(-0)\colon \mathbb{N} \to \mathbb{Z}$ definiert durch $(-0)(x) = x - 0$. Damit wird eine Teilmenge von $\mathbb{Z}$ definiert, nämlich $\{\, x - 0 \mid x\colon \mathbb{N} \,\}\colon \mathscr{P}\mathbb{Z}$, die positiven ganzen Zahlen. Die Teilmenge der negativen ganzen Zahlen erhält man analog mit $(0-)\colon \mathbb{N} \to \mathbb{Z}$, $(0-)(x) = 0 - x$ als $\{\, 0 - x \mid x\colon \mathbb{N} \,\}\colon \mathscr{P}\mathbb{Z}$. Die Quadratzahlen erhalten wir mit $\{x^2 \mid x\colon \mathbb{N}\}$, aber auch z.B. mit $\{x^2 \mid x\colon \mathbb{Z}\}$. Die Abbildung (-0) ist besonders natürlich, insbesondere ist sie eine injektive Funktion, d.h. es gilt

$$x = y : \mathbb{N} \iff x - 0 = y - 0 : \mathbb{Z};$$

außerdem ist sie mit der Addition verträglich:

$$(x - 0) + (y - 0) = (x + y) - 0,$$

und auch mit der Multiplikation:

$$(x - 0)(y - 0) = xy - 0.$$

Da es nur eine derart natürliche Funktion von $\mathbb{N}$ nach $\mathbb{Z}$ gibt, sind wir geneigt, $\mathbb{N}$ selbst als Teilmenge von $\mathbb{Z}$ anzusehen. Die dabei unterstellte natürliche Funktion gibt dann an, wie die Elemente dabei zu *identifizieren* sind, nämlich als

$$x = x - 0 : \mathbb{Z}, \qquad \text{für } x \in \mathbb{N}.$$

Man schreibt daher auch $\mathbb{N} = \{\, x - 0 \mid x\colon \mathbb{Z} \,\}$; oder $x\colon \mathbb{Z}$, falls $x\colon \mathbb{N}$, und daher auch $\mathbb{N}\colon \mathscr{P}\mathbb{Z}$. Zu Mißverständnissen kann es dabei kaum kommen, weil die Funktion (-0) mit allen Operationen verträglich ist.

Auch die Funktion $(0-)$ ermöglicht es, $\mathbb{N}$ mit den negativen ganzen Zahlen zu identifiziern. Dies geht jedoch nur so lange gut, als keine Multiplikation vorkommt: werden $x, y\colon \mathbb{N}$ als natürliche Zahlen multipliziert und dann mittels $(0-)$ mit einer ganzen Zahl identifiziert, erhält man $0 - xy$; multipliziert man dagegen die entsprechenden ganzen Zahlen, so erhält man $(0 - x)(0 - y) = xy - 0$, was im allgemeinen nicht dasselbe ist. Diese Identifikation ist daher nicht natürlich.

Auf analoge Weise kann man $\mathbb{Z}$ als Teilmenge von $\mathbb{Q}$ auffassen, mit $z = \frac{z}{1} : \mathbb{Q}$. Diese Identifikation mag (so wie die vorherige) recht trivial erscheinen. Man beachte aber, daß in vielen Programmiersprachen sehr wohl zwischen der ganzen Zahl 3 und der rationalen Zahl $\frac{3}{1}$, dargestellt als Paar von ganzen Zahlen oder als Gleitkommazahl, zu unterscheiden ist. Die Identifikation geschieht dann durch implizite oder explizite Typumwandlung.

2.4 Gleichmächtigkeit

Es ist für eine Menge eigentlich völlig irrelevant, welche konkreten Objekte sie enthält. So besteht die Menge der natürlichen Zahlen konzeptionell aus den Objekten $O, SO, S(SO), S(S(SO)), \ldots$. Praktischerweise werden sie aber meist als Dezimalzahlen $0, 1, 2, 3, \ldots, 756, \ldots$ notiert, also durch bestimmte Zeichenketten. In der Informatik wählt man statt dessen gerne das Binärsystem, d.h. man stellt Zahlen durch Zeichenketten über $\{0, 1\}$ dar, welche im Computer wiederum durch Folgen elektrischer oder magnetischer Zustände realisiert werden. Dabei ist stets klar, wie zwischen diesen Darstellungen hin und her gewechselt werden kann.

2.4.1 Definition. Eine Funktion $f\colon A \to B$ heißt *bijektiv* (oder eine *Bijektion*), wenn es eine Funktion $g\colon B \to A$ gibt, sodaß

$$g \circ f = \mathrm{id}_A \qquad\qquad f \circ g = \mathrm{id}_B\,.$$

Die Funktion g ist, wenn sie existiert, durch diese Eigenschaft eindeutig bestimmt, heißt die *Umkehrfunktion* von f und wird oft mit f^{-1} bezeichnet.

Eine Bijektion bewirkt damit so etwas wie eine andere Darstellung (Umbenennung, Codierung) der Elemente von A durch die Elemente von B. Wegen der Umkehrbarkeit kann man jederzeit wieder zu den ursprünglichen Elementen von A zurückkehren. So ist z.B. die Darstellung von abstrakten oder realen Objekten in einem Computer eine Bijektion zwischen der Menge der Objekte, die man eigentlich betrachtet, und einer bestimmten Menge von (irgendwie realisierten) Bitfolgen.

2.4.2 Definition. Zwei Mengen A, B heißen *gleichmächtig* wenn es eine Bijektion $f\colon A \to B$ gibt. Man schreibt dann

$$|A| = |B|\,.$$

2.4.3 Satz. *Gleichmächtigkeit ist eine Äquivalenzrelation für Mengen.*

Beweis. Wir müssen überprüfen, ob sie reflexiv, symmetrisch und transitiv ist:

Reflexivität: $\mathrm{id}\colon A \to A$ ist bijektiv, denn

$$\mathrm{id}_A^{-1} = \mathrm{id}_A\,.$$

Daher ist $|A| = |A|$.

Symmetrie: Sei $|A| = |B|$. Dann gibt es laut Definition eine Bijektion $f\colon A \to B$, und auch $f^{-1}\colon B \to A$ ist ebenfalls bijektiv, denn es gilt

$$(f^{-1})^{-1} = f.$$

Daher ist $|B| = |A|$.

Transitivität: Sei $|A| = |B|$ und $|B| = |C|$. Laut Definition gibt es daher Bijektionen $f\colon A \to B$ und $g\colon B \to C$. Dann ist aber auch $g \circ f\colon A \to C$ eine Bijektion, denn es gilt

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}.$$

Daher ist $|A| = |C|$.

□

Damit ist die Gleichmächtigkeit ein zulässiger Gleichheitsbegriff für Mengen.

2.4.4 Satz. *Sind A und A' gleichmächtig, und ebenso B und B', also $|A| = |A|'$ und $|B| = |B|'$, so gelten auch*

$$\begin{aligned} |A \times B| &= |A' \times B'|\,; \\ |A + B| &= |A' + B'|\,; \\ |A \to B| &= |A' \to B'|\,; \\ |\mathscr{P}(A)| &= |\mathscr{P}(A')|\,. \end{aligned}$$

Damit ist die Gleichmächtigkeit mit den gängigen Operationen für Mengen verträglich und scheint tatsächlich ein sinnvoller Gleichtheitsbegriff für Mengen zu sein. Allerdings bezeichnet man die Mengen mit diesem Gleichheitsbegriff üblicherweise als *Kardinalzahlen*, damit es zu keiner Verwechslung mit dem Gleichheitbegriff für Teilmengen kommt, die nur dann als gleich betrachtet werden, wenn sie dieselben Elemente enthalten (was aber nur dann Sinn macht, wenn klar ist, auf welche Weise sie in einer gemeinsamen Grundmenge eingebettet sind; siehe Abschnitt über Potenzmengen).

2.4.5 Definition. Eine Funktion $f\colon A \to B$ heißt

1. *injektiv* falls für alle $x, y\colon A$ gilt

$$fx = fy : B \implies x = y : A;$$

2. *surjektiv* falls es zu jedem $y\colon B$ ein $x\colon A$ gibt sodaß

$$y = fx : B.$$

Man beachte, daß diese beiden Eigenschaften in einem gewissen Sinne dual zu den beiden Eigenschaft sind, die jede Funktion erfüllen muss: die Injektivität ist dual zur Wohldefiniertheit ($x = y : A \implies fx = fy : B$) ist; und die Surjektivität ist dual zur Totalität von Funktionen (zu jedem $x\colon A$ gibt es ein $y\colon B$, sodaß $fx = y : B$.

2.4.6 Satz. *Sei $f\colon A \to B$. Dann gilt*

1. *f ist injektiv wenn es ein $g\colon B \to A$ gibt, sodaß $g \circ f = \mathrm{id}\, A$.*
2. *f ist surjektiv wenn es ein $g\colon B \to A$ gibt, sodaß $f \circ g = \mathrm{id}\, B$.*
3. *f ist genau dann bijektiv wenn sie injektiv und surjektiv ist.*

Man nennt daher eine injektive Funktion auch *links-invertierbar* und eine surjektive Funktion auch *rechts-invertierbar*. Bijektive Funktionen sind hiermit (beidseitig) *invertierbar*.

2.4.7 Definition. Sei A eine Menge. Dann heißt A

1. *endlich* falls es ein $n\colon \mathbb{N}$ gibt sodaß

$$|A| = |\{0, \ldots, n-1\}|\,;$$

2. *unendlich* falls es eine Injektion $\mathbb{N} \to A$ gibt (oder, wenn es zu jeder endlichen Teilmenge ein weiteres Element gibt).

3. *abzählbar unendlich* falls

$$|A| = |\mathbb{N}|\,;$$

4. in B einbettbar ($|A| \leq |B|$) falls es eine Injektion von A nach B gibt.

Statt $|\{0,\dots,n-1\}|$ kann man auch einfach n schreiben. Weiters werden gelegentlich verwendet: $|\mathbb{N}| = \aleph_0$ und $|\mathbb{R}| = c$.

2.4.8 Satz.

1. Für $n, m\colon \mathbb{N}$ gilt

$$|\{0,\dots,n-1\}| = |\{0,\dots,m-1\}| \iff n = m;$$

2. Für endliche Mengen A, B gilt

$$\begin{aligned} |A \times B| &= |A| \cdot |B| \\ |A + B| &= |A| + |B|\,, \\ |A \to B| &= |B|^{|A|}\,; \end{aligned}$$

daher verwendet man auch die Notation B^A für $A \to B$.

3. Ist die Menge A in die unendliche Menge B einbettbar, so gilt

$$\begin{aligned} |B| &= |A \times B|\,, \qquad \textit{(falls } |A| \geq 1\textit{)}, \\ |B| &= |A + B|\,, \\ |B| &\leq |A \to B|\,; \end{aligned}$$

4. Falls B mindestens 2 verschiedene Elemente hat, dann gibt es keine Surjektion von A nach $A \to B$.

2.4.9 Beispiel. $0 < 1 < \cdots < n = |\{0,\dots,n-1\}| = |\{1,\dots,n\}| = |\{2,4,\dots,2n\}| < |\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}| = |\mathbb{Q} \times \mathbb{Q}| = |\mathbb{Q}^n| = |\mathbb{Q}^*| = |\mathbb{A}|$. Die Mengen $\mathbb{R}$, $\mathbb{Q} \to \mathbb{B}$, $\mathbb{R} \to \mathbb{R}$ sind dagegen nicht abzählbar.

Weiters: $|A^{\varnothing}| = 1$, sogar wenn $A = \varnothing$, d.h. $|\varnothing^{\varnothing}| = 1$. Dagegen ist $\left|\varnothing^A\right| = 0$, sofern $|A| \geq 1$.

Um zu sehen, daß $\mathbb{N} \to \mathbb{B}$ nicht abzählbar ist, gehen wir folgendermaßen vor: Sei $f\colon \mathbb{N} \to (\mathbb{N} \to \mathbb{B})$. Dann definieren wir $g\colon \mathbb{N} \to \mathbb{B}$ durch

$$g(n) = \neg(f(n)(n)).$$

Wenn f surjektiv ist, dann gibt es ein $n\colon \mathbb{N}$, sodaß $g = f(n)$. Dann ist aber auch $g(n) = f(n)(n)$, was genau der Definition von g widerspicht. Es kann daher keine Surjektion von $\mathbb{N}$ nach $\mathbb{N} \to \mathbb{B}$ geben.

2.5 Konstruktion der Zahlenmengen

Ganze Zahlen

Es gibt zwei unterschiedliche Möglichkeiten, die ganzen Zahlen auf die natürlichen Zahlen zurückzuführen.

1. Variante

Die erste funktioniert so ähnlich wie das Geldwesen: alle Geldscheine sind zwar positiv, können aber in beide Richtungen wandern (Einnahmen/Ausgaben).

2.5.1 Axiom (ℤ-Introduktionsregel).

$$\frac{x\colon \mathbb{N} \qquad y\colon \mathbb{N}}{x-y\colon \mathbb{Z}}\mathbb{Z}\mathcal{I} \tag{2.5}$$

Der Ausdruck $x - y$ ist hier rein formal zu verstehen (Konstruktor, keine Rechnung) und bedeutet bei der obigen Interpretation: Einnahmen in der Höhe von x Euro stehen Ausgaben in der Höhe von y Euro gegenüber.

Diese Regel hat dieselbe Form wie die Introduktionsregel für das direkte Produkt. Eine ganze Zahl wird durch zwei natürliche Zahlen dargestellt. Entsprechend der Tatsache, daß höhere Einnahmen durch gleichviel höhere Ausgaben ausgeglichen werden (den selben Gewinn ergeben), verwenden wir allerdings als Gleichheitsbegriff nicht den für Paare, sondern den folgenden.

2.5.2 Axiom. *Für $x_1, y_1, x_2, y_2\colon \mathbb{N}$ definieren wir*

$$x_1 - y_1 = x_2 - y_2 : \mathbb{Z} \iff x_1 + y_2 = x_2 + y_2 : \mathbb{N}. \tag{2.6}$$

Damit wurde der Gleichheitsbegriff für ganze Zahlen auf den für die natürlichen Zahlen zurückgeführt. Freilich muß nachgewiesen werden, daß dieser Gleichheitsbegriff tatsächlich eine Äquivalenzrelation ist, was aber nicht schwer ist.

Die Addition läßt sich ebenfalls leicht auf $\mathbb{Z}$ definieren (wie bei einer Zusammenlegung von 2 Konten):

2.5.3 Definition. Für $x_1 - y_1\colon \mathbb{Z}$ und $x_2 - y_2\colon \mathbb{Z}$ definieren wir

$$(x_1 - y_1) + (x_2 - y_2) = (x_1 + x_2) - (y_1 + y_2) : \mathbb{Z}.$$

Damit wird $(\mathbb{Z}, +)$ nicht nur zu einer Halbgruppe oder einem Monoid (mit $0 - 0$ als neutralem Element), sondern auch zu einer Gruppe, mit inversem Element:

$$-(x - y) = y - x.$$

Und mit der üblichen Multiplikation

$$(m_1 - n_1)(m_2 - n_2) = (m_1 m_2 - n_1 n_2) - (m_1 n_2 + m_2 n_1) : \mathbb{Z}$$

erhält man sogar einen Ring.

Eine echte Rechnung ergibt sich, wenn man einen Kontoausgleich vornimmt:

2.5.4 Satz. *Zu jedem $z\colon \mathbb{Z}$ gibt es ein eindeutiges $n\colon \mathbb{N}$ sodaß*

$$z = n - 0 : \mathbb{Z} \vee z = 0 - n : \mathbb{Z}.$$

Damit erhält jede ganze Zahl eine kanonische Darstellung. Das hier konstruierte n wird üblicherweise mit $|z|$ bezeichnet. Den Ausdruck $0 - n$ kürzt man üblicherweise mit $-n$ ab, während $n - 0$ als $+n$ oder, noch einfacher, als n geschrieben wird. Die ganze Zahl $0 - 0$ kann auf beide Arten interpretiert werden,

daher ist $-0 = +0 : \mathbb{Z}$ Das Vorzeichen gibt an, welcher der Fälle für ein $z\colon \mathbb{Z}$ eintritt:

$$\operatorname{sgn} z = \begin{cases} +1 & \text{falls } z = n - 0, \text{ aber nicht } z = 0 - n; \\ -1 & \text{falls } z = 0 - n, \text{ aber nicht } z = n - 0; \\ 0 & \text{falls } z = n - 0 \text{ und } z = 0 - n. \end{cases}$$

Es gilt stets $z = |z| \cdot \operatorname{sgn} z$.

2. Variante

Die andere Art der Konstruktion der ganzen Zahlen aus den natürlichen geht von der zuletzt erwähnten Notation aus und verwendet eine Vorgangsweise wie bei der direkten Summe:

2.5.5 Axiom ($\mathbb{Z}$-Introduktionsregeln).

$$\frac{n\colon \mathbb{N}}{+n\colon \mathbb{Z}}\mathbb{Z}\mathcal{I} \qquad\qquad \frac{n\colon \mathbb{N}}{-n\colon \mathbb{Z}}+\mathcal{I}$$

Die Gleichheit wird dann ähnlich wie bei der direkten Summe festgelegt

$$+m = +n : \mathbb{Z} \iff m = n : \mathbb{N}$$
$$-m = -n : \mathbb{Z} \iff m = n : \mathbb{N}$$

aber

$$+m = -n : \mathbb{Z} \iff m = 0 = n : \mathbb{N}.$$

(im Gegensatz zu $\operatorname{Left} x = \operatorname{Right} y : A + B \iff \bot$).

Die üblichen Operationen lassen sich dann durch Fallunterscheidung (either) definieren. Dabei ist zu beachten, daß für $f, g\colon \mathbb{N} \to C$ die Konstruktion either $f\ g$ nur dann eine Funktion $\mathbb{Z} \to C$ ergibt, wenn $f(0) = g(0)$. Beispiel:

$$|.| = \operatorname{either} \operatorname{id} \operatorname{id},$$

wobei $\operatorname{id}\colon \mathbb{N} \to \mathbb{N}$ die identische Funktion bezeichnet. Oder, ausgeschrieben:

$$|+n| = n : \mathbb{N},$$
$$|-n| = n : \mathbb{N},$$

für alle $n\colon \mathbb{N}$.

Ähnlich ergibt sich die Negation

$$(-) = \operatorname{either}(-)(+) : \mathbb{Z} \times \mathbb{Z}$$

Dabei bezeichnet das $(-)$ auf der linken Seite die unäre Operation, während die Zeichen rechts die Konstruktoren bezeichnen. Ausgeschrieben wird es ein wenig deutlicher:

$$-(+n) = -n : \mathbb{Z}$$
$$-(-n) = +n : \mathbb{Z}$$

Bei der Addition hat man vier Fälle zu unterscheiden. Die binäre Subtraktion ergibt sich dann als $a - b = a + (-b) : \mathbb{Z}$.

