Studienbücherei



D. Ilse · I. Lehmann · W. Schulz

Gruppoide und Funktionalgleichungen



VEB Deutscher Verlag der Wissenschaften

Mathematik für Lehrer
Band 20

Herausgegeben von:
W. Engel, S. Brehmer, M. Schneider, H. Wussing

Unter Mitarbeit von:
G. Asser, J. Böhm, J. Flachsmeyer, G. Geise, T. Glocke,
K. Härtig, G. Kasdorf, O. Krötenheerdt, H. Lugowski,
P. H. Müller, G. Porath

Studienbücherei

Gruppoide und Funktionalgleichungen

D. Ilse
I. Lehmann
W. Schulz

Mit 77 Abbildungen

VEB Deutscher Verlag
der Wissenschaften
Berlin 1984

Verlagslektor: Erika Arndt
Verlagshersteller: Birgit Burkhardt
Umschlaggestaltung: Rudolf Wendt
© 1984 VEB Deutscher Verlag der Wissenschaften, DDR - 1080 Berlin, Postfach 1216
Lizenz-Nr. 206 · 435/106/84
Printed in the German Democratic Republic
Gesamtherstellung: VEB Druckhaus „Maxim Gorki", 7400 Altenburg
LSV 1034
Bestellnummer: 571 190 6
01980

Vorwort

Die Tatsache, daß Abbildungen oder Funktionen in der Mathematik und ihren Anwendungen eine sehr wichtige Rolle spielen, spiegelt sich auch im Mathematikunterricht wider. Eindeutige Zuordnungen treten nämlich sowohl im Arithmetikunterricht als auch im Geometrie- und Analysisunterricht auf. Von der Klasse 1 an beschäftigen sich die Schüler mit den Rechenoperationen in verschiedenen Zahlbereichen. In der Klasse 4 beginnt die Untersuchung geometrischer Abbildungen, und von Klasse 8 an werden elementare Funktionen recht gründlich untersucht. Insgesamt enthält der Mathematikunterricht der allgemeinbildenden Schule vielfältige Möglichkeiten, das funktionale Denken der Lernenden zu schulen und die Fähigkeit zu entwickeln, funktionale Abhängigkeiten in Natur und Technik auch selbständig erfassen zu können. Wir sind der Meinung, daß diese Möglichkeiten noch zu wenig genutzt werden. Von dieser Erkenntnis ausgehend, beschäftigen wir uns im vorliegenden Buch mit eindeutigen Zuordnungen, und zwar mit binären Operationen (Kap. 1 bis 5) und mit elementaren Funktionen (Kap. 6 bis 13). Dabei handelt es sich im ersten Fall um ausgewählte Teile der Theorie der Gruppoide und im zweiten Fall um ausgewählte Teile der Theorie der Funktionalgleichungen (im Sinne von J. Aczél [3]).

Bei der Stoffauswahl spielte eine wesentliche Rolle, daß sich die Motivationen für die jeweiligen Untersuchungen dem Mathematikunterricht entnehmen lassen. Dadurch werden mathematische Hintergründe im Schulstoff aufgehellt, deren Kenntnis den Lehrer befähigt, wichtige Zusammenhänge zu erkennen. Darüber hinaus werden bei der Darstellung des Stoffes insbesondere folgende Ziele verfolgt:

- Der Mathematiklehrer lernt über die vier Grundrechenarten hinaus eine Vielfalt binärer Operationen kennen. Dabei kommen neben Assoziativität und Kommutivität weitere Eigenschaften binärer Operationen zur Sprache. Viele dieser Operationen und Eigenschaften sind schon im Mathematiklehrgang angelegt.
- Durch vielfältige (z. T. recht elementare) axiomatische Untersuchungen werden Beziehungen zwischen den betrachteten Eigenschaften binärer Operationen bzw. elementarer Funktionen hergestellt. Dabei wird das Wissen über die Grundrechenarten und die elementaren Funktionen abgerundet.
- Axiomatische Untersuchungen über Eigenschaften elementarer Funktionen werden bis zur funktionalen Charakterisierung dieser Funktionen geführt. Diese Betrachtungen beinhalten auch, daß das im Mathematikunterricht praktizierte Verfahren, Eigenschaften elementarer Funktionen abzuleiten, ergänzt wird

durch die entgegengesetzte Sichtrichtung, nämlich zu gegebenen Eigenschaften Funktionen zu finden, die diese Eigenschaften besitzen. Der zuletzt genannte Aspekt ist auch bei Anwendungen der Mathematik wichtig.

- Sowohl im algebraisch orientierten ersten Teil als auch im analytisch orientierten zweiten Teil des Buches werden Beziehungen zu anderen mathematischen Gebieten hergestellt, so z. B. zur euklidischen und Lobačevskijschen Geometrie.
- Wichtige mathematische Grundbegriffe (wie z. B. Gruppe, Äquivalenzrelation, Isomorphie, Stetigkeit, Monotonie) werden gefestigt und vertieft.

Im Zusammenhang mit den genannten Zielen ergeben sich Potenzen für die Weiterentwicklung des Mathematikunterrichts. Darauf gehen wir in diesem Buch nicht weiter ein, verweisen jedoch gelegentlich auf entsprechende schulmathematische Literatur. Wir haben weder eine vollständige noch eine systematische Darstellung der Theorie der Gruppoide bzw. der Funktionalgleichungen angestrebt (vgl. dazu die Monographien von J. Aczél [3], V. D. Belousov [21], R. H. Bruck [32] sowie J. Dénes und A. D. Keedwell [44]). Vielmehr standen die genannten Gesichtspunkte der Mathematiklehrerausbildung im Vordergrund. Wir haben jedoch versucht, dieses Anliegen so zu verwirklichen, daß zwei systematisch aufgebaute und in sich geschlossene Teile entstehen, die zudem auch untereinander viele Beziehungen haben. So treten z. B. Loops, geometrische Schließungsbedingungen und Isotopismen in beiden Teilen auf (siehe Kap. 1, 4, 11). Überhaupt ist eine enge Verwandtschaft bereits dadurch gegeben, daß auch in dem Teil über Funktionalgleichungen binäre Operationen untersucht werden, wobei hier aber wesentlich die Stetigkeit hinzukommt.

Bei der Arbeit mit diesem Buch ist es möglich, in sich geschlossene Lehrveranstaltungen anhand ausgewählter Teile durchzuführen. Bewährt haben sich z. B. folgende Zusammenstellungen:

a) Kapitel 1, 4, 7 und 11,

b) Kapitel 1 bis 5,

c) Kapitel 6 bis 13.

Bei der Auswahl von Seminarthemen empfiehlt es sich, die Schlüsselstellung der Kapitel 1 und 7 zu berücksichtigen. Die teilweise recht umfangreichen Literaturhinweise sollen auch die Auswahl von Themen für weiterführende Seminare und für Diplomarbeiten erleichtern. Dabei haben wir auch die bereits erwähnte schulmathematische Literatur einbezogen sowie uns bemüht, leicht zu beschaffende Literatur besonders zu berücksichtigen.

Wir danken den Herausgebern, insbesondere Herrn Prof. Dr. S. Brehmer, für zahlreiche konstruktive und kritische Hinweise. Unser Dank gilt weiterhin der Verlagslektorin Frau E. Arndt für gute Zusammenarbeit und für die sachkundige redaktionelle Bearbeitung des Manuskripts. Dem VEB Druckhaus „Maxim Gorki" danken wir für die sorgfältige Arbeit.

Berlin, im Winter 1983/84

Dieter Ilse
Ingmar Lehmann
Wolfgang Schulz

Inhalt

Gruppoide

1. Algebraische Strukturen mit einer binären Operation 13

1.1. Gruppoide, Strukturtafeln und Cayley-Diagramme 14
1.2. D-Gruppoide und K-Gruppoide 19
1.3. Quasigruppen, lateinische Quadrate und Quasigruppen-Graphen 22
1.4. Loops 29
1.5. Halbgruppen und Gruppen 32
1.6. Assoziativitätstest 35
1.7. Translationen in Gruppoiden 42

2. Parastrophie von Quasigruppen 47

2.1. Parastrophe Quasigruppen 47
2.2. Parastrophe lateinische Quadrate 56
2.3. Die Anzahl p der Parastrophen einer Quasigruppe 59
2.4. Parastrophe Identitäten 62

3. Homomorphie 75

3.1. Äquivalenzrelationen 75
3.2. Kongruenzrelationen und Faktorgruppoide 81
3.3. Homomorphe und isomorphe Abbildungen 88

4. Isotopie 99

4.1. Isotope und Hauptisotope 99
4.2. Isotope von Gruppen 108

5. Spezielle Gruppoide und Quasigruppen 116

5.1. Bisymmetrische Gruppoide und Quasigruppen 116
5.2. Zusammenhänge zwischen ausgewählten Identitäten 121
5.3. Totalsymmetrische Quasigruppen und Steiner-Tripel-Systeme 130

Funktionalgleichungen

6. Über gegenseitige Ableitbarkeit bekannter Eigenschaften einiger elementarer Funktionen 141

6.1. Aus den Theorien einiger elementarer Funktionen 141

7. Funktionale Charakterisierungen der linearen Funktionen $f: \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x$ mit Hilfe der Funktionalgleichung $f(x + y) = f(x) + f(y)$. 147

7.1. Die stetigen Lösungen der Funktionalgleichung $f(x + y) = f(x) + f(y)$ 147
7.2. Weitere funktionale Charakterisierungen der linearen Funktionen $f: \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x$ 149
7.3. Die unstetigen Lösungen der Funktionalgleichung $f(x + y) = f(x) + f(y)$ . . . 154
7.4. Zu Lösungen von $f(x + y) = f(x) + f(y)$ über eingeschränktem Definitionsbereich 156
7.5. Bemerkungen zur direkten Proportionalität 158
7.6. Anwendungen 161

8. Funktionale Charakterisierungen der Exponential- und Logarithmusfunktionen mit Hilfe von Funktionalgleichungen 165

8.1. Funktionale Charakterisierungen der Exponentialfunktionen mit Hilfe der Funktionalgleichung $f(x + y) = f(x) \cdot f(y)$ 165
8.2. Die unstetigen Lösungen der Funktionalgleichung $f(x + y) = f(x) \cdot f(y)$ . . . 171
8.3. Funktionale Charakterisierungen der Logarithmusfunktionen mit Hilfe der Funktionalgleichung $f(x \cdot y) = f(x) + f(y)$ 172
8.4. Zusammenfassung 177

9. Verallgemeinerung der Exponential- und Logarithmusfunktionen unter Beibehaltung der Stetigkeitsforderung 179

9.1. Verallgemeinerung der Exponential- und Logarithmusfunktionen auf algebraischem Weg mit Hilfe der Funktionalgleichung $f(x + y) = F(f(x), f(y))$ . . . 179
9.2. Verallgemeinerung der Exponential- und Logarithmusfunktionen auf analytischem Weg mit Hilfe der Funktionalgleichung $f(F(x, y)) = f(x) + f(y)$ . . . 185

10. Verallgemeinerung der Exponentialfunktionen ohne Stetigkeitsforderung . . . 190

10.1. Dividierbare Gruppen 190
10.2. Allgemeine Lösung der Funktionalgleichung $f(x + y) = F(f(x), f(y))$ 191

11. Die Funktionalgleichung $h(F(x, y)) = f(x) + g(y)$ 198

11.1. Rechenstäbe, Netztafeln, Fluchtlinientafeln und die Schließungsbedingung von Brianchon 198
11.2. Charakterisierung einer Klasse zweistelliger reellwertiger Funktionen F mit $h(F(x, y)) = f(x) + g(y)$ 208

12. Funktionale Charakterisierungen von Winkelfunktionen 221

12.1. Die Additions- und Subtraktionstheoreme der Sinus- und Kosinusfunktionen . . 221
12.2. Äquivalente Regularitätsforderungen bei der funktionalen Charakterisierung der Sinus- und Kosinusfunktionen 232
12.3. Die Funktionalgleichungen $f(x + y) + f(x - y) = 2f(x) f(y)$ und $f(x + y) f(x - y) = f(x)^2 - f(y)^2$ 239
12.4. Anwendungen der d'Alembertschen Funktionalgleichung in der Geometrie und in der Physik 248

18. Lösungen der Funktionalgleichung $f(x + yf(x)) = f(x) \cdot f(y)$ 270

13.1. Einige Eigenschaften der Lösungen von $f(x + yf(x)) = f(x) \cdot f(y)$ 270
13.2. Reellwertige Lösungen von $f(x + yf(x)) = f(x) \cdot f(y)$ 271
13.3. Die allgemeine Lösung von $f(x + yf(x)) = f(x) \cdot f(y)$ 274
13.4. Die stetigen Lösungen von $f(x + yf(x)) = f(x) \cdot f(y)$ für einen Hilbertraum . . . 277

Literatur 281

Namenverzeichnis 288

Sachverzeichnis 291

Gruppoide

1. Algebraische Strukturen mit einer binären Operation

Strukturen spielen in der heutigen Mathematik eine zentrale Rolle. Wir beschäftigen uns im folgenden mit besonders einfachen mathematischen (algebraischen) Strukturen, nämlich solchen, die durch eine Trägermenge und eine zweistellige Operation festgelegt sind (vgl. MfL Bd. 1, 2.7., und MfL Bd. 3, 11.3.). Sie werden *Gruppoide* genannt. Halbgruppen und Gruppen sind spezielle Gruppoide. Besonders Gruppen werden wir mitunter heranziehen, um den Einstieg in ein Problem zu motivieren.

Läßt sich auch eine Reihe von Resultaten, die man für Gruppen gewonnen hat, auf Gruppoide (Quasigruppen, Loops) übertragen, so kann man die Theorie der Gruppoide heute keineswegs mehr als eine bloße „verallgemeinerte Gruppentheorie" bezeichnen.

Auch im gegenwärtigen Mathematikunterricht treten viele Operationen mit z. T. ganz unterschiedlichen Eigenschaften auf. Neben der Assoziativität und der Kommutativität, die den Schülern geläufig sind, sind im Mathematiklehrgang z. B. auch solche Eigenschaften wie Umkehrbarkeit, Kürzbarkeit oder die Existenz eines neutralen Elements angelegt. Sucht man z. B. Operationen $\circ$, die die Eigenschaften

(A) $(a \circ b) \circ c = a \circ (b \circ c)$ (Assoziativität),

(K) $a \circ b = b \circ a$ (Kommutativität)

oder

(N) $a \circ n = n \circ a = a$ (Existenz eines neutralen Elements)

besitzen, wobei n ein festes Element ist und a, b, c beliebige Elemente der zugrunde gelegten Menge sind, so ist die Addition (z. B. in $\mathbf{Z}$) ein Modell für alle drei Eigenschaften, während die sogenannte „Mittelpunktskonstruktion" in einer Ebene ε ($P \circ Q = R :\Leftrightarrow R$ ist Mittelpunkt der Strecke $\overline{PQ}$ und $P \circ P := P$ für alle $P, Q \in \varepsilon$) wohl eine Realisierung von (K), nicht aber von (A) und auch nicht von (N) darstellt. Anhand des letzten Beispiels ergibt sich auch schon, daß aus der Kommutativität einer Operation nicht deren Assoziativität folgen kann.

Ebenso einfach läßt sich zeigen, daß die Assoziativität ihrerseits nicht die Kommutativität nach sich zieht. Gelingt es, aus gegebenen Eigenschaften (hier zweistelliger Operationen) eine weitere Eigenschaft abzuleiten, dann gilt die so gewonnene Eigenschaft in jedem Modell der Ausgangseigenschaften.

Beispielsweise folgt aus (A) und (K)

(BS) $(a \circ b) \circ (c \circ d) = (a \circ c) \circ (b \circ d)$ (Bisymmetrie)

für alle a, b, c und d, so daß die Addition notwendig auch ein Modell von (BS) ist. Daß sich umgekehrt die Assoziativität nicht aus der Bisymmetrie ableiten läßt, zeigt die zuvor definierte Mittelpunktskonstruktion. Die Bisymmetrie dieser Operation spiegelt sich in dem bekannten Satz wider, daß die Mittelpunkte der Seiten eines Vierecks gerade die Eckpunkte eines Parallelogramms sind. Diese Operation ist aber nicht assoziativ. Auch die Kommuntativität folgt nicht aus der Bisymmetrie. Das läßt sich nicht mit der Mittelpunktskonstruktion beweisen. Hier kann man die Substraktion als Gegenbeispiel heranziehen. Sie ist, wie man leicht sieht, bisymmetrisch, aber nicht kommutativ.

In dem hier skizzierten Sinne kann man also einerseits auf vielfältige Weise den Umgang mit Modellen (für gewisse Axiomensysteme) pflegen und andererseits auch die Frage der Ableitbarkeit bzw. der Nichtableitbarkeit von Eigenschaften aufwerfen (vgl. auch Kap. 6ff.).

1.1. Gruppoide, Strukturtafeln und Cayley-Diagramme

Es sei G eine beliebige Menge. Dann wollen wir unter einer binären (d. h. zweistelligen) Operation in G eine Abbildung[1]) $\circ$ von der Produktmenge $G \times G$ in die Menge G verstehen:

1.1.1. Definition. $\circ$ heißt *binäre Operation* in $G :\Leftrightarrow \circ : G \times G \to G$.

Jedem geordneten Paar (x, y) von Elementen x und y aus G wird also durch $\circ$ genau ein Element z aus G zugeordnet. Das Element z nennen wir (in Anlehnung an die Multiplikation von Zahlen) das *Produkt* von x und y und schreiben deshalb auch für z anstelle von $\circ(x, y)$ einfach $x \circ y$. Die Menge G heißt die *Trägermenge* der binären Operation $\circ$.

Da wir uns ausschließlich mit binären (zweistelligen) Operationen beschäftigen werden, verzichten wir im folgenden auf die Angabe der Arität (Stelligkeit)[2]).

Wenn die Trägermenge G einer Operation $\circ$ nicht leer ist, erhält man bereits eine besonders einfache algebraische Struktur:

1.1.2. Definition. $\mathfrak{G} = (G, \circ)$ heißt *Gruppoid* $:\Leftrightarrow G \neq \emptyset$ und $\circ$ ist eine Operation in G.

Der Begriff „Gruppoid" geht auf H. Brandt (1886—1954) [28] zurück. Die im Rahmen seiner Untersuchungen über quadratische Formen betrachtete Struktur heißt heute *Brandtsches Gruppoid*. In dem hier definierten Sinne hat sich der Begriff „Gruppoid" erst später eingebürgert.

Im Mathematikunterricht wird der Begriff der Operation zwar benutzt, so wird z. B. schon in der Unterstufe von den vier Grundrechenoperationen gesprochen, eine präzise Definition des Begriffes erfolgt jedoch nicht.

[1]) Zur Definition des Begriffs Abbildung vgl. MfL Bd. 1, 2.4. Abbildung und Funktion benutzen wir als synonyme Begriffe.

[2]) Zur Definition k-stelliger Operationen mit $k \in \mathbf{N}^*$ vgl. MfL Bd. 1, 2.6.

Die Definition 1.1.1 sichert für alle $x, y \in G$ sowohl die Existenz des Produkts $x \circ y$ in G (Abgeschlossenheit von G bezüglich $\circ$; die Operation $\circ$ führt nicht aus G hinaus) als auch die Eindeutigkeit von $x \circ y$. Die Subtraktion in $\mathbf{N}$ ist in unserem Sinne also ebensowenig eine Operation wie alle vier Grundrechenarten in der Menge der Primzahlen. Auch das arithmetische Mittel $\left(x \circ y := \frac{x+y}{2}\right)$ liefert in $\mathbf{N}$ oder $\mathbf{Z}$ keine Operation. Natürlich gelingt es in all diesen Fällen, durch geeignet gewählte Trägermengen (Zahlbereichserweiterungen) Operationen zu erhalten. Im Fall des Skalarprodukts (in einem Vektorraum), das in unserem Sinne ebenfalls keine Operation liefert, ist das nicht möglich.

1.1.3. Beispiele für Gruppoide

1. $(\mathbf{N}^*, \circ)$ mit $x \circ y := x^y$.

2. $(\mathbf{Z}, \circ)$ mit $x \circ y := x^k + y^k$, wobei k eine beliebige, aber feste gerade natürliche Zahle mit $k > 1$ ist.

3. Es sei D die Menge aller nichtnegativen reellen Zahlen, die nach einer der üblichen Regeln auf m Dezimalstellen ($m \in \mathbf{N}^*$, beliebig, fest) gerundet sind. Die Operation $\circ$ ordne jedem Paar (x, y) derart gerundeter Zahlen das (nach dieser Regel) gerundete Produkt von x und y zu (R. H. Bruck [33]).

4. Über $G = \{a, b, c, d\}$ definieren wir die Operation $\circ$ durch

$$x \circ y := \begin{cases} a & \text{für} \quad x = a \text{ oder } y = a, \\ x & \text{für} \quad x = y, \\ z & \text{für} \quad x \neq y \text{ und } x \neq a \text{ und } y \neq a, \text{ wobei } z \neq x \text{ und } z \neq y \text{ und } z \neq a. \end{cases}$$

Unter der *Ordnung* $|\mathfrak{G}|$ eines Gruppoids $\mathfrak{G} = (G, \circ)$ verstehen wir (wie in der Gruppentheorie, vgl. MfL Bd. 3, 12.1.) die Anzahl der Elemente von G, wenn diese endlich ist. $\mathfrak{G}$ heißt dann ein *endliches* Gruppoid. Anderenfalls heißt $\mathfrak{G}$ ein *unendliches* Gruppoid. Das Beispiel 1.1.3.4 ist — im Gegensatz zu den vorangegangenen Gruppoiden — ein endliches Gruppoid.

Ist $\mathfrak{G} = (G, \circ)$ ein endliches Gruppoid, dann ist $\circ$ eine von $|\mathfrak{G}|^{(|\mathfrak{G}|^2)}$ möglichen Operationen in G. In diesem Zusammenhang sei vermerkt, daß zwei Operationen $\circ$ und $*$ genau dann identisch sind, wenn $x \circ y = x * y$ für alle x und y gilt.[1]) Sie müssen also notwendig in derselben (endlichen oder unendlichen) Trägermenge definiert sein.

Es sei $\mathfrak{G} = (G, \circ)$ ein endliches Gruppoid mit $G = \{a_1, a_2, \ldots, a_m\}$, also $|\mathfrak{G}| = m$. Wenn wir die in G definierte Operation $\circ$ in dem Sinne vollständig erfassen wollen, daß wir zu jedem geordneten Paar von Elementen x und y auch deren Produkt $x \circ y$ kennen, ist es oft vorteilhaft, wenn wir uns auf die zugehörige Strukturtafel (oder auch Operationstabelle, vgl. MfL Bd. 1, 2.6.) von $\mathfrak{G}$ stützen:

$\circ$	$\ldots$	a_j	$\ldots$
$\vdots$		$\vdots$	
a_i	$\ldots$	a_{ij}	
$\vdots$			

[1]) $\circ = * \Leftrightarrow \bigwedge_{(x,y)} ((x, y) \in G \times G \Rightarrow x \circ y = x * y)$ (vgl. MfL Bd. 1, 2.4.).

Für alle $a_i, a_j \in G$ mit $1 \leqq i, j \leqq m$ gilt

1.1.4. $a_{ij} := a_i \circ a_j$.

Für Gruppen, also eine spezielle Klasse von Gruppoiden, benutzte A. CAYLEY (1821—1895) als erster (1854) solche Tafeln; in diesem Fall heißen sie auch *Gruppentafeln* oder *Cayley-Tafeln*.

Die *Strukturtafel* eines endlichen Gruppoids $\mathfrak{G} = (G, \circ)$ besteht also neben der (vertikalen) Eingangs- oder Kopfspalte und der (horizontalen) Eingangs- oder Kopfzeile aus einem quadratischen Schema (Matrix) mit m^2 Feldern. Das Element $a_{ij} \in G$ steht im Schnittpunkt der i-ten Zeile und der j-ten Spalte. Die äußere Gestalt der Strukturtafel ist natürlich von der jeweils gewählten Reihenfolge der Elemente in den beiden Eingängen abhängig. Alle demnach möglichen Tafeln spiegeln aber dieselbe Operation wider, so daß wir von der Strukturtafel von $\mathfrak{G}$ sprechen können. Für die Kopfspalte und -zeile verabreden wir dieselbe Reihenfolge der Elemente der Trägermenge.

Zur Illustration betrachten wir die Strukturtafel des Gruppoids aus Beispiel 1.1.3.4:

$\circ$	a	b	c	d
a	a	a	a	a
b	a	b	d	c
c	a	d	c	b
d	a	c	b	d

Die definierenden Eigenschaften eines (endlichen) Gruppoids $\mathfrak{G} = (G, \circ)$ spiegeln sich in der zugehörigen Strukturtafel in der Form wider, daß jedes Feld mit genau einem Element aus G besetzt ist. Damit ist umgekehrt durch eine Tafel, die dieses Kriterium erfüllt, eine Operation in einer endlichen Menge definiert.

Wenn man einem endlichen Gruppoid $\mathfrak{G} = (G, \circ)$ auf eine noch zu beschreibende Weise einen speziellen Graphen zuordnet, gewährleistet dieser Graph (wie im Fall der Strukturtafel) das Ablesen und eine übersichtliche Anordnung der Produkte $x \circ y$ für alle $x, y \in G$. Darüber hinaus erlauben beide Formen der Darstellung oder Veranschaulichung, also sowohl Strukturtafeln als auch Graphen, bestimmte Rückschlüsse auf eventuell zu untersuchende Eigenschaften von Gruppoiden. In bestimmten Fällen erweist sich die graphentheoretische Darstellung der Strukturtafel überlegen.

Ein (endlicher) gerichteter Graph Γ ist durch zwei (endliche) nichtleere Mengen **K** und **B** festgelegt. Die Elemente von **K** heißen *Knoten*, die Elemente von **B** heißen *gerichtete Kanten* oder *Bögen*[1]) des Graphen. Ist $b \in \mathbf{B}$, so gibt es zwei Knoten $A(b)$ und $E(b)$ in **K**, die *Anfangs-* bzw. *Endpunkt* von b heißen. Neben *Schleifen*, d. h. Kanten, deren Anfangs- und Endpunkt zusammenfallen, lassen wir auch Mehrfachbögen, also mehrere Kanten gleicher Orientierung zwischen zwei Knoten, zu.[2]) Wir betrachten im folgenden nur endliche Graphen.

[1]) Die Bögen haben also einen Durchlaufssinn.

[2]) Da wir Schleifen und Mehrfachbögen nicht ausschließen, können wir uns nicht auf die Definition 1 in MfL Bd. 12, 7.2., stützen.

Ein gerichteter Graph $\Gamma = (\mathbf{K}, \mathbf{B})$ heißt *kantengefärbt*, wenn es eine Abbildung $\alpha: \mathbf{B} \to \mathbf{K}$ gibt, so daß keine zwei Bögen mit demselben Anfangspunkt dasselbe Bild bei α haben, d. h., wir fordern

$$\textbf{1.1.5.} \quad \bigwedge_{b_1, b_2 \in \mathbf{B}} \big(b_1 \neq b_2 \wedge A(b_1) = A(b_2) \Rightarrow \alpha(b_1) \neq \alpha(b_2)\big).$$

Ist $b \in \mathbf{B}$, $K \in \mathbf{K}$ und $\alpha(b) = K$, so sagen wir, der Bogen b hat die *Farbe* K (ist K-*gefärbt*). Die Menge aller Farben ist dann als Wertebereich von α eine Teilmenge von $\mathbf{K}$: $W(\alpha) \subseteq \mathbf{K}$.

Wegen (1.1.5) sind voneinander verschiedene Bögen, die von demselben Knoten ausgehen, unterschiedlich gefärbt. Gleichfarbige Böden mit demselben Anfangspunkt kann es also nicht geben.

Es sei jetzt $\mathfrak{G} = (G, \circ)$ ein endliches Gruppoid. Ein durch eine surjektive[1]) Abbildung $\alpha: \mathbf{B} \to \mathbf{K}$ kantengefärbter gerichteter Graph $\Gamma = (\mathbf{K}, \mathbf{B})$ heißt *Cayley-Diagramm* von $\mathfrak{G}$, wenn es eine bijektive[2]) Abbildung $\beta: G \to \mathbf{K}$ gibt, so daß für alle $x, y, z \in G$ folgende Eigenschaft gilt:

Ist $x \circ y = z$, dann existiert (genau) ein Bogen b der Farbe $Y \big(= \beta(y)\big)$, der vom Knoten $X \big(= \beta(x)\big)$ zum Knoten $Z \big(= \beta(z)\big)$ führt (siehe Abb. 1), d. h., wir fordern

$$\textbf{1.1.6.} \quad \bigwedge_{x,y,z \in G} \big((x \circ y = z \wedge \beta(x) = X \wedge \beta(y) = Y \wedge \beta(z) = Z)$$
$$\Rightarrow \bigvee_{b \in \mathbf{B}} (A(b) = X \wedge \alpha(b) = Y \wedge E(b) = Z)\big).$$

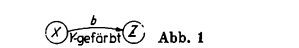


Abb. 1

Das Cayley-Diagramm von $\mathfrak{G}$ bezeichnen wir mit $\Gamma(\mathfrak{G})$.

Die gerichtete Kante $b \in \mathbf{B}$ in (1.1.6) ist wegen (1.1.5) eindeutig bestimmt. Das schließt natürlich nicht aus, daß zwischen den Knoten X und Z weitere (gleichgerichtete) Bögen existieren, die dann aber jeweils paarweise verschieden gefärbt sein müssen.

Jeder Knoten tritt auch als Farbe auf (Surjektivität von α). Damit gehen von jedem Knoten wegen (1.1.6) Bögen aller $|\mathfrak{G}|$ Farben aus. In einem Cayley-Diagramm eines Gruppoids $\mathfrak{G}$ stimmen also die Anzahl der Knoten, die Anzahl der von jedem Knoten abgehenden Bögen, die Anzahl der Farben und die Ordnung von $\mathfrak{G}$ überein.

A. Cayley benutzte als erster (1878) solche Graphen zur Illustration von Gruppen. Das Cayley-Diagramm einer Gruppe heißt *Gruppengraph*, mitunter auch *Dehnsches Gruppenbild* nach M. Dehn (1878–1952). Die Ausdehnung auf beliebige Gruppoide geht auf R. Artzy (1961) zurück (vgl. [17]).

---

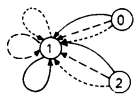
[1]) α ist eine Abbildung auf $\mathbf{K}$.

[2]) β ist eine 1-1-Abbildung von G auf $\mathbf{K}$.

1.1.7. Beispiele[1]

1. $(G, \circ)$ mit $G = \{0, 1, 2\}$ und $x \circ y := 1$ (siehe Abb. 2).

Cayley-Diagramm



0-gefärbt
1-gefärbt
2-gefärbt

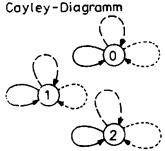
Strukturtafel

∘	0	1	2
0	1	1	1
1	1	1	1
2	1	1	1

Abb. 2

2. $(G, \circ)$ mit $G = \{0, 1, 2\}$ und $x \circ y := x$ (siehe Abb. 3).

Cayley-Diagramm



0-gefärbt
1-gefärbt
2-gefärbt

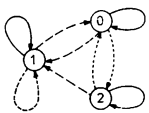
Strukturtafel

∘	0	1	2
0	0	0	0
1	1	1	1
2	2	2	2

Abb. 3

3. $(G, \circ)$ mit $G = \{0, 1, 2\}$ und $x \circ y := |x - y|$ (siehe Abb. 4).

Cayley-Diagramm



0-gefärbt
1-gefärbt
2-gefärbt

Strukturtafel

∘	0	1	2
0	0	1	2
1	1	0	1
2	2	1	0

Abb. 4

Wie im Fall der Strukturtafeln besteht auch hier umgekehrt die Möglichkeit, durch einen geeigneten Graphen eine Operation in einer (endlichen) Menge zu definieren.

1.1.8. Aufgaben

1. Welche der Zahlbereiche $\mathbf{N}^*$, $\mathbf{N}$, $\mathbf{Z}$, $\mathbf{Q}_+^*$, $\mathbf{Q}_+$, $\mathbf{Q}^*$, $\mathbf{Q}$, $\mathbf{R}_+^*$, $\mathbf{R}_+$, $\mathbf{R}^*$, $\mathbf{R}$ sind bezüglich der Bildung des

arithmetischen Mittels $\quad x \circ_1 y := \dfrac{x + y}{2}$,

geometrischen Mittels[2] $\quad x \circ_2 y := \sqrt{x \cdot y}$,

harmonischen Mittels $\quad x \circ_3 y := \dfrac{2xy}{x + y}$,

quadratischen Mittels $\quad x \circ_4 y := \sqrt{\dfrac{x^2 + y^2}{2}}$

[1] Wegen der Bijektivität von β identifizieren wir die Elemente eines Gruppoids $\mathfrak{G}$ und die Knoten des zugehörigen Cayley-Diagramms $\Gamma(\mathfrak{G})$.

[2] Im ersten Kapitel steht das Zeichen „$\cdot$" allein für die Multiplikation. Dabei benutzen wir auch solche Schreibweisen wie „$2x$" statt „$2 \cdot x$".

abgeschlossen, welche nicht? Man zeige, daß für die Mittelwerte folgende Ungleichung gilt:

$$\bigwedge_{x,y\in\mathbf{R}_+} x \circ_0 y \leqq x \circ_2 y \leqq x \circ_1 y \leqq x \circ_4 y.$$

Wann gilt das Gleichheitszeichen?

2. Man überlege sich, welche der 16 Operationen, die sich in der Zweiermenge $G = \{W, F\}$ definieren lassen, die klassischen Wahrheitsfunktionen repräsentieren.

3. Die Bildung des größten gemeinsamen Teilers (ggT) sowie des kleinsten gemeinsamen Vielfachen (kgV) liefert mit den natürlichen Zahlen als Trägermenge ebenfalls Beispiele für Gruppoide: $(\mathbf{N}, \sqcap)$ mit $x \sqcap y := \mathrm{ggT}(x, y)$ und $(\mathbf{N}, \sqcup)$ mit $x \sqcup y := \mathrm{kgV}(x, y)$. Man untersuche, wie sich die Zahlen 0 und 1 in den Gruppoiden $(\mathbf{N}, \sqcap)$ und $(\mathbf{N}, \sqcup)$ verhalten.

4. Man zeige, daß sich Durchschnitt und Vereinigung (mit der Potenzmenge $\mathfrak{P}(M)$ einer beliebigen nichtleeren Menge M als Trägermenge) ähnlich verhalten wie die Operationen $\sqcap$ und $\sqcup$. Welche Elemente in $\mathfrak{P}(M)$ übernehmen dabei die Rolle der 0 bzw. 1 aus Aufgabe 3?

5. Man gebe Strukturtafel und Cayley-Diagramm des Gruppoids $(G, \sqcap)$ mit $G = \{0, 1, 2, 3\}$ und $x \sqcap y := \mathrm{ggT}(x, y)$ an. Warum ist das kleinste gemeinsame Vielfache (kgV) in G keine Operation? Man mache sich klar, daß entweder $0 \sqcap 0 = 0$ oder $0 \sqcap 0 = 3$ gesetzt werden kann vgl. I. Lehmann und W. Schulz [104]).

1.2. D-Gruppoide und K-Gruppoide

Den noch sehr umfassenden Begriff des Gruppoids schränken wir im folgenden in der Weise ein, daß wir bestimmte Forderungen an die Operationen stellen bzw. bestimmte Elemente der Trägermenge auszeichnen.

1.2.1. Definition. $(G, \circ)$ heißt *D-Gruppoid* (Gruppid mit Division) $:\Leftrightarrow (G, \circ)$ ist ein Gruppoid und

(D) die Gleichungen $a \circ x = b$ und $y \circ a = b$ sind für alle $a, b \in G$ lösbar.

(D) heißt in Zeichen: $\bigwedge_{a,b\in G} \left(\bigvee_{x\in G} a \circ x = b \wedge \bigvee_{y\in G} y \circ a = b \right)$.

Die Forderung (D), die also die (nicht notwendig eindeutige) Existenz von Lösungen x, y der Gleichungen $a \circ x = b$ und $y \circ a = b$ garantiert, erinnert an die Division von Zahlen, deshalb auch der Name-D-Gruppoid (vgl. etwa R. H. Bruck [33]). Wenn $\mathfrak{G} = (G, \circ)$ ein D-Gruppoid ist, sagen wir auch, die Operation $\circ$ ist *umkehrbar* (in G).

Nach Definition des Operationsbegriffs zieht jede Änderung der Trägermenge (Einschränkung oder Erweiterung) eine Änderung der Operation selbst nach sich. Demzufolge ist auch bei jeder solchen Änderung die Umkehrbarkeit neu zu untersuchen. So ist die Multiplikation in $\mathbf{Q}$ z. B. nicht umkehrbar, wohl aber in $\mathbf{Q}^*$.

An dieser Stelle möchten wir ausdrücklich darauf hinweisen, daß das Potenzieren $(x \circ y := x^y)$ gemäß Definition 1.2.1 nicht umkehrbar ist. In $\mathbf{N}^*$ hat z. B. weder $1 \circ x = 4$ noch $y \circ 3 = 4$ eine Lösung. Wählen wir die Menge $\mathbf{R}_+^*$ der positiven reellen Zahlen als Trägermenge von $\circ$, so hat die Gleichung $1 \circ x = 4$ noch immer keine Lösung. Aber auch für $2 \circ x = 1$ scheidet $x = 0$ als Lösung aus $(0 \notin \mathbf{R}_+^*)$.

Der naheliegende Ausschluß der 1 bewirkt nun eine neue Störung: jetzt hat die Gleichung $a \circ x = b$ zwar die Lösung $x = \log_a b$, aber „leider" nur für den Fall $a \neq b$. Für $a = b$ wäre die Lösung die zuvor ausgeschlossene 1. Obwohl diese Schwierigkeiten für die Gleichung $y \circ a = b$ (mit der Lösung $y = \sqrt[a]{b}$) nicht auftreten,[1]) bleibt dem Potenzieren als Operation die Umkehrbarkeit versagt. Aber auch die anderen Operationen der Beispiele 1.1.3.1 bis 1.1.3.4 sind nicht umkehrbar.

1.2.2. Beispiele für D-Gruppoide

1. $(\mathbf{N}, \circ)$ mit $x \circ y := |x - y|$.

Die Gleichung $a \circ x = b$ (und analog $y \circ a = b$) hat für alle $a, b \in \mathbf{N}$ stets eine Lösung, nämlich $x = a + b$. Für $a \geqq b$ ist die Lösung nicht eindeutig: $x_1 = a + b$ und $x_2 = a - b$.

2. $(\mathbf{C}^*, \circ)$ mit $x \circ y := x^3 \cdot y^{-2}$ (T. N. Tracewell [174]).

Die Gleichung $a \circ x = b$ hat stets zwei Lösungen $\left(x^2 = \frac{a^3}{b}\right)$, die Gleichung $y \circ a = b$ sogar drei $(y^3 = a^2 \cdot b)$.

3. Es sei c ein Kegelschnitt[2]) in der reellen projektiven Ebene π, g eine Gerade, die mit c keinen Punkt gemeinsam hat. Für alle Punkte $P, Q \in c$ definieren wir $P \circ Q$ als den Punkt $R \in c$ mit $R \neq Q$, so daß die Geraden g, $g(QR)$ und die Tangente $g(PP)$ einen gemeinsamen Schnittpunkt S besitzen: $P \circ Q = R :\Leftrightarrow g(PP) \cap g = g(QR) \cap g$. Für $P = Q$ setzen wir $P \circ P := P$. Ist $g(QS)$ ebenfalls Tangente an c (in Q), so setzen wir $P \circ Q := Q$ (vgl. S. K. Stein [164]) (siehe Abb. 5).

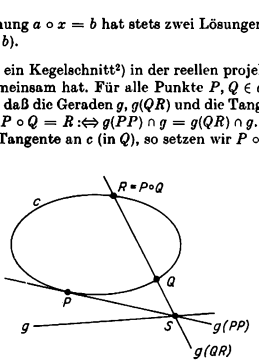


Abb. 5

$(c, \circ)$ ist ein D-Gruppoid, in dem die Gleichung $A \circ X = B$ für alle $A, B \in c$ sogar eindeutig lösbar ist, die Gleichung $Y \circ A = B$ hingegen i. allg. zwei Lösungen (zwei Tangenten!) hat.

Neben der Umkehrbarkeit einer Operation wird uns auch häufig die *Kürzbarkeit* oder *Kürzungsregel*[3]) begegnen.

[1]) Wenn man links- und rechtsseitige Umkehrbarkeit unterscheidet, ist das Potenzieren eine rechtsseitig umkehrbare Operation in $\mathbf{R}_+^*$.

[2]) In der projektiven Geometrie kann man nicht mehr zwischen Ellipse, Parabel und Hyperbel unterscheiden; man spricht vom sogenannten einteiligen Kegelschnitt. Wir verabreden folgende Bezeichnungen: $g(PQ)$ ist die Gerade, die durch die Punkte P und Q geht, $g(PP)$ ist die Tangente an eine Kurve (Kegelschnitt) im Punkt P.

[3]) Man kann — wie im Fall der Umkehrbarkeit — zwischen links- und rechtsseitiger Kürzbarkeit unterscheiden, vgl. etwa MfL Bd. 1, 2.6.

1.2.3. Definition. $(G, \circ)$ heißt *K-Gruppoid* (Gruppoid mit Kürzungsregel) $:\Leftrightarrow (G, \circ)$ ist ein Gruppoid und

(KB) die Gleichungen $a \circ x = b$ und $y \circ a = b$ haben für alle $a, b \in G$ höchstens eine Lösung $x \in G$ bzw. $y \in G$.

(KB) heißt in Zeichen: $\bigwedge_{a,x,y \in G} (a \circ x = a \circ y \vee x \circ a = y \circ a) \Rightarrow x = y$.

Mitunter werden wir die Kürzbarkeit auch in der (kontraponierten) Form benutzen:

$$\bigwedge_{a,x,y \in G} (x \neq y \Rightarrow a \circ x \neq a \circ y \wedge x \circ a \neq y \circ a).$$

Weder die D-Gruppoide der Beispiele 1.2.2.1 bis 1.2.2.3 noch die Gruppoide der Beispiele 1.1.3.1 bis 1.1.3.4 sind kürzbar.

1.2.4. Beispiele für K-Gruppoide

1. $(\mathbf{N}^* \setminus \{1\}, \circ)$ mit $x \circ y := x^y$.

2. $(\mathbf{N}, \circ)$ mit $x \circ y := x^k + y^k$, wobei k eine beliebige, aber feste natürliche Zahl mit $k > 0$ ist.[1])

3. $(\mathbf{N}^*, \circ)$ mit $x \circ y := x \cdot y$.

4. $(\mathbf{R}, \circ)$ mit $x \circ y := \begin{cases} x + y & \text{für} \quad x \geqq 0 \text{ oder } y \geqq 0, \\ 2(x + y) & \text{für} \quad x < 0 \text{ und } y < 0. \end{cases}$

Anhand der ersten zwei Beispiele wird deutlich, daß die Wahl der Trägermenge einen Einfluß darauf hat, ob die in ihr definierte Operation kürzbar ist (vgl. die Beispiele 1.1.3.1 und 1.1.3.2). So ist das Potenzieren in $\mathbf{N}^*$ nicht kürzbar, durch den Ausschluß der 1 z. B. läßt sich das aber korrigieren.

In der Gruppentheorie wird der Satz bewiesen, daß eine endliche Halbgruppe, in der die Kürzungsregel gilt, bereits eine Gruppe ist. Dieses Ergebnis läßt sich auf die folgende Weise erweitern:

1.2.5. Satz. *Ein endliches Gruppoid $(G, \circ)$ ist genau dann ein K-Gruppoid, wenn es ein D-Gruppoid ist.*

Beweis. 1. Es sei $(G, \circ)$ ein K-Gruppoid der endlichen Ordnung m und $a \in G$. Wir bilden die m Produkte $a \circ x$, wobei x alle Elemente von G durchläuft. Wegen der Kürzbarkeit sind diese m Produkte alle voneinander verschieden, d. h., sie müssen selbst die gesamte Menge G durchlaufen. Damit kann jedes Element $b \in G$ in der Gestalt $a \circ x$ mit geeignetem $x \in G$ angegeben werden. Für alle $a, b \in G$ ist also die Gleichung $a \circ x = b$ mit $x \in G$ lösbar. Analog zeigt man die Lösbarkeit der Gleichung $y \circ a = b$.

2. Es sei $(G, \circ)$ ein D-Gruppoid der endlichen Ordnung m. Dann ist für beliebige, im folgenden aber feste Elemente $a, b \in G$ die Gleichung $a \circ x = b$ lösbar. Sie habe

[1]) k kann jetzt auch ungerade sein, vgl. Beispiel 1.1.3.2. Für $k = 1$ erhalten wir die Addition natürlicher Zahlen.

die Lösungen x_1 und x_2 mit $x_1 \neq x_2$. Wir bilden wieder die m Produkte $a \circ x$, indem x alle Elemente aus G durchläuft. Wenigstens zwei dieser m Produkte sind dann nicht voneinander verschieden, nämlich $a \circ x_1$ und $a \circ x_2$. Damit können diese m Produkte $a \circ x$ ihrerseits nicht alle Elemente von G durchlaufen, d. h. aber, es gibt ein $c \in G$, so daß $a \circ x = c$ keine Lösung in G besitzt. Das ist ein Widerspruch zur vorausgesetzten Umkehrbarkeit der Operation $\circ$. Analog zeigt man, daß auch die Gleichung $y \circ a = b$ für alle $a, b \in G$ höchstens (und damit genau) eine Lösung hat.

Im Fall unendlicher Ordnung sind Umkehrbarkeit und Kürzbarkeit einer Operation natürlich nicht äquivalent. Das zeigen die Beispiele 1.2.2 und 1.2.4.

1.2.6. Aufgaben

1. In welchen der Zahlbereiche $\mathbf{N}^*$, $\mathbf{N}$, $\mathbf{Z}$, $\mathbf{Q}_+^*$, $\mathbf{Q}_+$, $\mathbf{Q}^*$, $\mathbf{Q}$, $\mathbf{R}_+^*$, $\mathbf{R}_+$, $\mathbf{R}^*$, $\mathbf{R}$ sind die vier Grundrechenarten Addition, Subtraktion, Multiplikation und Division umkehrbare oder (und) kürzbare Operationen?

2. Man zeige, daß die D-Gruppoide der Beispiele 1.2.2.1 bis 1.2.2.3 nicht kürzbar, die K-Gruppoide der Beispiele 1.2.4.1 bis 1.2.4.4 nicht umkehrbar und die Gruppoide der Beispiele 1.1.3.1 bis 1.1.3.4 weder kürzbar noch umkehrbar sind.

3. Man zeige, daß $(\mathbf{R}_+^*, \circ)$ mit $x \circ y := \dfrac{2xy}{x+y}$ (harmonisches Mittel) ein K-, aber kein D-Gruppoid ist.

1.3. Quasigruppen, lateinische Quadrate und Quasigruppen-Graphen

Fordern wir von einer Operation $\circ$ eines Gruppoids $(G, \circ)$ Umkehrbarkeit und Kürzbarkeit, so sind sowohl die Existenz als auch die Einzigkeit der Lösungen von $a \circ x = b$ und $y \circ a = b$ für alle $a, b \in G$ gesichert; $(G, \circ)$ ist D-Gruppoid und K-Gruppoid:

1.3.1. Definition. $(Q, \circ)$ heißt *Quasigruppe* $:\Leftrightarrow$ $(Q, \circ)$ ist ein Gruppoid und

(Q) für alle $a, b \in Q$ sind die Gleichungen $a \circ x = b$ und $y \circ a = b$ eindeutig lösbar.

Ist $(Q, \circ)$ ein endliches Gruppoid, so genügt wegen Satz 1.2.5 eine der beiden Forderungen (D) oder (KB), damit $(Q, \circ)$ Quasigruppe ist.

Insbesondere brauchen Quasigruppen nicht assoziativ zu sein. Den Anstoß zur Untersuchung nichtassoziativer algebraischer Strukturen gaben die Arbeiten der deutschen Mathematikerin R. Moufang (1905–1977). Sie entdeckte 1935 den Zusammenhang zwischen nichtdesarguesschen projektiven Ebenen und (den nach ihr benannten) *Moufang-Loops*. Das sind spezielle Quasigruppen. Die Definition 1.3.1 geht auf B. A. Hausmann und O. Ore [67] zurück. Aus der umfangreichen Literatur über Quasigruppen und ihre Anwendungen sei vor allem auf die Monographien von R. H. Bruck [32], V. D. Belousov [21] sowie J. Dénes und A. D. Keedwell [44] hingewiesen.

Keines der D- bzw. K-Gruppoide der Beispiele 1.2.2 und 1.2.4 ist eine Quasigruppe (vgl. Aufgabe 1.2.6.2).

1.3.2. Beispiele für Quasigruppen

1. Subtraktion und Division liefern u. a. folgende (nichtassoziative) Quasigruppen: $(\mathbf{Z}, -)$, $(\mathbf{Q}, -)$, $(\mathbf{R}, -)$, $(\mathbf{Q}^*, :)$, $(\mathbf{Q}_+^*, :)$, $(\mathbf{R}^*, :)$, $(\mathbf{R}_+^*, :)$.

2. $(\mathbf{Q}, \circ)$ mit $x \circ y := 2x + y$ bzw. $x \circ y := 3x - 2y$ bzw. $x \circ y := \frac{1}{2}(x + y)$ (arithmetisches Mittel),

$(\mathbf{R}_+^*, \circ)$ mit $x \circ y := x^2 \cdot y$ bzw. $x \circ y := \sqrt{x \cdot y}$ (geometrisches Mittel).

3. Für alle Punkte P, Q einer Ebene ε definieren wir $P \circ Q$ als den Punkt R der Geraden $g(PQ)$, so daß $\overline{PR} : \overline{RQ} = a : b$ gilt mit $a, b \in \mathbf{Z}$, beliebig, fest, $a \neq 0$, $b \neq 0$, $a : b \neq -1$. Für $P = Q$ setzen wir $P \circ P := P$. Im Fall $a : b = 1$ liefert dieses Teilverhältnis gerade die schon erwähnte „Mittelpunktskonstruktion": R ist Mittelpunkt der Strecke $\overline{PQ}$. Die Fälle $a : b = -2$ und $a : b = -\frac{1}{2}$ führen auf die Spiegelpunkte von P an Q bzw. von Q an P.

4. In der Menge $M = \{a, b, c\}$ definieren wir die Operation $\circ$ durch

$$x \circ y := \begin{cases} x & \text{für} \quad x = y, \\ z & \text{für} \quad x \neq y, \text{ wobei } z \neq x \text{ und } z \neq y. \end{cases}$$

Man überzeuge sich insbesondere davon, daß keine der hier angeführten Quasigruppen eine Gruppe[1]) ist.

Es sei $(Q, \circ)$ eine Quasigruppe. Dann ermöglicht die Eigenschaft (Q) die Konstruktion weiterer Operationen über derselben Trägermenge Q:

1. Zu beliebigen vorgegebenen Elementen $a, b \in Q$ liefert die Gleichung $a \circ x = b$ eine eindeutig bestimmte Lösung $x \in Q$, d. h., es liegt eine Abbildung $\backslash$ von $Q \times Q$ in Q vor: $(a, b) \mapsto x$ mit $x := a \backslash b$.

2. Analog ordnet die Gleichung $y \circ a = b$ jedem geordneten Paar $(a, b) \in Q \times Q$ genau ein $y \in Q$ zu: $(a, b) \mapsto y$ mit $y := b / a$.

Damit erhalten wir über derselben Trägermenge zwei weitere Operationen, die sogenannten *Umkehroperationen* $\backslash$ und $/$ von $\circ$. Speziell heißt $\backslash$ *Linksdivision* und $/$ *Rechtsdivision* von $\circ$. Aus der Definition der beiden Umkehroperationen folgt unmittelbar, daß die Gleichungen $a \circ b = c$, $a \backslash c = b$ und $c / b = a$ äquivalent sind.

An dieser Stelle sei noch ausdrücklich darauf hingewiesen, daß die Definition des Begriffs „Umkehroperation" in der Literatur nicht einheitlich ist. Insbesondere folgt unser Mathematiklehrgang in bezug auf diesen Terminus den Vorstellungen von E. Schröder (1841—1902), ohne daß dieser Begriff im Unterricht jedoch selbst definiert wird (I. Lehmann [101]). G. Asser definiert in MfL Bd. 1, 2.6., die links- und rechtsseitige Umkehrung ($\circ_l$ bzw. $\circ_r$) einer kürzbaren Operation $\circ$ in gleicher Weise wie E. Schröder [158] ($\overline{\ }_l$ bzw. $\overline{\ }_r$). Setzen wir auch Umkehrbarkeit voraus, so gilt folgender Zusammenhang: $x \overline{\ }_l y = x \circ_l y = x / y$ und $x \overline{\ }_r y = x \circ_r y = y \backslash x$.

Die oben angegebene Konstruktion der Links- und Rechtsdivision hat sich etwa seit 1950 durchgesetzt.[2])

[1]) Vgl. Definition 1.5.2 mit Aufgabe 1.5.6.4.

[2]) E. Schönhardt (1930), H. Boggs und G. Y. Rainich (1937), P. Lorenzen (1940), I. M. H. Etherington (1945), T. Evans (1949), H. A. Thurston (1952), B. H. Neumann (1953), A. I. Mal'cev (1954), G. Pickert (1955), R. H. Bruck (1958), J. Aczél (1960) und V. D. Belousov (1967).

1.3.3. Beispiele für Umkehroperationen

1. Die Umkehroperationen von Addition, Subtraktion, Multiplikation und Division:

$x \circ y$	Trägermenge	$x \backslash y$	x/y
$x + y$	$\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$	$-x + y$	$x - y$
$x - y$	$\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$	$x - y$	$x + y$
$x \cdot y$	$\mathbf{Q}^*$, $\mathbf{R}^*$, $\mathbf{R}_+^*$	$\frac{1}{x} \cdot y$	$x : y$
$x : y$	$\mathbf{Q}^*$, $\mathbf{R}^*$, $\mathbf{R}_+^*$	$x : y$	$x \cdot y$

Auch die beiden kommutativen Operationen, Addition und Multiplikation, haben also (gemäß obiger Konstruktion) zwei Umkehroperationen; z. B. gilt $3\backslash 4 \neq 3/4$.

2. Die Umkehroperationen der Operation $\circ$ mit $x \circ y := 2x + y$ in $\mathbf{Q}$ (Beispiel 1.3.2.2) sind durch $x\backslash y = -2x + y$ und $x/y = \frac{1}{2}(x - y)$ gegeben.

3. Es seien P und Q beliebige Punkte einer Ebene ε. Dann sind die beiden Operationen in ε, die durch die Konstruktion der Spiegelpunkte von P an Q bzw. von Q an P definiert werden, gerade die beiden Umkehroperationen der Mittelpunktskonstruktion (vgl. Beispiel 1.3.2.3).

In einer Quasigruppe $(Q, \circ)$ sind die beiden Identitäten[1])

1.3.4. $a \circ (a\backslash b) = b$

und

1.3.5. $(b/a) \circ a = b$

erfüllt. Davon überzeugen wir uns, indem wir in den Gleichungen $a \circ x = b$ und $y \circ a = b$ ihre (eindeutig bestimmten) Lösungen $x = a\backslash b$ bzw. $y = b/a$ einsetzen.

Es seien jetzt $(G, \circ)$, $(G, \square)$ und $(G, *)$ Gruppoide, deren Operationen den Identitäten $a \circ (a \square b) = b$ und $(b * a) \circ a = b$ genügen mögen. Dann muß $(G, \circ)$ zwar keine Quasigruppe sein, wohl aber ein D-Gruppoid. Die Identität $a \circ (a \square b) = b$ bedeutet nämlich für das Gruppoid $(G, \circ)$, daß jede Gleichung der Gestalt $a \circ x = b$ für alle $a, b \in G$ wenigstens eine Lösung hat, nämlich $x = a \square b$. Analog zieht die Identität $(b * a) \circ a = b$ für alle $a, b \in G$ die Lösbarkeit der Gleichung $y \circ a = b$ nach sich. Die Kürzbarkeit muß deshalb jedoch keineswegs erfüllt sein, wie das D-Gruppoid $(\mathbf{N}, \circ)$ aus Beispiel 1.2.2.1 lehrt.

Unter zusätzlichen Voraussetzungen läßt sich allerdings erreichen, auch eine Quasigruppe mittels Identitäten zu charakterisieren (T. Evans [52]).

1.3.6. Satz. *Ein Gruppoid $(Q, \circ)$ ist genau dann eine Quasigruppe, wenn in Q zwei*

[1]) Es seien f und g m- bzw. n-stellige Operationen in derselben Menge M (vgl. MfL Bd. 1, 2.6.). Dann verstehen wir unter einer *Identität* eine Gleichung der Form $f(x_1, x_2, \ldots, x_n) = g(y_1, y_2, \ldots, y_m)$, die für beliebige $x_i, y_j \in M$ mit $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$ wahr ist. Die Generalisierung der in einer solchen Gleichung auftretenden Variablen soll dabei schon durch das Wort „Identität" zum Ausdruck gebracht werden. Im Fall der Assoziativität ($m = n = 3$), der Kommutativität ($m = n = 2$) und der Distributivität ($m = n = 3$) spricht man in der Schule auch von *Rechengesetzen*.

weitere Operationen \ und / existieren, so daß neben (1.3.4) *und* (1.3.5) *die folgenden Identitäten gelten:*

1.3.7. $a\backslash(a \circ b) = b,$

1.3.8. $(b \circ a)/a = b.$

Beweis. 1. Es sei $(Q, \circ)$ eine Quasigruppe. Wir haben bereits festgestellt, daß dann die Operation $\circ$ gemeinsam mit ihren beiden Umkehroperationen \ und / die Identitäten (1.3.4) und (1.3.5) erfüllt. Es sei nun $a\backslash(a \circ b) = c$. Dann gilt $a \circ \big(a\backslash(a \circ b)\big) = a \circ c$. Wenden wir (1.3.4) auf die linke Seite an, so erhalten wir $a \circ b = a \circ c$, so daß $b = c$ gilt wegen der Kürzbarkeit von $\circ$, also die Identität (1.3.7) erfüllt ist. Analog zeigt man die Gültigkeit der Identität (1.3.8).

2. Es seien $(Q, \circ)$, $(Q, \backslash)$ und $(Q, /)$ Gruppoide, deren drei Operationen den Identitäten (1.3.4) bis (1.3.8) genügen. Infolge von (1.3.4) und (1.3.5) ist $(Q, \circ)$ dann bereits ein D-Gruppoid. Es ist also nur noch die Kürzbarkeit von $\circ$ zu zeigen. Es sei $a \circ x_1 = a \circ x_2 = b$. Dann gilt (wegen der Eindeutigkeit von \) $a\backslash(a \circ x_1) = a\backslash(a \circ x_2)$, woraus wegen (1.3.7) sofort $x_1 = x_2$ folgt, d. h., die Gleichung $a \circ x = b$ hat für alle $a, b \in Q$ wegen (1.3.4) wenigstens und wegen (1.3.7) höchstens eine Lösung. Analog bewirkt (1.3.8) die Eindeutigkeit der nach (1.3.5) existierenden Lösung von $y \circ a = b$ für alle $a, b \in Q$. Damit ist $(Q, \circ)$ eine Quasigruppe.