Der Minusoperator ist hier schon ziemlich überladen. Er hat mehrere Bedeutungen, abhängig von den Datentypen der Objekte, auf die er angewandt wird:

- $(-)\colon \mathbb{N} \to \mathbb{N} \to \mathbb{Z}$: binärer Konstruktor, gemäß 1. Konstruktionsmethode;
- $(-)\colon \mathbb{Z} \to \mathbb{Z}$: unäre Operation;
- $(-)\colon \mathbb{Z} \to \mathbb{Z} \to \mathbb{Z}$: binäre Operation;
- $(-)\colon \mathbb{N} \to \mathbb{Z}$: unäre Funktion bzw. Konstruktor nach der zweiten Konstruktionsmethode.

Die Überladung ist hier kein Problem, weil sie mit der natürlichen Einbettung der von $\mathbb{N}$ in $\mathbb{Z}$ stets kompatibel ist.

Rationale Zahlen

Die Konstruktion der rationalen Zahlen aus den ganzen erfolgt fast genauso wie die der ganzen aus den natürlichen (erste Variante). Der einzige wesentliche Unterschied besteht darin, daß 0 nicht als Nenner auftreten darf.

2.5.6 Axiom ($\mathbb{Q}$-Introduktionsregel).

$$\frac{p\colon \mathbb{Z} \qquad q\colon \mathbb{Z} \qquad q \neq 0 : \mathbb{Z}}{\frac{p}{q}\colon \mathbb{Z}}\,\mathbb{Q}\mathcal{I} \tag{2.7}$$

Die Gleichheit wird dann genau analog definiert.

2.5.7 Definition. Für $\frac{p_1}{q_1}\colon \mathbb{Q}$ und $\frac{p_2}{q_2}\colon \mathbb{Q}$ definieren wir

$$\frac{p_1}{q_1} = \frac{p_2}{q_2}\colon \mathbb{Q} \iff p_1 q_2 = p_2 q_1 : \mathbb{Z}.$$

Die Definition der Multiplikation entspricht dann exakt der Definition der Addition für ganze Zahlen, während für die Addition die etwas kompliziertere Version mit dem gemeinsamen Nenner erforderlich ist.

Dem Kontoausgleich bei den ganzen Zahlen entspricht bei den rationalen Zahlen das Kürzen. Entsprechend erhält man für jede rationale Zahl durch Kürzen eine kanonische Form,

2.5.8 Satz. *Zu jeder rationalen Zahl $r\colon \mathbb{Q}$ gibt es genau ein Paar von ganzen Zahlen $p, q\colon \mathbb{Z}$ sodaß*

$$r = \frac{p}{q} : \mathbb{Q} \qquad\qquad q > 0 \qquad\qquad p, q \text{ sind teilerfremd.}$$

Auch eine der 2. Konstruktionsmethode für ganze Zahlen entsprechende Variante gibt es hier: dabei rechnet man ausschließlich mit gekürzten Brüchen.

Ähnlich wie das Minuszeichen bei den ganzen Zahlen, ist hier der Bruchstrich überladen: Als $\mathbb{Z} \to \mathbb{Z} \to \mathbb{Q}$ ist er ein Konstruktor, als $\mathbb{Q} \to \mathbb{Q} \to \mathbb{Q}$ ist er eine binäre Operation (Doppelbruch).

Die rationalen Zahlen sind das Standardbeispiel für einen Körper; d.h. alle vier Grundrechnungsarten können wie gewohnt ausgeführt werden.

Irrationale Zahlen

Schon seit der Antike ist bekannt: es läßt sich kein $r\colon \mathbb{Q}$ finden, sodaß $r^2 = 2$, was geometrisch bedeutet, daß sich kein Zahlenverhältnis gibt, welches die Diagonale eines Quadrats mit dessen Seitenlänge in Verbindung bringt. Dies ist leicht zu erkennen. Sei etwa $r = \frac{p}{q}$, dann ist $r^2 = 2$ gleichbedeutend zu $p^2 = 2q^2$, $q \neq 0$. Die höchste Zweierpotenz, welche p^2 teilt, ist gerade; dasselbe gilt für q^2, weshalb sie für $2q^2$ ungerade sein muß. Weil aber $2q^2 = p^2$ ist, folgt eine Widerspruch.

Man kann das Problem lösen, indem man $\sqrt{2}$ als weitere (irrationale) Zahl zuläßt, unter Fortführung der üblichen Rechengesetze, was durchaus möglich ist. Allerdings fehlen dann immer noch die Zahlen $\sqrt{3}$, $\sqrt[3]{2}$. Man kann auch als Zahlen alle Lösungen von Gleichungen beliebig hohen Grades mit rationalen Koeffizienten zulassen. Die Gesamtheit aller Lösungung solcher Gleichungen bildet den Körper $\mathbb{A}$ der algebraischen Zahlen. Auch in diesem können alle Grundrechnungsarten durchgeführt werden, allerdings nur mit großem Aufwand. Aber selbst unter diesen gibt es z. B. keine Zahl, welche das Verhältnis des Umfangs eines Kreises zu dessen Diagonale beschreibt. D.h., π ist keine algebraische Zahl, sondern *transzendent* (konkret: $a_n\pi^n + \ldots + a_0 = 0$ ist nur möglich, wenn alle Koeffizienten $a_i = 0$). Dasselbe gilt für die Eulersche Zahl e und viele weitere Zahlen, welche bei der Bildung von Grenzwerten auftreten. Es gibt also immer noch „Lücken“ auf der Zahlengeraden.

Reelle Zahlen

Alle Lücken auf der Zahlengeraden lassen sich schließen, wenn man die Lücken selbst als Zahlen auffaßt. Dies führt zur Konstruktion der reellen Zahlen $\mathbb{R}$. Auf die Details sei hier verzichtet. Wesentlich ist, daß man jede reelle Zahl mit rationalen Zahlen beliebig genau approximieren kann. D.h.: Sei $x\colon \mathbb{R}$; dann gibt es zu jedem $\varepsilon > 0$ ein $q\colon \mathbb{Q}$, sodaß

$$|x - q| \leq \varepsilon$$

Ist etwa $\varepsilon = 10^{-5}$, so hat man eine Approximation auf 5 Nachkomma-Dezimalstellen. Es bietet sich daher an, reelle Zahlen als Folge immer besser werdender Approximationen darzustellen. Auf diese Weise entsteht erneut ein Körper, d.h. auch innerhalb der reellen Zahlen können alle Grundrechnungsarten uneingeschränkt und mit den gewohnten Rechengesetzen ausgeführt werden. Ebenso die diversen Grenzwertübergänge, die in der Infinitesimalrechnung unerläßlich sind. Freilich muß man sich diesen Fortschritt teuer erkaufen: für reelle Zahlen ist die Gleichheit unentscheidbar, d.h. es gibt kein allgemeines Verfahren, welche von zwei reellen Zahlen feststellt, ob sie gleich sind oder nicht; oder: die Aussage

$$x = y : \mathbb{R} + x \neq y : \mathbb{R}$$

ist ohne Prinzip des ausgeschlossenen Dritten nicht mehr allgemeingültig (d.h. die Gleichheit von reellen Zahlen ist nicht entscheidbar), und das exakte Rechnen mit reellen Zahlen ist sehr rechenaufwendig (aber nicht unmöglich!!!). Computer verwenden daher heutzutage immer noch Gleitkommazahlen (floating point) statt denn echten reellen Zahlen. Diese bestehen nur aus einer einzigen Approximation (etwa auf 15 Stellen). Auf diese Weise kann schnell gerechnet werden. Allerdings können sich die Rundungsfehler sehr übel auswirken, was zu

bösen Überraschungen führen kann. Insbesondere erfüllen die Gleitkommazahlen *nicht* die üblichen Rechengesetze, vor allem nicht das Assoziativgesetz. Die Nicht-Entscheidbarkeit der Gleichheit für reelle Zahlen wirkt sich so aus, daß Gleitkommazahlen, die sehr nahe beisammen liegen (innerhalb des Rundungsfehlers, mit dem zu rechnen ist), überhaupt nicht verglichen werden können (bzw. dürften). Jeder Test der Form $x = y$ für Gleitkommazahlen ist sinnlos; stattdessen ist stets festzustellen, ob $|x - y| < \varepsilon$, was bedeutet, daß x und y gleich sein *könnten* (wobei ε dem zu erwartenden Rundungsfehler entspricht). Um die Gleichheit definitiv feststellen zu können, müßten unendlich viele Stellen verglichen werden.

2.6 Kombinatorik

Permutationen

2.6.1 Definition. Sei L eine Liste der Länge n, deren Elemente alle verschieden sind. Eine Liste die daraus durch Umordnen der Elemente entsteht, heißt eine *Permutation* von L. Die Anzahl aller möglichen Permutationen von L hängt nur von n ab (nicht von den konkreten Elementen in der Liste). Sie heißt *Fakultät* von n und wird meist mit $n!$ bezeichnet.

In der leeren Liste kann nicht viel umgeordnet werden: es gibt genau eine Permutation davon. Von einer Liste mit $n+1$ Elementen bilden wir zuerst alle Permutationen der ersten n Elemente. Für das letzte Element bleiben dann genau $n+1$ Stellen, an denen es eingefügt werden kann. Damit erhalten wir die Beziehungen

$$0! = 1,$$
$$(n+1)! = (n+1) \cdot n!,$$

welche als rekursive Definition verwendet werden können.

Kombinationen

2.6.2 Definition. Sei A eine Menge mit n Elementen. Die Anzahl der k-elementigen Teilmengen (*Kombinationen*) von A hängt nur von n ab (nicht von A), heißt *Binomialkoeffizient* und wird meist mit $\binom{n}{k}$ bezeichnet.

Jede Menge hat genau eine 0-elementige Teilmenge. Die leere Menge hat sonst keine Teilmenge. Sei nun eine Menge mit $n+1$ Elementen gegeben. Wir zählen zuerst die $(k+1)$-elementigen Teilmengen, welche das letzte Element nicht enthalten; dies sind $\binom{n}{k+1}$. Dann zählen wir die $(k+1)$-elementigen Teilmengen, welche das letzte Element sehr wohl enthalten. Diese entsprechen genau den k-elementigen Teilmengen, welche das letzte Element nicht enthalten. Daraus ergibt sich

$$\binom{n}{0} = 1 \qquad \binom{0}{k+1} = 0 \qquad \binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k}$$

2.6.3 Satz. *Seien $n, k \colon \mathbb{N}$ mit $0 \le k \le n$. Dann gilt*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Beweis. Induktion nach n. □

2.6.4 THEOREM (Binomischer Lehrsatz). *Sei $n\colon \mathbb{N}$ und a, b aus irgendeiner Menge, in der die üblichen Rechenregeln für $+$ und $*$ gelten. Dann gilt*

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Beweis. Induktion nach n. □

Kapitel 3

Rekursion

3.1 Natürliche Zahlen

Konstruktoren

Eine grundlegende Eigenschaft der *natürlichen Zahlen* ist, daß es zu jeder natürlichen Zahl eine weitere (größere) gibt, deren *Nachfolger*. Ebenso grundlegend ist, daß es überhaupt eine natürliche Zahl gibt, etwa die *Null*. Das bedeutet:

Um eine natürliche Zahl zu konstruieren, kann man sie entweder als Nachfolger einer bereits konstruierten natürlichen Zahl konstruieren, oder man konstruiert die spezielle natürliche Zahl 0. Dies läßt sich formal durch die folgenden Einführungsregeln beschreiben.

3.1.1 Axiom.

$$\frac{k\colon \mathbb{N}}{Sk\colon \mathbb{N}}\mathbb{N}\mathcal{I}_S \qquad\qquad \frac{}{0\colon \mathbb{N}}\mathbb{N}\mathcal{I}_0$$

So wie bei der Disjunktion bzw. direkten Summe gibt es zwei Einführungsregeln und entsprechend die zwei Konstruktoren 0 und S:

$$\mathbb{N} \qquad\qquad S\colon \mathbb{N} \to \mathbb{N}.$$

Natürliche Zahlen sind somit: 0, $S0$, $S(S0)$, $S(S(S0))$, ... Üblicherweise verwendet man die Abkürzungen: $1 := S0$, $2 := S(S0)$, $3 := S(S(S0))$,.... Insbesondere gilt somit $2 = S1$, $3 = S2$, $4 = S3$, ...

Peano-Induktion

Eine weitere grundlegende Eigenschaft der natürlichen Zahlen ist, daß man mit den oben eingeführten Konstruktionsmethoden alle natürlichen Zahlen konstruieren kann. Dies läßt sich wieder durch eine entsprechende Eliminationsregel beschreiben, allerdings, so wie bei der Disjunktion, in einer etwas komplizierteren Form mit Fallunterscheidung. Dazu sei $A[n]$ irgendeine Ausage, in der die natürliche Zahl n vorkommen kann. Um zu zeigen, daß sie für jedes beliebige $n\colon \mathbb{N}$ gilt, kann man folgendermaßen vorgehen: Wir zeigen, daß, wenn die Aussage $A[k]$ für irgendein $k\colon \mathbb{N}$ gilt, dann gilt sie auch für dessen Nachfolger, d.h. es muß dann auch $A[Sk]$ gelten. Außerdem zeigen wir $A[0]$. Damit haben wir

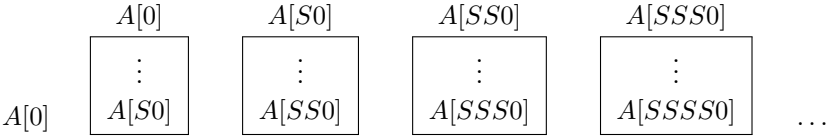
die Aussage für jedes n gezeigt, welches gemäß Einführungsregeln konstruiert wurde. Wenn es keine weiteren Möglichkeiten geben soll, natürliche Zahlen zu konstruieren, dann ist damit tatsächlich $A[n]$ für alle n bewiesen.

3.1.2 Axiom (Peano-Induktion).

$$\frac{\bigwedge_{k\colon \mathbb{N}} A[k] \implies A[Sk] \qquad A[0]}{\bigwedge_{n\colon \mathbb{N}} A[n]}\,\mathbb{N}\mathcal{E}$$

Die erste Voraussetzung heißt *Induktionsschluß*, die zweite *Induktionsanfang*.

Man kann das Induktionsprinzip auch folgendermaßen begründen: Wenn $A[n]$ für alle $n\colon \mathbb{N}$ gezeigt werden sollte, und wenn es außer $0, S0, SS0, SSS0, SSSS0, .$ keine weiteren natürlichen Zahlen gibt, dann reicht es, jede der Aussagen $A[0]$, $A[S0]$, $A[SS0]$, $A[SSS0]$, $A[SSSS0]$, usw beweisen. Haben wir aber $A[0]$ bereits bewiesen, so kann man diese Tatsache im Beweis von $A[S0]$ ausnutzen, d.h. man muß nur noch zeigen, daß man $A[S0]$ aus $A[0]$ herleiten kann. Ebenso kann man dann beim Beweis von $A[SS0]$ verwenden, daß man bereits $A[S0]$ gezeigt hat. Wir gehen also folgendermaßen vor:



Dabei sind immer noch unendlich viele Beweise zu führen. Diese sind aber alles spezielle Fälle des Induktionsschlusses.

Selektoren

Wir führen eine Bezeichnung für den Beweis der Konklusion in der Eliminationsregel ein.

3.1.3 Definition.

$$\frac{\varphi\colon \bigwedge_{k\colon \mathbb{N}} (A[k] \implies A[Sk]) \qquad a\colon A[0]}{\mathrm{rec}\,\varphi\,x\colon \bigwedge_{n\colon \mathbb{N}} A[n]}\,\mathbb{N}\mathcal{E}$$

Das φ ordnet jedem $k\colon \mathbb{N}$ und jedem $x\colon A[k]$ ein $\varphi_k(x)\colon A[Sk]$ zu. Der erste Parameter wird der Übersichtlichkeit halber als Index geschrieben. Man kann das auch als $\varphi_k\colon A[k] \implies A[Sk]$ lesen.

Gleichheit

Wie bei den allgemeinen Konstruktionen für Mengen, ergänzen wir auch hier die Regeln um den Gleichheitsbegriff.

3.1.4 Axiom ($\mathbb{N}^=$-Elimination).

$$\frac{\varphi = \psi : \prod_{k\colon \mathbb{N}} (A[k] \to A[Sk]) \qquad a = b : A[0]}{\operatorname{rec}\varphi\, a = \operatorname{rec}\varphi\, a : \prod_{n\colon \mathbb{N}} A[n]}\,\mathbb{N}\mathcal{E}^=$$

Dies bedeutet insbesondere, daß der binäre Operator

$$\operatorname{rec}\colon \prod_{k\colon \mathbb{N}} (A[k] \to A[Sk]) \to A[0] \to \prod_{n\colon \mathbb{N}} A[n]$$

wohldefiniert ist.

Durch verschiedenartige Anwendung dieser Regeln ergeben sich wieder die zugehörigen β- und η-Regeln:

3.1.5 Axiom.

$$\frac{\varphi\colon \prod_{k\colon \mathbb{N}} (A[k] \to A[Sk]) \qquad a\colon A[0] \qquad n\colon \mathbb{N}}{\operatorname{rec}\varphi\, a\, (Sn) = \varphi_n(\operatorname{rec}\varphi\, a\, n) : A[Sn]}\,\mathbb{N}\beta_S \qquad \frac{a\colon A[0]}{\operatorname{rec}\varphi\, a\, 0 = a : A[0]}\,\mathbb{N}\beta_0$$

3.1.6 Satz. *Sei $f = \operatorname{rec}\varphi\, a$, mit Datentypen wie oben. Dann gilt*

$$\begin{aligned} f(0) &= a, \\ f(Sn) &= \varphi_n(f(n)), \end{aligned}$$

und f ist die einzige Funktion diese beiden Gleichungen erfüllt.

Beweis. Wir setzen die Definition ein und vereinfachen mit β-Regeln:

$$\begin{aligned} f(0) &= \operatorname{rec}\varphi\, a\, 0 && \text{(Vorraussetzung)} \\ &= a && (\mathbb{N}\beta_0); \\ f(Sn) &= \operatorname{rec}\varphi\, a\, (Sn) && \text{(Vorraussetzung)} \\ &= \varphi_n(\operatorname{rec}\varphi\, a\, n) && (\mathbb{N}\beta_S) \\ &= \varphi_n(f(n)) && \text{(Voraussetzung).} \end{aligned}$$

Damit ist der erste Teil bewiesen.