Die in den Identitäten (1.3.4) bis (1.3.8) auftretenden Operationen sind sogar sämtlich Quasigruppenoperationen, mit anderen Worten, die Umkehroperationen einer umkehrbaren und kürzbaren Operation sind ihrerseits wieder umkehrbar und kürzbar:

1.3.9. Satz. *Ist* $(Q, \circ)$ *eine Quasigruppe, so sind auch* $(Q, \backslash)$ *und* $(Q, /)$ *Quasigruppen, wobei \ und / die Umkehroperationen*[1]) *von* $\circ$ *sind.*

Beweis. Wir zeigen, daß $(Q, \backslash)$ eine Quasigruppe ist. Für $(Q, /)$ verläuft der Beweis analog.

a) Die Gleichung $a\backslash x = b$ ist für alle $a, b \in Q$ eindeutig lösbar: Die Existenz einer Lösung wird durch (1.3.7) gesichert, nämlich $x = a \circ b$, während die Eindeutigkeit aus (1.3.4) folgt. Denn mit $a\backslash x_1 = a\backslash x_2 = b$ gilt $a \circ (a\backslash x_1) = a \circ (a\backslash x_2)$, also $x_1 = x_2$.

b) Die Gleichung $y\backslash a = b$ ist für alle $a, b \in Q$ eindeutig lösbar: Wegen $(a/b)\backslash a = (a/b)\backslash\big((a/b) \circ b\big) = b$ (nach (1.3.5) bzw. (1.3.7)) ist die Existenz einer Lösung von $y\backslash a = b$ gewährleistet, nämlich $y = a/b$. Da $a \circ (a\backslash b) = b$ für alle $a, b \in Q$ genau dann erfüllt ist, wenn $b/(a\backslash b) = a$ gilt, ist auch die Kürzbarkeit garantiert. Mit $y_1\backslash a = y_2\backslash a$ gilt nämlich auch $a/(y_1\backslash a) = a/(y_2\backslash a)$, also $y_1 = y_2$. Nebenher haben wir die Identitäten

1.3.10. $(a/b)\backslash a = b$

und

1.3.11. $a/(b\backslash a) = b$

von I. A. Mal'cev [110] erhalten.

[1]) Wir vereinbaren, daß in den ersten beiden Kapiteln mit \ und / immer nur die beiden Umkehroperationen einer Quasigruppenoperation $\circ$ bezeichnet werden.

Es sei $\mathfrak{Q} = (Q, \circ)$ eine endliche Quasigruppe. Dann wird die Strukturtafel von $\mathfrak{Q}$ durch folgenden Sachverhalt beschrieben:

1.3.12. Jedes Element der Trägermenge tritt in jeder Zeile und in jeder Spalte genau einmal auf.

Das heißt, alle Zeilen und Spalten der Strukturtafel sind Permutationen der Eingänge. Eine solche Strukturtafel heißt *lateinisches Quadrat*. Umgekehrt definiert jedes lateinische Quadrat eine Quasigruppenoperation.

Formal ließe sich zunächst unterscheiden, daß in jeder Zeile und in jeder Spalte der Strukturtafel wegen der Umkehrbarkeit der Operation $\circ$ alle Elemente der Trägermenge, infolge der Kürzbarkeit aber auch lauter verschiedene Elemente auftreten müssen. Da wir Strukturtafeln aber ohnehin nur für endliche Gruppoide eingeführt haben, sind wegen Satz 1.2.5 die Strukturtafeln von D- bzw. K-Gruppoiden ebenfalls schon lateinische Quadrate. Zum Beweis, daß die Strukturtafel einer endlichen Quasigruppe ein lateinisches Quadrat ist, nehmen wir an, in der i-ten Zeile komme ein Element $b \in Q$ gar nicht vor bzw. trete zweimal auf, etwa an k-ter und l-ter Stelle: $b = a_{ik} = a_{il}$ mit $k \neq l$. Im ersten Fall hat dann $a_i \circ x = b$ keine Lösung — im Widerspruch zur Umkehrbarkeit; im zweiten Fall folgt wegen (1.1.4) $b = a_l \circ a_k = a_i \circ a_l$, so daß die Kürzbarkeit $a_k = a_l$, also $k = l$ im Widerspruch zur Annahme liefert. In gleicher Weise zeigt man, daß in keiner Spalte ein Element fehlen bzw. doppelt stehen kann. Umgekehrt überzeugt man sich, daß Umkehrbarkeit und Kürzbarkeit vorliegen, wenn gleiche Elemente paarweise in verschiedenen Zeilen und in verschiedenen Spalten stehen.

1.3.13. Beispiele für lateinische Quadrate

1. Das lateinische Quadrat für Beispiel 1.3.2.4 hat folgendes Aussehen:

$\circ$	a	b	c
a	a	c	b
b	c	b	a
c	b	a	c

2. Das folgende lateinische Quadrat definiert in $Q = \{0, 1, 2\}$ eine Quasigruppenoperation. Es gilt $x \circ y = x - y \bmod 3$.

$\circ$	0	1	2
0	0	2	1
1	1	0	2
2	2	1	0

Die Umkehroperationen einer Quasigruppenoperation sind selbst wieder Quasigruppenoperationen (Satz 1.3.9). Im Fall endlicher Ordnung sind damit dann aber auch die Strukturtafeln dieser Umkehroperationen wieder lateinische Quadrate (vgl. dazu Kap. 2).

L. Euler (1707—1783), der übrigens auch Mathematiklehrbücher für die russischen Schulen schrieb, beschäftigte sich wohl als erster systematisch mit lateinischen Quadraten. Ausgangspunkt waren dabei stets Fragen aus der sogenannten Unterhaltungsmathematik. Da man hierbei anstelle von Zahlen große lateinische Buch-

staben zur Belegung der Felder benutzte, hat sich der Name „Lateinisches Quadrat" bis in unsere Zeit erhalten.

Neben Strukturtafeln haben wir Cayley-Diagramme zur Illustration bzw. Darstellung von Gruppoiden eingeführt. Wie spiegeln sich nun Umkehrbarkeit und Kürzbarkeit im zugehörigen Cayley-Diagramm eines (endlichen) Gruppoids wider? Dabei ist es wegen Satz 1.2.5 wieder gleichgültig, welche der Eigenschaften (D), (KB) oder (Q) zum Beweis der folgenden Charakterisierung herangezogen werden. Ein endliches Gruppoid $\mathfrak{Q} = (Q, \circ)$ ist genau dann eine Quasigruppe, wenn für das zugehörige Cayley-Diagramm $\Gamma(\mathfrak{Q})$ die beiden folgenden Bedingungen erfüllt sind:

1.3.14. Von jedem Knoten führt zu jedem Knoten (einschließlich zu sich selbst) genau ein Bogen,[1])

also formalisiert: $\bigwedge_{K_1, K_2 \in \mathbf{K}} \bigvee_{b \in \mathbf{B}}^{!} \big(A(b) = K_1 \wedge E(b) = K_2\big)$,

und

1.3.15. in jedem Knoten endet genau ein Bogen jeder Farbe,

in Zeichen: $\bigwedge_{K_1, K_2 \in \mathbf{K}} \bigvee_{b \in \mathbf{B}}^{!} \big(\alpha(b) = K_1 \wedge E(b) = K_2\big)$.

Es sei $\mathfrak{Q} = (Q, \circ)$ eine (endliche) Quasigruppe. Wir zeigen zunächst, daß dann im zugehörigen Cayley-Diagramm zwei beliebige Knoten K_1 und K_2 durch wenigstens einen Bogen verbunden sind. Die Gleichung $k_1 \circ x = k_2$ mit $\beta(k_1) = K_1$ und $\beta(k_2) = K_2$ hat eine eindeutige Lösung x_0, so daß es wegen (1.1.6) (genau) einen Bogen b der Farbe $X_0 = \beta(x_0) = \alpha(b)$ gibt, der von K_1 nach K_2 führt. Es kann nun aber auch keinen weiteren Bogen b' mit einer von X_0 verschiedenen Farbe geben, der in K_1 beginnt und in K_2 endet. Gäbe es einen solchen Bogen b' mit $\alpha(b') = X'$, so hätte im Widerspruch zur Kürzbarkeit $k_1 \circ x = k_2$ eine weitere Lösung $x' = \beta^{-1}(X')$. Die Gleichung $y \circ k_1 = k_2$ mit $\beta(k_1) = K_1$ und $\beta(k_2) = K_2$ hat ihrerseits für beliebige $K_1, K_2 \in \mathbf{K}$ eine eindeutige Lösung y_0, so daß es wegen (1.1.6) (genau) einen Bogen b der Farbe $K_1 = \beta(k_1) = \alpha(b)$ gibt, der von $Y_0 = \beta(y_0)$ nach K_2 führt. Es bleibt zu zeigen, daß es nicht mehrere Bögen der Farbe K_1 geben kann, die in K_2 enden. Gäbe es einen weiteren Bogen b' der Farbe K_1 mit $E(b') = K_2$, so hätte wieder im Widerspruch zur Kürzbarkeit $y \circ k_1 = k_2$ eine weitere Lösung y'. Es sei nun umgekehrt $\Gamma(\mathfrak{Q})$ ein Cayley-Diagramm eines (endlichen) Gruppoids $\mathfrak{Q} = (Q, \circ)$, das die Bedingungen (1.3.14) und (1.3.15) erfüllt. Für beliebige $k_1, k_2 \in Q$ mit $\beta(k_1) = K_1$ und $\beta(k_2) = K_2$ existiert dann einerseits genau ein Bogen b von K_1 nach K_2, andererseits endet in K_2 genau ein Bogen jeder Farbe. Die Farbe von b ist also eindeutig bestimmt: $\alpha(b) = X_0$. Damit ist $x_0 := \beta^{-1}(X_0)$ die eindeutige Lösung von $k_1 \circ x = k_2$. Unter den in K_2 endenden Bögen gibt es nun auch genau einen K_1-gefärbten Bogen b'. Der Anfangspunkt Y_0 von b' liefert uns die eindeutige Lösung $y_0 = \beta^{-1}(Y_0)$ der Gleichung $y \circ k_1 = k_2$.

[1]) Damit ist das Cayley-Diagramm einer Quasigruppe ein streng zusammenhängender Graph. Ein orientierter Graph heißt *streng zusammenhängend*, wenn je zwei Knoten durch eine Kette aufeinanderfolgender Bögen (gleicher Orientierung) verbunden sind.

Aufgrund dieser für eine Quasigruppe $(Q, \circ)$ charakteristischen Eigenschaften des zugehörigen Cayley-Diagramms läßt sich (im Unterschied zu beliebigen Gruppoiden) der graphentheoretische Begriffsapparat entlasten. Da nämlich Mehrfachbögen nicht mehr auftreten können, läßt sich jeder Bogen durch seinen Anfangs- und Endpunkt beschreiben; es gilt $\mathbf{B} = \mathbf{K} \times \mathbf{K}$. Die kantenfärbende Abbildung α ist dann sogar eine Quasigruppenoperation in $\mathbf{K}$.

1.3.16. Beispiele für Quasigruppen-Graphen

1. Das Cayley-Diagramm der Quasigruppe aus Beispiel 1.3.2.4 (vgl. auch Beispiel 1.3.13.1) hat folgendes Aussehen (siehe Abb. 6).

2. Das nebenstehende Cayley-Diagramm (vgl. auch Beispiel 1.3.13.2) definiert in $Q = \{0, 1, 2\}$ eine Quasigruppenoperation (siehe Abb. 7).

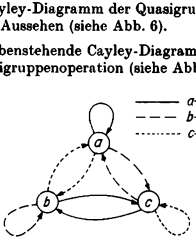


Abb. 6

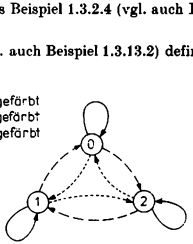


Abb. 7

Die Cayley-Diagramme der Umkehroperationen einer (endlichen) Quasigruppenoperation sind (wegen Satz 1.3.9) selbst wieder Quasigruppen-Graphen.

1.3.17. Aufgaben

1. Es sei G_n die Menge der Eckpunkte eines regelmäßigen n-Ecks mit $n = 2k + 1$ ($k \geq 1$, ganzzahlig). Für beliebige Punkte $P, Q \in G_n$ definieren wir $P \circ Q$ als den Eckpunkt des Polygons, der auf der Mittelsenkrechten der Strecke $\overline{PQ}$ liegt. Für $P = Q$ setzen wir $P \circ P := P$. Man zeige anhand der Strukturtafeln, daß $(G_3, \circ)$ und $(G_5, \circ)$ Quasigruppen sind, und gebe ihre Umkehroperationen an. Wie sehen die jeweiligen Quasigruppen-Graphen aus?

2. Man gebe die lateinischen Quadrate und Quasigruppen-Graphen der Umkehroperationen für die Beispiele 1.3.13.1 (1.3.16.1) und 1.3.13.2 (1.3.16.2) an und vergleiche sie.

3. Es sei $\triangle ABC$ ein vorgegebenes Dreieck in der Ebene ε. Für beliebige Punkte $P, Q \in \varepsilon$ definieren wir $P \circ Q$ als den Punkt $R \in \varepsilon$, daß die Dreiecke $\triangle ABC$ und $\triangle PQR$ gleichsinnig ähnlich sind; $P \circ P := P$. Man zeige, daß $(\varepsilon, \circ)$ eine Quasigruppe ist.

Wählt man in der komplexen Ebene für die Punkte des Dreiecks $\triangle ABC$ $a = 0$, $b = 1$ und $c = \alpha \cdot e^{i\beta}$, wobei α und β beliebige, aber feste reelle Zahlen mit $\alpha \neq 0$ und $\beta \neq k \cdot \pi$ (k ganzzahlig) sind, so erhält man für das Dreieck $\triangle PQR$

$$r = p \circ q = (1 - c)\, p + cq$$

(geometrische Darstellung der Multiplikation im Komplexen).

4. Man zeige, daß man anstelle von (1.3.14) und (1.3.15) auch

a) keine zwei verschiedenen Bögen führen von ein und demselben Knoten zu einem anderen (gemeinsamen) Knoten und

b) keine zwei Bögen derselben Farbe enden in demselben Knoten

fordern kann, um Umkehrbarkeit oder (und!) Kürzbarkeit einer Operation im zugehörigen Cayley-Diagramm zu beschreiben.

1.4. Loops

In einem Gruppoid kann es Elemente geben, die sich allen anderen Elementen gegenüber durch bestimmte Eigenschaften auszeichnen. Man denke z. B. an das neutrale oder das absorbierende Element in $(\mathbf{R}, \cdot)$, nämlich 1 und 0. Es gibt nun aber auch Gruppoide, in denen zwar kein neutrales, wohl aber einseitig-neutrale Elemente existieren. So ist die 0 in $(\mathbf{Z}, -)$ lediglich rechts-neutrales Element.

In den folgenden Definitionen seien n_l, n_r und n Elemente der Trägermenge G eines Gruppoids $(G, \circ)$.

1.4.1. Definition. n_l heißt *links-neutrales Element* von $(G, \circ) :\Leftrightarrow$

$$(N_l) \qquad \bigwedge_{x \in G} n_l \circ x = x,$$

n_r heißt *rechts-neutrales Element* von $(G, \circ) :\Leftrightarrow$

$$(N_r) \qquad \bigwedge_{x \in G} x \circ n_r = x,$$

n heißt *neutrales Element* von $(G, \circ) :\Leftrightarrow$

$$(N) \qquad \bigwedge_{x \in G} n \circ x = x \circ n = x.$$

Ein neutrales Element ist also sowohl links- als auch rechtsneutrales Element.

Das Gruppoid aus Beispiel 1.1.7.1 hat weder ein links- noch ein rechts-neutrales Element. In Beispiel 1.1.7.2 ist jedes Element rechtsseitig neutral, im Fall des Potenzierens (Beispiel 1.1.3.1) ist die 1 einziges rechts-neutrales Element; links-neutrale Elemente gibt es in beiden Fällen nicht.[1]) Schließlich ist in Beispiel 1.1.7.3 die 0 neutrales Element.

Wir wissen, daß jede Gruppe genau ein neutrales Element besitzt (MfL Bd. 3, 12.1.). Das gilt für den allgemeineren Fall eines Gruppoids schon nicht mehr, was die zuvor genannten Gegenbeispiele beweisen. Allerdings läßt sich zeigen, daß ein Gruppoid $(G, \circ)$ höchstens ein neutrales Element besitzt (MfL Bd. 1, 2.6.). Hat ein Gruppoid voneinander verschiedene einseitig neutrale Elemente, so sind sie sämtlich entweder links- oder rechts-neutral. Ist die Operation kürzbar, so kann es höchstens ein links- oder (und) rechts-neutrales Element geben (Beweis!).

Die Umkehrbarkeit einer Operation garantiert nicht die Existenz eines (einseitig) neutralen Elementes (vgl. Beispiel 1.3.13.1). Indessen hat jedes Element a einer Quasigruppe $(Q, \circ)$ ein eindeutig bestimmtes individuelles links-neutrales Element $n_l{}^a$ und ein eindeutig bestimmtes individuelles rechts-neutrales Element $n_r{}^a$, da ja die Gleichungen $y \circ a = a$ und $a \circ x = a$ für alle $a \in Q$ eindeutig lösbar sind, d. h., es gilt $n_l{}^a \circ a = a \circ n_r{}^a = a$. Im allgemeinen gilt $n_l{}^a \neq n_r{}^a$. Die individuellen links- und rechts-neutralen Elemente bedingen ihrerseits, daß jedes Element einer Quasigruppe $(Q, \circ)$ vier (i. allg. verschiedene) inverse Elemente besitzt.

---

[1]) Definieren wir $x \circ y := y$ in $G = \{0, 1, 2\}$, so ist umgekehrt jedes Element linksseitig neutral, und es gibt keine rechts-neutralen Elemente.

Sind in einer Quasigruppe alle individuellen links-neutralen Elemente miteinander identisch, so ist dieses Element linksseitig neutral (und umgekehrt); analog gilt das für individuelle rechts-neutrale Elemente. Quasigruppen, die ein neutrales Element besitzen, spielen in vielen Untersuchungen eine wichtige Rolle, ohne daß sie schon Gruppen sein müssen.

1.4.2. Definition. $(L, \circ)$ heißt *Loop* $:\Leftrightarrow$ $(L, \circ)$ ist eine Quasigruppe und besitzt ein neutrales Element.

Eine Quasigruppe mit einem links-(rechts-)neutralen Element bezeichnet man auch als eine *Links-(Rechts-)Loop*. Eine Quasigruppe ist also genau dann eine Loop, wenn sie Links- und Rechts-Loop ist.

Der Name „Loop" geht auf A. A. ALBERT [12] zurück. Diese Bezeichnung hat sich gegenüber den zuvor benutzten Begriffen „Normbereich" (G. BOL [26]) und „Union" (E. SCHÖNHARDT [157]) durchgesetzt. In der französischsprachigen Literatur stößt man mitunter auch auf den fragwürdigen[1]) Versuch „loop" mit „boucle" (= Schleife, Luppe) zu übersetzen.

Die Quasigruppe $(\mathbf{R}_+^*, \circ)$ mit $x \circ y := x^2 \cdot y$ ist eine Links-Loop, $(\mathbf{R}^*, :)$ eine Rechts-Loop. Keine der Quasigruppen der Beispiele 1.3.2.1 bis 1.3.2.4 ist eine Loop.

1.4.3. Beispiele für Loops

1. $(\mathbf{R}, \circ)$ mit $x \circ y := \begin{cases} x + y \text{ für } x \leqq 0 \text{ oder } y \leqq 0, \\ x + y + (x \cdot y)^2 \text{ für } x \geqq 0 \text{ und } y \geqq 0, \end{cases}$

bzw. $x \circ y := \begin{cases} x + y & \text{für } x \cdot y \geqq 0, \\ \dfrac{x^3 + y^3}{x^2 + y^2} & \text{für } x \cdot y < 0 \end{cases}$

(K. H. HOFMANN [68]).

Man überzeuge sich davon, daß keine der beiden Loops eine Gruppe[2]) ist.

2. Es sei $(Q, \circ)$ eine Quasigruppe. Für beliebige $a, b \in Q$ entsteht dann auf die folgende Weise eine Loop $(Q, *)$, in der $b \circ a$ das neutrale Elemente n ist: $x * y := (x/a) \circ (b \backslash y)$.

In Satz 1.3.6 haben wir Quasigruppen mit Hilfe dreier Operationen charakterisiert; ein analoger Versuch führt auch für Loops zum Erfolg. So kann man z. B. auf die Existenz eines neutralen Elements schließen, wenn in einer Quasigruppe $(Q, \circ)$ die Identität

$$x/x = y \backslash y \tag{1.4.4}$$

erfüllt ist; $\backslash$ und $/$ sind dabei wieder die beiden Umkehroperationen von $\circ$. Wir definieren $n := x/x$ für ein beliebiges, aber festes $x \in Q$. Wegen (1.4.4) gilt auch (für $y = x$) $x \backslash x = n$. Die Identitäten (1.3.4) und (1.3.5) erlauben dann die Umformungen

[1]) Fragwürdig deshalb, weil A. A. ALBERT diese Struktur nach einem Stadtteil von Chicago benannt hat.

[2]) Vgl. Definition 1.5.2 und Aufgabe 1.5.6.4.

$x \circ n = x \circ (x\backslash x) = x$ und $n \circ x = (x/x) \circ x = x$, d. h., n ist neutrales Element von $(Q, \circ)$. Ist andererseits $(Q, \circ)$ eine Loop mit dem neutralen Element n, so sind die Gleichungen $n \circ x = x$ und $y \circ n = n$ für alle $x, y \in Q$ eindeutig nach n auflösbar: wir erhalten $n = x/x$ bzw. $n = y\backslash y$. Folglich ist n sowohl von x als auch von y unabhängig. Es gilt (1.4.4). Damit haben wir den folgenden Satz bewiesen:

1.4.5. Satz. *Eine Quasigruppe $(Q, \circ)$ ist genau dann eine Loop, wenn die beiden Umkehroperationen von $\circ$ die Identität* (1.4.4) *erfüllen.*

Das neutrale Element n einer Loop $(L, \circ)$ ist zugleich links-neutrales Element in $(L, \backslash)$ und rechts-neutrales Element in $(L, /)$. Löst man nämlich die Gleichungen $n \circ x = x$ und $y \circ n = y$ nach x bzw. y auf, so ergeben sich $n\backslash x = x$ und $y/n = y$ für alle $x, y \in L$. Darüber hinaus sind beide Umkehroperationen einer Loopoperation unipotent. Dabei heißt ein Gruppoid $(G, *)$ oder eine Operation $*$ *unipotent*, wenn die folgende Identität erfüllt ist:

$$(U) \qquad \bigwedge_{x,y \in G} x * x = y * y.$$

In unserem Fall erhalten wir wegen (1.4.4) einerseits $x\backslash x = y\backslash y$ $(= n)$ und andererseits $x/x = y/y$ $(= n)$ für alle $x, y \in L$. Da in einer Loop $(L, \circ)$ alle individuellen links- und rechts-neutralen Elemente (mit dem neutralen Element n) zusammenfallen, besitzt jedes Element a nur zwei (i. allg. verschiedene) inverse Elemente. Die eindeutige Lösung von $y \circ a = n$ definiert das links-inverse Element ${}^{-1}a$ von a. Entsprechend liefert die eindeutige Lösung der Gleichung $a \circ x = n$ das rechts-inverse Element a^{-1} von a.

Es sei $(G, \circ)$ ein endliches Gruppoid. Dann spiegelt sich die Existenz eines links-neutralen Elements in der zugehörigen Strukturtafel in der Weise wider, daß es (wenigstens) eine Zeile gibt, die mit der Eingangszeile (= Kopfzeile) übereinstimmt. Die Strukturtafel in Beispiel 1.1.7.2 weist Übereinstimmung aller Spalten mit der Eingangsspalte (= Kopfspalte) auf; hier ist also jedes Element rechtsseitig neutral. Ein endliches Gruppoid $(G, \circ)$ besitzt also offenbar genau dann ein neutrales Element n, wenn in der Strukturtafel eine Zeile Z und eine Spalte S mit der Kopfzeile und -spalte identisch sind. n steht im Schnittpunkt von Z und S (vgl. Beispiel 1.1.7.3).

1.4.6. Beispiel für eine endliche (nichtassoziative) Loop

Das folgende lateinische Quadrat definiert in $L = \{a, b, c, d, e\}$ eine Loopoperation $\circ$:

$\circ$	a	b	c	d	e
a	a	b	c	d	e
b	b	a	e	c	d
c	c	d	a	e	b
d	d	e	b	a	c
e	e	c	d	b	a

Das neutrale Element ist a; $(L, \circ)$ ist unipotent.

Für die Strukturtafel eines unipotenten Gruppoids ist, wie wir sehen, charakteristisch, daß alle Felder der (von links oben nach rechts unten verlaufenden) Hauptdiagonalen dasselbe

Element enthalten. In diesem Zusammenhang sei zugleich erwähnt, daß ein (endliches) Gruppoid genau dann kommutativ ist, wenn die zugehörige Strukturtafel symmetrisch bezüglich der Hauptdiagonalen ist (vgl. Beispiel 1.3.13.1).

Im Cayley-Diagramm $\Gamma(\mathfrak{G})$ eines (endlichen) Gruppoids $\mathfrak{G} = (G, \circ)$ lassen sich diese Eigenschaften ebenfalls ablesen. Wir benutzen im folgenden die gleichen Bezeichnungen wie in 1.1. und 1.3., d. h., $\varkappa$ definiert die Kantenfärbung und β ordnet den Elementen von G die entsprechenden Knoten des Graphen $\Gamma = (\mathbf{K}, \mathbf{B})$ zu.

Links-neutrales Element: Es gibt einen Knoten N_l, von dem zu jedem Knoten X ein Bogen der Farbe X ausgeht.

Rechts-neutrales Element: Es gibt eine(n) Farbe (Knoten) N_r, so daß jeder Knoten X eine Schleife der Farbe N_r besitzt.

Unipotenz: Es gibt einen Knoten $A = \beta(a)$, in dem von jedem Knoten $X = \beta(x)$ (als Anfangspunkt) ein Bogen der Farbe X endet ($x \circ x = y \circ y = a$).

Kommutativität: Für beliebige Knoten $X = \beta(x)$, $Y = \beta(y)$, $Z = \beta(z)$ gilt: Wenn in Z ein Bogen endet, der den Anfangspunkt X und die Farbe Y hat, dann endet in Z auch ein Bogen mit dem Anfangspunkt Y und der Farbe X ($x \circ y = z \Rightarrow y \circ x = z$).

Man vgl. hierzu Abb. 2 bis 5, 9 und 10. Die Beweise seien dem Leser überlassen.

1.4.7. Aufgaben

1. Man führe den Nachweis, daß die beiden Gruppoide im Beispiel 1.4.3.1 tatsächlich Loops sind. (Hinweis: Die Quasigruppeneigenschaft kann man auch mit Hilfe der Translationen (vgl. 1.7.) zeigen. Da die partielle Ableitung $\dfrac{\partial R_a}{\partial x} = \dfrac{\partial}{\partial x}(x \circ a)$ stets positiv ist, schließt man zunächst auf strenge Monotonie von R_a, woraus die Eindeutigkeit der Translationen R_a folgt.)

2. Man zeige, daß keine der Quasigruppen der Beispiele 1.3.2.1 bis 1.3.2.4 eine Loop ist.

3. Eine nichtkommutative Loop besitzt wenigstens fünf Elemente (vgl. auch Aufgabe 1.5.6.5).

4. Man zeige, daß es in einer Loop $(L, \circ)$ mit dem neutralen Element n voneinander verschiedene Elemente $a, b \in L$ geben kann, für die $a \circ c = c \circ b = n$ mit $c \in L$ gilt. Warum darf $(L, \circ)$ dabei weder kommutativ noch unipotent sein? (Man vgl. auch Aufgabe 1.5.6.6.)

5. Man zeige, daß in einer Loop $(L, \circ)$ für alle $x \in L$ zwar $({}^{-1}x)^{-1} = {}^{-1}(x^{-1}) = x$, i. allg. aber weder $(x^{-1})^{-1} = x$ noch ${}^{-1}({}^{-1}x) = x$ gilt.

1.5. Halbgruppen und Gruppen

Unter den vielen Identitäten, die in solchen algebraischen Strukturen wie Gruppoid, Quasigruppe oder Loop erfüllt sein können, nimmt die Assoziativität eine Sonderstellung ein. Das drückt sich u. a. auch schon darin aus, daß assoziative Strukturen einen eigenen Namen erhalten haben. Insbesondere spielen Gruppen und Halbgruppen eine große Rolle für die Anwendung der Mathematik.

An dieser Stelle sei aber auch darauf hingewiesen, daß sich ein besseres Verständnis für die Identitäten wie Assoziativität oder auch Kommutativität erst erreichen läßt, wenn sie — gerade auch in der Schule — den Charakter von etwas „Selbstver-

ständlichem" verlieren. Deshalb sollte der Schüler z. B. auch einige nichtassoziative Operationen kennenlernen.[1])

1.5.1. Definition. $(H, \circ)$ heißt *Halbgruppe* $:\Leftrightarrow$ $(H, \circ)$ ist ein Gruppoid, in dem die Assoziativität erfüllt ist, d. h., es gilt:

$$(A) \qquad \bigwedge_{x,y,z \in H} (x \circ y) \circ z = x \circ (y \circ z).$$

Man beachte, daß eine klammerfreie Schreibweise, wie sie für assoziative Operationen üblich ist, auch im Fall nichtassoziativer Operationen auftreten kann. Es wird sogar von beiden Möglichkeiten, in $x \circ y \circ z$ Klammern zu setzen, Gebrauch gemacht. So versteht man unter $x - y - z$ üblicherweise $(x - y) - z$, unter x^{y^z} aber $x^{(y^z)}$.

Eine Halbgruppe mit neutralem Element wird *Monoid* genannt.

Bekanntlich kann man den Begriff der Gruppe auf verschiedene Weisen einführen. Die Gleichwertigkeit all dieser Definitionen wie auch die Beweise für einige der nachfolgenden Überlegungen findet man in der einschlägigen Literatur zur Gruppentheorie. Wir stützen uns auf die folgende Möglichkeit (vgl. MfL Bd. 3, 12.1.):

1.5.2. Definition. $(G, \circ)$ heißt *Gruppe* $:\Leftrightarrow$ $(G, \circ)$ ist ein D-Gruppoid und eine Halbgruppe.

Eine kommutative Gruppe heißt *abelsch*.