Um die Eindeutigkeit zu zeigen, sei f irgendeine Funktion, welche diese Gleichungen erfüllt. Dazu zeigen wir, daß $f = \operatorname{rec}\varphi\, a$. Gemäß Extensionalität heißt das, daß wir für jedes $n\colon \mathbb{N}$ zu zeigen haben: $f(n) = \operatorname{rec}\varphi\, a\, n$. Dies erledigen wir mit Induktion nach n.

Induktionsanfang: zu zeigen ist: $f(0) = \operatorname{rec}\varphi\, a\, 0$, was sofort folgt.

Induktionsschritt: Wir nehmen an, daß $f(n) = \operatorname{rec}\varphi\, a\, n$, und haben zu zeigen: $f(Sn) = \operatorname{rec}\varphi\, a\, (Sn)$. Auch dieser Schritt besteht in einfachem Nachrechen:

$$\begin{aligned} f(Sn) &= \varphi_n(f(n)) && \text{(Annahme)} \\ &= \varphi_n(\operatorname{rec}\varphi\, a\, n) && \text{(Induktionshypothese)} \\ &= \operatorname{rec}\varphi\, a\, (Sn) && (\mathbb{N}\mathcal{I}\beta_S). \end{aligned}$$

□

Der erste Parameter des Rekursionsoperators (φ) hat einen *abhängigen Datentyp*: Der Typ von φ_k hängt von k ab. So wichtig diese Form beim Beweisen mit Induktion ist, so wenig unterstützen heutige Programmiersprachen diesen Datentyp. Für den Fall, daß in der Aussage A gar kein k vorkommt, erhalten wir eine deutlich einfachere Variante der Rekursion, insbesondere wenn wir uns auch noch auf den Fall beschränken, daß auch φ nicht von k abhängt. Dann kann man den Rekursionsoperator auch als

$$\mathrm{rec}\colon (A \to A) \to A \to \mathbb{N} \to A$$

betrachten. Die β- und η-Regeln lauten dann:

$$\frac{\varphi\colon A \to A \qquad a\colon A \qquad n\colon \mathbb{N}}{\mathrm{rec}\,\varphi\,a\,(Sn) = \varphi(\mathrm{rec}\,\varphi\,a\,n) : A}\mathbb{N}\beta_S \qquad\qquad \frac{a\colon A}{\mathrm{rec}\,\varphi\,a\,0 = a : A}\mathbb{N}\beta_0$$

$$\frac{n\colon \mathbb{N}}{\mathrm{rec}\,S\,0\,n = n : \mathbb{N}}\mathbb{N}\eta$$

Intuitiv bedeutet $\mathrm{rec}\,\varphi\,a\,n$ damit einfach, daß die Funktion φ auf a n-mal angewendet wird, d.h.

$$\mathrm{rec}\,\varphi\,a\,n = \underbrace{\varphi(\ldots\varphi(\varphi}_{n\text{-mal}}(a))\ldots)$$

Mit dieser Betrachtungsweise ist auch die allgemeinere Variante besser zu verstehen:

$$\mathrm{rec}\,\varphi\,a\,n = \varphi_{n-1}(\ldots\varphi_1(\varphi_0(a))\ldots)$$

Vorgänger

3.1.7 Lemma. *Für alle $m, n\colon \mathbb{N}$ gilt*

$$Sm = Sn \implies m = n.$$

Beweis. Wir definieren die *Vorgänger-Funktion* (*Predecessor*). Sie sollte die Gleichungen

$$\begin{aligned} P(Sn) &= n \\ P\ \ 0\ \ &= 0 \end{aligned}$$

erfüllen. Tatsächlich läßt sich eine solche Funktion darstellen als

$$P = \mathrm{rec}(n \mapsto x \mapsto n)\,0.$$

Wir nehmen nun $Sm = Sn$ an und wenden auf beide Seiten P an. Damit haben wir $P(Sm) = P(Sn)$ hergeleitet. Wegen $P(Sm) = m$ und $P(Sn) = n$ erhalten wir daher $m = n$, womit die Behauptung $Sm = Sn \implies m = n$ bewiesen ist. □

3.1.8 Satz. *Für alle $m\colon \mathbb{N}$ gilt*

$$0 = Sm \implies 0 = S0.$$

Beweis. Wir beweisen mit Induktion nach m.

Induktionsanfang: Zu zeigen ist: $0 = S0 \implies 0 = S0$, was trivialerweise erfüllt ist.

Induktionshypothese: $0 = Sm \implies 0 = S0$. Zu zeigen: $0 = S(Sm) \implies 0 = S0$.

Wir nehmen also $0 = S(Sm)$ an und haben $0 = S0$ zu zeigen. Dazu wenden wir die Vorgänger-Funktion P auf beide Seiten der Voraussetzung an, und erhalten

$$P0 = P(S(Sm)).$$

Da $P0 = 0$ und $P(S(Sm)) = Sm$ ist, erhalten wir somit $0 = Sm$, was gerade die Voraussetzung in der Induktionshypothese ist. Daher gilt auch $0 = S0$. □

Notation: Wir verwenden die Abkürzung $1 := S0$.

3.1.9 Satz. *Für alle $m, n \colon \mathbb{N}$ gilt*

$$m = n \vee (m = n \implies 0 = 1).$$

Beweis. Wir führen eine Induktion nach m.

Induktionsanfang: Wir haben zu zeigen: $0 = n \vee (0 = n \implies 0 = 1)$. Dies wird im Lemma 3.1.10 unten gezeigt.

Induktionshypothese: $m = n \vee (m = n \implies 0 = 1)$

Zu zeigen: $Sm = n \vee (Sm = n \implies 0 = 1)$.

Wir zeigen wieder mit Induktion, diesmal nach n: Induktionsanfang: Zu zeigen: $Sm = 0 \vee (Sm = 0 \implies 0 = 1)$. Der zweite Teil dieser Aussage wurde im Satz 3.1.8 gezeigt.

Induktionshypothese: $Sm = n \vee (Sm = n \implies 0 = 1)$.

Zu zeigen: $Sm = Sn \vee (Sm = Sn \implies 0 = 1)$.

Die neue Induktionshypothese brauchen wir nicht, wohl aber die vorige. Da diese eine Disjunktion ist, gehen wir mit Fallunterscheidung vor.

Fall $m = n$: Dann ist auch $Sm = Sn$, und damit der erste Teil der Behauptung bewiesen.

Fall $m = n \implies 0 = 1$: Dann versuchen wir den zweiten Teil zu beweisen. Da dieser eine Implikation ist, nehmen wir $Sm = Sn$ an und versuchen damit $0 = 1$ herzuleiten. Wir wenden beiderseits die Vorgänger-Funktion an $P(Sm) = P(Sn)$, woraus sich durch deren Definition $m = n$ ergibt, und die Voraussetzung für diesen Fall angewandt werden kann. □

3.1.10 Lemma. *Sei $n \colon \mathbb{N}$; dann gilt*

$$0 = n \vee (0 = n \implies 0 = 1).$$

Beweis. Induktion nach n.

Induktionsanfang:

Zu zeigen: $0 = 0 \vee (0 = 0 \implies 0 = 1)$.

Offensichtlich gilt der linke Teil.

Induktionshypothese: $m = 0 \vee (m = 0 \implies 0 = 1)$

Zu zeigen: $Sm = 0 \vee (Sm = 0 \implies 0 = 1)$.

Wegen Satz 3.1.8 gilt der rechte Teil. (Die Induktionshypothese wurde nicht gebraucht.) □

3.1.11 Satz. *Seien $P, Q\colon \Omega$ beliebige Aussagen. Dann gilt*

$$0 = 1 \implies (P \iff Q)$$

Beweis. Wenn $0 = 1$, so gilt wegen der Wohldefiniertheit von rec auch

$$\operatorname{rec} \varphi\, P\, 0 = \operatorname{rec} \varphi\, P\, 1 : \Omega,$$

wobei $\varphi\colon \Omega \to \Omega$ eine konstante Funktion ist: $\varphi(R) = Q$. Da aber $\operatorname{rec} \varphi\, P\, 0 = P : \Omega$ und $\operatorname{rec} \varphi\, P\, 1 = Q : \Omega$ ist, muß auch $P = Q : \Omega$ sein, d.h. $P \iff Q$. □

Addition

Mittels Rekursion kann man Funktionen auf den natürlichen Zahlen definieren. Die einfachste ist $\operatorname{rec} S$. Dies ist eine Funktion in zwei Variablen. Bei $\operatorname{rec} S\, m\, n$ wird von n ausgehend m mal der Nachfolger gebildet, was offensichtlich der Summe von m und n entspricht. Die β-Reduktionen Regeln ergeben damit

$$\begin{aligned} \operatorname{rec} S\, m\, 0 &= n, \\ \operatorname{rec} S\, m\, (Sn) &= S(\operatorname{rec} S\, m\, n). \end{aligned}$$

Dies läßt sich leichter lesen, wenn statt $\operatorname{rec} S$ der binäre Operator $(+)$ verwendet wird:

$$\begin{aligned} m + \ 0 \ &= n, \\ m + Sn &= S(m + n). \end{aligned}$$

3.1.12 Definition. Die lineare Ordnung für natürliche Zahlen $m, n\colon \mathbb{N}$ wird festgelegt durch

$$m \leq n \iff \bigvee_{p\colon \mathbb{N}} m + p = n.$$

3.1.13 Satz. *Die lineare Ordung $\leq$ auf den natürlichen Zahlen ist eine Ordungsrelation.*

3.1.14 Satz. *Seien $a, b, c\colon \mathbb{N}$; dann gilt*

$$a \leq b \implies a + c \leq b + c$$

3.1.15 Satz. *Für $a, b\colon \mathbb{N}$ gilt*

$$\begin{aligned} 0 &\leq b \\ \neg(Sa &\leq 0) \\ Sa \leq Sb &\iff a \leq b \end{aligned}$$

3.1.16 Satz. *Für alle $a, b\colon \mathbb{N}$ gilt*

$$a \leq b \vee \neg(a \leq b)$$

3.1.17 Satz. *Für alle $a, b\colon \mathbb{N}$ gilt*

$$a \leq b \vee b \leq a$$

3.2 Teilbarkeit

3.2.1 Definition. Eine Zahl $d\colon \mathbb{Z}$ heißt ein *Teiler* von $n\colon \mathbb{Z}$ wenn es ein $q\colon \mathbb{Z}$ gibt sodaß $n = qd$. Man sagt dann auch d *teilt* n, und schreibt $d \mid n$. Es gilt also

$$d \mid n \iff \bigvee_{q\colon \mathbb{Z}} n = qd. \tag{3.1}$$

Man beachte die Analogie zur Definition von $m \leq n$.

3.2.2 Satz. *Teilbarkeit ist transitiv, d.h. für alle $a, b, c\colon \mathbb{Z}$ gilt*

$$a \mid b \wedge b \mid c \implies a \mid c$$

Beweis. Seien $a \mid b$ und $b \mid c$. Gemäß der Definition der Teilbarkeit gibt es daher $x, y\colon \mathbb{Z}$, sodaß $b = x \cdot a$ und $c = y \cdot b$. Einsetzen der ersten Gleichheit in die zweite ergibt damit $c = y \cdot (x \cdot a) = (y \cdot x) \cdot a$ (Assoziativgesetz), was wieder genau er Definition von $a \mid c$ entspricht. □

3.2.3 Satz. *Seinen $d, n, m, z\colon \mathbb{Z}$ und $d \mid n$, $d \mid m$. Dann gelten auch $d \mid n + m$, $d \mid n - m$, und $d \mid zn$.*

3.2.4 Bemerkung. Die Teilbarkeitsrelation in $\mathbb{Z}$ ist nicht definit: für jede ganze Zahl $a\colon \mathbb{Z}$ gelten zwar $a \mid (-a)$ und $(-a) \mid a$, aber nicht $a = -a$ (außer wenn $a = 0$).

3.2.5 Satz. *Seine $a, b\colon \mathbb{Z}$; dann gilt*

$$a \mid b \wedge b \mid a \iff |a| = |b|\,.$$

Eigentlich kann man sich bei der Teilbarkeit auf die von positiven Zahlen beschränken, den es gilt:

3.2.6 Satz. *Seien $a, b\colon \mathbb{Z}$. Dann gilt*

$$a \mid b \iff |a| \mid |b|$$

Schränkt man die Teilbarkeit auf die natürlichen Zahlen ein, so erhält man tatsächlich eine Ordungsrelation.

3.2.7 Bemerkung. Teilbarkeit für natürliche Zahlen $m, n\colon \mathbb{N}$ wird am natürlichsten als

$$d \mid n \iff \bigvee_{q\colon \mathbb{N}} n = q \cdot d. \tag{3.2}$$

Identifiziert man aber $m, n\colon \mathbb{N}$ mit den ganzen Zahlen $m, n\colon \mathbb{Z}$, erhält man jedoch aus der Definition der Teilbarkeit für ganze Zahlen

$$d \mid n \iff \bigvee_{q\colon \mathbb{Z}} n = q \cdot d,$$

was ein wenig anders aussieht (man beachte, daß hier $q\colon \mathbb{Z}$ statt $q\colon \mathbb{N}$). Erst wenn nachgewiesen ist, daß beide Definition übereinstimmen (was in diesem Fall leicht ist) kann man natürliche Zahlen auch im Zusammenhang mit Teilbarkeit in der üblichen Weise mit ganzen Zahlen identifizieren.

3.2.8 Definition. Man nennt d einen *gemeinsamen Teiler* von $a, b\colon \mathbb{Z}$, wenn $d \mid n$ und $d \mid m$ gilt.

3.2.9 Bemerkung. Man beachte: Jede Zahl teilt 0, denn es gilt stets $0 = x \cdot 0$, woraus unmittelbar $x \mid 0$ folgt. Es gilt sogar $0 \mid 0$. Andere Zahlen teilt 0 jedoch nicht, denn $0 \mid a$ bedeutet ja gerade daß $a = x \cdot 0$, woraus aber $a = 0$ folgt.

0 ist damit die größte Zahl, wenn diese nach Teilbarkeit geordnet werden. Die Zahl 1 dagegen teilt jede andere Zahl, und ist damit die kleinste.

3.2.10 Satz. *Ist d ein gemeinsamer Teiler von m und n, und t ein Teiler von d, dann ist auch t ein gemeinsamer Teiler von m und n.*

Beweis. Gemäß Voraussetzung gibt es Zahlen $a, b, c\colon \mathbb{Z}$, sodaß

$$m = adn = bdd = ct;$$

einfaches Einsetzen ergibt dann mit der Assoziativität

$$m = (ac)tn = (bc)t,$$

womit die Behauptung unmittelbar folgt. □

3.2.11 Definition. Ein gemeinsamer Teiler d von a und b heißt ein *größter gemeinsamer Teiler*, wenn jeder andere gemeinsame Teiler t von a und b bereits ein Teiler von d ist.

Man beachte, daß sich das *größer* in dieser Definition auf die Teilbarkeitsrelation bezieht, und vorerst nichts mit der normalen linearen Größerrelation zu tun hat. Es ist damit auch nicht von vorneherein klar, ob es überhaupt größte gemeinsame Teiler gibt, und ob diese eindeutig bestimmt sind. Letzteres ist aber einfach:

3.2.12 Satz. *Seien $a, b\colon \mathbb{Z}$. Zwei größte gemeinsame Teiler von a und b stimmen bis auf das Vorzeichen überein.*

Als *den* größten gemeinsamen Teiler bezeichnet man dann meistens die positive Variante.

Manchmal ist der größte gemeinsame Teiler leicht zu bestimmen.

3.2.13 Satz. *Ist d ein Teiler von a, so ist d ein größter gemeinsamer Teiler von d und a.*

Beweis. Weil $d \mid d$ (Teilbarkeit ist reflexiv) ist d klarerweise ein gemeinsamer Tiler von d und a. Ist nun t ein weiterer gemeinsamer Teiler von d und a, dann ist insbesondere t ein Teiler von d, weshalb d tatsächlich ein größter gemeinsame Teiler ist. □

3.2.14 Lemma. *Seien $m, n\colon \mathbb{Z}$ und $m = qn + r$, mit $q, r\colon \mathbb{Z}$. Dann ist ein $d\colon \mathbb{Z}$ genau dann ein gemeinsamer Teiler von von m und n wenn d ein gemeinsamer Teiler von n und r ist.*

Beweis. Wir nehmen an, daß $d \mid m$ und $d \mid n$, und zeigen damit daß $d \mid n$ und $d \mid r$. Der erste Teil steht schon in der Voraussetzung. Wegen $r = m - qn$, und weil $d \mid m$ und $d \mid n \mid qn$, gilt wegen Satz 3.2.3 auch $d \mid r$.

Die Umkehrung funktioniert ganz ähnlich. □

3.2.15 Satz. *Seien $a, b\colon \mathbb{Z}$ und $b \neq 0$. Dann gibt es es $q, r\colon \mathbb{Z}$ sodaß*

$$a = q \cdot b + r$$

mit $|r| < |b|$.

q und r sind sogar eindeutig bestimmt, wenn der zulässige Bereich für r auf $|b|$ aufeinanderfolgende Zahlen beschränkt wird, z.B. $0 \leq r < |b|$ oder $-\frac{|b|}{2} < r \leq \frac{|b|}{2}$.

3.2.16 Theorem (Euclidischer Algorithmus). *Seien $m, n\colon \mathbb{Z}$. Dann gibt es genau einen (bis auf das Vorzeichen) größten gemeinsamen Teiler $d\colon \mathbb{N}$. Ferner gibt es ganzzahlige Koeffizienten $x, y\colon \mathbb{Z}$ sodaß*

$$d = xm + yn.$$

Beweis. Wir beweisen mit Induktion nach $|n|$.

Ist $n = 0$, so ist m ein größter gemeinsamer Teiler von m und n. Wegen $d = m = 1m + 0n$ sind auch geeignete Koeffizienten leicht gefunden.

Ist $|n| > 0$, dann gibt es Zahlen q, r mit $m = qn + r$ und $0 \leq r < |n|$. Laut Induktionsvoraussetzung haben n und r einen größten gemeinsamen Teiler d, und dieser läßt sich als $d = xn + yr$ darstellen, mit passenden Koeffizienten $x, y\colon \mathbb{Z}$. Wegen Lemma 3.2.14 ist aber d auch ein größter gemeinsamer Teiler von m und n. Weiters gilt $d = xn + yr = xn + y(m - qn) = ym + (x - yq)n$. Damit sind auch für n und m passende Koeffizienten gefunden. □

Dieser Beweis liefert ein einfaches und trotzdem recht effizientes Verfahren, um den größten gemeinsamen Teiler, und auch passende Koeffizienten zu finden. Die einfachere Variante berechnet nur den ggT:

$$\begin{aligned} &\gcd m\, 0 = m; \\ &\gcd m\, n = \gcd n\, r \quad \text{wobei } m = qn + r,\ 0 \leq r < |n|. \end{aligned}$$

Die erweiterte Variante berechnet gleichzeitig auch passende Koeffizienten:

$$\begin{aligned} &\operatorname{xgcd} m\, 0 = m; \\ &\operatorname{xgcd} m\, n = (d, y, x - yq) \quad \text{wobei } m = qn + r,\ 0 \leq r < |n| \\ &\qquad\qquad\qquad\qquad\quad \text{und } (d, x, y) = \operatorname{xgcd} n\, r. \end{aligned}$$

3.2.17 Bemerkung. Die Einschränkung für r bei diesem Algorithmus ist nicht die einzig mögliche. Eine sinnvolle Alternative ist es zum Beispiel, den betragsmäßig kleinsten Rest zu verwenden, d.h. $r \leq \left|\frac{n}{2}\right|$.

3.3 Modulare Arithmetik

3.3.1 Definition. Sei $m\colon \mathbb{N}$; dann heißen ganze Zahlen $a, b\colon \mathbb{Z}$ *kongruent modulo m* (Schreibweise: $a \equiv_m b$) falls m ein Teiler von deren Differenz $a - b$ ist.

Neben der erwähnten sind noch die Schreibweisen: $a \equiv b \pmod m$, oft auch abgekürzt zu $a \equiv b \ (m)$, und auch $a = b \pmod m$ bzw $a = b \ (m)$ üblich. Also:

$$\begin{aligned} m \mid (a-b) \iff a &\equiv_m b \\ a &\equiv b \pmod m \\ a &\equiv b \ (m) \\ a &= b \pmod m \\ a &= b \ (m). \end{aligned}$$

Wenn das m aus dem Zusammenhang klar ist, schreibt man oft auch nur $a \equiv b$ oder sogar $a = b$.

Es ist ist leicht nachzurechnen, daß die Kongruenz modulo m stets eine Äquivalenzrelation ist. Man kann also den Gleichheitsbegriff auf $\mathbb{Z}$ durch die Kongruenz modulo m ersetzen.

3.3.2 Definition. Sei $m\colon \mathbb{N}$. Dann bezeichnet $\mathbb{Z}_m = \mathbb{Z}/\equiv_m$ die Menge der Restklassen modulo m.

Damit erhalten wir eine weitere Notation für $a \equiv_m b$, nämlich: $a = b : \mathbb{Z}_m$.

3.3.3 Satz. *Die Kongruenz modulo m ist mit der Addition, der Subtraktion und der Multiplikation verträglich, d.h. sind $a \equiv_m b$ und $c \equiv_m d$, so gelten auch*

$$a + c \equiv_m b + d \qquad a - c \equiv_m b - d \qquad a \cdot c \equiv_m b \cdot d.$$

Dieser Satz garantiert, daß diese drei Grundrechnungsarten auch für Restklassen wohldefiniert sind. Man beachte, daß dies für die Division nicht funktioniert: z.B.: $40 \equiv_5 30$ und $10 \equiv_5 15$, aber es gilt nicht $40/10 \equiv_5 30/15$. Dafür kann man auch Potenzieren (mit ganzzahligen Exponenten):

3.3.4 Satz. *Sei $m\colon \mathbb{Z}$, $a = b : \mathbb{Z}_m$, $n\colon \mathbb{N}$; dann gilt*

$$a^n = b^n : \mathbb{Z}_m$$

Somit kann man modulo m ganz normal rechnen, falls keine Divisionen dabei sind.

3.3.5 Satz. *Sei $m\colon \mathbb{N}$ und $a\colon \mathbb{Z}_m$. Dann gibt es genau dann ein $b\colon \mathbb{Z}_m$ mit $ab = 1 : \mathbb{Z}_m$ falls $ggT(a, m) = 1$.*

Beweis. Gilt $ggT(a, m) = 1$, so liefert der erweiterte Euklidische Algorithmus Lösungen $x, y\colon \mathbb{Z}$, sodaß $ax + my = 1$. Diese Gleichung gilt natürlich auch modulo m, und vereinfacht sich in diesem Fall zu $ax = 1 : \mathbb{Z}_m$, womit jedes $b = x : \mathbb{Z}_m$ eine passende Lösung ist.

Für die Umkehrung sei $ab = 1 : \mathbb{Z}_m$. Das heißt aber, daß es ein y gibt, sodaß $ab - 1 = my$, bzw $ab - my = 1$. Jeder gemeinsame eiler von a und m ist damit auch ein Teiler von 1, d.h. $ggT(a, m) = 1$. □

3.3.6 Satz. *Die Berechnung von $a^n : \mathbb{Z}_m$ ist durch sukzessives Quadrieren (und sofortiges Reduzierem modulo m) effizient möglich.*

Beweis. Wir verwenden die Gleichungen;

$$\begin{aligned} a^0 &= 1; \\ a^{2n} &= (a^n)^2; \\ a^{2n+1} &= a(a^n)^2. \end{aligned}$$

Für ganze Zahlen bringt das nichts, weil sie bei jedem Quadrieren doppelt so lange werden. Wenn wir aber in jedem Schritt modulo m reduzieren können, bleiben alle Zwischenergebnisse durch m beschränkt. Ist etwa n in der Größenordung von 2^{1000} (also etwa 300 Dezimalstellen), so muß man damit nur etwa 1000 mal Quadrieren und etwa ebensooft multiplizieren. Mit dem normalen Verfahren würde man dagegen $2^{1000} \approx 10^{300}$ Multiplikationen benötigen, was selbst mit Computern in astronomischer Größe unmöglich wäre. □

3.4 Primzahlen

3.4.1 Definition. Ein natürliche Zahl $n > 1$ heißt *Primzahl* wenn sie keine echten Teiler hat (alsonur 1 und n selbst. 1 ist per Definition keine Primzahl.

3.4.2 Lemma. *Eine natürliche Zahl $p > 1$ ist genau dann eine Primzahl, wenn gilt:*

$$\bigwedge_{n,m\colon \mathbb{N}} \big(p \mid nm \implies p \mid n \vee p \mid m\big).$$

Ein bekanntes Verfahren zur Bestimmung aller Primzahlen ist das **Sieb des Eratosthenes**: Man schreibt alle natürlichen Zahlen ab 2 in einer Liste an. (Freilich ist die Liste eigentlich unendlich lange, aber man kann ja nach Bedarf weiterschreiben.) Die erste Zahl (also 2 wird als Primzahl gekennzeichnet und all ihre nicht-trivialen Vielfachen werden aus der Liste gestrichen (denn diese sind keine Primzahlen). Die nächste (noch nicht gestrichene Zahl ist 3. Auch diese wird als Primzahl gekennzeichnet und ihre Vielfachen gestrichen. Die nächste Zahl nicht gestrichene Zahl ist daher 5, welche ebenfalls als Primzahl gekennzeichnet wird, und deren Vielfache gestrichen werden. Auf diese Weise fährt man unendlich lange fort und findet weiters die Primzahlen $7, 11, 13, 17, 19, 23, \ldots$.

Man beachte: dieses Verfahren läuft unendlich lange. Dies ist aber kein Problem, da dennoch jede einzelne Primzahl nach endlicher Zeit gefunden wird.

Um zu Testen, ob eine Zahl n prim ist, kann man nachsehen, ob sie in der vom Sieb des Eratosthenes erzeugten Liste aufscheint (nie gestrichen wird). Oder man testet alle Zahlen bis zu ca $\sqrt{n}$, ob sie n teilen (wenn nicht, ist n eine Primzahl). Beide Verfahren sind für große Zahlen (z.B. 300-stellige Zahlen) allerdings nicht brauchbar. In der Praxis verwendet man probabilistische Tests, welche bedeutend schneller sind, und für die ausreichend zuverlässig arbeiten (siehe unten).

3.4.3 Theorem (Hauptsatz der Arithmetik). *Jede natürliche Zahl n hat eine eindeutige Zerlegung in Primfaktoren*

$$n = p_1^{k_1} \ldots p_m^{k_m},$$

wobei alle p_i Primzahlen sind, mit $p_i < p_{i+1}$, und $k_i > 1$.

Leider ist das Auffinden von Primfaktor-Zerlegungen für große Zahlen ist im allgemeinen noch schwieriger als ein Primzahltest, und es gibt auch kein probabilistisches Verfahren. „Leider“ sollte man hier allerdings nicht sagen, denn gerade diese Tatsache, daß sich für große Zahlen die Primzahlzerlegung praktisch nicht bestimmen läßt, ist die Grundlage der gängigsten Verfahren zur Verschlüsselung von Daten (siehe unten) und daher von größter Bedeutung in der Praxis.

3.4.4 Satz. *Seien p, q verschiedene Primzahlen, und für $x, y\colon \mathbb{Z}$ gelte*

$$x \equiv y \pmod{p} \quad \textit{und} \quad x \equiv y \pmod{q}.$$

Dann gilt auch

$$x \equiv y \pmod{pq}.$$

Beweis. Laut Vorausetzung gilt $p \mid (x-y)$ und $q \mid (x,y)$, und daher auch $pq \mid (x-y)$. □

3.4.5 Definition. Sein $n\colon \mathbb{N}$. Dann bezeichnet $\varphi(n)$ die Anzahl der Zahlen $k\colon \mathbb{N}$, $k < n$, welche zu n teilerfremd sind. Die Funktion φ heißt die *Eulersche φ-Funktion*.

Ist p eine Primzahl, so gilt $\varphi(p) = p-1$, und allgemeiner, $\varphi(p^k) = p^{k-1}(p-1)$. Ferner gilt $\varphi(mn) = \varphi(m)\varphi(n)$, falls m und n teilerfremd sind. Damit kann man φ für alle Zahlen leicht berechnen, für die man die Primfaktorzerlegung kennt.

3.4.6 Theorem. *Für $n\colon \mathbb{N}$ und alle $a\colon \mathbb{Z}$ teilerfremd zu n gilt*

$$a^{\varphi(n)} \equiv 1 \pmod{n}.$$

3.4.7 Folgerung. *Seien p, q verschiedene Primzahlen, dann gilt für beliebige $a\colon \mathbb{Z}$*

$$a^p \equiv a \pmod{p} \quad \textit{und} \quad a^{\varphi(pq)+1} \equiv a \pmod{pq}.$$

Beweis. Der erste Teil ist trivial, falls $a \equiv 0 \pmod{p}$; ansonsten folgt er aus dem vorigen Satz mittels Multiplikation mit a.

Für den zweiten Teil zeigen wir zunächst daß die Kongruenz modulo p gilt: Wieder ist der Fall $a \equiv 0 \pmod{p}$ trivial; ansonsten folgt wieder mit dem vorigen Satz:

$$a^{\varphi(pq)+1} = a^{(p-1)(q-1)+1} = \left(a^{p-1}\right)^{q-1} a \equiv 1 \pmod{p}.$$

Analog zeigt man die Kongruenz modulo q, woraus mit Satz 3.4.4 auch die Kongruez modulo pq folgt. □

Die Eigenschaft $a^{p-1} \equiv 1 \pmod{p}$ kann als Primzahltest gebraucht werden: ist sie für ein beliebiges $a < p$ verletzt, so weiß man, daß p keine Primzahl sein kann. Ist sie dagegen für ein a erfüllt, so kann man vermuten, daß sie eine Primzahl ist, gilt sie für mehrere a, dann ist es schon ziemlich wahrscheinlich. Damit hat man einen einfachen brauchbaren Test; allerdings gibt es Nicht-Primzahlen, die nach diesem Test wie Primzahlen aussehen (Pseudoprimzahlen). Besser ist der Miller-Rabin-Test: Dieser funktioniert nach einem ähnlichen Prinzip, ermöglicht es aber die Fehlerwahrscheinlichkeit tatsächlich beliebig klein zu wählen.

RSA-Verfahren

Das RSA-Verfahren ist das bekannteste asymmetrische Verschlüsselungsverfahren. Der Grundgedanke dabei ist, daß jeder Teilnehmer an einem Kommunikationssystem zwei Schlüssel bekommt: einen öffentlichen Schlüssel P, der sowie eine Telefonnummer oder Email-Adresse allen anderen Teilnehmern bekannt gemacht wird, und einen geheimen Schlüssel S, welchen jeder Teilnehmer für sich behält. Ein Schlüssel ist dabei eine Funktion, welche eine Nachricht in eine andere Nachricht umwandelt; dabei sollte aber gelten

$$P \circ S = \mathrm{id} = S \circ P.$$

Wir nehmen jetzt an, daß Alice eine geheime Nachricht a an Bob senden will. Dazu verschlüsselt Alice die Nachricht a mit Bob's öffentlichem Schlüssel (der ist ja allgemein bekannt); dabei entsteht eine unleserliche Nachricht $P(a)$. Diese kann bedenkenlos z.B. via Internet verschickt werden. Bob kann sie aber lesen, denn er (und nur er) hat ja den dazupassenden geheimem Schlüssel S, und es gilt ja $S(P(a)) = a$.

Nun möchte Bob eine Nachricht b verschicken (egal an wen) und dabei sicherstellen, daß jeder garantiert weiß, daß diese tatsächlich von ihm stammt. Dazu verschlüsselt er mit seinem geheimen Schlüssel, d.h. er verschickt $S(b)$. Jeder kann überprüfen, daß diese Nachricht tatsächlich von Bob stammt, da die Anwendung seines öffentlichen Schlüssels auf diese Nachricht $P(S(b)) = b$ ergibt, und nur Bob eine solche Nachricht erzeugen kann.

Das ganze System funktioniert natürlich nur dann, wenn es praktisch unmöglich ist, den geheimen Schlüssel zu finden, selbst wenn man den öffentlichen Schlüssel und vielleicht auch noch zahlreiche mit dem geheimen Schlüssel verschlüsselte Nachrichten kennt.

Der Grundgedanke beim RSA-Verfahren ist es nun, sich zu Nutze zu machen, daß es zwar relativ leicht ist, zwei Zahlen miteinander zu multiplizieren (selbst wenn sie hunderte Stellen haben), es aber umgekehrt im allgemeinen sehr schwierig ist, die Primfaktoren einer solchen Zahl zu bestimmen, obwohl diese ja auch eindeutig bestimmt sind.

Dabei geht man (leicht vereinfacht) folgendermaßen vor:

Man wählt zwei große Primzahlen p und q (512 bit große Zahlen gelten heutzutage als sicher, das sind etwa 150 Dezimalstellen, Primzahltest!), und bilde deren Produkt $n = pq$, sowie $\varphi = (p-1)(q-1)$. Weiters wähle man eine zu φ teilerfremde Zahl e. Damit erhält man den öffentlichen Schlüssel

$$P(a) = a^e \bmod m.$$

Bekanntgegeben werden also m und e, nicht aber die Primzahlen p, q und auch nicht φ.

Um den dazupassenden geheimen Schlüssel zu erhalten, bestimmen wir ein d mit $de \equiv 1 \pmod{\varphi}$, was möglich ist, weil e zu φ teilerfremd gewählt wurde (Erweiterter Euklidischer Algorithmus). Der geheime Schlüssel ist dann

$$S(a) = a^d \bmod m.$$

Tatsächlich gilt wegen $de \equiv 1 \pmod{\varphi}$ und $a^{\varphi} \equiv 1 \pmod{m}$,

$$S(P(a)) \equiv (a^e)^d \equiv a^{ed} \equiv a^1 \equiv 1 \pmod{m}.$$

Ist also $a < m$, so kann a tatsächlich rekonstruiert werden.

Man beachte, daß wir, um d zu bestimmen, die Zahl φ gebraucht haben, welche wiederum leicht aus p und q berechnet werden kann. Es ist aber nicht ersichtlich, wie es direkt aus dem Produkt $m = pq$ bestimmt werden kann. Auf der Annahme, daß dies tatsächlich praktisch unmöglich ist, beruht das RSA-Verfahren. Bisher ist es nicht geknackt worden.

3.5 Listen

Sei im folgenden A ein beliebiger Datentyp.

Introduktion: Listen-Konstruktoren

3.5.1 Definition. Der Datentyp aller *Listen* über A wird mit $\operatorname{List} A$ oder A^* bezeichnet und folgendermaßen definiert: Um ein Element von A^* zu konstuieren, muß man entweder die leere Liste $[]$ konstruieren oder $\operatorname{cons} h\, t$, d.h. ein Element $h\colon A$ zusammen mit einer Liste $t\colon A^*$.

3.5.2 Notation. Der Ausdruck $\operatorname{cons} h\, t$ sollte bedeuten, daß h bei der Liste t vorne eingefügt wird. Wir verwenden dafür daher die suggestivere Notation $h \lhd t$; und ein Ausdruck der Form $a \lhd b \lhd \ell$ ist so wie $a \lhd (b \lhd \ell)$ zu verstehen (andersherum würden die Datentypen gar nicht passen). Weiters verwenden wir oft $[a, b, c, \ldots, z]$ als übersichtlichere Darstellung statt $a \lhd b \lhd c \lhd \ldots \lhd z \lhd []$; also insbesondere $[a] = a \lhd []$ und $[a, b] = a \lhd b \lhd []$.

Ein Beispiel für eine Liste über $\mathbb{N}$ ist $2 \lhd 6 \lhd 1 \lhd 17 \lhd []$ (oder einfacher notiert: $[2, 6, 1, 17]$).

Die Definition können wir formal durch die folgenden Einführungsregeln notieren:

3.5.3 Axiom.

$$\frac{}{[]\colon A^*}\operatorname{List}_{[]}\mathcal{I} \qquad\qquad \frac{h\colon A \qquad t\colon A^*}{h \lhd t\colon A^*}\operatorname{List}_{\lhd}\mathcal{I}$$

Oder wir legen einfach die Datentypen der betroffenen Konstruktoren fest:

$$\begin{aligned} &[]\colon A^*;\\ &\lhd\colon A \to A^* \to A^*. \end{aligned}$$

Elimination: Listen-Induktion

Sei $P\colon A^* \to \Omega$ eine beliebige Eigenschaft auf A^*, d.h. ein Prädikat, welches jeder Liste eine Aussage zuordnet. Wenn es gelingt, jede der Aussagen

$$\begin{aligned} &P([]),\\ &P([]) \implies P([17 \lhd []]),\\ &P(17 \lhd []) \implies P(1 \lhd 17 \lhd []]),\\ &P(1 \lhd 17 \lhd []) \implies P([6 \lhd 1 \lhd 17 \lhd []]) \end{aligned}$$

zu beweisen, dann ist klarerweise (mehrmalige Anwendung der $\Rightarrow\mathcal{E}$) auch die Aussage $P([6 \lhd 1 \lhd 17 \lhd []])$ bewiesen.