Durch N. H. Abel (1802—1829) und E. Galois (1811—1832) wurde das dem normalen Sprachgebrauch entnommene Wort „le groupe" zum feststehenden mathematischen Fachausdruck. Neben diesen beiden Mathematikern waren insbesondere L. Euler (1707—1783) J. L. Lagrange (1736—1813), P. Ruffini (1765—1822), C. F. Gauss (1777—1855), A. L. Cauchy (1789—1857), A. Cayley (1821—1895) C. Jordan (1835—1922), S. Lie (1842—1899) und F. Klein (1849—1925) Wegbereiter der Gruppentheorie. Ein abstraktes Axiomensystem für Gruppen, wenn auch nur für den Fall endlicher Ordnung, geht auf L. Kronecker (1823—1891) [92] zurück. Die Definition der Gruppe gemäß (1.5.2) findet man z. B. schon bei H. Weber (1842—1913) [178] und E. V. Huntington (1874—1952) [72]. E. V. Huntington diskutierte als erster die Frage der Unabhängigkeit der Gruppenaxiome. Zur Geschichte und Entwicklung des Gruppenbegriffs sei auf H. Wussing [184] verwiesen.

Sowohl die Umkehrbarkeit als auch die Assoziativität einer Operation lassen für sich allein noch nicht auf die Existenz eines neutralen Elementes schließen. In einer Gruppe, in der ja nach Voraussetzung beide Eigenschaften gelten, gibt es jedoch (genau) ein neutrales Element n. Während ferner in einem Monoid jedes Element a

[1]) Für die Kommutativität kann man hier auch auf Gegenbeispiele aus seiner Erfahrungswelt zurückgreifen, etwa das Verdünnen einer Säure: „Gieße Wasser in die Säure, dann passiert das Ungeheure!" oder das Zusammensetzen von Substantiven: Maul-Tier, Baum-Stamm, Hunde-Blut, Gift-Schlangen, Uhr-Armband usw. Im Fall der Assoziativität kann man dieselben Operationen heranziehen. So erkennt man beispielsweise beim Zubereiten eines Grogs den Dilettanten, der den Zucker z im Alkohol a statt im heißen Wasser w auflöst: $w \circ (z \circ a)$ statt $(w \circ z) \circ a$. Für das Zusammensetzen von Substantiven seien z. B. Mädchen-Handels-Schule und Doppel-Zeit-Wort genannt.

höchstens ein inverses Element $\bar{a}$ besitzt, ist in einer Gruppe $(G, \circ)$ infolge von (D) auch die Existenz von $\bar{a}$ $(= a^{-1})$ gesichert.[1])

$$\text{(IE)} \qquad \bigwedge_{a \in G} \bigvee_{a^{-1} \in G} a^{-1} \circ a = a \circ a^{-1} = n.$$

Damit ist jede Gruppe aber auch ein K-Gruppoid. Gilt nämlich $a \circ x_1 = a \circ x_2$ für ein beliebiges $a \in G$, so erhalten wir $a^{-1} \circ (a \circ x_1) = a^{-1} \circ (a \circ x_2)$, so daß wegen (A), (IE) und (N) $x_1 = x_2$ gilt. Analog ist $\circ$ auch rechtsseitig kürzbar. Die eindeutigen Lösungen der Gleichungen $a \circ x = b$ und $y \circ a = b$ $(a, b \in G$, beliebig) können wir jetzt auch mit Hilfe der inversen Elemente schreiben: $x = a \backslash b = a^{-1} \circ b$ bzw. $y = b / a = b \circ a^{-1}$, d. h., in einer Gruppe $(G, \circ)$ haben die beiden Umkehroperationen $\backslash$ und $/$ die Gestalt

1.5.3. $\bigwedge_{x,y \in G} (x \backslash y = x^{-1} \circ y \wedge x / y = x \circ y^{-1})$.

Zusammengefaßt sind in einer Gruppe folglich die Eigenschaften (A), (D), (KB), (Q), (N) und (IE) erfüllt. Das gestattet u. a., eine Gruppe als assoziative Quasigruppe oder als assoziative Loop zu definieren. Besonders häufig stützt man sich aber auf (A), (N) und (IE) als Axiome zur Einführung des Gruppenbegriffs.

In jeder Gruppe $(G, \circ)$ gilt auch die Beziehung

1.5.4. $\bigwedge_{x,y \in G} (x \circ y)^{-1} = y^{-1} \circ x^{-1}$.

Weiter ist eine endliche kürzbare Halbgruppe schon eine Gruppe, da im Fall endlicher Ordnung eine kürzbare Operation auch umkehrbar ist (Satz 1.2.5). Am Rande sei vermerkt, daß man in Anlehnung an Satz 1.3.6 und Satz 1.4.5 den Gruppenbegriff auch mit Hilfe der Umkehroperationen gewinnen kann; vgl. die Literaturhinweise in I. LEHMANN [99].

1.5.5. Beispiele für Halbgruppen und Gruppen[2])

1. $(G, \circ)$ mit $G = \{0, 1, 2\}$ und $x \circ y := x$ (vgl. Beispiel 1.1.7.2) ist eine Halbgruppe ohne neutrales Element. Sie ist nicht kürzbar und auch nicht kommutativ.

2. $(G, \circ)$ mit $G = \{0, 1, 2\}$ und $x \circ y := \min(x, y)$ ist ein kommutatives Monoid. Es ist nicht kürzbar; das neutrale Element ist $n = 2$.

3. $(Q, *)$ mit $Q = \{0, 1, 2\}$ und $x * y := x + y \bmod 3$ ist eine abelsche Gruppe. Die Operation $*$ ist gerade die Rechtsdivision von $\circ$ aus Beispiel 1.3.13.2.

4. Im Mathematikunterricht kommen u. a. folgende Gruppenoperationen vor
- Addition in $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, Multiplikation in $\mathbf{Q}^*$, $\mathbf{Q}^*$, $\mathbf{R}^*$, $\mathbf{R}^*$;
- $x \circ y := \frac{1}{2} xy$ in $\mathbf{R}_+^*$ (Flächeninhalt vom Dreieck), $n = 2$ (!), $x^{-1} = \frac{4}{x}$;
- Nacheinanderausführung von Verschiebungen (Bewegungen) einer Ebene;
- Addition von Vektoren.

Daneben treten auch Halbgruppen im Schulstoff auf, die keine Gruppen sind, etwa: $(\mathbf{N}, +)$, $(\mathbf{N}, \cdot)$, $(\mathbf{Z}, \cdot)$, $(\mathbf{Q}_+, +)$, $(\mathbf{Q}_+, \cdot)$, $(\mathbf{Q}, \cdot)$, $(\mathbf{R}, \cdot)$, $(\mathbf{N}, \mathrm{ggT})$, $(\mathbf{N}, \mathrm{kgV})$, $(\mathfrak{P}(M), \cap)$, $(\mathfrak{P}(M), \cup)$, $(\mathbf{N}, \min)$ und $(\mathbf{N}, \max)$.

[1]) In Gruppen bezeichnet man das zu a inverse Element üblicherweise mit a^{-1} $(\bar{a} = {}^{-1}a = a^{-1})$.

[2]) Für weitere Beispiele verweisen wir u. a. auf MfL Bd. 3, 11.3. und 12.1.

1.5.6. Aufgaben

1. Man zeige, daß die im Beispiel 1.5.5.4 genannten Gruppoide $(\mathbf{N}, \sqcap) = (\mathbf{N}, \mathrm{ggT})$, $(\mathbf{N}, \sqcup) = (\mathbf{N}, \mathrm{kgV})$, $(\mathfrak{P}(M), \cap)$, $(\mathfrak{P}(M), \cup)$, $(\mathbf{N}, \min)$ und $(\mathbf{N}, \max)$ tatsächlich assoziativ sind. Welche dieser Halbgruppen besitzen ein neutrales Element?

2. Welche der Zahlbereiche $\mathbf{N}^*$, $\mathbf{N}$, $\mathbf{Z}$, $\mathbf{Q}_+^*$, $\mathbf{Q}_+$, $\mathbf{Q}^*$, $\mathbf{Q}$, $\mathbf{R}_+^*$, $\mathbf{R}_+$, $\mathbf{R}^*$, $\mathbf{R}$ sind mit der Addition oder Multiplikation als Operation eine Halbgruppe, ein Monoid, ein kürzbares Monoid oder eine Gruppe?

3. Man zeige, daß

a) $(\mathbf{R} \setminus \{-1\}, \circ_1)$ mit $x \circ_1 y := x + y + xy$,

b) $(\mathbf{R} \setminus \{1\}, \circ_2)$ mit $x \circ_2 y := x + y - xy$,

c) $(\mathbf{R}_+^* \setminus \{1\}, \circ_3)$ mit $x \circ_3 y := x^{\ln y}$,

d) $(\mathfrak{P}(M), \circ_4)$ mit $X \circ_4 Y := (X \setminus Y) \cup (Y \setminus X)$ und

e) $(\mathfrak{P}(M), \circ_5)$ mit $X \circ_5 Y := (X \cup Y) \setminus (X \cap Y)$

eine abelsche Gruppe ist. Man gebe das neutrale Element sowie das zu einem Gruppenelement x gehörige inverse Element x^{-1} an.

4. Man zeige, daß die Gruppoide der Beispiele 1.1.3.1 bis 1.1.3.4, 1.1.7.3, 1.2.2.1 bis 1.2.2.3, 1.2.4.1, 1.2.4.2 (mit Ausnahme des Falls $k = 1$), 1.2.4.4, 1.3.2.1 bis 1.3.2.4, 1.3.13.1, 1.3.13.2, 1.4.3.1 und 1.4.6 keine Halbgruppen (und folglich auch keine Gruppen) sind. Insbesondere beachte man, daß die Nichtassoziativität der Operation aus Beispiel 1.1.3.3 eine bemerkenswerte Erscheinung im Zusammenhang mit Rechenmaschinen (Taschenrechnern) ist.

5. Eine nichtassoziative Loop besitzt wenigstens fünf Elemente (vgl. Aufgabe 1.4.7.3).

6. Man zeige, daß in einem Monoid $(M, \circ)$ mit dem neutralen Element n für alle Elemente $a, b, c \in M$ gilt: Wenn $a \circ c = c \circ b = n$, so $a = b$ (vgl. aber Aufgabe 1.4.7.4).

1.6. Assoziativitätstest

Es sei $(G, \circ)$ ein endliches Gruppoid der Ordnung m. Lassen sich solche Eigenschaften wie Kommutativität oder Unipotenz sehr schnell aus der zugehörigen Strukturtafel ablesen, so ist dies für die Assoziativität leider nicht mehr möglich. Anstatt m^3 Gleichungen der Form $(x \circ y) \circ z = x \circ (y \circ z)$ zu überprüfen, kann man zwar zur Assoziativität äquivalente Eigenschaften testen, aber selbst unter der zusätzlichen Voraussetzung, daß $(G, \circ)$ Quasigruppe oder sogar Loop ist, erfordert auch ein solcher Test noch immer einen erheblichen Aufwand.

Ein notwendiges und hinreichendes Kriterium für die Assoziativität einer Loop $(L, \circ)$ ist z. B. die sogenannte *Reidemeister-Bedingung* (vgl. J. Aczél [6]):

$$\text{(R)} \qquad \bigwedge_{\substack{x_i, y_i \in L \\ i,j = 1,\ldots,4}} (x_1 \circ y_2 = x_2 \circ y_1 \wedge x_1 \circ y_4 = x_2 \circ y_3 \wedge x_3 \circ y_2 = x_4 \circ y_1 \Rightarrow x_3 \circ y_4 = x_4 \circ y_3).$$

Es seien x_i und y_j $(i, j = 1, 2, 3, 4)$ beliebige Elemente aus L, n sei das neutrale Element, und es gelte die Prämisse von (R). Dann schließen wir mit Hilfe von (A), (N), (IE) und (1.5.4) auf die Konklusion

$$\begin{aligned} x_3 \circ y_4 &= x_3 \circ (y_2 \circ y_2^{-1}) \circ (x_1^{-1} \circ x_1) \circ y_4 = (x_3 \circ y_2) \circ (x_1 \circ y_2)^{-1} \circ (x_1 \circ y_4) \\ &= (x_4 \circ y_1) \circ (x_2 \circ y_1)^{-1} \circ (x_2 \circ y_3) = x_4 \circ (y_1 \circ y_1^{-1}) \circ (x_2^{-1} \circ x_2) \circ y_3 \\ &= x_4 \circ y_3, \end{aligned}$$

d. h., (R) ist erfüllt.

Umgekehrt folgt aus der Reidemeister-Bedingung in einer Loop $(L, \circ)$ mit dem neutralen Element n die Assoziativität. Indem wir in (R) x_1 und y_1 durch n ersetzen, erhalten wir zunächst die Bedingung

$$(R') \qquad \bigwedge_{\substack{x_i, y_j \in L \\ i,j=2,3,4}} (x_2 = y_3 \wedge x_2 \circ y_3 = y_4 \wedge x_3 \circ y_2 = x_4 \Rightarrow x_3 \circ y_4 = x_4 \circ y_3).$$

Mit $x_3 = x$, $x_2 = y$ und $y_3 = z$ geht die Prämisse von (R') in $y = y_2$, $y \circ z = y_4$ und $x \circ y_2 = x_4$ über, so daß aufgrund der Konklusion $x \circ (y \circ z) = x_3 \circ y_4 = x_4 \circ y_3 = (x \circ y_2) \circ y_3 = (x \circ y) \circ z$ für alle $x, y, z \in L$ erfüllt ist. Den hier bewiesenen Sachverhalt fassen wir in dem folgenden Satz zusammen:

1.6.1. Satz. *Eine Loop $(L, \circ)$ erfüllt genau dann die Reidemeister-Bedingung* (R), *wenn sie eine Gruppe ist.*

Interessanterweise kann man auch von der Assoziativität direkt auf die Bedingung (R') schließen. Es seien x, y und z beliebige Elemente einer Halbgruppe $(H, \circ)$. Mit $x_3 = x$, $x_2 = y_2 = y$ und $y_3 = z$ sowie $x_2 \circ y_3 = y_4$ und $x_3 \circ y_2 = x_4$ gilt dann

$$\begin{aligned} x_3 \circ y_4 &= x \circ (x_2 \circ y_3) = x \circ (y \circ z) = (x \circ y) \circ z \\ &= (x_3 \circ y_2) \circ y_3 = x_4 \circ y_3. \end{aligned}$$

Da umgekehrt beim Beweis der Assoziativität aus (R') die Voraussetzungen (Q) und (N) nicht herangezogen werden mußten, erhalten wir:

1.6.2. Satz. *Ein Gruppoid $(G, \circ)$ erfüllt genau dann die Bedingung* (R'), *wenn es eine Halbgruppe ist.*

Ausgangspunkt unserer Überlegungen war die Frage, wie sich die Assoziativität in der Strukturtafel widerspiegelt. Anstelle der Assoziativität können wir jetzt also prüfen, ob die Bedingungen (R) oder (R') erfüllt sind. Es sei deshalb $(L, \circ)$ eine endliche[1]) Loop der Ordnung m mit $L = \{a_1 = n, a_2, a_3, \ldots, a_m\}$. Unter Berücksichtigung von (1.1.4) und geeignet gewählten Indizes nimmt (R) dann die folgende Gestalt an:

Für alle $a_{i_r}, a_{j_s} \in L$ mit $1 \leqq r, s \leqq 4$ und $1 \leqq i_r, j_s \leqq m$ gilt:

$$a_{i_1 j_3} = a_{i_3 j_1} \wedge a_{i_1 j_2} = a_{i_2 j_1} \wedge a_{i_3 j_2} = a_{i_2 j_3} \Rightarrow a_{i_3 j_4} = a_{i_4 j_3} \text{ (siehe Abb. 8).}$$

Wählt man also im lateinischen Quadrat zwei Rechtecke derart aus, daß sie in drei einander entsprechenden Eckpunkten übereinstimmen (d. h., die durch eine gestrichelte Linie verbundenen Felder sind mit jeweils demselben Element besetzt), so müssen auch die beiden vierten Felder dasselbe Element enthalten. Durch $a_{i_1} = a_{j_1} = n$ verschiebt sich der „Quader" in Abb. 8 in der Weise, daß wir mit $a_{i_3} = x$, $a_{i_2} = a_{j_3} = y$ und $a_{j_2} = z$ die Assoziativität unmittelbar im lateinischen Quadrat durch (R) wiedererkennen (siehe Abb. 9).

[1]) Die Sätze 1.6.1 und 1.6.2 gelten auch im Fall unendlicher Ordnung. Strukturtafeln haben wir jedoch stets als endlich vorausgesetzt.

Die Bedingung (R) findet man schon bei M. Frolov [57]. Ihre Benennung nach K. Reidemeister (1893–1971) erfolgte aber erst später, als man sie im Rahmen der Theorie der Gewebe als Regularitätskriterium geometrischer Gewebe entdeckte (vgl. etwa W. Blaschke und G. Bol [24], J. Aczél [4]). J. Aczél [6] wies schließlich

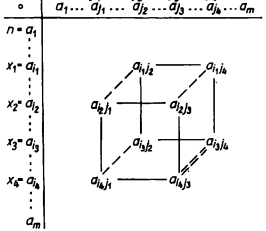


Abb. 8

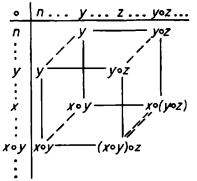


Abb. 9

nach, daß der Light-Test (1949), bei dem die erzeugenden Elemente des zu untersuchenden Gruppoids eine entscheidende Rolle spielen (vgl. D. J. Schadach [156], A. H. Clifford und G. B. Preston [35]), mit der Bedingung (R′) identisch ist.

Ohne Beweis sei noch mitgeteilt, daß aus der Reidemeister-Bedingung allein schon die Kürzbarkeit einer Operation folgt. Man kann auch — wie im Fall von (R′) — auf die Voraussetzung (Q) verzichten, will man die Assoziativität nachweisen; die Existenz eines neutralen Elementes muß jedoch gefordert werden (vgl. M. A. Taylor [168]).

Neben der Reidemeister-Bedingung läßt sich ein weiteres aus der Theorie der Gewebe bekanntes Regularitätskriterium für unsere Zwecke ausnutzen, die sogenannte *Thomsen-Bedingung* (vgl. auch Kap. 11):

$$(T) \qquad \bigwedge_{\substack{x_i,y_i\in L\\ i,j=1,2,3}} (x_1 \circ y_2 = x_2 \circ y_1 \wedge x_1 \circ y_3 = x_3 \circ y_1 \Rightarrow x_2 \circ y_3 = x_3 \circ y_2).$$

Wie sich zeigen wird, ist ein Assoziativitätstest mit Hilfe der Thomsen-Bedingung nur dann sinnvoll, wenn die Loop $(L, \circ)$ kommutativ ist. Sich hiervon vorab zu überzeugen erfordert allerdings, wie wir wissen, keine große Mühe. In Analogie zu Satz 1.6.1 gilt (auch für den Fall unendlicher Ordnung)

1.6.3. Satz. *Eine Loop $(L, \circ)$ erfüllt genau dann die Thomsen-Bedingung* (T), *wenn sie eine abelsche Gruppe ist.*

Zum Beweis, daß in einer abelschen Gruppe die Thomsen-Bedingung erfüllt ist, ziehen wir die Eigenschaften (A), (K), (N), (IE) und (1.5.4) heran:

$$\begin{aligned} x_2 \circ y_3 &= x_2 \circ (y_1 \circ y_1^{-1}) \circ (x_1^{-1} \circ x_1) \circ y_3 = (x_2 \circ y_1) \circ (x_1 \circ y_1)^{-1} \circ (x_1 \circ y_3) \\ &= (x_1 \circ y_2) \circ (x_1 \circ y_1)^{-1} \circ (x_3 \circ y_1) = (x_3 \circ y_1) \circ (x_1 \circ y_1)^{-1} \circ (x_1 \circ y_2) \\ &= x_3 \circ (y_1 \circ y_1^{-1}) \circ (x_1^{-1} \circ x_1) \circ y_2 = x_3 \circ y_2. \end{aligned}$$

Zum Beweis der umgekehrten Richtung setzen wir in (T) $x_1 = n$, wobei n das neutrale Element von $(L, \circ)$ sei. Damit folgt aus $y_2 = x_2 \circ y_1$ und $y_3 = x_3 \circ y_1$ die Gleichung $x_2 \circ y_3 = x_3 \circ y_2$, also $x_2 \circ (x_3 \circ y_1) = x_3 \circ (x_2 \circ y_1)$. Das ist die Identität der *Links-Permutabilität* (L-P). Sie liefert mit $y_1 = n$ die Kommutativität der Loop $(L, \circ)$. Unter Ausnutzung von (K) und (L-P) ergibt sich dann $x_2 \circ (y_1 \circ x_3) = x_2 \circ (x_3 \circ y_1) = x_3 \circ (x_2 \circ y_1) = (x_2 \circ y_1) \circ x_3$ für alle $x_2, x_3, y_1 \in L$, d. h., $(L, \circ)$ ist eine abelsche Gruppe.

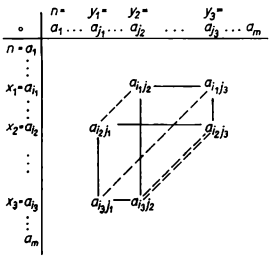


Abb. 10

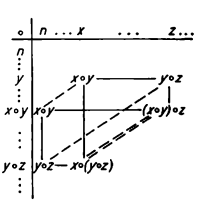


Abb. 11

Zur Illustration betrachten wir die beiden Abbildungen 10 und 11. Mit $a_{i_1} = n$ verschiebt sich die Thomsen-Figur in Abb. 10 so, daß für $a_{i_2} = x$, $a_{i_3} = y$ und $a_{j_3} = z$ die Assoziativität in Abb. 11 unmittelbar hervortritt.

Eine weitere Möglichkeit, ein lateinisches Quadrat (mit neutralem Element) auf Assoziativität hin zu untersuchen, geht auf H. Brandt [28] zurück. Dieses notwendige und hinreichende Kriterium wird mitunter auch *Rechteck-Regel* genannt (vgl. etwa A. Hammel [65], H. Siemon [162], O. Borůvka [27]).

Wir zeigen, daß diese Regel in einer (endlichen oder unendlichen) Loop mit der Reidemeister-Bedingung (R) äquivalent ist:

1.6.4. Satz. *Eine Loop $(L, \circ)$ mit dem neutralen Element n ist genau dann eine Gruppe, wenn sie die Rechteck-Regel*

$$\textbf{1.6.5.} \quad \bigwedge_{x_1, x_2, y_1, y_2 \in L} x_1 \circ y_1 = n \Rightarrow x_2 \circ y_2 = (x_2 \circ y_1) \circ (x_1 \circ y_2)$$

erfüllt.

Beweis. Da in einer Gruppe $(L, \circ)$ neben (N) und (A) wegen (IE) mit $x_1 \circ y_1 = n$ auch $y_1 \circ x_1 = n$ gilt, erhalten wir $x_2 \circ y_2 = x_2 \circ n \circ y_2 = x_2 \circ (x_1 \circ y_1) \circ y_2 = x_2 \circ (y_1 \circ x_1) \circ y_2 = (x_2 \circ y_1) \circ (x_1 \circ y_2)$, d. h., (1.6.5) ist erfüllt. Umgekehrt gelte (1.6.5) in einer Loop $(L, \circ)$, und es seien x, y, z und y_1 beliebige Elemente aus L. Dann gibt es wegen (Q) eindeutig bestimmte Elemente $x_2, x_3, y_2 \in L$ mit $x_2 \circ y_1 = n$, $x_3 \circ y_1 = x$ und $x_2 \circ y_2 = y$. Wegen (1.6.5) gilt $x_3 \circ y_2 = (x_3 \circ y_1) \circ (x_2 \circ y_2) = x \circ y$. Weiterhin existieren wegen (Q) eindeutig bestimmte Elemente $x_1, y_3 \in L$ mit $x_1 \circ y_2$

$= n$ und $x_1 \circ y_3 = z$, so daß mit $x_2 \circ y_2 = y$ wegen (1.6.5) $x_2 \circ y_3 = (x_2 \circ y_2) \circ (x_1 \circ y_3)$ $= y \circ z$ folgt. Wenden wir jetzt die Rechteck-Regel auf $n = x_2 \circ y_1$, $x = x_3 \circ y_1$ und $y \circ z = x_2 \circ y_3$ an, so erhalten wir einerseits $x_3 \circ y_3 = (x_3 \circ y_1) \circ (x_2 \circ y_3) = x \circ (y \circ z)$; für $n = x_1 \circ y_2$, $x \circ y = x_3 \circ y_2$ und $z = x_1 \circ y_3$ liefert sie andererseits $x_3 \circ y_3$ $= (x_3 \circ y_2) \circ (x_1 \circ y_3) = (x \circ y) \circ z$. Damit ist die Loop $(L, \circ)$ assoziativ, also eine Gruppe.

Zur Illustration der Rechteck-Regel in einem lateinischen Quadrat mit $L = \{a_1 = n, a_2, a_3, \ldots, a_m\}$ setzen wir in (1.6.5) $x_1 = a_{i_1}$, $x_2 = a_{i_2}$, $y_1 = a_{j_1}$ und $y_2 = a_{j_2}$, so daß sich das „Rechteck" n-x-y-$x \circ y$ in Abb. 12 ergibt.

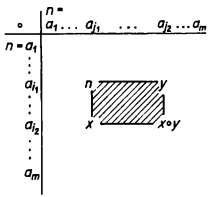


Abb. 12

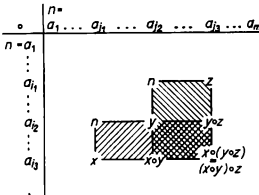


Abb. 13

Wählt man also zwei beliebige Elemente $x\,(= a_{i_2 j_1})$ und $y\,(= a_{i_1 j_2})$ so aus, daß im Schnittpunkt der j_1-ten Spalte S_{j_1} (in der also x aufgesucht worden ist) und der i_1-ten Zeile Z_{i_1} (in der sich das gewählte y befindet) gerade das neutrale Element $n\,(= a_{i_1 j_1})$ steht, dann muß im Schnittpunkt der i_2-ten Zeile Z_{i_2} und der j_2-ten Spalte S_{j_2} das Produkt $x \circ y$ stehen.

Anhand der Abb. 13 wird noch einmal deutlich, wie durch viermaliges Anwenden der Rechteck-Regel der Nachweis für die Assoziativität geführt worden ist: n-x-y-$x \circ y$, n-y-z-$y \circ z$, n-x-$y \circ z$-$x \circ (y \circ z)$ und n-$x \circ y$-z-$(x \circ y) \circ z$.

Daß wir bei diesem Test nicht auf die Quasigruppeneigenschaft verzichten können, zeigt die folgende Strukturtafel. In ihr ist zwar die Rechteck-Regel erfüllt, nicht aber die Assoziativität:

$\circ$	0	1	2
0	0	1	2
1	1	2	1
2	2	1	1

Wie immer man die Assoziativität in einer Strukturtafel prüfen will, der Aufwand läßt sich aber im Vergleich zu den m^3 zu untersuchenden Gleichungen $(x \circ y) \circ z$ $= x \circ (y \circ z)$ nur geringfügig mindern (vgl. F. D. Hammer [66]).

Im zugehörigen Cayley-Diagramm spiegelt sich diese Eigenschaft erfreulicherweise etwas übersichtlicher wider. So weist das Cayley-Diagramm $\Gamma(\mathfrak{H})$ einer Halbgruppe $\mathfrak{H} = (H, \circ)$ die folgende charakteristische Eigenschaft auf (vgl. R. Artzy [17]): Wenn zwei aufeinanderfolgende Bögen b_1 und b_2 mit den Farben X bzw. Y vom Knoten K_1 zum Knoten K_2 führen, so gibt es auch einen Bogen b_3 der Farbe $X \circ Y$[1]) von K_1 nach K_2 (siehe Abb. 14). Es gilt also

1.6.6. $\bigwedge_{b_1, b_2 \in \mathbf{B}} \Big(A(b_2) = E(b_1) \Rightarrow \bigvee_{b_3 \in \mathbf{B}} A(b_3) = A(b_1) \wedge E(b_3) = E(b_2) \wedge \alpha(b_3) = \beta\big(\beta^{-1}\alpha(b_1) \circ \beta^{-1}\alpha(b_2)\big)\Big)$.

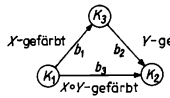


Abb. 14

Zum Beweis der ersten Richtung sei $\mathfrak{H} = (H, \circ)$ eine (endliche) Halbgruppe, b_1 und b_2 seien zwei (beliebige) aufeinanderfolgende Bögen im zugehörigen Cayley-Diagramm, für die die genannten Voraussetzungen zutreffen mögen: $\alpha(b_1) = X$, $\alpha(b_2) = Y$, $A(b_1) = K_1$, $E(b_1) = A(b_2) = K_3$ und $E(b_2) = K_2$. Mit $\beta(k_i) = K_i$ $(i = 1, 2, 3)$, $\beta(x) = X$ und $\beta(y) = Y$ erhalten wir dann einerseits die Gleichungen $k_1 \circ x = k_3$ und $k_3 \circ y = k_2$, also $(k_1 \circ x) \circ y = k_2$, andererseits ist aufgrund der Assoziativität auch $k_1 \circ (x \circ y) = k_2$ erfüllt, so daß es wegen (1.1.6) einen Bogen b_3 mit dem Anfangspunkt K_1, der Farbe $\alpha(b_3) = \beta(x \circ y) = X \circ Y$ und dem Endpunkt K_2 geben muß.

Es sei umgekehrt $\Gamma(\mathfrak{H})$ ein Cayley-Diagramm, das die Bedingung (1.6.6) erfüllt. Mit den gleichen Bezeichnungen wie zuvor liefert (1.6.6) dann die Gleichungen $k_1 \circ x = k_3$, $k_3 \circ y = k_2$ und $k_1 \circ (x \circ y) = k_2$, so daß wir auf $(k_1 \circ x) \circ y = k_1 \circ (x \circ y)$ für alle $k_1, x, y \in H$ stoßen.

Eine andere Variante, die Assoziativität mit Hilfe des Cayley-Diagramms zu testen, hat gegenüber (1.6.6) den Vorteil, daß sie mit rein graphentheoretischen Mitteln auskommt. So kann man auf die zuvor benötigte Kenntnis von Produkten $X \circ Y$ der Farben X und Y verzichten, muß allerdings diese Vereinfachung durch andere zusätzliche Voraussetzungen erkaufen: Das Cayley-Diagramm $\Gamma(\mathfrak{L})$ einer Loop $\mathfrak{L} = (L, \circ)$ ist genau dann ein Gruppen-Graph, wenn die folgende Bedingung erfüllt ist: Wenn eine Kette von Bögen $b_1, b_2, \ldots, b_k$ mit den (nicht notwendig voneinander verschiedenen) Farben $F_1, F_2, \ldots, F_k$ ein Kreis ist, d. h. $A(b_{i+1}) = E(b_i)$ für alle $i = 1, 2, \ldots, k-1$ und $E(b_k) = A(b_1)$ gilt, dann ist jede Kette aus Bögen dieser Farbfolge ein

[1]) Der Einfachheit halber identifizieren wir wieder die Elemente aus H und die (vermöge β zugeordneten) Knoten aus $\mathbf{K}$, so daß wir $X \circ Y$ anstelle von

$$\alpha(b_3) = Z = \beta(z) = \beta(x \circ y) = \beta(\beta^{-1}(X) \circ \beta^{-1}(Y)) = \beta\big(\beta^{-1}(\alpha(b_1)) \circ \beta^{-1}(\alpha(b_2))\big)$$

schreiben können.

Kreis. Der Beweis sei dem Leser überlassen. (Hinweis: In einer Loop, die diese Bedingung erfüllt, gilt für alle x, y die sogenannte Rechts-Invers-Eigenschaft $(x \circ y) \circ y^{-1} = x$, woraus dann auf $(x^{-1})^{-1} = x$ geschlossen werden kann.)

Die Beschreibung von Gruppen mittels Cayley-Diagrammen kann man erheblich vereinfachen. Dabei spielen dann die erzeugenden Elemente einer Gruppe (vgl. MfL Bd. 3, 12.2.) eine entscheidende Rolle. Ein orientierter Graph bleibt nämlich trotz Weglassens aller Bögen bestimmter Farben genau dann streng zusammenhängend[1]), wenn die Elemente, die den verbleibenden Farben entsprechen, gerade ein Erzeugendensystem der Gruppe bilden (R. Artzy [17], H. S. M. Coxeter und W. O. J. Moser [37], J. Dénes und A. D. Keedwell [44]).

Die bisher betrachteten algebraischen Strukturen mit einer binären Operation sind in Abb. 15 zusammengestellt.

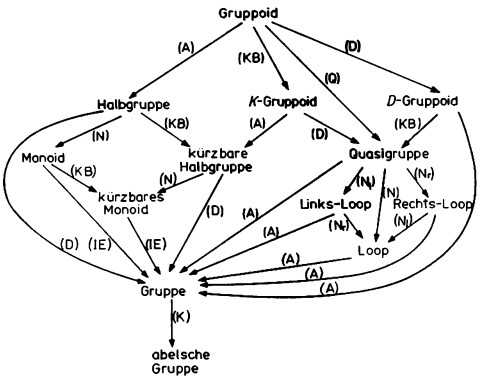


Abb. 15

1.6.7. Aufgaben

1. Man zeige, daß in der angegebenen Strukturtafel zwar die Reidemeister-Bedingung (R), nicht aber die Bedingung (R′) gilt. Ist die Operation assoziativ?

$\circ$	1	2	3	4
1	1	2	4	3
2	4	3	1	2
3	2	4	3	1
4	3	1	2	4

2. Man prüfe anhand der Strukturtafel oder des Cayley-Diagramms, ob die Quasigruppen aus Aufgabe 1.3.17.1 Gruppen sind.

[1]) Vgl. die Fußnote auf S. 27.

3. Man teste die angegebenen lateinischen Quadrate (oder die zugehörigen Quasigruppen-Graphen) auf Assoziativität und Kommutativität.

∘	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	3	1	6	4	5
3	3	1	2	5	6	4
4	4	5	6	1	2	3
5	5	6	4	3	1	2
6	6	4	5	2	3	1

*	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	3	4	5	6	1
3	3	4	1	6	2	5
4	4	5	6	1	3	2
5	5	6	2	3	1	4
6	6	1	5	2	4	3

4. Man gebe den vollständigen Gruppen-Graphen der nach F. KLEIN benannten Vierergruppe $\mathfrak{V}_4$ an und überzeuge sich, daß die auf die Farben der erzeugenden Elemente reduzierten Cayley-Diagramme streng zusammenhängende Graphen sind.

1.7. Translationen in Gruppoiden

Der folgende Abschnitt setzt voraus, daß der Leser die Grundbegriffe der Abbildungstheorie beherrscht. Dabei folgen wir in unseren Ausführungen der Terminologie von MfL 1 Bd. 1.

Die bisher betrachteten algebraischen Strukturen (vgl. Abb. 15) lassen sich auch mit Hilfe spezieller Abbildungen, den sogenannten *Translationen*, beschreiben.

1.7.1. Definition. Es sei $(G, \circ)$ ein Gruppoid, a ein beliebiges, aber festes Element aus G. Dann versteht man unter der *Links-Translation* L_a bezüglich a die folgende Abbildung von G in sich: $L_a\colon x \mapsto a \circ x$ für alle $x \in G$, d. h., jedes Element des Gruppoids wird von links mit a verknüpft. Analog heißt $R_a\colon G \to G$ mit $R_a(x) := x \circ a$ für alle $x \in G$ *Rechts-Translation* bezüglich a.

Die Operation $\circ$ garantiert die Eindeutigkeit von L_a bzw. R_a. Aus der Gleichheit der Urbilder $(x = y)$ folgt nämlich wegen der Eindeutigkeit von $\circ$ die Gleichheit der Bilder $\big(L_a(x) = L_a(y)$ bzw. $R_a(x) = R_a(y)\big)$. Für den Fall, daß wir gleichzeitig mehrere Operationen in derselben Trägermenge betrachten, vermerken wir die zugehörige Operation als hochgestellten Index an den Translationen, also z. B. $L_a^{(\circ)}$, $L_a^{(*)}$ usw. Beziehen wir die Umkehroperationen \ und / einer Quasigruppenoperation $\circ$ mit in unsere Untersuchungen ein, so werden wir für die Ausgangsoperation diese zusätzliche Kennzeichnung meist unterlassen, d. h., wir schreiben dann weiterhin L_a statt $L_a^{(\circ)}$ bzw. R_a statt $R_a^{(\circ)}$.

Mit den folgenden Sätzen heben wir, wie angekündigt, spezielle Klassen von Gruppoiden mit Hilfe der Translationen voneinander ab. So ist die Surjektivität derartiger Abbildungen allein schon (notwendig und) hinreichend für die Umkehrbarkeit, die Injektivität dagegen für die Kürzbarkeit einer Operation.

1.7.2. Satz. *Ein Gruppoid $(G, \circ)$ ist genau dann ein D-Gruppoid, wenn alle Links- und Rechts-Translationen Abbildungen von G auf sich sind.*

Beweis. 1. Es sei $(G, \circ)$ ein D-Gruppoid, L_a eine beliebige Links-Translation. Wir haben $W(L_a) = G$ zu zeigen. Nehmen wir an, $L_a(x) = a \circ x$ durchlaufe nicht ganz G.

Dann existiert ein Element $b \in G$, so daß die Gleichung $a \circ x = b$ keine Lösung hat im Widerspruch zur Voraussetzung. Entsprechend verläuft der Beweis für $W(R_a) = G$. Alle Translationen sind folglich surjektiv.

2. In einem Gruppoid $(G, \circ)$ gelte $W(L_a) = W(R_a) = G$ für alle $a \in G$. Da also alle Elemente von G als Bilder unter den Links- und Rechts-Translationen auftreten, können wir für jedes Element $b \in G$ solche Elemente (Urbilder) x und y aus G finden, so daß $L_a(x) = a \circ x = b$ bzw. $R_a(y) = y \circ a = b$ gilt. Damit ist $(G, \circ)$ ein D-Gruppoid.

1.7.3. Satz. *Ein Gruppoid $(G, \circ)$ ist genau dann ein K-Gruppoid, wenn alle Links- und Rechts-Translationen 1–1-Abbildungen von G in sich sind.*

Beweis. 1. Es sei $(G, \circ)$ ein K-Gruppoid. In der Sprache der Translationen hat die Kürzbarkeit folgende Gestalt: Aus der Gleichheit der Bilder $\big(L_a(x) = L_a(y)$ oder $R_a(x) = R_a(y)\big)$ folgt die Gleichheit der Urbilder ($x = y$), d. h., L_a und R_a sind für alle $a \in G$ injektiv.

2. In einem Gruppoid $(G, \circ)$ seien alle Links- und Rechts-Translationen L_a und R_a 1–1-Abbildungen von G in sich. Dann sind auch die zu L_a und R_a inversen Korrespondenzen L_a^{-1} bzw. R_a^{-1} eindeutig umkehrbar, und es gilt $L_a^{-1} \bullet L_a = I_G$ bzw. $R_a^{-1} \bullet R_a = I_G$, wobei $\bullet$ die Nacheinanderausführung von Korrespondenzen (insbesondere von Abbildungen) und I_G die durch $I_G := \{(x, x): x \in G\}$ definierte identische Abbildung von G ist.[1]) Es gilt also

$$\mathbf{1.7.4.} \quad \bigwedge_{a,x \in G} \Big(L_a^{-1}\big(L_a(x)\big) = x \wedge R_a^{-1}\big(R_a(x)\big) = x\Big).$$

(Man mache sich klar, daß L_a^{-1} bzw. R_a^{-1} in einem Gruppoid stets eine Korrespondenz, in einem K-Gruppoid sogar eine Abbildung ist, aber keine Translation zu sein braucht.)

Es sei nun $a \circ x = a \circ y$ für beliebige $a, x, y \in G$, also $L_a(x) = L_a(y)$. Dann gilt auch $L_a^{-1}\big(L_a(x)\big) = L_a^{-1}\big(L_a(y)\big)$, so daß wir mit (1.7.4) auf $x = y$ schließen können. Analog wird $x \circ a = y \circ a$ mit Hilfe von R_a^{-1} auf $x = y$ abgebildet. $(G, \circ)$ ist also ein K-Gruppoid.

Als unmittelbare Folgerung dieser beiden Sätze erhalten wir damit, daß bijektive Translationen die Quasigruppeneigenschaft (Q) widerspiegeln, d. h., es gilt

1.7.5. Korollar. *Ein Gruppoid $(G, \circ)$ ist genau dann eine Quasigruppe, wenn alle Links- und Rechts-Translationen 1–1-Abbildungen von G auf sich sind.*

Für eine Quasigruppe $(Q, \circ)$ und ihre Umkehroperationen $\backslash$ und $/$ halten wir noch fest, daß zwischen den Translationen in $(Q, \backslash)$ bzw. $(Q, /)$ einerseits sowie den Translationen in $(Q, \circ)$ andererseits ein sehr einfacher Zusammenhang besteht. Sind näm-

[1]) Die Nacheinanderausführung $g \bullet f$ (gelesen: g nach f) von Abbildungen $f: M \to N$ und $g: N \to P$ ist eine Abbildung von M in P, d. h., es gilt $(g \bullet f)(x) := g(f(x))$ für alle $x \in M$. Will man sich für die umgekehrte Reihenfolge, also für $(g \bullet f)(x) := f(g(x))$ entscheiden, so sollte man die Abbildungen als Exponenten schreiben: $x^{g \bullet f} := (x^g)^f$ — wie das besonders im englischen Sprachraum oft anzutreffen ist.

lich x, y und z beliebige Elemente aus Q mit $x \circ y = z$ und lösen wir diese Gleichung nach x bzw. y auf, so erhalten wir $y = x\backslash z$ bzw. $x = z/y$. Unter Benutzung von (1.7.4) finden wir damit

$$L_x^{-1}(z) = L_x^{-1}(x \circ y) = L_x^{-1}\big(L_x(y)\big) = y = x\backslash z$$

und

$$R_y^{-1}(z) = R_y^{-1}(x \circ y) = R_y^{-1}\big(R_y(x)\big) = x = z/y,$$

so daß (bei Umbenennung der Variablen) in einer Quasigruppe $(Q, \circ)$ die Translationen in bezug auf die beiden Umkehroperationen $\backslash$ und $/$, also $L_a^{(\backslash)}$ und $R_a^{(/)}$, sich durch die inversen Abbildungen der Translationen $L_a = L_a^{(\circ)}$ bzw. $R_a = R_a^{(\circ)}$ ersetzen lassen:

1.7.6. $\bigwedge_{a,x \in Q} \big(L_a^{(\backslash)}(x) = L_a^{-1}(x) = a\backslash x \wedge R_a^{(/)}(x) = R_a^{-1}(x) = x/a\big)$.

Das heißt also zugleich, daß jede inverse Abbildung einer Translation in einer Quasigruppe $(Q, \circ)$ selbst wieder eine Translation bezüglich desselben Elementes ist, jetzt aber in der Quasigruppe $(Q, \backslash)$ bzw. $(Q, /)$.

Gibt es in einem Gruppoid $(G, \circ)$ einseitig-neutrale Elemente, so sind die Translationen mit diesen Elementen gerade die identische Abbildung I_G (und umgekehrt). Das liest man unmittelbar aus der Definition 1.4.1 ab: $n_l \circ x = x$ genau dann, wenn $L_{n_l} = I_G$ und $x \circ n_r = x$ genau dann, wenn $R_{n_r} = I_G$. Unter Einbeziehung von Korollar 1.7.5 können wir damit auch eine Loop allein mit Hilfe von Translationen charakterisieren.

1.7.7. Korollar. *Eine Quasigruppe $(Q, \circ)$ ist genau dann eine Loop, wenn es ein Element $n \in Q$ gibt, so daß die Links-Translation L_n, die Rechts-Translation R_n und die identische Abbildung I_Q zusammenfallen, d. h., es gilt $L_n = R_n = I_Q$.*

Formuliert man solche Identitäten wie z. B. Assoziativität oder Kommutativität in der Sprache der Translationen, so lassen sich weitere Strukturen, in unserem Fall (kommutative) Halbgruppen oder (abelsche) Gruppen beschreiben. Interessanterweise entspricht dabei der Assoziativität einer Operation die Vertauschbarkeit von Links- und Rechts-Translationen.

1.7.8. Satz. *Ein Gruppoid $(G, \circ)$ ist genau dann eine Halbgruppe, wenn $L_a \bullet R_b = R_b \bullet L_a$ für alle $a, b \in G$ gilt.*

Beweis. Für alle $a, b, x \in G$ gilt

$$a \circ (x \circ b) = (a \circ x) \circ b \Leftrightarrow a \circ R_b(x) = L_a(x) \circ b \Leftrightarrow L_a\big(R_b(x)\big) = R_b\big(L_a(x)\big),$$

also $L_a \bullet R_b = R_b \bullet L_a$.

Ein Monoid $(M, \circ)$ läßt sich folglich durch $L_n = R_n = I_M$ und $L_a \bullet R_b = R_b \bullet L_a$, eine Gruppe $(G, \circ)$ durch surjektive Translationen und ebenfalls $L_a \bullet R_b = R_b \bullet L_a$ charakterisieren.

Es sei a ein beliebiges Element einer Gruppe $(G, \circ)$, a^{-1} sein inverses Element. Dann ist die inverse Abbildung einer Translation bezüglich a selbst wieder eine

Translation in $(G, \circ)$, nämlich bezüglich a^{-1}, d. h., es gilt $L_a{}^{-1} = L_{a^{-1}}$ und $R_a{}^{-1} = R_{a^{-1}}$ für alle $a \in G$. Zum Beweis ziehen wir (1.7.6) und (1.5.3) heran. Für alle $x \in G$ gilt nämlich

$$L_a{}^{-1}(x) = L_a{}^{(\backslash)}(x) = a \backslash x = a^{-1} \circ x = L_{a^{-1}}(x)$$

und

$$R_a{}^{-1}(x) = R_a{}^{(/)}(x) = x/a = x \circ a^{-1} = R_{a^{-1}}(x).$$

In der Schreibweise der Translationen nimmt die Kommutativität eines Gruppoids $(G, \circ)$ schließlich die folgende Gestalt an: Für alle $a, x \in G$ gilt $a \circ x = x \circ a \Leftrightarrow L_a(x) = R_a(x)$, also $L_a = R_a$, d. h., jede Links-Translation fällt mit der Rechts-Translation bezüglich desselben Elements zusammen. In einer kommutativen Halbgruppe $(G, \circ)$, speziell also auch in jeder abelschen Gruppe, gelten deshalb auch $L_a \bullet L_b = L_b \bullet L_a$ und $R_a \bullet R_b = R_b \bullet R_a$ für alle $a, b \in G$.[1])

Tabelle 1

$(G, \circ)$	Eigenschaften der Translationen					
	surjektiv	injektiv	$L_n = I_G$	$R_n = I_G$	$L_a \bullet R_b = R_b \bullet L_a$	$L_a = R_a$
K-Gruppoid	–	×	–	–	–	–
D-Gruppoid	×	–	–	–	–	–
Quasigruppe	×	×	–	–	–	–
Links-Loop	×	×	×	–	–	–
Rechts-Loop	×	×	–	×	–	–
Loop	×	×	×	×	–	–
Halbgruppe	–	–	–	–	×	–
Monoid	–	–	×	×	×	–
kürzbare Halbgruppe	–	×	–	–	×	–
kürzbares Monoid	–	×	×	×	×	–
Gruppe	×	×	×	×	×	–
abelsche Gruppe	×	×	×	×	×	×

In Tabelle 1 sind noch einmal die (für die hier betrachteten Klassen von Gruppoiden) charakteristischen Eigenschaften der Translationen zusammengefaßt.

Ist $\mathfrak{G} = (G, \circ)$ ein Gruppoid endlicher Ordnung, so lassen sich die Links- und die Rechts-Translationen unmittelbar aus den Zeilen bzw. aus den Spalten der zugehörigen Strukturtafel ablesen. Wenn $\mathfrak{G}$ eine Quasigruppe ist, dann sind nach Korollar 1.7.5 alle Translationen sogar Permutationen von G. Das deckt sich mit der Aussage in (1.3.12).

[1]) Diesen beiden Beziehungen entsprechen im übrigen zwei der 16 möglichen Identitäten, die man aus der Assoziativität durch alle möglichen Klammerungen und Vertauschungen der drei Variablen erhält (vgl. T. Faragó [54] und M. Hosszú [70]). Sie stehen für die Identitäten $a \circ (b \circ c) = b \circ (a \circ c)$ (Links-Permutabilität) bzw. $(a \circ b) \circ c = (a \circ c) \circ b$ (Rechts-Permutabilität).

Im Cayley-Diagramm eines (endlichen) Gruppoids $\mathfrak{G} = (G, \circ)$ sind die Rechts-Translationen gegenüber den Links-Translationen ausgezeichnet. Der Grund für diese Asymmetrie liegt in der Definition des Cayley-Diagramms selbst. In (1.1.6) wird nämlich das Element x in $x \circ y \ (= z)$ mit einem Knoten, das Element y aber mit einer Farbe in Verbindung gebracht. Das heißt mit anderen Worten, den Anfangspunkten der Bögen (Urbilder) werden vermöge der Rechts-Translationen die Endpunkte dieser Bögen (Bilder) zugeordnet (vgl. Abb. 2 bis 6).

1.7.9. Aufgaben

1. Man beschreibe die folgenden Identitäten mit Hilfe von Translationen:

$(x \circ y) \circ x = x \circ (y \circ x)$	(Elastizität),
$x \circ (x \circ y) = y$	(Links-Systemregel),
$(x \circ y) \circ y = x$	(Rechts-Systemregel),
$x \circ (y \circ x) = y$	(Links-Halbsymmetrie),
$(x \circ y) \circ x = y$	(Rechts-Halbsymmetrie),
$x \circ (y \circ z) = (x \circ y) \circ (x \circ z)$	(Links-Selbstdistributivität),
$(x \circ y) \circ z = (x \circ z) \circ (y \circ z)$	(Rechts-Selbstdistributivität).

2. Wie spiegelt sich die für (endliche) Quasigruppen charakteristische Eigenschaft der Translationen, nämlich bijektiv zu sein, im Quasigruppen-Graph wider?

3. Man gebe die Links- und Rechts-Translationen der Gruppoide aus den Beispielen 1.1.7.1 bis 1.1.7.3, 1.3.13.1, 1.3.13.2 und 1.4.6 an. Welche dieser Translationen sind Permutationen?

4. Man zeige, daß ein Gruppoid $(G, \circ)$ genau dann eine Halbgruppe ist, wenn $L_a \bullet L_b = L_{a \circ b}$ (bzw. $R_a \bullet R_b = R_{b \circ a}$) für alle $a, b \in G$ gilt.

2. Parastrophie von Quasigruppen

Ist in einem Gruppoid $(Q, \circ)$ neben der Existenz auch die Eindeutigkeit der Lösungen der Gleichungen $a \circ x = b$ und $y \circ a = b$ für alle $a, b \in Q$ gesichert, so besitzt die Operation $\circ$ zwei (i. allg. voneinander verschiedene) Umkehroperationen, die Linksdivision $\backslash$ und die Rechtsdivision $/$. Die beiden Umkehroperationen sind in Q selbst wieder umkehrbar und kürzbar, d. h., neben $(Q, \circ)$ sind auch $(Q, \backslash)$ und $(Q, /)$ Quasigruppen (vgl. 1.3.). Diese Tatsache erlaubt, von den bereits bekannten Umkehroperationen $\backslash$ und $/$ wieder deren jeweilige Umkehroperationen zu konstruieren. Zugleich stellt sich die Frage, ob dieser Prozeß überhaupt noch neue Operationen liefert. Das ist i. allg. der Fall. Im Höchstfall finden wir allerdings noch drei weitere Operationen, die von der Ausgangsoperation $\circ$ sowie deren Umkehroperationen $\backslash$ und $/$ verschieden sind. Mit diesen so miteinander verbundenen (maximal sechs) Operationen wollen wir uns in diesem Kapitel beschäftigen.

2.1. Parastrophe Quasigruppen

Es sei $(Q, \circ)$ eine Quasigruppe. Dann sind $(Q, \backslash)$ und $(Q, /)$ wegen Satz 1.3.9 ebenfalls Quasigruppen. Zwischen der Ausgangsoperation $\circ$ und ihren beiden Umkehroperationen $\backslash$ und $/$ besteht dabei der Zusammenhang

$$\bigwedge_{x,y,z \in Q} x \circ y = z \Leftrightarrow x \backslash z = y \Leftrightarrow z / y = x.$$

Um die weiteren Betrachtungen zu erleichtern, führen wir vorübergehend neue Bezeichnungen für die Links- und Rechtsdivision einer Operation ein (vgl. V. D. Belousov [21][1])):

$$x \circ^{-1} y := x \backslash y \quad \text{und} \quad x {}^{-1}\!\circ y := x / y.$$

Die obigen drei äquivalenten Gleichungen lauten demnach jetzt

2.1.1. $\bigwedge_{x,y,z \in Q} x \circ y = z \Leftrightarrow x \circ^{-1} z = y \Leftrightarrow z {}^{-1}\!\circ y = x.$

Wenn wir in dem lawinenartig anwachsenden Prozeß der Konstruktion der jeweiligen Umkehroperationen die Links- und Rechtsdivision von $\circ$ als Operationen der

[1]) Man beachte, daß die Linksdivision $\backslash$ bei V. D. Belousov [21] rechte Umkehroperation von $\circ$ und die Rechtsdivision $/$ als linke Umkehroperation von $\circ$ genannt wird.

ersten Generation bezeichnen, liefert die nächste Generation interessanterweise statt vier höchstens zwei neue Operationen:

$(Q, \backslash) = (Q, \circ^{-1})$: Die beiden Umkehroperationen von $\circ^{-1}$ sind $(\circ^{-1})^{-1}$ und ${}^{-1}(\circ^{-1})$.

$(Q, /) = (Q, {}^{-1}\circ)$: Die beiden Umkehroperationen von ${}^{-1}\circ$ sind $({}^{-1}\circ)^{-1}$ und ${}^{-1}({}^{-1}\circ)$.

Dabei gilt aber

2.1.2. $(\circ^{-1})^{-1} = {}^{-1}({}^{-1}\circ) = \circ$,

d. h., die Linksdivision der Linksdivision von $\circ$ ist ebenso wie die Rechtsdivision der Rechtsdivision von $\circ$ wieder die Ausgangsoperation $\circ$. Zum Beweis ersetzen wir $\circ$ in (1.3.4) und (1.3.5) durch $\circ^{-1}$ bzw. ${}^{-1}\circ$. Der Vergleich mit (1.3.7) bzw. (1.3.8) liefert dann aufgrund der Kürzbarkeit die Behauptung.

Wegen Satz 1.3.9 sind nun auch $\left(Q, {}^{-1}(\circ^{-1})\right)$ und $\left(Q, ({}^{-1}\circ)^{-1}\right)$ wieder Quasigruppen, so daß wir die jeweiligen Umkehroperationen konstruieren können:

$\left(Q, {}^{-1}\backslash\right) = \left(Q, {}^{-1}(\circ^{-1})\right)$: Die beiden Umkehroperationen von ${}^{-1}(\circ^{-1})$ sind $\left({}^{-1}(\circ^{-1})\right)^{-1}$ und ${}^{-1}\left({}^{-1}(\circ^{-1})\right) = \circ^{-1}$ (vgl. (2.1.2)).

$\left(Q, /^{-1}\right) = \left(Q, ({}^{-1}\circ)^{-1}\right)$: Die beiden Umkehroperationen von $({}^{-1}\circ)^{-1}$ sind $\left(({}^{-1}\circ)^{-1}\right)^{-1} = {}^{-1}\circ$ (vgl. (2.1.2)) und ${}^{-1}\left(({}^{-1}\circ)^{-1}\right)$.

In dieser dritten Generation tritt aber tatsächlich nur eine neue Operation auf. Es gilt nämlich

2.1.3. $\bigwedge_{x,y\in Q} x\left({}^{-1}(\circ^{-1})\right)^{-1} y = x\,{}^{-1}\left(({}^{-1}\circ)^{-1}\right) y = y \circ x.$

Den Beweis führen wir, indem wir wiederholt (2.1.1) anwenden: Für alle $x, y, z \in Q$ gilt einerseits

$$x\left({}^{-1}(\circ^{-1})\right)^{-1} y = z \Leftrightarrow x\,{}^{-1}(\circ^{-1})\,z = y \Leftrightarrow y \circ^{-1} z = x \Leftrightarrow y \circ x = z$$

und andererseits

$$x\,{}^{-1}\left(({}^{-1}\circ)^{-1}\right) y = z \Leftrightarrow z\,({}^{-1}\circ)^{-1}\, y = x \Leftrightarrow z\,{}^{-1}\circ\, x = y \Leftrightarrow y \circ x = z.$$

Wegen (2.1.3) sind die beiden Operationen $\left({}^{-1}(\circ^{-1})\right)^{-1}$ und ${}^{-1}\left(({}^{-1}\circ)^{-1}\right)$ also identisch. Deshalb setzen wir $\circ^* := \left({}^{-1}(\circ^{-1})\right)^{-1} = {}^{-1}\left(({}^{-1}\circ)^{-1}\right)$. Folglich gilt

2.1.4. $\bigwedge_{x,y\in Q} x \circ^* y = y \circ x.$

In der vierten Generation stoßen wir schon auf keine einzige neue Operation mehr. Um das zu zeigen, genügt es, die Quasigruppe $(Q, \circ^*)$ zu betrachten:

$(Q, \circ^*)$: Die beiden Umkehroperationen von $\circ^*$ sind $(\circ^*)^{-1} = \left(\left({}^{-1}(\circ^{-1})\right)^{-1}\right)^{-1}$
$= {}^{-1}(\circ^{-1})$ (vgl. (2.1.2)) und ${}^{-1}(\circ^*) = {}^{-1}\left({}^{-1}\left(({}^{-1}\circ)^{-1}\right)\right) = ({}^{-1}\circ)^{-1}$ (vgl. (2.1.2)).

Die Ausgangsoperation $\circ$ ist somit im Rahmen der fortgesetzten Bildung der Umkehroperationen mit maximal fünf weiteren (von ihr verschiedenen) Operationen

verbunden, d. h., wir erhalten folgendes 6-Tupel von Operationen in Q: $(\circ, \circ^{-1}, {}^{-1}\circ, {}^{-1}(\circ^{-1}), ({}^{-1}\circ)^{-1}, \circ^*)$. Zu jeder dieser sechs Operationen haben wir die beiden zugehörigen Umkehroperationen konstruiert. Damit sind wir auch in der Lage, die drei äquivalenten Gleichungen aus (2.1.1) auf diese Operationen auszudehnen.