Gelingt es nun, für alle $h\colon A$ und $t\colon A^*$ die Aussage

$$P(t) \implies P(h \lhd t)$$

nebst $P([])$ zu beweisen, so sind auch alle Aussagen der Form $P(a \lhd b \lhd \ldots \lhd z \lhd [])$ (bzw. $P[a, b, \ldots, z]$) für beliebige $a, b, \ldots, z\colon A$ bewiesen, d.h. für alle Listen, die sich auch $[]$ und $\lhd$ aufbauen lassen. Die folgende Eliminationsregel besagt daher gerade, daß tatsächlich alle Listen derart entstehen.

3.5.4 Axiom (Primitive Induktion über Listen).

$$\frac{\bigwedge\limits_{t\colon A^*} \big(P(t) \implies \bigwedge\limits_{h:A} P(h \lhd t)\big) \qquad P[[]]}{\bigwedge\limits_{\ell\colon A^*} P[\ell]}\text{List-}\mathcal{E}$$

Das bedeutet: Kann man zeigen, daß eine Aussage für die leere Liste gilt und, wenn sie für eine Liste t gilt, dann auch für jede Liste der Form $h \lhd t$, dann gilt sie für jede Liste.

Selektor: Rekursion

Wir ergänzen die Listen-Eliminationsregel durch den passenden Selektor:

3.5.5 Axiom (Primitive Rekusion über Listen).

$$\frac{\varphi\colon \bigwedge\limits_{t\colon A^*} \big(P(t) \implies \bigwedge\limits_{h:A} P(h \lhd t)\big) \qquad a\colon P([])}{\operatorname{listrec}\varphi\, a\colon \bigwedge\limits_{\ell\colon A^*} P[\ell]}\text{List-}\mathcal{E}$$

$P(t)$ kann dabei durchaus ein beliebiger Datentyp sein (nicht nur eine Aussage). Dieser Selektor hat somit den Typ:

$$\operatorname{listrec}\colon \prod_{t\colon A^*} \big(P(t) \to \prod_{h:A} P(h \lhd t)\big) \to A \to \prod_{\ell\colon A^*} P(\ell),$$

wobei P als beliebiger durch A^* parametrisierter Datentyp zu verstehen ist. Wenn P nicht parametrisiert ist, dann verschwinden die abhängigen Typen (Allaussagen werden zu Implikationen) und alles vereinfacht sich zu

$$\operatorname{listrec}\colon (A^* \to P \to A \to P) \to A \to A^* \to P.$$

Dies ist die übliche Form, wie Funktionen auf Listen rekursiv definiert werden.

Gleichheit

Wenn es für A einen vernünftigen Gleichheitsbegriff gibt, dann sollte dieser auch auf A^* übertragen werden.

Zwei Listen sollten genau dann gleich sein, wenn sie auf dieselbe Weise konstruiert wurden. Aus den Introduktions- und Eliminationsregeln ergibt sich damit:

3.5.6 Axiom.

$$\frac{}{[] = []\colon A^*} \qquad \frac{\ell_1 = \ell_2\colon A^* \qquad a_1 = a_2\colon A}{a_1 \lhd \ell_1 = a_2 \lhd \ell_2\colon A^*}$$

$$\frac{\varphi_1 = \varphi_2\colon \bigwedge\limits_{t\colon A^*} \big(P(t) \implies \bigwedge\limits_{h:A} P(h \lhd t)\big) \qquad a_1 = a_2\colon P([])}{\operatorname{listrec}\varphi_1\, a_1, \operatorname{listrec}\varphi_2\, a_2\colon \bigwedge\limits_{\ell\colon A^*} P[\ell]}\text{List-}\mathcal{E}$$

Wir eruieren nun, wie die Konstruktoren [] und $\lhd$ mit dem Selektor listrec zusammenhängen. Wir verwenden dazu die Notation wie in der Regel für die Listen-Rekursion. Wurde $\bigwedge_{\ell\colon A^*} P(\ell)$ mit Listenrekursion bewiesen, dann bezeichnet der Ausdruck $\mathrm{listrec}\,\phi\,p\,\ell$ stets einen Beweis von $P(\ell)$. Für die leere Liste [] haben wir zwei Beweise: $p\colon P[]$ und $\mathrm{listrec}\,\varphi\,p\,[]$. Ebenso für eine Liste der Form $h \lhd t$: $\mathrm{listrec}\,\varphi\,p\,t$ und $\varphi\,t\,q\,a$, wenn $q\colon P(t)$ ist. Tatsächlich entspricht es gerade der Intention, daß diese jeweils gleich sind.

3.5.7 Axiom (β-Regel für Listen-Rekursion).

$$\begin{aligned}\mathrm{listrec}\,\varphi\,p\,[] &= p\\ \mathrm{listrec}\,\varphi\,p\,(h \lhd t) &= \varphi\,h\,t\,(\mathrm{listrec}\,\varphi\,p\,t)\end{aligned}$$

Ist nun $f = \mathrm{listrec}\,\varphi\,p$, dann gilt somit

$$\begin{aligned}f\,[] &= p\\ f\,(h \lhd t) &= \varphi\,h\,t\,(f\,t)\end{aligned}$$

Dies ist im Prinzip eine Definition mit Fallunterscheidung; allerdings kommt das zu definierende f in der zweiten Gleichung auf beiden Seiten for. Es ist daher nicht selbstverständlich, daß es eine solche Funktion überhaupt gibt. Genau das besagt aber die Regel für die Listen-Rekursion. Wichtig dabei ist, daß das f auf der rechten Seiten nicht an beliebiger Stelle vorkommen darf, sondern nur in der Kombination ft, also ein bestimmter (leichter zu bestimmender) Funktionswert (Rekursionsprinzip).

3.5.8 Notation. Das φ ist eine dreistellige Operation; dessen Anwendung daher etwas unübersichtlich. Aber für jedes $\ell\colon A^*$ ist $\varphi\ell$ eine binäre Operation, welche man als Infixoperator schreiben kann. Das ℓ wird dabei sinnvollerweise als Index geschrieben. Wir verwenden also statt $\varphi\,\ell\,a\,q$ die Notation:

$$a\;\varphi_\ell\;q,$$

wobei die Klammern in einem Ausdruck wie $a\;\varphi_\ell\;(b\;\varphi_\ell\;q)$ weggelassen werden dürfen.

3.5.9 Beispiel. Sei $f = \mathrm{listrec}\,\varphi\,p$ wir berechnen den Funktionswert an der Stelle $[1,2,3]$:

$$\begin{aligned}f[1,2,3] &= f(1 \lhd 2 \lhd 3 \lhd [])\\ &= 1\;\varphi_{[2,3]}\;f(2 \lhd 3 \lhd [])\\ &= 1\;\varphi_{[2,3]}\;2\;\varphi_{[3]}\;f(3 \lhd [])\\ &= 1\;\varphi_{[2,3]}\;2\;\varphi_{[3]}\;3\;\varphi_{[]}\;f([])\\ &= 1\;\varphi_{[2,3]}\;2\;\varphi_{[3]}\;3\;\varphi_{[]}\;p\end{aligned}$$

3.5.10 Satz. *Für $h, h'\colon A$ und $t, t'\colon A^*$ sowie jede beliebige Aussage P gelten*

$$\begin{aligned}h \lhd t = h' \lhd t' &\implies h = h' \wedge t = t';\\ h \lhd t = [] &\implies P.\end{aligned}$$

Länge

3.5.11 Beispiel. Sei A ein beiliebiger Datentyp. Die Länge einer Liste, $\mathrm{length}\colon A^* \to \mathbb{N}$, wird definiert als $\mathrm{length} = \mathrm{listrec}\, S\, 0$, d.h.

$$\begin{aligned} \mathrm{length}[] &= 0; \\ \mathrm{length}(h \lhd t) &= S(\mathrm{length}\, t). \end{aligned}$$

3.5.12 Notation. Statt $\mathrm{length}\, \ell$ schreiben wir einfacher $|\ell|$. Damit lesen sich die obigen Gleichungen als

$$\begin{aligned} |[]| &= 0; \\ |h \lhd t| &= S\,|t|\,. \end{aligned}$$

Hintenanfügen und Umkehren

Die Funktion cons wird als Einfügen eines Elements am Beginn der Liste interpretiert (z.B. $5 \lhd [1,2,3] = [5,1,2,3]$). Wir definieren nun eine Funktion, die dementsprechend am hinteren Ende einer Liste ein Element anhängt, z.B. $[1,2,3] \rhd 5 = [1,2,3,5]$.

3.5.13 Definition. Sei A ein beiliebiger Datentyp. Die Operation $\rhd\colon A^* \to A \to A^*$ wird definiert durch

$$\begin{aligned} [] \rhd a &= a \lhd [] \\ (h \lhd t) \rhd a &= h \lhd (t \rhd a) \end{aligned}$$

Auch dies ist tatsächlich eine gültige rekursive Definition. Man kann sie, wenn auch wegen dem zweiten Parameter etwas umständlicher, auch mit dem Rekursionsoperator schreiben:

$$\lhd = \mathrm{listrec}(t \mapsto q \mapsto h \mapsto a \mapsto h \lhd (t \rhd a))\,(a \mapsto a \lhd [])$$

3.5.14 Bemerkung. Man beachte, daß durch Vertauschen der Seiten auch die Definition von $\lhd$ durch $\rhd$ entsteht:

$$\begin{aligned} h \lhd &= [] \rhd h \\ h \lhd (t \rhd a) &= (h \lhd t) \rhd a \end{aligned}$$

3.5.15 Satz. *Für $a\colon A$ und $\ell\colon A^*$ gilt*

$$|\ell \rhd a| = S\,|\ell|\,.$$

Beweis. Wir beweisen mit Listenindution nach ℓ.

Induktionsanfang: Zu zeigen ist

$$|[] \rhd a| = S\,|[]|\,.$$

Tatsächlich gilt

$$\begin{aligned} |[] \rhd a| &= |a \lhd []| && \text{(Definition von } \rhd) \\ &= S\,|[]| && \text{(Definition von length).} \end{aligned}$$

Induktionsschritt: Für $t\colon A^*$ verwenden wir die Induktionshypothese

$$|t \triangleright a| = S\,|t|$$

und zeigen, daß dann für alle $h\colon A$ auch

$$|(h \triangleleft t) \triangleright a| = S\,|h \triangleleft t|$$

gilt. Tatsächlich ergibt sich

$$\begin{aligned}
|(h \triangleleft t) \triangleright a| &= |h \triangleleft (t \triangleright a)| && \text{(Definition von } \triangleright) \\
&= S\,|t \triangleright a| && \text{(Definition von length)} \\
&= S(S\,|t|) && \text{(Induktionshypothese)} \\
&= S\,|h \triangleleft t| && \text{(Def. von length, von rechts nach links).}
\end{aligned}$$

□

3.5.16 Definition. Wir definieren eine Funktion reverse$\colon A^* \to A^*$ die die Reihenfolge der Elemente in einer Liste umkehrt, z.B. reverse$[1,2,3,4] = [4,3,2,1]$. Sei A ein beiliebiger Datentyp. Die rekursiven Gleichungen sind leicht hinzuschreiben:

$$\begin{aligned}
\text{reverse}[] &= [], \\
\text{reverse}(h \triangleleft t) &= \text{reverse}\,t \triangleright h.
\end{aligned}$$

Auch diese Definition läßt sich mittels Rekursionsoperator ausdrücken (Übung).

3.5.17 Satz. *Für alle $\ell\colon A^*$ gilt*

$$|\text{reverse}\,\ell| = |\ell|\,.$$

Beweis. Einfache Übung für einen Beweis mit Listeninduktion. □

Listenverkettung

3.5.18 Beispiel. Wir definieren die *Listenverkettung* als eine binäre Operation $\diamond\colon A^* \to A^* \to A^*$, welche zwei Listen zusammenhängt, z.B. $[1,2,3] \diamond [5,6] = [1,2,3,5,6]$. Sie erfüllt die Gleichungen

$$\begin{aligned}
[] \diamond \ell &= \ell \\
(h \triangleleft t) \diamond \ell &= h \triangleleft (t \diamond \ell)
\end{aligned}$$

und ist damit durch eine primitive Rekursion über Listen wohldefiniert.

3.5.19 Satz. *Für alle $u, v, w\colon A^*$ gilt*

$$\begin{gathered}
(u \diamond v) \diamond w = u \diamond (v \diamond w); \\
[] \diamond u = u = u \diamond [].
\end{gathered}$$

Beweis. Übung (Induktion nach u). □

Die Listenverkettung ist somit eine assoziative Operation und die leere Liste ist ein neutrales Element. Man sagt auch, sie bildet ein Monoid.

3.5.20 Satz. *Sei A eine beliebige Menge. Dann gilt stets*

$$|\ell \diamond r| = |\ell| + |r|\,.$$

Beweis. Wir verwenden Induktion nach ℓ.

Induktionsanfang: Zu zeigen ist $|[] \diamond r| = |[]| + |r|$, was wegen der Definitionen klar ist.

Induktionshypothese: $|\ell \diamond r| = |\ell| + |r|$. Zu zeigen: $|(a \triangleleft \ell) \diamond r| = |a \triangleleft \ell| + |r|$. Wir rechnen:

$$\begin{aligned} |(a \triangleleft \ell) \diamond r| &= |a \triangleleft (\ell \diamond r)| \\ &= S\ |\ell \diamond r| \\ &= S(|\ell| + |r|) \\ &= (S\ |\ell|) + |r| \\ &= |a \triangleleft \ell| + |r|\,. \end{aligned}$$

□

3.5.21 Definition. Seien A und B beliebige Datentypen und $f\colon A \to B$. Dann läßt sich f durch „elementweises“ Anwenden auf eine Funktion $f^*\colon A^* \to B^*$ hochziehen:

$$\begin{aligned} f^*\,[] &= [] \\ f^*\,(x \triangleleft \ell) &= (fx) \triangleleft f^*\,\ell. \end{aligned}$$

Die Funktion wird oft auch mit map bezeichnet und hat den Typ $\mathrm{map}\colon (A \to B) \to (A^* \to B^*)$.

3.5.22 Notation. Eine weitere Notation für map ist

$$[fx \mid x \leftarrow \ell] := \mathrm{map}\, f\, \ell = f^*\ell,$$

welche besonders dann praktisch ist, wenn f duch einen Ausdruck wie z.B. in $f = (x \mapsto x^2)$ gegeben ist, weil man dann einfach $[x^2 \mid x \leftarrow [2,3,4]\,]$ schreiben kann. Diese Notation (list comprehension) ist auf vielfältige Weise erweiterbar.

3.5.23 Satz (Funktor-Eigenschaft). *Es seien A, B, C beliebige Mengen. Dann gilt für alle $f\colon A \to B$ und $g\colon B \to C$:*

$$\begin{aligned} (\mathrm{id}_A)^* &= \mathrm{id}_{A^*} \\ (g \circ f)^* &= g^* \circ f^* \end{aligned}$$

Beweis. Es geht um die Gleichheit von Funktionen. Wir setzten auf beiden Seiten ein beliebiges Element ein (vgl. Extensionalität) und rechnen nach.

Für die erste Gleichung haben wir zu zeigen:

$$\bigwedge_{\ell\colon\, A^*} (\mathrm{id}_A)^*\,\ell = \mathrm{id}_{A^*}\,\ell$$

Wir beweisen mit Induktion nach ℓ. Für den Induktionsanfang ist zu zeigen: $(\mathrm{id}_A)^*[] = \mathrm{id}_{A^*}[]$; die linke Seite ist gemäß der Definition von map gleich $[]$, die rechte Seite gemäß der Definition von id. Für den Induktionsschritt verwenden

wir die Induktionshypothese $\mathrm{id}_{A^*}\, t = \mathrm{id}_{A^*}\, t$ und zeigen, daß dann auch, für beliebiges $h\colon A$, gilt $(\mathrm{id}_A)^*(h \lhd t) = \mathrm{id}_{A^*}(h \lhd t)$:

$$\begin{aligned}(\mathrm{id}_A)^*(h \lhd t) &= \mathrm{id}_A\, h \lhd (\mathrm{id}_A)^* t && \text{(Definition von map)}\\ &= \mathrm{id}_A\, h \lhd \mathrm{id}_{A^*}\, t && \text{(Induktionshypothese)}\\ &= h \lhd t && \text{(Definition von id)}\\ &= \mathrm{id}_{A^*}(h \lhd t) && \text{(Definition von id)}\end{aligned}$$

Für die zweite Gleichung haben wir zu zeigen:

$$\bigwedge_{\ell\colon A^*} (g \circ f)^*\, \ell = (g^* \circ f^*)\, \ell,$$

was wir wiederum mit Induktion nach ℓ zeigen. (**Übung!**) □

Aufgrund der obigen Eigenschaft nennt man map einen *Funktor*.

3.5.24 Satz. *Ein paar weitere nette Eigenschaften von* map*:*

$$\begin{aligned}\mathrm{length} \circ f^* &= \mathrm{length};\\ f^*(\ell \diamond r) &= f^*\ell \diamond f^* r;\\ f^* \circ \mathrm{concat} &= \mathrm{concat} \circ (f^*)^*.\end{aligned}$$

Beweis. **Übung!** □

Auswahl

3.5.25 Definition. Sei $b\colon A \to \mathbb{B}$. Dann definieren wir

$$\begin{aligned}\mathrm{filter}\, b\, [] &= []\\ \mathrm{filter}\, b\, (a \lhd \ell) &= \mathrm{if}\ ba\ \mathrm{then}\ a \lhd \mathrm{filter}\, b\, \ell\ \mathrm{else}\ \mathrm{filter}\, b\, \ell\end{aligned}$$

Der Funktor map verändert normalerweise die Elemente einer Liste, läßt aber deren Länge unverändert. filter dagegen verändert normalerweise die Länge einer Liste, ändert aber nichts an den Elementen selbst.

3.5.26 Notation.

$$[a \mid a \leftarrow \ell,\, ba] := \mathrm{filter}\, b\, \ell$$

Diese Notation kann mit der ähnlichen für map nach Belieben kombiniert werden, insbesondere:

$$[fa \mid a \leftarrow \ell,\, ba] := \mathrm{map}\, f\, (\mathrm{filter}\, b\, \ell)$$

und noch deutlich allgemeinere Varianten, mit stets offensichtlicher Bedeutung.

Beispiel: $[x^2 \mid x \leftarrow [1,2,3,4,5,6,7,8,9],\ x \text{ ist Primzahl }] = [4,9,25,49]$.