Für alle $x, y, z \in Q$ gilt

a) $\quad x \circ y = z \Leftrightarrow \quad x \circ {}^{1}z = y \Leftrightarrow \quad z\,{}^{-1}\circ y = x,$

b) $\quad x \circ^{-1} z = y \Leftrightarrow \quad x\,(\circ^{-1})^{-1} y = z \Leftrightarrow \quad y\,{}^{-1}(\circ^{-1})\, z = x,$

c) $\quad z\,{}^{-1}\circ y = x \Leftrightarrow \quad z\,({}^{-1}\circ)^{-1} x = y \Leftrightarrow \quad x\,{}^{-1}({}^{-1}\circ)\, y = z,$

d) $\quad y\,{}^{-1}(\circ^{-1})\, z = x \Leftrightarrow \quad y\left({}^{-1}(\circ^{-1})\right)^{-1} x = z \Leftrightarrow \quad x\,{}^{-1}\left({}^{-1}(\circ^{-1})\right) z = y,$

e) $\quad z\,({}^{-1}\circ)^{-1} x = y \Leftrightarrow \quad z\left(({}^{-1}\circ)^{-1}\right)^{-1} y = x \Leftrightarrow \quad y\,{}^{-1}\left(({}^{-1}\circ)^{-1}\right) x = z,$

f_1) $\quad y\left({}^{-1}(\circ^{-1})\right)^{-1} x = z \Leftrightarrow y\left(\left({}^{-1}(\circ^{-1})\right)^{-1}\right)^{-1} z = x \Leftrightarrow z\,{}^{-1}\left(\left({}^{-1}(\circ^{-1})\right)^{-1}\right) x = y,$

f_2) $\quad y\,{}^{-1}\left({}^{-1}(\circ^{-1})\right) x = z \Leftrightarrow y\left({}^{-1}\left({}^{-1}(\circ^{-1})\right)\right)^{-1} z = x \Leftrightarrow z\,{}^{-1}\left({}^{-1}\left(({}^{-1}\circ)^{-1}\right)\right) x = y.$

Berücksichtigen wir die Beziehungen (2.1.2), (2.1.3) und (2.1.4), so bleiben hiervon insgesamt sechs zueinander äquivalente Gleichungen übrig:

$$x \circ y = z \Leftrightarrow x \circ^{-1} z = y \Leftrightarrow z\,{}^{-1}\circ y = x$$
$$\Leftrightarrow y\,{}^{-1}(\circ^{-1})\, z = x \Leftrightarrow z\,({}^{-1}\circ)^{-1} x = y \Leftrightarrow y \circ^* x = z.$$

Unter nochmaliger Einbeziehung von (2.1.4) folgen hieraus

$$y\,{}^{-1}(\circ^{-1})\, z = z\,{}^{-1}\circ y = y\,({}^{-1}\circ)^*\, z$$

und

$$z\,({}^{-1}\circ)^{-1} x = x \circ^{-1} z = z\,(\circ^{-1})^*\, x$$

für alle $x, y, z \in Q$, also ${}^{-1}(\circ^{-1}) = /^*$ und $({}^{-1}\circ)^{-1} = \backslash^*$.

Kehren wir zu unseren „alten" Bezeichnungen zurück, so nimmt das erhaltene 6-Tupel von Operationen folgende Gestalt an:

2.1.5. $(\circ, \backslash, /, /^*, \backslash^*, \circ^*)$,

während die sechs äquivalenten Gleichungen in ihrer endgültigen Form dann folgendes Aussehen haben:

2.1.6. $\bigwedge_{x,y,z \in Q} x \circ y = z \Leftrightarrow x \backslash z = y \Leftrightarrow z/y = x \Leftrightarrow y/^* z = x \Leftrightarrow z \backslash^* x = y$
$$\Leftrightarrow y \circ^* x = z.$$

2.1.7. Definition. Die sechs Operationen aus (2.1.5), also einschließlich $\circ$, heißen die *parastrophen Operationen* der Ausgangsoperation $\circ$. Entsprechend nennt man $(Q, \circ)$, $(Q, \backslash)$, $(Q, /)$, $(Q, /^*)$, $(Q, \backslash^*)$ und $(Q, \circ^*)$ die *parastrophen Quasigruppen* von $(Q, \circ)$.

Der Name „parastroph" tritt erstmalig bei J. B. Shaw [161] auf; eingebürgert hat sich diese Bezeichnung jedoch erst durch die Arbeiten A. Sades (vgl. u. a. [153]).

Den wiederholt durchgeführten Prozeß der Konstruktion der Umkehroperationen einer Operation $\circ$, der — wie wir gesehen haben — letztmalig in der dritten Generation eine neue Operation liefern kann, können wir jetzt wie in Abb. 16 veranschaulichen. Die hervorgehobenen Konstruktionen liefern gerade die sechs parastrophen Operationen von $\circ$, d. h., für das 6-Tupel (2.1.5) läßt sich eine reduzierte Darstellung angeben (siehe Abb. 17).

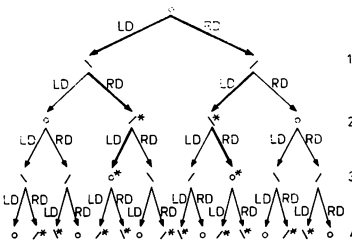


Abb. 16

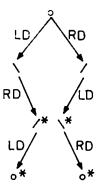


Abb. 17

2.1.8. Beispiel für sechs voneinander verschiedene parastrophe Quasigruppen Als Ausgangs-Quasigruppe wählen wir $(\mathbb{Q}, \circ)$ mit $x \circ y := 2x + y$ (vgl. Beispiel 1.3.3.2). Für die fünf weiteren parastrophen Operationen des 6-Tupels (2.1.5) erhalten wir dann

$$x \backslash y := -2x + y, \quad x / y := \frac{1}{2}(x - y),$$

$$x /^* y := -\frac{1}{2}(x - y), \quad x \backslash^* y := x - 2y$$

und

$$x \circ^* y := x + 2y.$$

Wählen wir die Operationen aus (2.1.5) der Reihe nach als Ausgangsoperationen, um die jeweils zugehörigen parastrophen Operationen zu konstruieren, so erhalten wir stets wieder diese sechs Operationen, mit anderen Worten: Je zwei Operationen aus (2.1.5) sind zueinander parastroph. Deshalb bezeichnet man den Übergang zwischen zwei Operationen aus (2.1.5) auch als *Parastrophismus*. Behalten wir die von uns in (2.1.5) gewählte Reihenfolge der parastrophen Operationen bei (die sich gemäß Abb. 16 ergibt, wenn man die Generationen von oben nach unten und jeweils von links nach rechts durchläuft), so erhalten wir insgesamt die folgenden 6-Tupel:[1])

2.1.9. $(\circ, \backslash, /, /^*, \backslash^*, \circ^*)$,
$(\backslash, \circ, /^*, /, \circ^*, \backslash^*)$,
$(/, \backslash^*, \circ, \circ^*, \backslash, /^*)$,
$(/^*, \circ^*, \backslash, \backslash^*, \circ, /)$,
$(\backslash^*, /, \circ^*, \circ, /^*, \backslash)$,
$(\circ^*, /^*, \backslash^*, \backslash, /, \circ)$.

[1]) Im Fall von $/^*$, $\backslash^*$ bzw. $\circ^*$ muß man den „Baum" in Abb. 16 natürlich fortsetzen.

Neben der Symmetrie erfüllt die Parastrophie auch die weiteren Eigenschaften einer Äquivalenzrelation, die Reflexivität und die Transitivität. Beides kann man aus (2.1.9) ablesen.

Es seien jetzt $(Q, \circ)$ und $(Q, \square)$ parastrophe Quasigruppen. Dann sind wegen (2.1.6) die Gleichungen $x \circ y = z$ und $x' \square y' = z'$ äquivalent, wobei $\begin{pmatrix} x & y & z \\ x' & y' & z' \end{pmatrix}$ eine Permutation von $\{x, y, z\}$ ist. Folglich lassen sich die sechs parastrophen Quasigruppen von $(Q, \circ)$ auch mit Hilfe aller Permutationen der Dreiermenge $\{x, y, z\}$ gewinnen (siehe Tab. 2).

Tabelle 2

$(Q, \circ)$	$(Q, \backslash)$	$(Q, /)$	$(Q, /^*)$	$(Q, \backslash^*)$	$(Q, \circ^*)$
$x \circ y = z$	$x \backslash z = y$	$z / y = x$	$y /^* z = x$	$z \backslash^* x = y$	$y \circ^* x = z$
(x, y, z)	(x, z, y)	(z, y, x)	(y, z, x)	(z, x, y)	(y, x, z)
$\pi_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$ $= I$	$\pi_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ $= (2\ 3)$	$\pi_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ $= (1\ 3)$	$\pi_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$ $= (1\ 3\ 2)$	$\pi_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ $= (1\ 2\ 3)$	$\pi_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$ $= (1\ 2)$

Die lexikographisch geordneten Permutationen π_i $(i = 1, 2, \ldots, 6)$ geben also jeweils an, in welcher Weise die Komponenten des (Ausgangs-) Tripels (x, y, z) zu vertauschen sind.

Beispiel: $(Q, \square) = (Q, /^*)$, $\pi_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = (1\ 3\ 2)$.

$$x \circ y = z \overset{\pi_5}{\longrightarrow} y /^* z = x$$

bzw.

	(x, y, z)	(y, z, x)
x	1. Komponente	3. Komponente
y	2. Komponente	1. Komponente
z	3. Komponente	2. Komponente

Für eine parastrophe Operation von $\circ$ schreiben wir jetzt anstelle von $\square$ zweckmäßigerweise gleich $\pi(\circ)$. Im einzelnen gilt also

$$\pi_1(\circ) = \circ, \quad \pi_2(\circ) = \backslash, \quad \pi_6(\circ) = /,$$
$$\pi_5(\circ) = /^*, \quad \pi_4(\circ) = \backslash^*, \quad \pi_3(\circ) = \circ^*.$$

Da also vermöge der Permutationen π_i $(i = 1, 2, \ldots, 6)$ eine Quasigruppe $(Q, \circ)$ in eine (zu ihr) parastrophe Quasigruppe $(Q, \pi_i(\circ))$ übergeht, ist π_i ein Parastrophismus von $(Q, \circ)$ auf $(Q, \pi_i(\circ))$.[1]) Die Parastrophismen π_i $(i = 1, 2, \ldots, 6)$ bilden bekannt-

[1]) Ein Parastrophismus ist also — im Unterschied zu einem (Homo-, Mono-, Iso-, Auto-, Endo-, Epi-) Morphismus — keine Abbildung von Q in oder auf sich.

Tabelle 3

π_1 $(Q, \circ)$	π_2 $(Q, \backslash)$	π_6 $(Q, /)$	π_5 $(Q, /^*)$	π_4 $(Q, \backslash^*)$	π_3 $(Q, \circ^*)$		
$+$ $f_1(a, b)$ $= a \cdot b = g$ $+$		$\overline{l}$ $f_3(b, g)$ $= \frac{g}{b} = a$ $-$		$\overline{r}$ $f_2(g, a)$ $= g : a = b$ $\div$		SCHRÖDER	1873 1881 1887
$M_1 - \gamma_{ijk}$	$M_3 - \gamma_{ikj}$	$M_4 - \gamma_{kji}$	$M_5 - \gamma_{jki}$	$M_6 - \gamma_{kij}$	$M_2 - \gamma_{jik}$	SHAW	1915
$\mathfrak{G}(\times)$		$\mathfrak{G}(\triangle)$		$\mathfrak{G}(\triangledown)$		SUSCHKEWITSCH	1929
Q	Q_2	Q_1	Q_{13}	Q_{23}	Q_3	SCHÖNHARDT	1930
$\{\mathfrak{S}, F\}$ $\{\mathfrak{S}, \circ\}$			$\{\mathfrak{S}, H\}$ $\{\mathfrak{S}, =\}$	$\{\mathfrak{S}, G\}$ $\{\mathfrak{S}, \triangle\}$		WARD	1930
$F(x, y, e)$	$F(x, e, y)$	$F(e, y, x)$	$F(y, e, x)$	$F(e, x, y)$	$F(y, x, e)$	BRUCK	1944
X	X^l	$X^\backslash$	$X^\vee$	$X^\wedge$	X^-	ETHERINGTON	1945
Q	Q_r	Q_l				ELLIS-UTZ	1951
μ	λ	ϱ				EVANS-NEUMANN	1953
$(\mathfrak{R}_0, +)$	$(\mathfrak{R}_0, D_l)$	$(\mathfrak{R}_0, D_r)$				PICKERT	1955
G	$(2\ 3)\ G$	$(1\ 3)\ G$	$(1\ 3\ 2)\ G$	$(1\ 2\ 3)\ G$	$(1\ 2)\ G$	STEIN	1957
$\triangle_{123} = \triangle$	$\triangle_{321}$ $= \triangle_{kr}$ $= \triangle_{lk}$	$\triangle_{213}$ $= \triangle_{kl}$ $= \triangle_{rk}$	$\triangle_{231}$ $= \triangle_r$	$\triangle_{312}$ $= \triangle_l$	$\triangle_{132}$ $= \triangle_k$	ČULIK	1958
$\times$ $Q = E(\cdot)$	$\odot$ $A = E(\odot)$ $= QP^{23}$ $S = E(\odot)$	$\ominus$ $U = E(\ominus)$ $= QP^{13}$ $T = E(\ominus)$	$\otimes$ $T' = E(\otimes)$ $= QP^{132}$ $V = E(\otimes)$	$\oplus$ $T = E(\oplus)$ $= QP^{123}$ $U = E(\oplus)$	$*$ $C = E(*)$ $= QP^{12}$ $R = E(\times)$	SADE	1959 1965 1969

f			$\tilde{f}$	$\bar{f}$		RADÓ	1959
$Q(\cdot)$	$Q(\varrho\tau)$	$Q(\tau\varrho)$	$Q(\varrho)$ $\varrho = \begin{pmatrix} x & y & z \\ y & z & x \end{pmatrix}$ [1]	$Q(\varrho^2)$	$Q(\tau)$ $\tau = \begin{pmatrix} x & y & z \\ y & x & z \end{pmatrix}$	ARTZY	1963
$Q(\cdot)$	$Q(\cdot)^{-1}$ $= Q(\backslash)$	$Q^{-1}(\cdot)$ $= Q(/)$	$Q^{-1}((\cdot)^{-1})$	$Q(^{-1}(\cdot))^{-1}$	$Q(\cdot)^*$	BELOUSOV	1967
$f(a, b) = c$		$g(c, b) = a$		$h(c, a) = b$		MÜLLER	1972
$(M, \circ)$		$(M, \circ_l)$		$(M, \circ_r)$		ASSER (MfL Bd. 1)	1973
(Q, ϑ)	(Q, ϑ^{-1})	$(Q, {}^{-1}\vartheta)$	$(Q, (^{-1}\vartheta)^*)$ $= (Q, {}^{-1}(\vartheta^{-1}))$	$(Q, (\vartheta^{-1})^*)$ $= (Q, (^{-1}\vartheta)^{-1})$	(Q, ϑ^*)	DÉNES-KEEDWELL	1974
$(Q, \otimes(1, 2, 3))$	$(Q, \otimes(1, 3, 2))$	$(Q, \otimes(3, 2, 1))$	$(Q, \otimes(2, 3, 1))$ [2]	$(Q, \otimes(3, 1, 2))$	$(Q, \otimes(2, 1, 3))$	LINDNER-STEEDLEY	1975
$(M, \circ)$	$(M, \backslash)$	$(M, /)$	$(M, \ominus)$	$(M, \oplus)$	$(M, \otimes)$	LEHMANN	1978

[1]) Der Permutation ϱ entspricht π_5 und nicht etwa π_4, da (ϱ) durch $x \cdot y = z \Rightarrow \varrho(x)\,(\varrho)\,\varrho(y) = \varrho(z)$ definiert wird, also $\varrho(x) = y$, $\varrho(y) = z$, $\varrho(z) = x$ und $(\varrho) = /^*$.

[2]) $(2, 3, 1)$ ist demnach so zu interpretieren, daß die zweite Komponente (y) mit der dritten (z), diese mit der ersten (x) und diese dann mit der zweiten (y) zu vertauschen ist. In dieser von unserer Lesart abweichenden Interpretation der Permutationen π_1 bis π_6 vertauschen aber nur π_4 und π_5 ihre Rollen.

lich (vgl. MfL Bd. 3, 12.1.) bezüglich der Nacheinanderausführung eine Gruppe, die zur symmetrischen Gruppe $\mathfrak{S}_3$ oder zu einer ihrer Untergruppen isomorph[1]) ist.

Betrachten wir jetzt den eingangs wiederholt durchgeführten Prozeß der Konstruktion der Umkehroperationen einer Quasigruppenoperation $\circ$ aus gruppentheoretischer Sicht, so erweist sich dieses Vorgehen als eine von neun gleichwertigen Möglichkeiten, das 6-Tupel (2.1.5) zu gewinnen. Dem Übergang zu den beiden Umkehroperationen $\backslash$ und $/$ entsprechen nämlich gerade die beiden Parastrophismen π_2 und π_6, d. h., es gilt für die erste Generation $\backslash = \pi_2(\circ)$ und $/ = \pi_6(\circ)$. In der zweiten Generation erhalten wir einerseits

$$\pi_2(\backslash) = \pi_2\big(\pi_2(\circ)\big) = (\pi_2 \bullet \pi_2)\,(\circ) = \pi_1(\circ) = \circ,$$
$$\pi_6(\backslash) = \pi_6\big(\pi_2(\circ)\big) = (\pi_6 \bullet \pi_2)\,(\circ) = \pi_5(\circ) = /^*$$

und andererseits

$$\pi_2(/) = \pi_2\big(\pi_6(\circ)\big) = (\pi_2 \bullet \pi_6)\,(\circ) = \pi_4(\circ) = \backslash^*,$$
$$\pi_6(/) = \pi_6\big(\pi_6(\circ)\big) = (\pi_6 \bullet \pi_6)\,(\circ) = \pi_1(\circ) = \circ.$$

Für die dritte Generation liefert schließlich $\pi_2 \bullet \pi_6 \bullet \pi_2$ bzw. $\pi_6 \bullet \pi_2 \bullet \pi_6$ den Parastrophismus π_3, also die Operation $\circ^*$. Sämtliche Permutationen der symmetrischen Gruppe $\mathfrak{S}_3$ lassen sich folglich allein durch die beiden Elemente π_2 und π_6 erzeugen. Man sagt, die Menge $\{\pi_2, \pi_6\}$ ist ein Erzeugendensystem der $\mathfrak{S}_3$ (vgl. MfL Bd. 3, 12.2.). Neben $\{\pi_2, \pi_6\}$ sind auch $\{\pi_2, \pi_3\}$, $\{\pi_2, \pi_4\}$, $\{\pi_2, \pi_5\}$, $\{\pi_3, \pi_4\}$, $\{\pi_3, \pi_5\}$, $\{\pi_3, \pi_6\}$, $\{\pi_4, \pi_6\}$ und $\{\pi_5, \pi_6\}$ (minimale) Erzeugendensysteme der $\mathfrak{S}_3$. Für die Gewinnung parastropher Quasigruppen gilt folglich der

2.1.10. Satz. *Es sei $(Q, \circ)$ eine Quasigruppe. Um das 6-Tupel (2.1.5) der parastrophen Quasigruppen von $(Q, \circ)$ zu erzeugen, genügt anstelle der sechs Permutationen π_1 bis π_6 schon irgendeine der folgenden neun Kombinationen von zwei Permutationen: $\{\pi_2, \pi_3\}$, $\{\pi_2, \pi_4\}$, $\{\pi_2, \pi_5\}$, $\{\pi_2, \pi_6\}$, $\{\pi_3, \pi_4\}$, $\{\pi_3, \pi_5\}$, $\{\pi_3, \pi_6\}$, $\{\pi_4, \pi_6\}$ oder $\{\pi_5, \pi_6\}$.*

Will man unter den sechs parastrophen Operationen von $\circ$ zwei als (die beiden) Umkehroperationen auszeichnen, so stehen somit neun Operationspaare zur Auswahl. Klammert man die Ausgangsoperation $\circ$ selbst aus, so fällt auf, daß allein die Kombination $\{/^*, \backslash^*\}$ ausscheidet, da eben $\{\pi_4, \pi_5\}$ kein Erzeugendensystem der $\mathfrak{S}_3$ ist. Viele Autoren haben sich ohnehin neben der Ausgangsoperation $\circ$ zumeist nur mit zwei der fünf parastrophen Operationen $\backslash$, $/$, $/^*$, $\backslash^*$ und $\circ^*$ beschäftigt und diese dann die Umkehroperationen von $\circ$ genannt. Vorrangig hat man von diesen genannten Operationen jedoch die Links- und Rechtsdivision gewählt und in der gleichen Weise wie wir hier bezeichnet (vgl. H. Boggs und G. Y. Rainich [25], P. Lorenzen [107], T. Evans [52], H. A. Thurston [171], A. I. Mal'cev [110], R. H. Bruck [32], J. Aczél [4], J. Aczél, G. Pickert und F. Radó [11] u. a.).

So vielfältig die Arbeiten über parastrophe Quasigruppen sind, so vielfältig ist leider auch die Terminologie. Für einige Arbeiten geben wir deshalb die jeweils gewählten Bezeichnungen in Tabelle 3 (vgl. S. 52/53) wieder.

[1]) Diese Einschränkung wird in 2.3. verständlich werden. (Zu den Begriffen „Untergruppe" und „isomorph" vgl. 3.2. und 3.3.)

Die Bezeichnung „parastroph" benutzen u. a. A. SADE [153], R. ARTZY [18], J. ACZÉL [4], V. D. BELOUSOV [22] sowie J. DÉNES und A. D. KEEDWELL [44]. Synonyme dieses Begriffs sind „koordiniert" (E. SCHÖNHARDT [157]), „adjugate" (R. A. FISHER und F. YATES [55]), „associated" (R. H. BRUCK [29]), „transposed" (I. M. H. ETHERINGTON [50]) und „conjugate(d)" (S. K. STEIN [164], J. DÉNES und A. D. KEEDWELL [44]).

Die Vielfalt der Bezeichnungen wird noch dadurch vergrößert, daß einige der parastrophen Operationen ihren eigenen Namen erhalten haben.

Anmerkung zum Begriff der Umkehroperation im Mathematikunterricht

Man kann sicher zu Recht unterstellen, daß der gegenwärtige Mathematiklehrgang in bezug auf den Begriff „Umkehroperation(en)" den Vorstellungen von E. SCHRÖDER (1873) folgt. Ein Indiz für diese Behauptung[1]) findet man in der Formulierung: „Die Subtraktion ist die Umkehroperation der Addition." Der bestimmte Artikel ist jedoch nur gerechtfertigt, wenn beide Umkehroperationen der Addition zusammenfallen. Das ist — im Unterschied zur Links- und Rechtsdivision — für die Schröderschen (Umkehr-) Operationen / und * aber der Fall. Um das einzusehen, zeigen wir zunächst, daß in einer kommutativen Quasigruppe $(Q, \circ)$

2.1.11. $\bigwedge_{x,y \in Q} x \backslash y = y/x$

gilt. Es seien a, b und c beliebige Elemente aus Q mit $a \circ b = c$. Aufgrund der Kommutativität gilt dann auch $b \circ a = c$. Mit (2.1.6) erhalten wir damit einerseits $a \backslash c = b$ und andererseits $c/a = b$, also $a \backslash c = c/a$. Die Quasigruppeneigenschaft sichert, daß mit b auch c ganz Q durchläuft, d. h., die Beziehung (2.1.11) ist bewiesen.

Unabhängig davon, ob die Ausgangsoperation $\circ$ kommutativ ist oder nicht, gilt wegen (2.1.4) für alle $x, y \in Q$ auch $x \backslash y = y \backslash^* x$, so daß für eine kommutative Operation $\circ$ die parastrophen Operationen / und * identisch sind.

Wir haben bereits festgestellt, daß beide Varianten, nämlich einerseits die Operationen \ und / oder andererseits die Operationen $\overline{\ }_l = / = \pi_6(\circ)$ und $\overline{\ }_r = \backslash^* = \pi_4(\circ)$ als Umkehroperationen von $\circ$ auszuzeichnen, gleichberechtigt neben sieben weiteren derartigen Möglichkeiten stehen (Satz 2.1.10). Das Erzeugendensystem $\{\pi_4, \pi_6\}$ ist

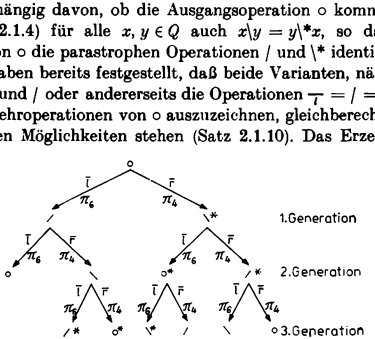


Abb. 18

[1]) Man berücksichtige hierbei allerdings, daß ja weder der Begriff „Operation" noch der Begriff „Umkehroperation" explizit definiert wird.

insofern sogar effektiver, als man bereits mit der zweiten Generation alle parastrophen Operationen erzeugt hat (siehe Abb. 18).

Die Kommutativität der Ausgangsoperation reduziert natürlich nicht nur das in Abb. 18 fixierte Schema, sondern auch den Prozeß, dem das Erzeugendensystem $\{\pi_2, \pi_6\}$ zugrunde liegt. Für die Addition (etwa in $\mathbf{Z}$) haben wir die Konstruktion der parastrophen Operationen für die beiden hier diskutierten Wege einander gegenübergestellt (siehe Abb. 19).

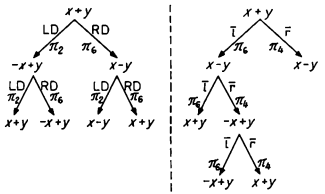


Abb. 19

2.1.12. Aufgaben

1. Man konstruiere die parastrophen Operationen der vier Grundrechenarten Addition, Subtraktion, Multiplikation und Division (in geeigneten Trägermengen) — vgl. Beispiel 1.3.3.1.

2. Man gebe die parastrophen Operationen für die Quasigruppen aus Beispiel 1.3.2.2 an.

3. Man konstruiere die Parastrophen einer Quasigruppe $(Q, \circ)$ allein mit Hilfe der Parastrophismen π_3 und π_4 bzw. π_3 und π_5 (K. Čulik [38], R. Artzy [18]).

4. Man zeige, daß die Parastrophie eine Äquivalenzrelation ist.

2.2. Parastrophe lateinische Quadrate

Es sei $\mathfrak{Q} = (Q, \circ)$ eine endliche Quasigruppe mit $Q = \{a_1, a_2, \ldots, a_m\}$. Dann ist die zugehörige Strukturtafel ein lateinisches Quadrat. Wegen Korollar 1.7.5 wissen wir, daß dann (und nur dann) die Translationen L_{a_i} und R_{a_j} (für alle i, j mit $1 \leqq i, j \leqq m$) 1–1-Abbildungen von Q auf sich sind. Fassen wir die Eingangs- oder Kopfzeile Z_0 und die Eingangs- oder Kopfspalte S_0 als das m-Tupel $(a_1, a_2, \ldots, a_m)$ auf, so können wir die Zeilen Z_i und die Spalten S_j des lateinischen Quadrates folgendermaßen mit Hilfe der Links- bzw. Rechts-Translationen von $\mathfrak{Q}$ darstellen:

$$\begin{aligned} Z_i &= (a_{i1}, a_{i2}, \ldots, a_{im}) = \big(L_{a_i}(a_1), L_{a_i}(a_2), \ldots, L_{a_i}(a_m)\big) \\ &= L_{a_i}\big((a_1, a_2, \ldots, a_m)\big) = L_{a_i}(Z_0), \\ S_j &= (a_{1j}, a_{2j}, \ldots, a_{mj}) = \big(R_{a_j}(a_1), R_{a_j}(a_2), \ldots, R_{a_j}(a_m)\big) \\ &= R_{a_j}\big((a_1, a_2, \ldots, a_m)\big) = R_{a_j}(S_0). \end{aligned}$$

Wir stellen uns jetzt das Ziel, für die parastrophen Quasigruppen von $(Q, \circ)$ das jeweils zugehörige lateinische Quadrat zu konstruieren. Dazu genügt es, wenn wir alle Zeilen oder alle Spalten des zu bestimmenden lateinischen Quadrates kennen.

Konstruktion des lateinischen Quadrates von $(Q, \backslash)$: Nach Voraussetzung steht in der Ausgangstafel im Schnittpunkt der i-ten Zeile $Z_i^{(\circ)} = Z_i$ und der j-ten Spalte $S_j^{(\circ)} = S_j$ das Element $a_{ij} \in Q$: $a_{ij} := a_i \circ a_j$. Im lateinischen Quadrat von $(Q, \backslash)$ sei dieses Feld mit dem Element $b_{ij} \in Q$ besetzt: $b_{ij} := a_i \backslash a_j$, d. h., aufgrund der Beziehung (1.7.6) gilt $b_{ij} = a_i \backslash a_j = L_{a_i}^{(\backslash)}(a_j) = L_{a_i}^{-1}(a_j)$ bzw., erweitert für die ganze i-te Zeile $Z_i^{(\backslash)}$ der neuen Tafel, $Z_i^{(\backslash)} = L_{a_i}^{(\backslash)}(Z_0) = L_{a_i}^{-1}(Z_0)$ für alle i mit $1 \leqq i \leqq m$.

Die Inversen der Links-Translationen der Ausgangsquasigruppe erzeugen folglich die Zeilen des lateinischen Quadrates von $(Q, \backslash)$.

Konstruktion des lateinischen Quadrates von $(Q, /)$: Für die zu bestimmenden Elemente $c_{ij} \in Q$ des lateinischen Quadrates von $(Q, /)$ gilt $c_{ij} := a_i / a_j$ für alle $i, j = 1, 2, \ldots, m$. In Analogie zur vorhergehenden Konstruktion erhalten wir $c_{ij} = a_i / a_j = R_{a_j}^{(/)}(a_i) = R_{a_j}^{-1}(a_i)$ bzw. für die j-te Spalte $S_j^{(/)}$ der neuen Tafel $S_j^{(/)} = R_{a_j}^{(/)}(S_0) = R_{a_j}^{-1}(S_0)$ für alle j mit $1 \leqq j \leqq m$. Die Spalten der neuen Tafel werden durch die Inversen der Rechts-Translationen bezüglich $\circ$ erzeugt.