3.5.27 Definition. Sei A ein Menge auf der eine Addition definiert ist. Dann wird die *Summe* aller Elemente einer Liste definiert durch:

$$\begin{aligned}\textstyle\sum [] &= 0\\ \textstyle\sum (a \lhd \ell) &= a + \sum \ell.\end{aligned}$$

Anders geschrieben: $\sum = \mathrm{listrec} + 0$. Die Notation wird üblicherweise folgendermaßen modifiziert:

$$\sum_{k=1}^{n} a_k = \sum [a_k \mid k \leftarrow [1..n]]$$

Wir haben daher insbesondere:

$$\sum_{k=1}^{0} = 0$$
$$\sum_{k=1}^{Sn} a_k = \sum_{k=1}^{n} a_k + a_{Sn}$$

3.5.28 Satz. *Einige elementare Beziehungen für Summen:*

$$\sum_{k=1}^{n}(a_k + b_k) = \sum_{k=1}^{n} a_k + \sum_{k=1}^{n} b_k$$
$$\sum_{k=1}^{n} a_k \cdot \sum_{\ell=1}^{m} b_k = \sum_{k=1}^{n}\sum_{\ell=1}^{m} a_k b_\ell$$
$$\sum_{k=1}^{n} a_k = \sum_{k=0}^{n-1} a_{k+1}$$

Beweis. Alle derartigen Aussagen lassen sich recht einfacht mittels Induktion nach n aus den elementaren Eigenschaften der Grundrechnungsarten gewinnen. □

Geordnete Listen

Sei A eine beliebige Menge, auf der eine Ordnung $\preceq\colon A^* \to A^* \to \Omega$ definiert ist.

3.5.29 Definition. Das Prädikat $\mathrm{issorted}\colon A^* \to \Omega$ sollte ausdrücken, daß eine Liste sortiert ist. Es sei induktiv durch die folgenden Regeln definiert:

$$\mathrm{issorted}[]$$
$$\mathrm{issorted}(a \lhd [])$$
$$\mathrm{issorted}(h \lhd t) \implies a \preceq h \implies \mathrm{issorted}(a \lhd h \lhd t)$$

Dies ist so zu verstehen, daß umgekehrt eine Liste nur dann sortiert sein sollte, wenn dies aus einer dieser Regeln folgt. Daraus ergibt sich auf natürliche Weise wieder ein Induktionsprinzip:

$$P([]) \wedge \bigwedge_{a\colon A} \wedge \bigwedge_{t\colon A^*} \bigwedge_{h \lhd t} (\mathrm{issorted}(h \lhd t) \implies \bigwedge_{a\colon A} (a \preceq h \implies P(a \lhd h \lhd t))) \implies \bigwedge_{\ell\colon A^*} (\mathrm{is}$$

3.5.30 Definition. Die Relation $A^* \to A^* \to \Omega$ sollte ausdrücekn, daß zwei Listen bis auf die Reihenfolge ihrer Elemente übereinstimmen. Sie sei induktiv

durch die folgenden Regeln definiert. (bezeichnet mit $\sim$), definieren wir drei Regeln:

$$\begin{aligned} [] &\sim [] \\ t_1 \sim t_2 &\implies z \triangleleft t_1 \sim z \triangleleft t_2 \\ t_1 \sim t_2 &\implies n \triangleleft p \triangleleft t_1 \sim p \triangleleft n \triangleleft t_2 \end{aligned}$$

3.5.31 Satz. *Sei A eine beliebige Menge, auf der eine entscheidbare lineare Ordnung $\preceq$ definiert ist, und $\ell\colon A^*$. Dann gibt es genau eine Liste ℓ' sodaß $\ell \sim \ell'$ und* issorted ℓ'.

Quicksort. Die folgende Funktion liefert einen konstruktiven Beweis:

$$\begin{aligned} \operatorname{sort}[] &= [] \\ \operatorname{sort}(a \triangleleft \ell) &= \operatorname{sort}[x \mid x \leftarrow \ell,\, x \preceq a] \diamond [a] \diamond \operatorname{sort}[x \mid x \leftarrow \ell,\, x \succ a] \end{aligned}$$

Die Rekursion führt hier über die Länge der Liste. □

Kapitel 4

Algebra

4.1 Halbgruppen, Monoide

4.1.1 Definition. Sei A eine Menge und $\circ\colon A \to A \to A$ eine zweistellige Operation darin. Dann heißt A zusammen mit $\circ$ eine *Halbgruppe* genau dann wenn $\circ$ assoziativ ist, d.h. wenn für alle $x, y, z\colon A$ gilt

$$(x \circ y) \circ z = x \circ (y \circ z).$$

4.1.2 Beispiel. Halbgruppen sind

1. alle Zahlenmengen, sowohl mit der Addition als auch mit der Multiplikation: $(\mathbb{N}, +)$, $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{A}, +)$, $(\mathbb{R}, +)$; $(\mathbb{N}, \cdot)$, $\ldots$, $(\mathbb{R}, \cdot)$;
2. die logischen Aussagen zusammen mit $\wedge$ oder $\vee$, also $(\Omega, \wedge)$ und $(\Omega, \vee)$;
3. die Potenzmenge einer Menge X zusammen mit Durchschnitt oder Vereinigung, also $(\mathscr{P}(X), \cap)$ und $(\mathscr{P}(X), \cup)$; Für eine beliebige Menge X, die Listen über diesem Alphabet zusammen mit der Listenverkettung, also $(X^*, \diamond)$;
4. die Menge aller Funktionen einer Menge X in sich zusammen mit der Hintereinanderausführung, $(X \to X, \circ)$.
5. Die Menge aller Relationen auf einer Menge X zusammen mit dem Relationenprodukt.

Keine Halbgruppe erhält man dagegen mit dem Potenzieren, und mit der Subtraktion (Assoziativität verletzt); ferner mit dem Listeneinfügen (hier handelt es sich nicht einmal um eine zweistellige Operation in der Menge), bei der logischen Implikation und Äquivalenz (ebenfalls nicht assoziativ). Bei der Implikation beachte man besonders, daß wir gezeigt haben, daß eine Aussage der Form $A \implies (B \implies C$ stets zu $(A \wedge B) \implies C$ äquivalend ist, was im allgemeinen nicht dasselbe ist wie $(A \implies B) \implies C$. Bei der Äquivalenz ist zu beachten, daß die Notation $A \iff B \iff C$ als Abkürzung für $(A \iff B) \wedge (B \iff C)$ zu verstehen ist, und daher weder zu $(A \iff B) \iff C$ noch zu $A \iff (B \iff C)$ äquivalent ist.

Ebenfalls keine Halbgruppe bilden etwa die ungeraden Zahlen zusammen mit der Addition (weil die Summe zweier ungerader Zahlen nicht wieder ungerade ist), wohl aber mit der Multiplikation.

4.1.3 Definition. Sei $(H, \circ)$ eine Halbgruppe und $e\colon H$. Dann heißt e ein *neutrales Element* wenn für alle $h\colon H$ gilt

$$e \circ h = h = h \circ e.$$

Man nennt dann H zusammen mit $\circ$ und e, also $(H, \circ, e)$, ein *Monoid*.

Sind $e_1, e_1\colon H$ zwei neutrale Elemente einer Halbgruppe, dann gilt einerseits $e_1 \circ e_2 = e_2$ (weil e_1 neutral ist) und andererseits $e_1 \circ e_2 = e_1$ (weil auch e_2 neutral ist. Daher muß $e_1 = e_2$ sein, d.h. in jeder Halbgruppe gibt es höchstens ein neutrales Element. Man kann daher gefahrlos auch einfach vom Monoid $(H, \circ)$ reden, da das neutrale Element eindeutig bestimmt ist. Alle erwähnten Beispiele von Halbgruppen enthalten ein neutrales Element und können daher als Monoide betrachtet werden. Das neutrale Element für die Addition ist stets 0, das für die Multiplikation ist 1. Bei den logischen Verknüpfungen sind $\top$ bzw. $\bot$ neutral, bei der Potenzmenge ist es die leere Menge bzw. die Grundmenge. Bei der Listenverkettung ist es die leere Liste und bei der Hintereinanderausführung von Funktionen die identische Funktion.

Kein neutrales Element haben dagegen etwa die strikt positiven Zahlen zusammen mit der Addition: hier fehlt die Null; oder die geraden Zahlen, mit der Multiplikation (weil 1 nicht in der Menge ist).

In einem Monoid H kann man auch Potenzieren: ist $h\colon H$ und $n\colon \mathbb{N}$, dann definiert man rekursiv

$$\begin{aligned} h^0 &= 1 \\ h^{n+1} &= h \circ h^n. \end{aligned}$$

(Dies funktioniert natürlich auch für Halbgruppen ohne neutrales Element, wenn man mit $h^1 = h$ anfängt.) Es gilt dann das bekannte Rechengesetz

$$h^m \circ h^n = h^{m+n}, \tag{4.1}$$

im allgemeinen aber nicht $(g \circ h)^n = g^n \circ h^n$, weil dies Kommutativität von $\circ$ voraussetzt. In jedem Fall aber gilt $h^m \circ h^n = h^n \circ h^m$. Man beachte, daß wir hier nur von positiven Exponenten reden.

4.2 Gruppen

4.2.1 Definition. Sei $(G, \circ, 1)$ ein Monoid. Dann heißen $g, h\colon G$ zueinander invers, wenn $g \circ h = 1 = h \circ g$ gilt. Besitzt jedes Element ein inverses, dann ist dieses eindeutig bestimmt und wird mit g^{-1} bezeichnet (oder mit $-g$, bei additiver Schreibweise). Das Monoid zusammen mit dieser zusätzlichen (einstelligen) Operation heißt dann eine Gruppe. D.h., $(G, \circ, 1, (^{-1}))$ ist eine Gruppe falls $(G, \circ, 1)$ ein Monoid ist und für alle $g\colon G$ gilt

$$g \circ g^{-1} = 1 = g^{-1} \circ g.$$

Die Notation für das Inverse passt gut zum Potenzieren, da damit tatsächlich die Beziehung in Gleichung 4.1 auf beliebige ganzzahlige Exponenten erweitert wird.

Einige unserer Standardbeispiele für Halbgruppen führen tatsächlich zu Gruppen: $(\mathbb{Z}, +, 0, -)$, $(\mathbb{Q}, \cdot, (^{-1}))$, $(\mathbb{R}, \cdot, (^{-1}))$. Keine inversen Elemente gibt es dagegen bei der Listenverkettung. Auch die erwähnten logischen Operationen erlauben keine Inversen ($\neg A$ ist *nicht* invers zu A, denn $A \wedge \neg A = \bot$, neutral ist aber $\top$). Auch $(X \to X, \circ)$ führt zu keiner Gruppe, da Funktionen nur dann ein Inverses haben, wenn sie bijektiv sind. Allerdings ergibt die Hintereinanderausführung von bijektiven Funktionen wieder eine bijektive Funktion (es gilt konkret: $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$, man beachte die Umkehrung der Reihenfolge), weshalb die Menge der bijektiven Funktionen von $X \to X$, oft auch *Permutationen* von X genannt und mit $P(X)$ bezeichnet, sehr wohl eine Gruppe bildet. Im Gegensatz zu den erwähnten Gruppen mit Zahlen, ist die Gruppe der Permutationen nicht kommutativ.

4.3 Ringe und Körper

Bei Gruppen betrachtet man nur zwei der üblichen vier Grundrechnungsarten.

4.3.1 Definition. Sei $(R, +, 0, -)$ eine kommutative Gruppe und $(R, \cdot, 1)$ ein Monoid, dann heißt $(R, +, 0, -, \cdot, 1)$ ein *Ring*, falls zusätzlich die Distributivgesetze

$$(a + b) \cdot c = a \cdot c + b \cdot c$$
$$a \cdot (b + c) = a \cdot b + a \cdot c$$

gelten.

Da sich die jeweiligen neutralen Elemente und die Operation für das Inverse eindeutig ergeben, spricht man kürzer vom Ring $(R, +, \cdot)$, oder einfach vom Ring R, falls auch die Operationen aus dem Zusammenhang klar sind.

Man beachte, daß man beide Distributivgesetze benötigt, da die Multiplikation nicht immer als kommutativ vorausgesetzt wird. Die bekannten Beispiele $(\mathbb{Z}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ sind allerdings allsamt kommutative Ringe. Nichtkommutative Ringe ergeben sich in natürlicher Weise in der Linearen Algebra duch lineare Abbildungen oder Matrizen.

Eine weitere wichtige Klasse von kommutativen Ringen bilden die Restklassenringe $\mathbb{Z}_m$, welche im wesentlichen dem Ring $\mathbb{Z}$ entsprechen, allerdings werden zwei Elemente als gleich betrachtet, wenn die modulo m denselben Rest ergeben, d.h. $\mathbb{Z}_m = \{0, 1, \ldots, m-1\}$.

Weiters sind die Polynomringe zu erwähnen: ist R ein kommutativer Ring, dann bezeichnen wir mit $R[x]$ den Ring aller *Polynome* in der Variablen x mit Koeffizienten aus R, d.h.

$$R[x] = \{\sum_{k=0}^{n} a_k x^k \mid n \colon \mathbb{N},\ a_1 \ldots, a_n \colon R\}.$$

Kapitel 5

Lineare Algebra

5.1 Affine Räume

Ein Punkt in einer Ebene (oder im 3-dimensionalen Raum unserer Anschauung) bezeichnet eine ganz bestimmte Stelle. Er hat keine Ausdehnung, Farbe oder sonstige weitere Eigenschaften.

Es macht keinen geometrischen Sinn, einen Punkt mit einer Zahl zu multiplizieren oder zwei Punkte zu addieren. Wohl aber läßt sich zwei Punkten P und Q derjenige Punkt M zuordnen, welcher genau in der Mitte dieser beiden liegt. Es liegt nahe, M mit $\frac{1}{2}P + \frac{1}{2}Q$ zu bezeichnen, obwohl wir gerade bemerkt haben, daß die Hälfte eines Punktes geometrisch keinen Sinn ergibt.

Die Idee, den Mittelpunkt zu betrachten, läßt sich verallgemeinern. Sei V der Mittelpunkt von P und M, also

$$\begin{aligned} V &= \tfrac{1}{2}P + \tfrac{1}{2}M && \text{(Definition von } V\text{)} \\ &= \tfrac{1}{2}P + \tfrac{1}{2}(\tfrac{1}{2}P + \tfrac{1}{2}Q) && \text{(Definition von } M \text{ eingesetzt)} \\ &= \tfrac{3}{4}P + \tfrac{1}{4}Q && \text{(übliche Rechenregeln)} \end{aligned}$$

Obwohl die Anwendung der üblichen Rechenregeln sich hier bloß aus der gewählten Notation für den Mittelpunkt ergab und überhaupt nicht geometrisch gerechtfertigt wurde, kann das Ergebnis sehr gut geometrisch interpretiert werden, denn V liegt gerade bei einem Viertel des Weges auf der Strecke von P nach Q. Mit dieser Interpretation ist auch klar, daß etwa der Punkt $\frac{2}{3}P + \frac{1}{3}Q$ auf einem Drittel des Weges zu liegen hat. Allgemeiner bezeichnet $(1-\lambda)P + \lambda Q$ denjenigen Punkt, welcher einen Anteil von λ an der Strecke von P nach Q bezeichnet. Dies ist allgemein verständlich, falls $0 \leq \lambda \leq 1$ gilt. Für $\lambda = 0$ erhalten wir insbesondere $1P + 0Q = P$. Schwieriger wird die Interpretation, wenn etwa $\lambda = -1$ ist. Das bedeutet, daß die Strecke von P nach Q minus einmal zurückgelegt wurde, was man am besten so interpretiert, daß man an einem Punkt N angelangt ist, nachdem die volle Strecke in der entgegengesetzten Richtung zurückgelegt wurde. Es gilt dann $P = \frac{1}{2}N + \frac{1}{2}Q$, und eine Umformung gemäß den üblichen Rechenregeln ergibt dann $N = 2P - Q$, was tatsächlich genau dem Fall $\lambda = -1$ entspricht.

Sind P und Q Punkte, und $\lambda, \mu \in \mathbb{R}$ mit $\lambda + \mu = 1$, dann heißt $\lambda P + \mu Q$ eine *Affinkombination* der Punkte P und Q. Mittels der obigen Überlegungen haben wir jeder Affinkombination von Punkten in einer Ebene (oder im Raum)

eine geometrische Bedeutung gegeben, d.h. einen passenden Punkt auf der Verbindungsgeraden zugeordnet.

Es können auch mehrere Punkte affin kombinert werden.

5.1.1 DEFINITION. Für beliebige Zahlen $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ mit $\sum_{k=1}^{n} \lambda_k$ bezeichnet der Ausdruck $\sum_{k=1}^{n} \lambda_k P_k$ eine *Affinkombination* von $P_1, \ldots, P_n$. Die Zahlen λ_k heißen *Koeffizienten* oder *Gewichte*.

Die Bezeichnung *Gewichte* kann man durchaus wörtlich nehmen: Sind in den Punkten P_k die Gewichte λ_k verteilt, so ist $\sum_{k=1}^{n} \lambda_k P_k$ gerade deren Schwerpunkt. Das gilt auch für *negative* Gewichte, also Kräfte, die nach *oben* wirken.

5.1.2 DEFINITION. Ist jeder Affinkombination von Elementen einer Menge wieder ein Element derselben Menge zugeordnet, sodaß für die Affinkombinationen die der Notation entsprechenden üblichen Rechenregeln gelten, dann spricht man von einem *affinen Raum*.

Beispiele: Ebene, Raum, aber auch z.B. unendliche Folgen von Punkten (hier erfolgen die Affinkombinationen komponentenweise).

5.1.3 DEFINITION. Sind $P_1, \ldots, P_n$ Punkte eines affinen Raumes, dann heißt die Gesamtheit aller Affinkombinationen dieser Punkte $\{\sum_{k=1}^{n} \lambda_k P_k \mid \lambda_1, \ldots, \lambda_n \in \mathbb{R},\ \sum_{k=1}^{n} \lambda_k = 1\}$ die *affine Hülle* von $P_1, \ldots, P_n$.

Jede affine Hülle bildet einen affinen Raum.

Die affine Hülle von zwei verschiedenen Punkten ist eine Gerade.

Die affine Hülle von drei Punkten ist eine Ebene, außer sie liegen zufällig alle auf einer Geraden.

5.1.4 DEFINITION. Eine Menge von Punkten heißt *affin unabhängig*, falls die Darstellung jedes Punktes in deren affiner Hülle eindeutig ist. Ansonsten ist sie *affin abhängig*.

5.1.5 SATZ. *Eine Menge von Punkten ist affin abhängig falls sich einer dieser Punkte als Affinkombination der anderen darstellen läßt, oder falls sich trotz Weglassen eines geeigneten Punktes dieselbe affine Hülle ergibt.*

5.1.6 DEFINITION. Eine Menge von Punkten eines affinen Raumes heißt *aufspannend*, falls sich jeder Punkt des Raumes als Affinkombination von Punkten aus dieser Menge darstellen läßt, d.h. wenn die affine Hülle dieser Menge der ganze Raum ist.

5.1.7 DEFINITION. Eine affin unabhängige aufspannende Menge von Punkten eines affinen Raumes heißt *affine Basis* dieses Raumes.

Es läßt sich zeigen, daß die Anzahl der Elemente zweier affiner Basen desselben Raumes stets gleich ist. Die *Dimension* eines affinen Raums ist die um eins verminderte Anzahl der Elemente einer affinen Basis.

Beispiel: Jede Ebene hat eine affine Basis bestehend aus 3 Punkten. Sie hat daher die Dimension 2.

5.1.8 DEFINITION. Eine Affinkombination heißt Konvexkombination, wenn alle Koeffizienten ≥ 0 sind.

Entsprechend ergibt sich der Begriff der *konvexen Hülle*.

5.1.9 Beispiel. Der statistische Mittelwert ist eine Konvexkombination aller beobachteten Werte. Wurden n verschiedene Werte $x_1, \ldots, x_n$ beobachtet, dann wird jeder Wert mit dessen relativer Häufigkeit $h_n(x_k)$ gewichtet, was zur Konvexkombination $\sum_{k=1}^{n} h_n(x_k)x_k$ führt. Man beachte, daß die relativen Häufigkeiten stets ≥ 0 sind und in Summe stets 1 ergeben, was gerade die Bedingungen für eine Konvexkombination sind.

Der Erwartungswert $E(X)$ einer Zufallsgröße X ist eine Konvexkombination aller möglichen Werte $x_1, \ldots, x_n$, und die Koeffizienten entsprechen gerade den Wahrscheinlichkeiten $P(X = x_k)$ für das Auftreten der entsprechenden Werte. Also $E(X) = \sum_{x_k}^{n} P(X = x_k)x_k$. Anmerkung: Dies ist genaugenommen natürlich nur dann richtig, wenn X eine diskrete Zufallsgröße ist, die nur endlich viele Werte annehmen kann. Ansonsten muß man unendliche Konvexkombinationen bzw. ein Integral (bei kontinuierlichen Zufallsgrößen) betrachten. Darauf wollen wir hier aber nicht eingehen.

Alle statistischen Methoden, die mit dem Mittelwert arbeiten, funktionieren daher nur dann sinnvoll, wenn die Werte der Zufallsgrößen nach Belieben zumindest konvex kombiniert werden können. In der Statistik spricht man dann von einem *metrischen Merkmal*.

5.2 Lineare Räume (Vektorräume)

Daß bei einer Affinkombination die Summe der Koeffizienten stets gleich 1 sein muß, ist ein lästiger Schönheitsfehler. Insbesondere die Notation kann sehr verwirrend sein, da etwa $\frac{1}{2}P + \frac{1}{2}Q$ eine Bedeutung hat, obwohl $\frac{1}{2}P$ und $\frac{1}{2}Q$ sinnlos sind.

Zur Schönheitskorrektur wählt man irgendeinen Punkt O als besonderen Punkt (*Nullpunkt*, *Ursprung*), und betrachtet alle Punkte relativ zu diesem. Damit wird jeder Punkt P mit dem *Vektor* von O nach P identifiziert. Ein Vektor wird üblicherweise als Pfeil dargestellt, ist aber als Bewegungsrichtung zu verstehen, d.h. parallelverschobene Pfeile bezeichnen denselben Vektor.

Ein Vektor $\mathbf{v}$ kann mit beliebigen Zahlen *skaliert* werden: $\frac{1}{2}\mathbf{v}$ bezeichnet den Vektor, der in dieselbe Richtung zeigt wie $\mathbf{v}$, aber nur halb so lang ist; $-\mathbf{v}$ zeigt in die umgekehrte Richtung.

Sind $\mathbf{v}, \mathbf{w}$ Vektoren, so köennen diese „Bewegungsrichtungen“ auch zusammengesetzt werden: man bewegt sich zuerst entlang $\mathbf{v}$ und von dort ausgehend entlang $\mathbf{w}$. Dieser zusammengesetzte Vektor wird mit $\mathbf{v} + \mathbf{w}$ bezeichnet und bildet tatsächlich eine sinnvolle Addition. Insbesondere ist sie assoziativ, der Nullvektor $\mathbf{o}$ (der Vektor von O nach O) verhält sich neutral, es gibt Inverse ($-\mathbf{v} + \mathbf{v} = \mathbf{o} = \mathbf{v} - \mathbf{v}$), und sie ist auch kommutativ: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ (Kräfteparallelogramm). Die Addition von Vektoren führt daher zu einer (additiv geschriebenen) kommutativen Gruppe.

Mit Vektoren können somit beliebige *Linearkombinationen* gebildet werden. D.h., sind $\mathbf{v}_1, \ldots, \mathbf{v}_n$ Vektoren und $\lambda_1, \ldots, \lambda_n$ beliebige Zahlen, dann ist auch $\sum_{k=1}^{n} \lambda_k \mathbf{v}_k$ ein Vektor, und es gelten die üblichen Rechengesetze.

Da Vektoren neben der Addition auch noch die Skalierung zulassen, haben sie eine reichere algebraische Struktur als die einer kommutative Gruppe. Sie bilden einen Vektorraum (auch: linearer Raum). Genauer:

5.2.1 Definition. Wird für eine kommutative Gruppe V jede Zahl λ mit einer

Abbildung $V \to V$ identifiziert, sodaß die folgenden Rechengesetze gelten,

$$\lambda(\mathbf{v}+\mathbf{w}) = \lambda\mathbf{v} + \lambda\mathbf{w}) \qquad \lambda(\mu\mathbf{v}) = (\lambda\mu)\mathbf{v} \qquad 1\mathbf{v} = \mathbf{v} \qquad (\lambda+\mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$$

dann spricht man vom *Vekorraum* (oder *linearen Raum*) V.

Analog zur affinen Hülle, besteht die *lineare Hülle* einer Menge von Vektoren aus all denjenigen Vektoren, die man daraus mittels Linearkombinationen erhält. Eine solche Menge von Vektoren identifiziert man dabei am besten mit der Menge aller zugehörigen Punkte, also derjenigen, die mit einem der Vektoren aus der Menge vom Ursprung aus erreicht werden. Man beachte, daß dann die lineare Hülle eines einzigen Vektors bereits eine Gerade ergibt, und die lineare Hülle von zwei Vektoren eine Ebene (außer einer der Vektoren ist ein Vielfaches des anderen). Ist T eine Menge von Vektoren, dann bezeichnen wir mit $L(T)$ deren lineare Hülle.

Eine Menge von Vektoren heißt *linear unabhängig*, falls die Darstellung jedes Vektores in deren linearen Hülle eindeutig ist. Ansonsten ist sie *linear abhängig*.

Eine Menge von Vektoren ist linear abhängig falls sich einer dieser Vektoren als Linearkombination der anderen darstellen läßt, oder falls sich trotz Weglassen eines geeigneten Vektors dieselbe lineare Hülle ergibt.

Eine Menge von Vektoren ist bereits dann *linear unabhängig*, wenn sich der Nullvektor nur auf die eine Weise (nämlich alle Koeffizienten 0) darstellen läßt.

Eine Menge von Vektoren eines linearen Raumes heißt *aufspannend*, falls sich jeder Vektor des Raumes als Linearkombination von Vektoren aus dieser Menge darstellen läßt, d.h. wenn die lineare Hülle dieser Menge der ganze Raum ist.

5.3 Basis

Eine linear unabhängige aufspannende Menge von Vektoren eines linearen Raumes heißt *lineare Basis* dieses Raumes (oder einfach *Basis*).

Es läßt sich zeigen, daß die Anzahl der Elemente zweier Basen desselben Raumes stets gleich ist. Die *Dimension* eines linearen Raums ist die Anzahl der Elemente einer Basis.

Beispiel: Jede Ebene hat eine Basis bestehend aus 2 Vekoren. Sie hat daher die Dimension 2.

Selbstverständlich ist jeder lineare Raum auch ein affiner Raum (weil jede Affinkombination eine Linearkombination ist). Umgekehrt kann man jeden affinen Raum als linearen Raum betrachten, sobald ein Ursprung gewählt ist. Der Unterschied besteht daher nur in der Betrachtungsweise: ist ein besonderer Punkt ausgezeichnet oder nicht?

Auch Vektoren sind immer noch eher geometrische Gebilde, mit denen man nicht rechnen kann. Wir wählen daher zusätzlich zum Ursprung noch eine Basis $\mathbf{B}$. Ist der Vektorraum n-dimensional, so besteht diese aus n Vektoren, welche wir mit $\mathbf{b}_1, \ldots, \mathbf{b}_n$ bezeichnen. Sei nun $\mathbf{v}$ ein beliebiger Vektor, dann läßt sich dieser eindeutig als Linearkombination der Basiselemente darstellen, d.h. es gibt eindeutig bestimmte Koeffizienten $\lambda_1, \ldots, \lambda_n$, sodaß $\mathbf{v} = \sum_{k=1}^{n} \lambda_k \mathbf{b}_k$. Diese Koeffizienten heißten die *Koordinaten* von $\mathbf{v}$ bezüglich $\mathbf{b}_1, \ldots, \mathbf{b}_n$.

Damit ist es uns gelungen, jedem Punkt eindeutig eine endliche Folge von Zahlen zuzuordnen. Und damit kann man effektiv rechnen, denn, wie nicht

schwer zu erkennen, findet man die Koeffizienten einer Linearkombination einfach, indem man dieselbe Linearkombination komponentenweise für alle Koeffizenten durchführt. Z.B.

$$\sum_{k=1}^{n} \lambda_k \mathbf{b}_k + \sum_{k=1}^{n} \mu_k \mathbf{b}_k = \sum_{k=1}^{n} (\lambda_k + \mu_k)\mathbf{b}_k;$$
$$\alpha \sum_{k=1}^{n} \lambda_k \mathbf{b}_k = \sum_{k=1}^{n} (\alpha\lambda_k)\mathbf{b}_k.$$

Zu beachten ist auch, daß diese Operationen für die einzelnen Koeffizienten unabhängig voneinander sind, d.h. parallel ausgeführt werden können. Dies ist insbesondere dann von Vorteil, wenn man einen Parallelrechner mit n Prozessoren zur Verfügung hat, da dann eine Addition von Vektoren genausolange dauert wie die Addition von Zahlen.

Da Vektoren als Zahlenfolgen (fixer Länge) dargestellt werden können, bilden auch diese selbst, mit komponentenweisen Linearkombinationen, einen Vektorraum. Damit erhalten Zahlenfolgen eine geometrische Interpretation.

5.3.1 Beispiel. Der einfachste Vektorraum ist $\mathbb{R}$ selbst. Geometrisch interpretiert entspricht er einer Geraden. Entsprechend sind $\mathbb{R}^2$ und $\mathbb{R}^2$, mit komponentenweisen Linearkombinationen, Vektorräume und als Ebene bzw. Raum zu interpretieren. Dasselbe gilt natürlich auch für höherdimensionale $\mathbb{R}^n$, wobei dann natürlich die menschliche Vorstellungskraft etwas überfordert ist, was aber nicht viel ausmacht, weil sich ohnehin nicht viel gegenüber dem $\mathbb{R}^2$ ändert.

Selbst unendlich lange Folgen bilden einen Vektorraum. Dies läßt sich nochmals verallgemeinern: Ist V ein Vektorraum und X eine beliebige Menge, so liegt auch auf der Menge aller Funktionen $X \to V$ eine Vektorraumstruktur vor, wobei die Operationen wieder komponentenweise (oder punktweise) definiert werden, konkret: sind $f, g\colon X \to V$ und $\lambda \in \mathbb{R}$, dann werden $f + g$ und λf definiert als

$$(f + g)(x) = f(x) + g(x),$$
$$(\lambda f)(x) = \lambda(f(x)).$$

Auch auf dem direkten Produkt $V \times W$ von zwei Vektorräumen läßt sich in natürlicher Weise (d.h. hier wieder: komponentenweise) eine Vektorraumstruktur definieren. Die Räume $\mathbb{R}^n$ können als Spezialfälle sowohl dieser als auch der vorigen Konstruktion angesehen werden.

Ist U eine Teilmenge eines Vektorraums V, welche gegenüber Linearkombinationen abgeschlossen ist, dann ist U selbst ein Vektorraum, und man schreibt dann $U \leq V$ und bezeichnet U als *Untervektorraum*. Dies gilt insbesondere für alle linearen Hüllen: Ist T eine beliebige Teilmenge von V, dann ist deren lineare Hülle $L(T) \leq V$. Beispiel: für jedes $k \in \mathbb{R}$ ist $\{(x, kx) \mid x \in \mathbb{R}$ ein Untervektorraum von $\mathbb{R}^2$, ebenso $\{(0, x) \mid x \in \mathbb{R}\}$. Dies sind tatsächlich alle: sie entsprechen genau den Geraden duch den Ursprung. Alle anderen Geraden sind nur affine Unterräume, d.h. gegenüber Affinkombinationen abgeschlossen.

Betrachen wir irgendein Zufallsexperiment. Dann bildet die Menge aller dazugehörigen Zufallsgrößen, welche Werte in einem bestimmen Vektorraum V annehmen können, ebenfalls einen Vektorraum.

5.4 Lineare Abbildungen

5.4.1 Definition. *Lineare Abbildungen* sind jene, welche mit Linearkombinationen verträglich sind. Genauer: Seien V, W Vektorräume; dann heißt eine Abbildung $h\colon V \to W$ *linear*, falls für alle Vektoren $\mathbf{v}_1, \ldots, \mathbf{v}_n$ und Skalare $\lambda_1, \ldots, \lambda_n$ gilt

$$h(\sum_{k=1}^{n} \lambda_k \mathbf{v}_k) = \sum_{k=1}^{n} \lambda_k h(\mathbf{v}_k).$$

Diese Forderung ist äquivalent dazu, daß für alle Vektoren $\mathbf{v}, \mathbf{w}$ und Skalare λ, μ gilt

$$h(\lambda\mathbf{v} + \mu\mathbf{w}) = \lambda h(\mathbf{v}) + \mu h(\mathbf{w}).$$

5.4.2 Beispiel. Wichtie Typen linearer Abbildungen sind:

Spiegelung Der einfachste Typ einer linearen Abbildung ist (neben der identischen Abbildung) die Spiegelung im Nullpunkt: Dabei wird jedem Vektor derjenige zugeordnet, der in die entgegengesetzte Richtung zeigt. Klarerweise gelten

$$\begin{aligned} -(\mathbf{v} + \mathbf{w}) &= -\mathbf{v} + -\mathbf{w}, \\ -(\lambda\mathbf{v}) &= \lambda(-\mathbf{v}), \end{aligned}$$

was gerade der Definition einer linearen Abbildung entspricht.

Skalierung Sie α ein beliebiger Skalar. Dann ist die Abbildung $s_\alpha\colon V \to V$, welche jeden Vektor ver-α-facht, also

$$s_\alpha(\mathbf{v}) = \alpha\mathbf{v},$$

linear, denn

$$\begin{aligned} s_\alpha(\lambda\mathbf{v} + \mu\mathbf{w}) &= \alpha(\lambda\mathbf{v} + \mu\mathbf{w}) \\ &= \alpha(\lambda\mathbf{v}) + \alpha(\mu\mathbf{v}) \\ &= \lambda(\alpha\mathbf{v}) + \mu(\alpha\mathbf{v}) \\ &= \lambda s_\alpha(\mathbf{v}) + \mu s_\alpha(\mathbf{w}). \end{aligned}$$

Auch anschaulich ist dieser Sachverhalt unmittelbar einsichtig.

Jedem Skalar α wird somit eine lineare Abbildung s_α zugeordnet. Diese Zuordnung ist nicht nur bijektiv, sondern auch sehr natürlich, denn es gelten

$$\begin{aligned} s_\alpha + s_\beta &= s_{\alpha+\beta}, \\ s_\alpha \cdot s_\beta &= s_{\alpha\cdot\beta}. \end{aligned}$$

Man kann also tatsächlich einen Skalar mit der entsprechenden Skalierung identifizieren. Positive Zahlen entsprechen dabei echten Streckungen; -1 entspricht der Punktspiegelung im Nullpunkt, und bei anderen negativen Zahlen wird gespiegelt und gestreckt.

Projektion Eine Projektion P_U auf eine Gerade (bzw eine Ebene) U ordnet jedem Punkt $\mathbf{x}$ der Ebene (bzw. des Raumes) den nächstgelegenen Punkt $P_U(\mathbf{x})$ auf der Geraden (bzw. der Ebene) U zu. Wenn U ein linearer Unterraum ist, so ist P_U eine Abbildung. Sie erfüllt unter anderem

$$P_U \circ P_U = P_U.$$

Drehung Für jeden Winkel φ ist d_φ, die Drehung eines Vektors in der Ebene um den Winkel φ (im Gegenuhrzeigersinn), eine lineare Abbildung. Weiters gilt:

$$d_\varphi \circ d_\psi = d_{\varphi+\psi}.$$

Bei Drehungen im Raum ist zusätzlich noch die Drehachse zu beachten.

Drehstreckung Da d_{180° tatsächlich eine Skalierung mit -1 ist, identifizieren wir sie mit dem Skalar -1, d.h.

$$d_{180^\circ} = -1.$$

Wir setzen nun $i := d_{180^\circ}$; dann gilt

$$i^2 = -1$$

(genauer: $i \circ i = s_{-1}$), sodaß wir eine Wurzel aus -1 gefunden haben. Diese entspricht zwar keiner reellen Zahl, aber man kann damit dennoch wie mit normalen Zahlen rechnen. Die Menge aller *Drehstreckungen* (das sind lineare Abbildungen, die aus einer Drehung und einer Streckung zusammengesetzt sind) heißt dann auch die Menge der *Komplexen Zahlen* $\mathbb{C}$. Diese ist (wie die rationalen Zahlen $\mathbb{Q}$ oder reellen Zahlen $\mathbb{R}$) gegenüber allen Grundrechnungsarten abgeschlossen, und es gelten die üblichen Rechengesetze.

Scherung Dieser Typ von linearer Abbildung bildet ein Quadrat auf ein Parallelogramm ab.

Differentialoperator Ein weniger geometrisches Beispiel ist die Abbildung, welche eine Funktion auf deren Ableitung abbildet, denn auch hier gilt

$$(f+g)' = f' + g' \qquad\qquad (\lambda f)' = \lambda(f').$$

Auch der Integraloperator ist eine lineare Abbildung.