Konstruktion der lateinischen Quadrate von $(Q, \circ^*)$, $(Q, \backslash^*)$ *und* $(Q, /^*)$: Wegen (2.1.4) erhalten wir die lateinischen Quadrate dieser parastrophen Quasigruppen, indem wir die Tafeln von $(Q, \circ)$, $(Q, \backslash)$ bzw. $(Q, /)$ an ihrer Hauptdiagonalen spiegeln. Es gilt nämlich $x \circ^* y = y \circ x$, $x \backslash^* y = y \backslash x$ und $x /^* y = y / x$ für alle $x, y \in Q$.

Tabelle 4

	$(Q, \circ)$	$(Q, \backslash)$	$(Q, /)$	$(Q, /^*)$	$(Q, \backslash^*)$	$(Q, \circ^*)$
	$a_{ij} = a_i \circ a_j$ $= L_{a_i}(a_j)$ $= R_{a_j}(a_i)$	$b_{ij} = a_i \backslash a_j$ $= L_{a_i}^{-1}(a_j)$	$c_{ij} = a_i / a_j$ $= R_{a_j}^{-1}(a_i)$	$d_{ij} = a_i /^* a_j$ $= a_j \backslash a_i$ $= R_{a_i}^{-1}(a_j)$	$e_{ij} = a_i \backslash^* a_j$ $= a_j /^* a_i$ $= L_{a_j}^{-1}(a_i)$	$f_{ij} = a_i \circ^* a_j$ $= a_j \circ a_i$ $= R_{a_i}(a_j)$ $= L_{a_j}(a_i)$
$Z_i =$	$L_{a_i}(Z_0)$	$L_{a_i}^{-1}(Z_0)$	–	$R_{a_i}^{-1}(Z_0)$	–	$R_{a_i}(Z_0)$
$S_j =$	$R_{a_j}(S_0)$	–	$R_{a_j}^{-1}(S_0)$	–	$L_{a_j}^{-1}(S_0)$	$L_{a_j}(S_0)$

(Für alle i, j mit $1 \leqq i, j \leqq m$.)

Natürlich lassen sich die lateinischen Quadrate von $(Q, \backslash^*)$ und $(Q, /^*)$ auch ohne den Umweg über die Tafeln von $(Q, \backslash)$ bzw. $(Q, /)$ konstruieren. In Tabelle 4 stellen wir deshalb die Konstruktion aller parastrophen lateinischen Quadrate einer Quasigruppe $(Q, \circ)$ der Ordnung m in der Weise zusammen, daß sich die zu findenden Zeilen oder Spalten allein aus der Ausgangstafel gewinnen lassen.

2.2.1. Beispiel parastropher lateinischer Quadrate

Nach einem Vorschlag von E. SCHÖNHARDT [157] lassen sich die Parastrophen einer (endlichen) Quasigruppe auch räumlich veranschaulichen, indem man die sechs lateinischen Quadrate in geeigneter Weise als die Seiten eines Würfels auffaßt.

∘	1	2	3	4	5	6
1	6	1	2	5	3	4
2	4	5	3	2	6	1
3	1	2	5	6	4	3
4	5	3	6	4	1	2
5	3	6	4	1	2	5
6	2	4	1	3	5	6

$L_2(Z_0)$

$R_3(S_0)$

\	1	2	3	4	5	6
1	2	3	5	6	4	1
2	6	4	3	1	2	5
3	1	2	6	5	3	4
4	5	6	2	4	1	3
5	4	5	1	3	6	2
6	3	1	4	2	5	6

$L_2^{-1}(Z_0)$

/	1	2	3	4	5	6
1	3	1	6	5	4	2
2	6	3	1	2	5	4
3	5	4	2	6	1	3
4	2	6	5	4	3	1
5	4	2	3	1	6	5
6	1	5	4	3	2	6

$R_3^{-1}(S_0)$

∘*	1	2	3	4	5	6
1	6	4	1	5	3	2
2	1	5	2	3	6	4
3	2	3	5	6	4	1
4	5	2	6	4	1	3
5	3	6	4	1	2	5
6	4	1	3	2	5	6

$R_3(Z_0)$

$L_2(S_0)$

*	1	2	3	4	5	6
1	2	6	1	5	4	3
2	3	4	2	6	5	1
3	5	3	6	2	1	4
4	6	1	5	4	3	2
5	4	2	3	1	6	5
6	1	5	4	3	2	6

$L_2^{-1}(S_0)$

/*	1	2	3	4	5	6
1	3	6	5	2	4	1
2	1	3	4	6	2	5
3	6	1	2	5	3	4
4	5	2	6	4	1	3
5	4	5	1	3	6	2
6	2	4	3	1	5	6

$R_3^{-1}(Z_0)$

$$L_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 3 & 2 & 6 & 1 \end{pmatrix}, \quad L_2^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 4 & 3 & 1 & 2 & 5 \end{pmatrix},$$

$$R_3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 5 & 6 & 4 & 1 \end{pmatrix}, \quad R_3^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 1 & 2 & 5 & 3 & 4 \end{pmatrix}.$$

2.2.2. Aufgaben

1. Man gebe die Parastrophen der folgenden lateinischen Quadrate an:

$\circ_1$	0	1	2
0	0	2	1
1	2	1	0
2	1	0	2

$\circ_2$	0	1	2
0	0	2	1
1	1	0	2
2	2	1	0

$\circ_3$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\circ_4$	0	1	2
0	0	1	2
1	2	0	1
2	1	2	0

2. Man vergleiche für die folgenden lateinischen Quadrate die Anzahl ihrer voneinander verschiedenen Parastrophen:

$\circ_1$	1	2	3	4
1	1	2	3	4
2	2	1	4	3
3	3	4	1	2
4	4	3	2	1

$\circ_2$	1	2	3	4
1	1	2	3	4
2	2	3	4	1
3	3	4	1	2
4	4	1	2	3

$\circ_3$	1	2	3	4
1	1	4	2	3
2	3	1	4	2
3	2	3	1	4
4	4	2	3	1

$\circ_4$	1	2	3	4
1	2	1	4	3
2	3	2	1	4
3	4	3	2	1
4	1	4	3	2

3. Man konstruiere die Cayley-Diagramme der parastrophen Quasigruppen aus Beispiel 2.2.1.

4. Es sei $(Q, \circ)$ eine beliebige (endliche) Quasigruppe.

a) Man überlege sich, warum das Cayley-Diagramm von $(Q, /)$ unmittelbar aus dem von $(Q, \circ)$ abgelesen werden kann.

b) Man zeige, daß die parastrophen Cayley-Diagramme von $(Q, \circ)$ und $(Q, /)$, von $(Q, \backslash)$ und $(Q, /^*)$ sowie von $(Q, \circ^*)$ und $(Q, \backslash^*)$ sich einzig und allein durch die Orientierung ihrer Kanten unterscheiden (Hinweis: $/ = \pi_6(\circ)$, $/^* = \pi_6(\backslash)$, $\backslash^* = \pi_6(\circ^*)$ (vgl. (2.1.9))).

2.3. Die Anzahl p der Parastrophen einer Quasigruppe

Schon für die Addition (etwa in $\mathbf{Z}$) haben wir festgestellt, daß die Anzahl ihrer (voneinander verschiedenen) Parastrophen kleiner als 6 ist. Wegen der Kommutativität der Addition und (2.1.4) sind $+ = \pi_1(+)$ und $\pi_3(+)$, wegen (2.1.11) und wieder (2.1.4) aber auch $- = \pi_6(+)$ und $\pi_4(-)$ identisch. Darüber hinaus fallen auch $\pi_2(-)$ und $\pi_5(+)$ zusammen, so daß insgesamt nur drei voneinander verschiedene parastrophe Operationen auftreten (vgl. auch die Aufgaben 2.2.2.1 und 2.2.2.2).

Wir werden im folgenden zeigen, daß neben der Anzahl $p = 6$ und $p = 3$ nur noch die Fälle $p = 2$ und $p = 1$ möglich sind. Daß tatsächlich nur diese vier Fälle für p auftreten können, überlegen wir uns anhand der 6-Tupel aus (2.1.9).

Für $p = 6$ verweisen wir auf das Beispiel 2.2.1. Hier sind für alle π_i mit $i = 1, 2, \ldots, 6$ die parastrophen Operationen $\pi_i(\circ)$ paarweise verschieden.

Wenn es den Fall $p = 5$ geben soll, müssen zwei der (zunächst formal konstruierten) sechs parastrophen Operationen zusammenfallen, etwa $\circ^*$ und $\backslash$. Dann gilt natürlich auch $\pi_i(\circ^*) = \pi_i(\backslash)$ für alle $i = 1, 2, \ldots, 6$. Im einzelnen heißt das:

$$i = 1: \circ^* = \backslash, \qquad i = 2: /^* = \circ, \qquad i = 6: \backslash^* = /^*,$$
$$i = 5: \backslash = /, \qquad i = 4: / = \circ^*, \qquad i = 3: \circ = \backslash^*,$$

d. h., insgesamt fallen nicht nur $\circ^*$ und $\backslash$ zusammen, sondern $\circ^*$, $\backslash$ und $/$. Darüber hinaus sind aber auch die parastrophen Operationen $/^*$, $\circ$ und $\backslash^*$ miteinander identisch. Das folgt unmittelbar aus (2.1.9). Da nämlich das letzte und zweite 6-Tupel in (2.1.9) nach Voraussetzung gleich sind, liefert der Vergleich der einzelnen Komponenten, daß statt der erwarteten fünf Parastrophen nur zwei voneinander verschiedene parastrophe Operationen auftreten können.

In analoger Weise kann man alle denkbaren Fälle des Zusammenfallens parastropher Operationen durchgehen. Im Ergebnis einer solchen systematischen Fallunterscheidung erhalten wir dann den

2.3.1. Satz. *Es sei $(Q, \circ)$ eine Quasigruppe. Die Anzahl p der voneinander verschiedenen parastrophen Quasigruppen ist entweder 1, 2, 3 oder 6. Dabei sind folgende Fälle möglich:*

$p = 6$: $\circ, \backslash, /, /^*, \backslash^*, \circ^*$ *sind paarweise verschieden.*

$p = 3$: (a) $\circ = \backslash, / = /^*, \backslash^* = \circ^*$, *aber* $\circ, /, \backslash^*$ *sind paarweise verschieden.*

(b) $\circ = /, \backslash = \backslash^*, /^* = \circ^*$, *aber* $\circ, \backslash, /^*$ *sind paarweise verschieden.*

(c) $\circ = \circ^*, \backslash = /^*, / = \backslash^*$, *aber* $\circ, \backslash, /$ *sind paarweise verschieden.*

$p = 2$: $\circ = /^* = \backslash^*, \backslash = / = \circ^*$, *aber* $\circ$ *und* $\backslash$ *sind verschieden.*

$p = 1$: $\circ = \backslash = / = /^* = \backslash^* = \circ^*$.

Beispiele für die genannten Fälle:

$p = 6$: Beispiel 2.2.1 und o_3 in Aufgabe 2.2.2.2,

$p = 3$: (a) o_4 in Aufgabe 2.2.2.2,

(b) o_4 in Aufgabe 2.2.2.1,

(c) o_2 in Aufgabe 2.2.2.2 (zyklische Gruppe der Ordnung 4),

$p = 2$:

$\circ$	1	2	3	4	5
1	1	2	3	4	5
2	2	1	4	5	3
3	3	5	1	2	4
4	4	3	5	1	2
5	5	4	2	3	1

$p = 1$: o_1 in Aufgabe 2.2.2.1 und 2.2.2.2 (Kleinsche Vierergruppe).

Die Anzahl p der Parastrophen einer Quasigruppe läßt sich auch folgendermaßen charakterisieren:

2.3.2. Satz. *Es sei $(Q, \circ)$ eine Quasigruppe.*

a) *Der Fall $p = 3$ (a) liegt genau dann vor, wenn die Operation $/$, nicht aber die Operation $\circ$ kommutativ ist.*

b) *Der Fall $p = 3$ (b) liegt genau dann vor, wenn die Operation $\backslash$, nicht aber die Operation $\circ$ kommutativ ist.*

c) *Der Fall $p = 3$ (c) liegt genau dann vor, wenn die Operation $\circ$, nicht aber die Operation $\backslash$ (und auch nicht $/$) kommutativ ist.*

d) *Der Fall $p = 2$ liegt genau dann vor, wenn die Operation $\circ$ (und $\backslash$) halbsymmetrisch*[1]*), nicht aber kommutativ ist.*

e) *Der Fall $p = 1$ liegt genau dann vor, wenn die Operationen $\circ$ und $\backslash$ (und $/$) kommutativ sind oder die Operation $\circ$ kommutativ und halbsymmetrisch ist.*

Beweis. a) Es sei $p = 3$ (a) erfüllt. Dann folgt aus $x/y = x/^*y = y/x$ für alle $x, y \in Q$ (vgl. (2.1.4)) die Kommutativität der Operation $/$. Wäre auch $\circ$ kommutativ, so fiele sie im Widerspruch zur Voraussetzung mit $\circ^*$ zusammen. Umgekehrt sei $(Q, /)$ eine kommutative Quasigruppe. Dann gilt $x/y = y/x = x/^*y$ für alle $x, y \in Q$ (vgl. (2.1.4)), also $/ = /^*$. Wegen $\pi_i(/) = \pi_i(/^*)$ für $i = 1, 2, \ldots, 6$ gilt dann auch $\circ = \backslash$ und $\backslash^* = \circ^*$. Der Fall $p = 1$ kann dabei ausgeschlossen werden. Fiele nämlich $/$ mit allen Parastrophen (also auch $\circ$) zusammen, so wäre $\circ$ im Widerspruch zur Voraussetzung ebenfalls kommutativ.

Der Beweis von b), c) und e) erfolgt analog.

d) Es sei $p = 2$ erfüllt. Dann sind wegen (2.1.6) und $\circ = /^* = \backslash^*$ für beliebige $x, y, z \in Q$ die Gleichungen $x \circ y = z$, $y \circ z = x$ und $z \circ x = y$ äquivalent. Durch Einsetzen erhalten wir die Identitäten $(x \circ y) \circ x = y$ und $y \circ (x \circ y) = x$, d. h., die

[1]) Das heißt, die Identitäten $x \circ (y \circ x) = y$ (Links-Halbsymmetrie) und $(x \circ y) \circ x = y$ (Rechts-Halbsymmetrie) sind erfüllt. In 2.4. werden wir sehen, daß beide Identitäten gleichwertig sind.

Rechts- bzw. Links-Halbsymmetrie. Die Operation $\circ$ kann nicht kommutativ sein, da sie sonst im Widerspruch zur Voraussetzung mit $\circ^*$ zusammenfiele. Umgekehrt sei $(Q, \circ)$ eine halbsymmetrische, aber nichtkommutative Quasigruppe. Dann gilt für alle $x, y, z \in Q$ mit $x \circ y = z$ auch $y \circ z = y \circ (x \circ y) = x$ (wegen L-HS) und $z \circ x = (x \circ y) \circ x = y$ (wegen R-HS). Über den Vergleich mit den zu $x \circ y = z$ äquivalenten Gleichungen $y/^* z = x$ und $z\backslash^* x = y$ aus (2.1.6) erhalten wir folglich $\circ = /^* = \backslash^*$. Damit gilt dann auch $\pi_i(\circ) = \pi_i(/^*) = \pi_i(\backslash^*)$ für $i = 1, 2, \ldots, 6$, was für die restlichen parastrophen Operationen $\backslash = / = \circ^*$ nach sich zieht. Da nach Voraussetzung $\circ$ nicht kommutativ sein soll, kann nur der Fall $p = 2$, nicht aber der Fall $p = 1$ vorliegen.

2.3.3. Korollar. *Der Fall $p = 6$ liegt genau dann vor, wenn keine der sechs parastrophen Operationen kommutativ oder halbsymmetrisch ist.*

In 2.4. wird sich zeigen, daß Satz 2.3.2 mit dem von Ch. C. Lindner und D. Steedley (vgl. [106]) erzielten Resultat gleichwertig ist.

Betrachten wir jetzt die Parastrophismen π_i, die in den Fällen $p = 1, 2, 3$ und 6 zur Erzeugung der dort auftretenden Operationen nötig sind, so stoßen wir gerade auf die Trägermengen der Untergruppen der symmetrischen Gruppe $\mathfrak{S}_3$:

p	Parastrophismen	Untergruppe der $\mathfrak{S}_3$
6	$\pi_1, \pi_2, \pi_6, \pi_5, \pi_4, \pi_3$	$\mathfrak{S}_3$
3	$\pi_1 \vee \pi_2, \pi_6 \vee \pi_5, \pi_4 \vee \pi_3$	$(\{\pi_1, \pi_4, \pi_5\}, \bullet)$
3	$\pi_1 \vee \pi_6, \pi_2 \vee \pi_4, \pi_5 \vee \pi_3$	$(\{\pi_1, \pi_4, \pi_5\}, \bullet)$
3	$\pi_1 \vee \pi_3, \pi_2 \vee \pi_5, \pi_6 \vee \pi_1$	$(\{\pi_1, \pi_4, \pi_5\}, \bullet)$
2	$\pi_1 \vee \pi_5 \vee \pi_4, \pi_2 \vee \pi_6 \vee \pi_3$	$(\{\pi_1, \pi_2\}, \bullet)$ oder $(\{\pi_1, \pi_3\}, \bullet)$ oder $(\{\pi_1, \pi_6\}, \bullet)$
1	$\pi_1 \vee \pi_2 \vee \pi_6 \vee \pi_5 \vee \pi_4 \vee \pi_3$	$(\{\pi_1\}, \bullet)$

Das erklärt jetzt auch unsere in 2.1. vorweggenommene Feststellung, daß die Parastrophismen bezüglich der Nacheinanderausführung eine Gruppe bilden, die zur $\mathfrak{S}_3$ oder zu einer ihrer Untergruppen isomorph ist.

Am Beispiel der Multiplikation (etwa in $\mathbb{Q}^*$) untersuchen wir abschließend, wieviel voneinander verschiedene parastrophe Operationen existieren. Aus $x \circ y := x \cdot y$ folgt für die beiden Umkehroperationen $x\backslash y := \frac{1}{x} \cdot y$ und $x/y := \frac{x}{y}$. Damit gilt für alle $x, y \in \mathbb{Q}^*$

$$x \circ^* y = y \circ x = y \cdot x = x \cdot y = x \circ y, \quad \text{also} \quad \circ = \circ^*,$$

$$x/^* y = y/x = \frac{y}{x} = \frac{1}{x} \cdot y = x\backslash y, \quad \text{also} \quad \backslash = /^*,$$

$$x\backslash^* y = y\backslash x = \frac{1}{y} \cdot x = \frac{x}{y} = x/y, \quad \text{also} \quad / = \backslash^*.$$

Ferner wissen wir, daß die Multiplikation kommutativ ist, nicht aber ihre beiden Umkehroperationen. Es liegt also der Fall $p = 3$ (c) vor.

2.3.4. Aufgaben

1. Man konstruiere die Parastrophen der durch die folgenden lateinischen Quadrate gegebenen Quasigruppen und entscheide, welcher der Fälle $p = 1, 2, 3, 6$ vorliegt:

$\circ_1$	1	2	3	4
1	4	1	3	2
2	3	4	2	1
3	1	2	4	3
4	2	3	1	4

$\circ_2$	0	1	2
0	1	0	2
1	0	2	1
2	2	1	0

$\circ_3$	0	1	2
0	2	1	0
1	0	2	1
2	1	0	2

2. Man zeige, daß sich die Anzahl p der Parastrophen einer Quasigruppe $(Q, \circ)$ in der Sprache der Translationen folgendermaßen widerspiegelt:

$p = 6$: $\bigvee_{a \in Q} (L_a \neq L_a^{-1} \wedge L_a \neq R_a)$,

$p = 3$ (a): $\bigwedge_{a \in Q} L_a = L_a^{-1} \wedge \bigvee_{b \in Q} L_b \neq R_b$,

$p = 3$ (b): $\bigwedge_{a \in Q} R_a = R_a^{-1} \wedge \bigvee_{b \in Q} L_b \neq R_b$,

$p = 3$ (c): $\bigwedge_{a \in Q} L_a = R_a \wedge \bigvee_{b \in Q} R_b \neq R_b^{-1} \quad \left(\bigvee_{c \in Q} L_c = L_c^{-1}\right)$,

$p = 2$: $\bigwedge_{a \in Q} L_a = R_a^{-1} \wedge \bigvee_{b \in Q} L_b \neq R_b$,

$p = 1$: $\bigwedge_{a \in Q} L_a = L_a^{-1} = R_a = R_a^{-1}$. (Hinweis: Man beachte (1.7.6) sowie (2.1.4)).

2.4. Parastrophe Identitäten

Während die Parastrophen einer Quasigruppe selbst immer wieder Quasigruppen sind, bleiben solche Eigenschaften wie (N), (K) und (A), d. h., die Existenz eines neutralen Elementes, die Kommutativität bzw. die Assoziativität, beim Übergang von einer Quasigruppe $(Q, \circ)$ zu einer Parastrophen $(Q, \pi(\circ))$ i. allg. nicht erhalten. Das lehrt uns schon die Addition (etwa in $\mathbf{Z}$). So ist die Subtraktion (als π_6-Parastrophe der Addition) weder kommutativ noch assoziativ, auch existiert in $(\mathbf{Z}, -)$ kein neutrales Element. Wir werden jedoch im folgenden sehen, daß jeder Eigenschaft in $(Q, \circ)$ wieder eine Eigenschaft in $(Q, \pi(\circ))$ entspricht.

Betrachten wir z. B. die Parastrophen einer Loop $(L, \circ)$, so wissen wir bereits aus 1.4., daß das neutrale Element n von $(L, \circ)$ in $(L, \backslash)$ links- und in $(L, /)$ rechts-neutrales Element ist. Darüber hinaus sind beide Umkehroperationen unipotent. Wenn wir die restlichen drei parastrophen Operationen von $\circ$ mit in unsere Überlegungen einschließen, erhalten wir infolge von (2.1.6) folgende äquivalente Gleichungen:

$$n \circ x = x \Leftrightarrow n \backslash x = x \Leftrightarrow x / x = n \Leftrightarrow x /^* x = n$$
$$\Leftrightarrow x \backslash^* n = x \Leftrightarrow x \circ^* n = x$$

und

$$x \circ n = x \Leftrightarrow x \backslash x = n \Leftrightarrow x / n = x \Leftrightarrow n /^* x = x$$
$$\Leftrightarrow x \backslash^* x = n \Leftrightarrow n \circ^* x = x$$

für alle $x \in L$, d. h., $(L, \backslash)$ und L, $(*)$ sind unipotente Links-Loops, $(L, /)$ und $(L, \backslash^*)$ unipotente Rechts-Loops, während $(L, \circ^*)$ stets wieder eine Loop ist:

2.4.1. Satz. *Für die Parastrophen einer Loop $(L, \circ)$ gilt:*

$(L, \circ)$	$(L, \backslash)$	$(L, /)$	$(L, /^*)$	$(L, \backslash^*)$	$(L, \circ^*)$
(N_l)	(N_l)	(U)	(U)	(N_r)	(N_r)
(N_r)	(U)	(N_r)	(N_l)	(U)	(N_l)

Wir haben hier die Eigenschaft (N) der Ausgangsquasigruppe $(L, \circ)$ bewußt in (N_l) und (N_r) zerlegt. Das erlaubt nämlich, mit Hilfe der 6-Tupel aus (2.1.9) die sicher naheliegende Frage zu beantworten, welche Eigenschaften die Unipotenz ihrerseits in den einzelnen Parastrophen nach sich zieht. Wählen wir etwa $(L, /)$ als unipotente Ausgangsquasigruppe, so liefert das dritte 6-Tupel aus (2.1.9) wegen Satz 2.4.1 für die parastrophen Operationen $\pi_i(/)$ mit $i = 1, 2, 6, 5, 4, 3$ der Reihe nach (U), (N_r), (N_l), (N_r), (N_l) und (U). Will man dieses Resultat dagegen wieder mittels der äquivalenten Gleichungen (2.1.6) gewinnen, so müssen wir die Unipotenz $x \circ x = y \circ y$ einer Quasigruppe $(Q, \circ)$[1]) zunächst in geeigneter Weise umformen. Da $x \circ x$ von x selbst unabhängig ist, gibt es ein Element $c \in Q$, so daß $x \circ x = c$ für alle $x \in Q$ gilt. Umgekehrt kann man im Fall der Existenz eines solchen Elementes $c \in Q$ mit $x \circ x = c$ für alle $x \in Q$ wieder auf die Unipotenz von $\circ$ schließen. Mit (2.1.6) folgt mithin

$$x \circ x = c \Leftrightarrow x \backslash c = x \Leftrightarrow c/x = x \Leftrightarrow x/^* c = x$$
$$\Leftrightarrow c \backslash^* x = x \Leftrightarrow x \circ^* x = c,$$

d. h., c ist in $(Q, \backslash)$ und $(Q, /^*)$ rechts-neutrales, in $(Q, /)$ und $(Q, \backslash^*)$ aber links-neutrales Element. Damit haben wir die Schlußfolgerung aus Satz 2.4.1 bestätigt:

2.4.2. Korollar. *Für die Parastrophen einer unipotenten Quasigruppe $(Q, \circ)$ gilt:*

$(Q, \circ)$	$(Q, \backslash)$	$(Q, /)$	$(Q, /^*)$	$(Q, \backslash^*)$	$(Q, \circ^*)$
(U)	(N_r)	(N_l)	(N_r)	(N_l)	(U)

Aus dem bisher Bewiesenen folgt schließlich, daß die Parastrophen einer unipotenten Loop stets wieder unipotente Loops sind. Diese Parastrophen brauchen deshalb jedoch nicht zusammenzufallen — vgl. z. B. die Illustration des Falles $p = 2$ zu Satz 2.3.1.

Die eben skizzierten Abhängigkeiten zwischen Eigenschaften einer Operation und zugehörigen Operationen legen die Frage nahe, ob nicht auch im Fall solcher Identitäten wie z. B. der Assoziativität oder der Kommutativität eine Aussage über ihr Verhalten beim Übergang zu parastrophen Operationen gemacht werden kann. Wir werden im folgenden sehen, daß in jeder parastrophen Quasigruppe $(Q, \pi(\circ))$ von $(Q, \circ)$, in der nach Voraussetzung (A) oder (K) gelte, wieder eine Identität erfüllt ist. Um das einzusehen, stützen wir uns auf eine Methode von S. K. Stein [164].

[1]) Wir betrachten jetzt also wieder — wie üblich — $\circ$ als Ausgangsoperation.

Ihr liegt der Gedanke zugrunde, daß sich eine Identität durch eine Kette voneinander abhängiger Gleichungen, die alle die Form $a \circ b = c$ haben, gleichwertig ersetzen läßt.

Es sei also $(Q, \circ)$ eine assoziative Quasigruppe, d. h. eine Gruppe. Dann lösen wir die Assoziativität

$$(A) \qquad \bigwedge_{x,y,z \in Q} (x \circ y) \circ z = x \circ (y \circ z)$$

schrittweise in Gleichungen auf: $(x \circ y) \circ z = a \circ z = b$ und $x \circ (y \circ z) = x \circ c = b$. Damit können wir (A) durch die folgende *Bedingung*, wie S. K. Stein eine solche Gleichungskette nennt, ersetzen:

$$\bigwedge_{\substack{x,y,z \in Q \\ a,b,c \in Q}} (x \circ y = a \wedge a \circ z = b \wedge y \circ z = c \Rightarrow x \circ c = b).$$

Umgekehrt können wir aus der Gültigkeit dieser Implikation auch wieder auf die Assoziativität der Operation $\circ$ schließen. Nachdem wir die Assoziativität derart in Gleichungen zerlegt haben, bereitet es keine Mühe mehr, zu den gewünschten parastrophen Operationen von $\circ$ überzugehen. Wegen (2.1.6) nimmt die obige Gleichungskette nämlich jeweils folgende Gestalt an:
Für alle $x, y, z, a, b, c \in Q$ gilt

$(Q, \backslash)$: $\quad x\backslash a = y \wedge a\backslash b = z \wedge y\backslash c = z \Rightarrow x\backslash b = c,$

$(Q, /)$: $\quad a/y = x \wedge b/z = a \wedge c/z = y \Rightarrow b/c = x,$

$(Q, /^*)$: $\quad y/^*a = x \wedge z/^*b = a \wedge z/^*c = y \Rightarrow c/^*b = x,$

$(Q, \backslash^*)$: $\quad a\backslash^*x = y \wedge b\backslash^*a = z \wedge c\backslash^*y = z \Rightarrow b\backslash^*x = c,$

$(Q, \circ^*)$: $\quad y \circ^* x = a \wedge z \circ^* a = b \wedge z \circ^* y = c \Rightarrow c \circ^* x = b.$

Durch Gleich- und Einsetzen erhalten wir wieder Identitäten, die wir die zu (A) *parastrophen Identitäten* nennen. Beispielsweise finden wir $b/c = a/y$ für die Operation $/$, also $b/c = (b/z)/(c/z)$. Das ist die Identität der *Rechts-Transitivität*. Entsprechend erhalten wir, daß neben der π_6-Parastrophen $/$ auch die π_4-Parastrophe $\backslash^*$ rechts-transitiv ist. Die π_2-Parastrophe $\backslash$ und die π_5-Parastrophe $/^*$ sind links-transitiv,[1]) während die π_3-Parastrophe $\circ^*$ assoziativ ist.[2]) Im letzteren Fall erhalten wir also wieder eine Gruppe.

---

[1]) Man kann von den vier Gleichungen der Bedingung in $(Q, \circ)$ drei beliebige auswählen, um auf die verbleibende vierte Gleichung schließen zu können. Vertauschen wir beispielsweise die zweite und vierte Gleichung miteinander, so erhalten wir $x \circ y = a \wedge x \circ c = b \wedge y \circ z = c \Rightarrow a \circ z = b$. Angenommen, es sei $a \circ z = b'$. Dann folgt aufgrund der ursprünglichen Implikation $x \circ c = b'$. Nach Voraussetzung ist jedoch $x \circ c = b$, so daß $b' = b$ gelten muß. Anstelle der oben angegebenen Bedingung für $(Q, \backslash)$ ergibt sich damit die Implikation $x\backslash a = y \wedge x\backslash b = c \wedge y\backslash c = z \Rightarrow a\backslash b = z$, so daß wegen $a\backslash b = y\backslash c = (x\backslash a)\backslash(x\backslash b)$ die *Links-Transitivität* erfüllt ist. Da wir auf eine Definition des Begriffs „Bedingung" verzichtet haben, können wir hier die generelle Möglichkeit einer solchen Vertauschbarkeit der Gleichungen nicht beweisen. Man überzeuge sich aber, daß dies für jede Identität, die wir hier betrachten, erlaubt ist.