Erwartungswert Ein Beispiel für eine lineare Funktion in der Statistik liefert der Erwartungswert, wegen

$$E(X+Y) = E(X) + E(Y) \qquad\qquad E(\lambda X) = \lambda E(X).$$

Keine lineare Abbildung dagegen ist eine Translation; das ist eine Abbildung der Form $t_{\mathbf{c}} \colon \mathbf{v} \mapsto \mathbf{v} + \mathbf{c}$, wobei $\mathbf{c}$ ein beliebiger Vektor $\neq \mathbf{o}$ ist. So ist etwa $t_{\mathbf{c}}(2\mathbf{v}) = (2\mathbf{v} + \mathbf{c}) \neq 2(\mathbf{v} + \mathbf{c}) = 2t_{\mathbf{c}}(\mathbf{v})$. Translationen erhalten allerdings alle Affinkombinationen und bilden daher das typische Beispiel für eine *affine Abbildung*, die nicht linear ist.

Ganz entscheidend ist, daß die Koordinatenabbildung linear ist: Ist V ein n-dimensionaler Vektorraum und B eine Basis von V, dann ist die Abbildung $V \to \mathbb{R}^n$, welche jedem Vektor dessen Koordinaten zuordnet, eine lineare Abbildung. Dies erlaubt uns, mit den Koordinaten eines Vektors anstatt mit dem Vektor selbst zu rechnen. Genauer:

5.4.3 DEFINITION. Eine lineare Abbildung $V \to W$ heißt *Isomorphismus*, wenn sie außerdem bijektiv ist. Die Vektorräume V und W heißen dann *isomorph*.

Isomorphe Vektorräume haben exakt die selbe Struktur, da sich die Elemente zusammen mit den Rechenoperationen (Linearkombinationen) eins zu eins übertragen lassen. Insbesondere ist die Koordinatenabbildung ein Isomorphismus und alle n-dimensionalen Vektorräume sind daher isomorph zu $\mathbb{R}^n$. Dies ist gerade die Grundlage der Darstellung von Vektoren als Zahlentupel.

5.5 Matrizen

So wie die Vektorräume selbst, sind auch lineare Abbildungen zuerst einmal rein geometrische Objekte. Wird in den betroffenen Vektorräumen aber jeweils eine Basis festgelegt, dann ergibt sich auch für lineare Abbildung eine sehr konkrete numerische Darstellung.

Seien V, W Vektorräume, $h\colon V \to W$ eine lineare Abbildung und $\{b_1, \ldots, b_n\}$ eine Basis von V. Dann gilt bekanntlich

$$h(\sum_{j=1}^{n} \lambda_j b_j) = \sum_{j=1}^{n} \lambda_j h(b_j),$$

für beliebige Koeffizienten $\lambda_1, \ldots, \lambda_n$, und jeder Vektor $\mathbf{v}$ hat eine solche Darstellung als Linearkombination mit passenden Koeffizient. Dies hat weitreichende Konsequenzen: eine lineare Abbildung ist durch ihre Werte auf den Basiselementen, also durch die n Vektoren $h(b_1), \ldots, h(b_m)$, bereits festgelegt. Anstatt festzulegen, wie jeder der unendlich vielen Vektoren von V abzubilden ist, und hinterher nachzurechnen, daß das Ergebnis tatsächlich linear ist, brauchen nur noch n Vektoren angegeben, und erhalten automatisch eine lineare Abbildung. Sei nun weiters $c_1, \ldots, c_m$ eine Basis von W. Dann läßt sich jedes $h(b_j)$ (so wie jeder andere Vektor von W) eindeutig als Linearkombination darstellen:

$$h(b_j) = \sum_{i=1}^{m} a_{i,j} c_i,$$

sodaß die lineare Abbildung h schließlich durch nm Skalare spezifiziert ist. Diese werden zu einer Matrix, der *Abbildungmatrix* von h bezüglich B und C, zusammengefaßt, und man schreibt:

$$h_{B,C} = \begin{pmatrix} a_{1,1} & \ldots & a_{1,n} \\ \vdots & \ldots & \vdots \\ a_{m,1} & \ldots & a_{m,n} \end{pmatrix}$$

Damit entspricht jeder linearen Abbildung von n-dimensionalen Raum V in den m-dimensionalen Raum W genau eine Matrix mit m Zeilen und n Spalten.

Gilt $V = W$, ist also h eine Abbildung eines Vektorraumes in sich, so ist es üblich, auch dieselben Basen zu wählen, also $B = C$. Insbesonere ist die Abbildungsmatrix in diesem Fall stets quadratisch (wegen $n = m$.

5.5.1 Beispiel. Sei α ein Skalar und s_α die Streckung um den Faktor α. Egal, welche Basis man wählt (es handelt sich um eine Abbildung $V \to V$, daher nur eine Basis), als Abbildungsmatrix ergibt sich stets eine *Diagonalmatrix*, das ist eine Matrix der Gestalt,

$$\begin{pmatrix} \alpha & 0 & \dots & 0 \\ 0 & \alpha & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha \end{pmatrix}$$

denn jeder Basisvektor $\mathbf{b}_i$ wird auf $\alpha\mathbf{b}_i$ abgebildet. Die Anzahl der Zeilen und Spalten dieser Matrix entspricht dabei der Dimension von V.

Sei s die Spiegelung an einer Geraden durch den Ursprung. Als Basis wählen wir einen Vektor $\mathbf{b}_1$ in Richtung der Geraden und einen Vektor $\mathbf{b}_2$, welcher auf $\mathbf{b}_1$ im rechten Winkel steht. Dann gilt $s(\mathbf{b}_1) = \mathbf{b}_1$ und $s(\mathbf{b}_2) = -\mathbf{b}_2$. Die Abbildungsmatrix ist daher

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

Ist p die Projektion eines Vektors im dreidimensionalen Raum auf die von $\mathbf{b}_1$ und $\mathbf{b}_2$ aufgespannte Ebene. Als Basis wählen wir $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$, wobei b_3 ein Vektor ist, der senkrecht auf die Ebene steht. Dann gelten $p(\mathbf{b}_1) = \mathbf{b}_1$, $p(\mathbf{b}_2) = \mathbf{b}_2$ und $p(\mathbf{b}_2) = \mathbf{o}$. Die Abbildungsmatrix ist daher

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

Sei i die lineare Abbildung, welche jeden Vektor der Ebene um 90°. Als Basis wählen wir zwei gleichlange Vektoren, die aufeinander senkrecht stehen, sodaß $i(\mathbf{b}_1)$ und $i(\mathbf{b}_2) = -\mathbf{b}_1$ ist. Dann erhalten wir als Abbildungsmatrix

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

Bei diesen Beispielen waren die Abbildungsmatrizen immer besonders einfach, da die Basen passend gewählt wurden.

Sei ein Vekorraum V fixiert. Dann ist mit d_α jedem Skalar α eindeutig eine lineare Abbildung zugeordnet. Lineare Abbildungen, und daher auch Matrizen, können somit auch als „verallgemeinerte Zahlen“ aufgefaßt werden. Diese Betrachtungsweise macht nur dann Sinn, wenn auch die wichtigsten Rechenoperationen verallgemeinert werden können.

Zwei lineare Abbildungen $g, h\colon V \to W$ können punktweise addiert werden:

$$(g + h)\mathbf{v} = g\mathbf{v} + h\mathbf{v}.$$

Die entsprechenden Abbildungsmatrizen werden dabei komponentenweise addiert (sie haben dasselbe Format). Diese Addition führt tatsächlich zu einer

kommutativen Gruppe. Auch die Skalierung funktioniert punkt- bzw. komponentenweise. Damit bildet die Gesamtheit aller linearen Abbildungen zwischen zwei Vektorrämen wieder einen Vektorraum, ebenso die alle Matrizen eines bestimmten Formats. Die komponentenweise Multiplikation ist dagegen nicht in ähnlicher Weise mit einer Operation für lineare Abbildungen in Verbindung zu bringen und daher nicht interessant. Dennoch gibt es auch eine sinnvolle Multiplikation von linearen Abbildungen bzw. von Matrizen.

Seien $g\colon U \to V$, $h\colon V \to W$ lineare Abbildungen, dann können sie, so wie nicht-lineare Abbildungen auch, hintereinander ausgeführt werden. Dies ist ein sinnvoller Multiplikationsbegriff, denn es gelten beide Distributivgesetze:

$$(h_1 + h_2)g = h_1 g + h_2 g$$
$$h(g_1 + g_2) = hg_1 + hg_2$$

Man beachte, diese beiden Gesetze keineswegs äquivalent sind: das erste entspricht der Definition für die Addition und gilt auch für nicht-lineare Abbildungen; das zweite dagegen besagt gerade, daß h linear ist.

Seien A, B, C Basen von U, V, W. Analog zur Situation bei der Addition, sollte für die Multiplikation von Matrizen gelten.

$$h_{B,C} \cdot g_{A,B} = (h \circ g)_{A,C}.$$

Etwas Rechnung ergibt dann, daß Matrizen auf folgende Weise zu multiplizieren sind: Seien $\mathbf{A} = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \dots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} b_{1,1} & \dots & b_{1,p} \\ \vdots & \dots & \vdots \\ bn,1 & \dots & b_{n,p} \end{pmatrix}$, dann ist $\mathbf{A} \cdot \mathbf{B} = \begin{pmatrix} c_{1,1} & \dots & c_{1,p} \\ \vdots & \dots & \vdots \\ c_{m,1} & \dots & c_{m,p} \end{pmatrix}$, mit

$$c_{i,k} = \sum_{j=1} a_{i,j} b_{j,k}.$$

Damit bilden auch die quadratischen Matrizen (jeweils mit fixer Dimension) einen Ring (genauer: eine *lineare Algebra*). Man kann daher mit Matrizen oder linearen Abbildungen ganz ähnlich wie mit Zahlen rechnen. Hauptunterschied, neben dem Fehlen des Kommutativgesetzes, ist, daß es bei Matrizen (und linearen Abbildungen) verschiedene Abstufungen zwischen Null und Nicht-Null gibt, insbesondere muß eine von der Nullmatrix verschiedene Matrix nicht invertierbar sein, etwa die Projektionsmatrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, denn

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix},$$

was niemals mit dem neutralen Element $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ übereinstimmen kann. Insbesondere ist auch die Nullteilerfreiheit verletzt, d.h. das Produkt von zwei Matrizen kann durchaus die Nullmatrix ergeben, ohne daß einer der Faktoren die Nullmatrix ist:

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

Dies hat zur Folge, daß sich gegenüber Zahlen vor allem dann etwas ändert, wenn Divisionen ins Spiel kommen.

Erwähnenswert ist auch, daß sich damit auch Wurzeln negativer Zahlen konstruieren lassen. Sei etwa i, wie oben, eine Drehung um 90° in einer Ebene. Dann entspricht i^2 einer Drehung um 180°. Das ist aber dasselbe wie die Streckung s_{-1}, welche wir der Zahl -1 identifizieren können. Damit erhalten wir $i^2 = -1$, bzw. $i = \sqrt{-1}$. Die entsprechende Gleichheit für Matrizen ist

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

5.6 Lineare Gleichungen

Sind a, b Skalare, dann heißt

$$ax = b$$

eine lineare Gleichung. Diese heißt homogen, wenn $b = 0$, inhomogen sonst. Im homogenen Fall gibt es auf jeden Fall die Lösung $x = 0$. Wenn $a \neq 0$ ist dies auch die einzige Lösung, anderenfalls ist jede Zahl eine Lösung. Im inhomogenen Fall erhält man, wenn $a \neq 0$ ist, ebenfalls genau eine Lösung, anderenfalls aber keine Lösung. Für lineare Abbildungen und Matrizen ist die Situation ganz ähnlicjjfffßsßßeßeh, es gibt aber mehrere Abstufungen.

Sei h eine lineare Abbildung. Dann heißt $h\mathbf{x} = \mathbf{o}$ eine homogene Gleichung. Die Lösungsmenge dieser Gleichung ist dann $\{\mathbf{x}\colon V \mid h\mathbf{x} = \mathbf{o}\}$, besteht also aus all jenen Vektoren, welche durch h auf den Nullvektor abgebildet werden; sie heißt daher auch der *Nullraum* von h. Im eindimensionalen Fall gibt es nur die zwei Fälle: Nullraum besteht nur aus dem Nullvektor, und Nullraum besteht aus ganz V. Im allgemeinen kann der Nullraum ein beliebiger linearer Unterraum von V sein. Die möglichen Lösungen einer inhomogenen Gleichung $h\mathbf{x} = \mathbf{b}$ dagegen sind gerade die affinen Unterräume.

In der Matrizenform hat eine lineare Gleichung die Gestalt $\mathbf{Ax} = \mathbf{b}$, wobei $\mathbf{A}$ eine $m \times n$-Matrix ist und $\mathbf{b}$ ein Spaltenvektor mit m Zeilen. Für die Lösung $\mathbf{x}$ kommt damit nur ein Spaltenvektor mit n Zeilen in Frage. Man spricht in diesem Zusammenhang auch von einem *Gleichungssystem*, genauer von einem System von m Gleichungen in n Unbekannten, denn mit $\mathbf{A} = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \dots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ bn \end{pmatrix}$ und $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ liest sich das ausgeschrieben als

$$\begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & \dots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ bn \end{pmatrix}$$

und ausmultipliziert als

$$\begin{aligned} a_{1,1}x_1 + \dots a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + \dots a_{2,n}x_n &= b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots & \\ a_{m,1}x_1 + \dots a_{m,n}x_n &= b_m \end{aligned}$$

Eine Darstellung der Lösungsmenge (d.h. eine spezielle Lösung und eine Basis für den Nullraum) erhält man z.B. mit dem Gauss'schen Eliminationsverfahren (systematisches „Nullenerzeugen“ durch geeignetes Kombinieren der Zeilen).

5.7 Skalarprodukt

Um die Geometrie eines Vektorraumes genauer zu beschreiben, verwendet man oft ein Skalarprodukt (auch inneres Produkt).

5.7.1 Definition. Ein *Skalarprodukt* ist eine positiv definite symmetrische Bilinearform auf einem Vektorraum, d.h. eine Abbildung $\langle .\,,\,. \rangle \colon V \to V \to \mathbb{K}$, sodaß

$$\begin{aligned} \langle \lambda \cdot \mathbf{u} + \mu \cdot \mathbf{v}\,,\, \mathbf{w} \rangle &= \lambda \cdot \langle \mathbf{u}\,,\, \mathbf{w} \rangle + \mu \cdot \langle \mathbf{v}\,,\, \mathbf{w} \rangle, \\ \langle \mathbf{v}\,,\, \mathbf{w} \rangle &= \langle \mathbf{w}\,,\, \mathbf{v} \rangle, \\ \langle \mathbf{v}\,,\, \mathbf{v} \rangle &\geq 0, \\ \langle \mathbf{v}\,,\, \mathbf{v} \rangle = 0 &\iff \mathbf{v} = \mathbf{o}. \end{aligned}$$

Einen Vektorraum, für den ein Skalarprodukt definiert ist, nennt man eine *Hilbertraum*

Man beachte, daß das Skalarprodukt zweier Vektoren nicht wieder ein Vektor ist (sondern ein Skalar). Insbesondere gilt auch das Assoziativgesetz nicht.

5.7.2 Beispiel. Für Punkte in der Ebene (und ebenso im Raum) erhalten wir ein Skalarprodukt, indem wir festlegen: Das Skalarprodukt von zwei Vektoren sei das Produkt von deren Längen und dem Cosinus des eingeschlossenen Winkels.

Motiviert durch dieses Beispiel legen wir fest:

5.7.3 Definition. Zwei Vektoren $\mathbf{v}$, $\mathbf{w}$ heißen *orthogonal*, wenn $\langle \mathbf{v}\,,\, \mathbf{w} \rangle = 0$. $\sqrt{\langle \mathbf{v}\,,\, \mathbf{v} \rangle}$ wird mit $\|\mathbf{v}\|$ bezeichnet und heißt die *Länge* des Vektors $\mathbf{v}$.

Für orthogonale Vektoren gilt der Satz von Pythagoras:

5.7.4 Theorem. $\langle \mathbf{v}\,,\, \mathbf{w} \rangle = 0 \implies \|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$.

Für das Skalarprodukt von Punkten in der Ebene ist unmittelbar klar, daß es nie größer sein kann als das Produkt der Längen der einzelnen Vektoren. Dies gilt ganz allgemein und heißt *Cauchy-Schwarz-Ungleichung*:

5.7.5 Theorem. *Für beliebige Vektoren eines Hilbertraumes gilt:*

$$|\langle \mathbf{v}\,,\, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|\,.$$

Somit kann man für Vektoren eines beliebigen Hilbertraums einen Winkel definieren:

5.7.6 Definition. Der Winkel φ zwischen zwei Vekoren $\mathbf{v}, \mathbf{w}$ eines beliebigen Hilbertraumes ist definiert duch

$$\cos \varphi = \frac{\langle \mathbf{v}\,,\, \mathbf{w} \rangle}{\|\mathbf{v}\| \cdot \|\mathbf{w}\|}.$$

All diese Begriffe sind somit auch für Hilberträume sinnvoll, welche keine direkte geometrische Bedeutung haben.

5.7.7 Beispiel. 1. In $\mathbb{R}^2$ definiert man ein Skalarprodukt durch

$$\langle (x_1, y_1)\,,\, (x_2, y_2) \rangle := x_1 x_2 + y_1 y_2.$$

2. Allgemeiner definiert man im $\mathbb{R}^n$:

$$\langle (x_1, \ldots, x_n)\,,\, (y_1, \ldots, y_n) \rangle := \sum_{k=1}^{n} x_i \cdot y_i.$$

3. Für (ausreichend beschränkte) Funktionen definiert man

$$\langle f\,,\, g \rangle := \int_a^b f(x) \cdot g(x)\, dx.$$

4. Und für Zufallsvariable (ebenfalls ausreichend beschränkt) ist

$$\langle X\,,\, Y \rangle := E(X \cdot Y)$$

Damit wir wir weiterhin Punkte in der Ebene (oder im Raum) mit dem $\mathbb{R}^2$ (bzw. $\mathbb{R}^3$) identifizieren können muß die Koordinatenausbildung nicht nur einer linearer Isomorphismus (d.h. linear und bijektiv). sondern auch eine Isometrie sein, d.h. sie muß zusätzlich das Skalarprodukt unverändert lassen. Dies gilt genau dann wenn wir eine *Orthonormalbasis* wählen, d.h. alle Basisvektoren müssen aufeinander orthogonal stehen und Länge 1 haben.

Ein endlich-dimensionaler Hilbertraum heißt *Euklidischer Vektorraum*. Man erkennt leicht, daß jeder Euklidische Vektorraum zu einem $\mathbb{R}^n$ isomorph ist.

Für Matrizen erkennt man, daß der Nullraum genau aus jenen Vekoren besteht, welche zu den Zeilen der Matrix orthogonal sind; d.h. der Nullraum ist das orthogonale Komplement des Zeilenraums.