[2]) Wegen (2.1.4) ergeben sich die π_2-, π_4- und π_5-Parastrophen von (A) auch unmittelbar durch „Rückwärtslesen" der π_1-, π_2- bzw. π_6-Parastrophen von (A) — unter Beibehaltung der Klammerung.

Um die parastrophen Identitäten übersichtlicher bzw. „lesbarer" zu gestalten, benutzen wir in den folgenden Tabellen anstelle der Operationszeichen „$\circ$", „$\backslash$", „$/$", „$/^*$", „$\backslash^*$" und „$\circ^*$" einen Punkt und verabreden darüber hinaus die übliche Schreibweise zur Klammereinsparung, also xy statt $x \cdot y$.

2.4.3. Satz. *Die parastrophen Identitäten der Assoziativität sind:*

$(Q, \circ)$	$(Q, \backslash)$	$(Q, /)$	$(Q, /^*)$	$(Q, \backslash^*)$	$(Q, \circ^*)$
(A)	(L-T)	(R-T)	(L-T)	(R-T)	(A)
$xy \cdot z$	$xy \cdot xz$	$xy \cdot zy$	$xy \cdot xz$	$xy \cdot zy$	$xy \cdot z$
$= x \cdot yz$	$= yz$	$= xz$	$= yz$	$= xz$	$= x \cdot yz$

Hinter der Rechts-Transitivität (R-T) verbirgt sich ein bekanntes Rechengesetz der Subtraktion bzw. Division (als π_6-Parastrophe der Addition bzw. Multiplikation): $(x - y) - (z - y) = x - z$ und $(x : y) : (z : y) = x : z$. Aufgrund von Satz 2.4.3 wissen wir jetzt, daß sich diese Eigenschaft zwangsläufig ergibt; die Rechts-Transitivität der Subtraktion und Division folgt aus der Assoziativität der Addition bzw. Multiplikation. Daß die Umkehrung ebenfalls richtig ist, zeigen die folgenden Überlegungen.

Wegen $\pi_2(\pi_2(\circ)) = \pi_2(\backslash) = \circ$ ist die π_2-Parastrophe der Links-Transitivität gerade wieder die Assoziativität; (A) ist aber wegen $\pi_6(\pi_6(\circ)) = \pi_6(/) = \circ$ auch die π_6-Parastrophe von (R-T). Darüber hinaus folgt aus Satz 2.4.3 mit (2.1.9) für alle Parastrophen von (L-T) und (R-T) das

2.4.4. Korollar. *Die parastrophen Identitäten der Links- und Rechts-Transitivität sind:*

$(Q, \circ)$	$(Q, \backslash)$	$(Q, /)$	$(Q, /^*)$	$(Q, \backslash^*)$	$(Q, \circ^*)$
(L-T)	(A)	(L-T)	(R-T)	(A)	(R-T)
(R-T)	(R-T)	(A)	(A)	(L-T)	(L-T)

Daß die Assoziativität die π_2-Parastrophe der Links-Transitivität ist, veranlaßte S. K. Stein [164] zu der Feststellung, daß die Gruppentheorie zur Theorie der linkstransitiven Quasigruppen äquivalent ist. Erscheint eine solche Auffassung zunächst auch etwas ungewöhnlich, so ist sie letzten Endes doch die tiefere Ursache dafür, daß man Gruppen sowohl mit Hilfe der Links-Transitivität als auch der Rechts-Transitivität axiomatisch charakterisieren kann. Eine solche Einführung des Gruppenbegriffs ist auch verschiedentlich diskutiert worden (vgl. 5.2.).

In einer kommutativen Quasigruppe $(Q, \circ)$ fallen wegen (2.1.4) die Operationen $\circ$ und $\circ^*$ zusammen. Darüber hinaus gilt wegen Satz 2.3.1 auch stets $\backslash = /^*$ und $/ = \backslash^*$. Als parastrophe Identitäten der Kommutativität finden wir (nach der Methode von Stein) die *Rechts-* bzw. *Links-Systemregel*:

(R-SR) $(x\backslash y)\backslash y = x$

und

(L-SR) $x/(x/y) = y$.

2.4.5. Satz. *Die parastrophen Identitäten der Kommutativität sind:*

$(Q, \circ)\ (= (Q, \circ^*))$	$(Q, \backslash)\ (= (Q, /^*))$	$(Q, /)\ (= (Q, \backslash^*))$
(K)	(R-SR)	(L-SR)
$xy = yx$	$xy \cdot y = x$	$x \cdot xy = y$

Umgekehrt ist die Kommutativität auch eine parastrophe Identität sowohl der Links- als auch der Rechts-Systemregel. Unter Beachtung der Sätze 2.3.1 und 2.3.2 — $p = 3$ (a), (b) — ergibt sich nämlich

2.4.6. Korollar. *Die parastrophen Identitäten der Links- und Rechts-Systemregel sind:*

$(Q, \circ)\ (= (Q, \backslash))$	$(Q, /)\ (= (Q, /^*))$	$(Q, \backslash^*)\ (= (Q, \circ^*))$
(L-SR)	(K)	(R-SR)

und

$(Q, \circ)\ (= (Q, /))$	$(Q, \backslash)\ (= (Q, \backslash^*))$	$(Q, /^*)\ (= (Q, \circ^*))$
(R-SR)	(K)	(L-SR)

Damit lassen sich jetzt alle in den Sätzen 2.3.1 und 2.3.2 sowie Korollar 2.3.3 auftretenden Fälle für die Anzahl p der voneinander verschiedenen Parastrophen einer Quasigruppe $(Q, \circ)$ allein mit Hilfe der Operation $\circ$ charakterisieren (vgl. auch Ch. C. Lindner und D. Steedley [106]). Betrachten wir etwa den Fall $p = 3$ (a). Wegen Satz 2.3.2 ist die Operation $\circ$ nicht kommutativ, wohl aber ihre Rechtsdivision $/$. Dann muß infolge von Korollar 2.4.6 die Ausgangsoperation $\circ$ die Links-Systemregel erfüllen. Umgekehrt folgt aus der Links-Systemregel für $\circ$ die Kommutativität von $/$. Ist $\circ$ selbst nicht kommutativ, so können wir wieder mit Satz 2.3.2 auf den Fall $p = 3$ (a) zurückschließen. In analoger Weise lassen sich auch für die restlichen Fälle $p = 1, 2, 3, 6$ die Umkehroperationen eliminieren:

2.4.7. Satz. *Es sei $(Q, \circ)$ eine Quasigruppe.*

a) *Der Fall $p = 3$ (a) liegt genau dann vor, wenn die Operation $\circ$ die Links-Systemregel erfüllt, aber nicht kommutativ ist.*

b) *Der Fall $p = 3$ (b) liegt genau dann vor, wenn die Operation $\circ$ die Rechts-Systemregel erfüllt, aber nicht kommutativ ist.*

c) *Der Fall $p = 3$ (c) liegt genau dann vor, wenn die Operation $\circ$ kommutativ ist, aber keine der beiden Systemregeln erfüllt.*

d) *Der Fall $p = 2$ liegt genau dann vor, wenn die Operation $\circ$ halbsymmetrisch, aber nicht kommutativ ist.*

e) *Der Fall $p = 1$ liegt genau dann vor, wenn die Operation $\circ$ die Kommutativität und beide*[1]) *Systemregeln erfüllt (oder wenn die Operation $\circ$ kommutativ und halbsymmetrisch ist).*[2])

[1]) Es genügt natürlich schon die Annahme, daß $\circ$ neben (K) eine der beiden Systemregeln erfüllt. Dann gilt auch die zweite Systemregel (vgl. die Ausführungen im Anschluß an Korollar 2.4.9).

[2]) Wie eng die Identitäten (K), (L-SR), (R-SR), (L-HS) und (R-HS) über den hier durch die Parastrophie begrenzten Rahmen hinaus miteinander zusammenhängen, wird in 5.2. gezeigt.

f) *Der Fall $p = 6$ liegt genau dann vor, wenn die Operation $\circ$ keine der Identitäten* (K), (L-SR), (R-SR), (L-HS) *oder* (R-HS) *erfüllt.*

Beispiele für Operationen, die die Links- oder Rechts-Systemregel erfüllen, haben wir demnach bereits mehrfach kennengelernt. Bemerkenswert ist in diesem Zusammenhang sicher, daß in der Kleinschen Vierergruppe $\mathfrak{A}_4$ (vgl. $\circ_1$ in Aufgabe 2.2.2.2) neben der Assoziativität und Kommutativität auch deren sämtliche parastrophen Identitäten, neben (A) und (K) also (R-T), (L-T), (L-SR) und (R-SR), erfüllt sind. Darüber hinaus ist sie auch halbsymmetrisch.

Wir werden sehen, daß man mit Hilfe der Steinschen Methode auf diese Weise in sehr vielen Fällen Verbindungen herstellen kann zwischen Identitäten in einer Quasigruppe und Identitäten in den zugehörigen parastrophen Quasigruppen. Es kann jedoch vorkommen, daß man auf diese Weise keine parastrophe Identität erhält, so z. B. im Fall der Links- und Rechts-Selbstdistributivität.

Bei der Auswahl der Identitäten, für die wir im folgenden die zugehörigen parastrophen Identitäten angeben werden, haben wir uns auf eine Liste von insgesamt 46 Identitäten gestützt (vgl. J. DÉNES und A. D. KEEDWELL [44] sowie A. SADE [152]). Viele dieser Identitäten sind in der Literatur unter z. T. unterschiedlichen Gesichtspunkten untersucht worden. Wir werden auf einige dieser Identitäten ebenfalls wieder zurückkommen (vgl. auch Aufgabe 1.7.9.1).

Auswahl einiger Identitäten[1])

(1)	$x \circ x = x$	(Idempotenz),
(2)	$x \circ x = y \circ y$	(Unipotenz),
(3)	$x \circ y = y \circ x$	(Kommutativität),
(4)	$x \circ (x \circ y) = y$	(L-Systemregel),
(5)	$(x \circ y) \circ y = x$	(R-Systemregel),
(6)	$x \circ (y \circ x) = y$	(L-Halbsymmetrie),
(7)	$(x \circ y) \circ x = y$	(R-Halbsymmetrie),
(8)	$x \circ (x \circ y) = y \circ x$	(Stein-I-Identität),
(9)	$x \circ (x \circ y) = (x \circ y) \circ y$	(Schröder-I-Identität),
(10)	$x \circ (y \circ x) = (x \circ y) \circ x$	(Elastizität),
(11)	$x \circ (y \circ x) = (y \circ x) \circ y$	(Stein-II-Identität),
(12)	$(x \circ y) \circ (y \circ x) = x$	(Schröder-II-Identität),
(13)	$(x \circ y) \circ (y \circ x) = y$	(Stein-III-Identität),
(14)	$x \circ (y \circ z) = (x \circ y) \circ z$	(Assoziativität),
(15)	$x \circ (y \circ z) = z \circ (x \circ y)$	(Zyklische Assoziativität),
(16)	$x \circ (y \circ z) = (z \circ x) \circ y$,	
(17)	$x \circ (y \circ z) = z \circ (y \circ x)$	(Abel-Graßmann-Identität),
(18)	$x \circ (y \circ z) = (y \circ x) \circ z$,	
(19)	$x \circ (y \circ z) = y \circ (x \circ z)$	(L-Permutabilität),
(20)	$(x \circ y) \circ z = (x \circ z) \circ y$	(R-Permutabilität),

[1]) Es sei auch noch einmal daran erinnert, daß wir an früherer Stelle (vgl. die Fußnote auf S. 24) vereinbart haben, daß in einer Identität die auftretenden Variablen stets die gesamte Trägermenge durchlaufen.

(21) $(x \circ y) \circ z = x \circ (z \circ y)$ (Eingewandtes Produkt),

(22) $(x \circ y) \circ (x \circ z) = y \circ z$ (L-Transitivität),

(23) $(x \circ y) \circ (z \circ y) = x \circ z$ (R-Transitivität),

(24) $(x \circ y) \circ (y \circ z) = x \circ z$ (M-Transitivität),

(25) $(x \circ y) \circ (x \circ z) = z \circ y$ (L-Schweitzer-Identität),

(26) $(x \circ y) \circ (z \circ y) = z \circ x$ (R-Schweitzer-Identität),

(27) $x \circ (y \circ z) = (x \circ y) \circ (x \circ z)$ (L-Selbstdistributivität),

(28) $(x \circ y) \circ z = (x \circ z) \circ (y \circ z)$ (R-Selbstdistributivität),

(29) $x \circ (y \circ z) = (x \circ y) \circ (z \circ x)$ (L-abelsche Selbstdistributivität),

(30) $(x \circ y) \circ z = (z \circ x) \circ (y \circ z)$ (R-abelsche Selbstdistributivität),

(31) $x \circ \big(y \circ (z \circ x)\big) = z \circ y$ (Tarski-Identität),

(32) $x \circ \big((y \circ z) \circ (y \circ x)\big) = z$ (Neumann-Identität),

(33) $(x \circ y) \circ (u \circ v) = (x \circ u) \circ (y \circ v)$ (Bisymmetrie).

Einige dieser Identitäten treten in der Literatur unter verschiedenen Namen auf.

In Tabelle 5 stellen wir für die von uns ausgewählten Identitäten die zugehörigen parastrophen Identitäten zusammen. Die Identitäten (2), (3), (4), (5), (14), (22) und

Tabelle 5

$(Q, \circ)$	$(Q, \backslash)$	$(Q, /)$
$xx = x$ Idempotenz	$xx = x$ Idempotenz	$xx = x$ Idempotenz
$x \cdot yx = y$ $xy \cdot x = y$ Halbsymmetrie	$x \cdot yx = y$ $xy \cdot x = y$ Halbsymmetrie	$x \cdot yx = y$ $xy \cdot x = y$ Halbsymmetrie
$x \cdot xy = yx$ Stein-I	$(xy)\,(xy \cdot y) = x$ $(x \cdot xy)\,y = x$	$x(y \cdot yx) = yx$
$x \cdot xy = xy \cdot y$ Schröder-I	$x \cdot xy = xy \cdot y$ Schröder-I	$x \cdot xy = xy \cdot y$ Schröder-I
$x \cdot yx = xy \cdot x$ Elastizität	$(xy \cdot x)\,(xy \cdot y) = xy$	$(x \cdot xy)\,(y \cdot xy) = xy$
$x \cdot yx = yx \cdot y$ $xy \cdot x = y \cdot xy$ Stein-II	$(xy \cdot x)\,y = x$ $(xy \cdot y)\,(xy) = x$	$(xy)\,(x \cdot xy) = y$ $x(y \cdot xy) = y$
$xy \cdot yx = x$ Schröder-II	$xy \cdot yx = x$ Schröder-II	$xy \cdot yx = x$ Schröder-II
$xy \cdot yx = y$ Stein-III	$(xy)\,(y \cdot xy) = x$ $(x \cdot yx)\,y = x$	$(xy \cdot x)\,(xy) = y$ $x(yx \cdot y) = y$
$x \cdot yz = z \cdot xy$ $x \cdot yz = y \cdot zx$ Zyklische Assoziativität	$(xy \cdot xz)\,z = y$ $(x \cdot yz)\,(xz) = y$ $x((y \cdot xz)\,z) = y$	$x(y \cdot zx) = zy$ Tarski

(23), die weiter vorn ausführlich behandelt worden sind, werden dabei jedoch nicht noch einmal berücksichtigt. Für einige der verbleibenden Identitäten, nämlich (1), (12), (13), (16), (24), (27) und (33), hat bereits S. K. STEIN [164] die zugehörigen Parastrophen angegeben (vgl. auch V. D. BELOUSOV [20] sowie J. DÉNES und A. D. KEEDWELL [44]).

Um Tabelle 5 übersichtlicher gestalten zu können, setzen wir wieder anstelle der jeweiligen parastrophen Operation $\pi(\circ)$ einen Punkt als Operationszeichen und beachten die Regeln der Klammereinsparung. Darüber hinaus weisen wir darauf hin, daß Identitäten, die im selben Kästchen stehen, äquivalent sind. In vielen Fällen ergibt sich die Gleichwertigkeit solcher Identitäten einfach durch Umbenennung der Variablen (vgl. z. B. (11)). Die Äquivalenz der Links- und Rechts-Halbsymmetrie läßt sich dagegen folgendermaßen beweisen: $xy \cdot x = (xy) \cdot (y \cdot xy) = y$ und $x \cdot yx = (yx \cdot y) \cdot (yx) = y$. Dabei haben wir zunächst zweimal die Identität (6) und dann zweimal die Identität (7) angewendet. Es genügt also künftig, von Halbsymmetrie zu sprechen, auch wenn wir die Identitäten (6) und (7) weiterhin mit (L-HS) bzw. (R-HS) bezeichnen werden.

2.4.8. Satz.[1]) *Es gelten die in Tabelle 5 angegebenen Identitäten.*

[1]) Ein freies Feld bedeutet, daß die in dieser parastrophen Quasigruppe $(Q, \pi(\circ))$ geltende Bedingung zu keiner Identität führt, in der allein die Operation $\pi(\circ)$ auftritt.

$(Q, /^*)$	$(Q, \backslash^*)$	$(Q, \circ^*)$
$xx = x$ Idempotenz	$xx = x$ Idempotenz	$xx = x$ Idempotenz
$x \cdot yx = y$ $xy \cdot x = y$ Halbsymmetrie	$x \cdot yx = y$ $xy \cdot x = y$ Halbsymmetrie	$x \cdot yx = y$ $xy \cdot x = y$ Halbsymmetrie
$(xy \cdot y)\, x = xy$	$(x \cdot xy)\,(xy) = y$ $x(xy \cdot y) = y$	$xy \cdot y = yx$
$x \cdot xy = xy \cdot y$ SCHRÖDER-I	$x \cdot xy = xy \cdot y$ SCHRÖDER-I	$x \cdot xy = xy \cdot y$ SCHRÖDER-I
$(xy \cdot x)\,(xy \cdot y) = xy$	$(x \cdot xy)\,(y \cdot xy) = xy$	$x \cdot yx = xy \cdot x$ Elastizität
$(xy \cdot x)\, y = x$ $(xy \cdot y)\,(xy) = x$	$(xy)\,(x \cdot xy) = y$ $x(y \cdot xy) = y$	$x \cdot yx = yx \cdot y$ $xy \cdot x = y \cdot xy$ STEIN-II
$xy \cdot yx = x$ SCHRÖDER-II	$xy \cdot yx = x$ SCHRÖDER-II	$xy \cdot yx = x$ SCHRÖDER-II
$(xy)\,(y \cdot xy) = x$ $(x \cdot yx)\, y = x$	$(xy \cdot x)\,(xy) = y$ $x(yx \cdot y) = y$	$xy \cdot yx = y$ STEIN-III
$(xy \cdot z)\, x = zy$	$x(xy \cdot zy) = z$ $(xy)\,(x \cdot y) = z$ $(x(xy \cdot z))\, y = z$	$xy \cdot z = yz \cdot x$ $xy \cdot z = zx \cdot y$

Tabelle 5 (Fortsetzung)

$(Q, \circ)$	$(Q, \backslash)$	$(Q, /)$
$x \cdot yz = zx \cdot y$ $xy \cdot z = y \cdot zx$ (16)	$(xy \cdot zy)\, x = z$ $(xy)\,(zx \cdot y) = z$ $((xy \cdot z)\, x)\, y = z$	$x(yz \cdot yx) = z$ $(x \cdot yz)\,(xy) = z$ $x(y(z \cdot xy)) = z$ NEUMANN
$x \cdot yz = z \cdot yx$ ABEL-GRASSMANN	$x((x \cdot yz)\, z) = y$ $(x(xy \cdot z))\, z = y$	$x \cdot yz = z \cdot yx$ ABEL-GRASSMANN
$x \cdot yz = yx \cdot z$ $xy \cdot z = y \cdot xz$ (18)	$x(xy \cdot z) = yz$	$xy \cdot xz = zy$ L-SCHWEITZER
$x \cdot yz = y \cdot xz$ L-Permutabilität	$x \cdot yz = y \cdot xz$ L-Permutabilität	
$xy \cdot z = xz \cdot y$ R-Permutabilität		$xy \cdot z = xz \cdot y$ R-Permutabilität
$xy \cdot z = x \cdot zy$ $x \cdot yz = xz \cdot y$ Eingewandtes Produkt	$xy \cdot zy = zx$ R-SCHWEITZER	$(x \cdot yz)\, z = xy$
$xy \cdot yz = xz$ M-Transitivität	$xy \cdot yz = xz$ M-Transitivität	$xy \cdot yz = xz$ M-Transitivität
$xy \cdot xz = zy$ L-SCHWEITZER	$(x \cdot yz)\, z = xy$	$x \cdot yz = yx \cdot z$ $xy \cdot z = y \cdot xz$ (18)
$xy \cdot zy = zx$ R-SCHWEITZER	$xy \cdot z = x \cdot zy$ $x \cdot yz = xz \cdot y$ Eingewandtes Produkt	$x(xy \cdot z) = yz$
$x \cdot yz = xy \cdot xz$ L-Selbstdistributivität	$x \cdot yz = xy \cdot xz$ L-Selbstdistributivität	
$xy \cdot z = xz \cdot yz$ R-Selbstdistributivität		$xy \cdot z = xz \cdot yz$ R-Selbstdistributivität
$x \cdot yz = xy \cdot zx$ L-abelsche Selbstdistributivität	$(xy \cdot xz)\,(yz) = x$ $((x \cdot yz)\, y)\,((x \cdot yz)\, z) = x$	$(xy)\,(z(y \cdot xz)) = xz$
$xy \cdot z = zx \cdot yz$ R-abelsche Selbstdistributivität	$((xy \cdot z)\, x)\,(zy) = xy$	$(xy)\,(xz \cdot yz) = z$ $(x(xy \cdot z))\,(y(xy \cdot z)) = z$
$x(y \cdot zx) = zy$ TARSKI	$x(xy \cdot zy) = z$ $(xy)\,(zx \cdot y) = z$ $(x(xy \cdot z))\, y = z$	$x \cdot yz = z \cdot xy$ $x \cdot yz = y \cdot zx$ Zyklische Assoziativität
$x(yz \cdot yx) = z$ $(x \cdot yz)\,(xy) = z$ $x(y(z \cdot xy)) = z$ NEUMANN	$x(yz \cdot yx) = z$ $(x \cdot yz)\,(xy) = z$ $x(y(z \cdot xy)) = z$ NEUMANN	$x \cdot yz = zx \cdot y$ $xy \cdot z = y \cdot zx$ (16)
$xy \cdot uv = xu \cdot yv$ Bisymmetrie	$xy \cdot uv = xu \cdot yv$ Bisymmetrie	$xy \cdot uv = xu \cdot yv$ Bisymmetrie

$(Q, /^*)$	$(Q, \backslash^*)$	$(Q, \circ^*)$
$(xy \cdot zy)\, x = z$ $(xy)\,(zx \cdot y) = z$ $((xy \cdot z)\, x)\, y = z$	$x(yz \cdot yx) = z$ $(x \cdot yz)\,(xy) = z$ $x(y(z \cdot xy)) = z$ NEUMANN	$x \cdot yz = zx \cdot y$ $xy \cdot z = y \cdot zx$ (16)
$xy \cdot z = zy \cdot x$	$x((x \cdot yz)\, z) = y$ $(x(xy \cdot z))\, z = y$	$xy \cdot z = zy \cdot x$
$xy \cdot zy = zx$ R-SCHWEITZER	$(x \cdot yz)\, z = xy$	$x \cdot yz = xz \cdot y$ $xy \cdot z = x \cdot zy$ Eingewandtes Produkt
	$xy \cdot z = xz \cdot y$ R-Permutabilität	$xy \cdot z = xz \cdot y$ R-Permutabilität
$x \cdot yz = y \cdot xz$ L-Permutabilität		$x \cdot yz = y \cdot xz$ L-Permutabilität
$x(xy \cdot z) = yz$	$xy \cdot xz = zy$ L-SCHWEITZER	$x \cdot yz = yx \cdot z$ $xy \cdot z = y \cdot xz$ (18)
$xy \cdot yz = xz$ M-Transitivität	$xy \cdot yz = xz$ M-Transitivität	$xy \cdot yz = xz$ M-Transitivität
$xy \cdot z = x \cdot zy$ $x \cdot yz = xz \cdot y$ Eingewandtes Produkt	$x(xy \cdot z) = yz$	$xy \cdot zy = zx$ R-SCHWEITZER
$(x \cdot yz)\, z = xy$	$x \cdot yz = yx \cdot z$ $xy \cdot z = y \cdot xz$ (18)	$xy \cdot xz = zy$ L-SCHWEITZER
	$xy \cdot z = xz \cdot yz$ R-Selbstdistributivität	$xy \cdot z = xz \cdot yz$ R-Selbstdistributivität
$x \cdot yz = xy \cdot xz$ L-Selbstdistributivität		$x \cdot yz = xy \cdot xz$ L-Selbstdistributivität
$((xy \cdot z)\, x)\,(zy) = xy$	$(xy)\,(xz \cdot yz) = z$ $(x(xy \cdot z))\,(y(xy \cdot z)) = z$	$xy \cdot z = zx \cdot yz$ R-abelsche Selbstdistributivität
$(xy \cdot xz)\,(yz) = x$ $((x \cdot yz)\, y)\,((x \cdot yz)\, z) = x$	$(xy)\,(z(y \cdot xz)) = xz$	$x \cdot yz = xy \cdot zx$ L-abelsche Selbstdistributivität
$xy \cdot z = yz \cdot x$ $xy \cdot z = zx \cdot y$	$(xy \cdot xz)\, z = y$ $(x \cdot yz)\,(xz) = y$ $x((y \cdot xz)\, z) = y$	$(xy \cdot z)\, x = zy$
$x \cdot yz = zx \cdot y$ $xy \cdot z = y \cdot zx$ (16)	$(xy \cdot zy)\, x = z$ $(xy)\,(zx \cdot y) = z$ $((xy \cdot z)\, x) = z$	$(xy \cdot zy)\, x = z$ $(xy)\,(zx \cdot y) = z$ $((xy \cdot z)\, x)\, y = z$
$xy \cdot uv = xu \cdot yv$ Bisymmetrie	$xy \cdot uv = xu \cdot yv$ Bisymmetrie	$xy \cdot uv = xu \cdot yv$ Bisymmetrie

2.4.9. Korollar. *Die Idempotenz, die Halbsymmetrie, die Schröder-I-Identität, die Schröder-II-Identität, die Mittel-Transitivität und die Bisymmetrie sind parastrophieinvariante Identitäten, d. h., aus ihrer Gültigkeit in* $(Q, \circ)$ *folgt ihre Gültigkeit in allen parastrophen Quasigruppen* $(Q, \pi_i(\circ))$ *mit* $i = 1, 2, \ldots, 6$.

Nicht invariant unter Parastrophie sind dagegen die Kommutativität, die Links- oder die Rechts-Systemregel (Satz 2.4.5, Korollar 2.4.6). Interessant ist jedoch, daß je zwei dieser Identitäten, sofern sie gleichzeitig in einer Quasigruppe $(Q, \circ)$ erfüllt sind, auch in allen parastrophen Quasigruppen $(Q, \pi(\circ))$ gelten.

Fall 1: In $(Q, \circ)$ seien (K) und (L-SR) erfüllt. Dann fallen wegen Satz 2.4.5 und Korollar 2.4.6 sogar alle sechs parastrophen Operationen zusammen, d. h., der Fall $p = 1$ liegt vor.

Fall 2: (K) und (R-SR) ziehen mit derselben Begründung den Fall $p = 1$ nach sich.

Fall 3: Aus (L-SR) und (R-SR) folgt wegen Korollar 2.4.6 ebenfalls $\circ = \backslash = / = /^* = \backslash^* = \circ^*$.

Mit Satz 2.4.7e) muß dann auch stets die dritte dieser zueinander parastrophen Identitäten gelten. Aus je zwei der Identitäten (K), (L-SR) und (R-SR) folgt also die dritte.[1])

2.4.10. Definition. Sind in einem Gruppoid $(G, \circ)$ die Identitäten (K) und (L-SR) erfüllt, so sagen wir auch, das Gruppoid $(G, \circ)$ (bzw. die Operation $\circ$) ist *totalsymmetrisch*:

(TS) $\quad :\Leftrightarrow$ (K) $\wedge$ (L-SR).

Als Fazit unserer Überlegungen ergibt sich nun einerseits (TS) $\Leftrightarrow$ (K) $\wedge$ (L-SR) $\Leftrightarrow$ (K) $\wedge$ (R-SR) $\Leftrightarrow$ (L-SR) $\wedge$ (R-SR) und andererseits

2.4.11. Satz. *Die Totalsymmetrie ist parastrophieinvariant.*

Im Unterschied etwa zur Bisymmetrie (BS), die ja ebenfalls parastrophieinvariant ist, fallen aber infolge der Totalsymmetrie (TS) alle parastrophen Operationen zusammen, so daß die folgenden Gleichungen allesamt äquivalent sind (vgl. (2.1.6)):

$$x \circ y = z \Leftrightarrow x \circ z = y \Leftrightarrow z \circ y = x \Leftrightarrow y \circ z = x$$
$$\Leftrightarrow z \circ x = y \Leftrightarrow y \circ x = z.$$

Dieser vollständigen Symmetrie verdankt die Totalsymmetrie ihren Namen (vgl. R. H. Bruck [29]).

Zu einigen schulmathematischen Potenzen

Die Vielfalt an Operationen, die etwa ab Klasse 5 vorliegt und bis hin zu Klasse 12 ständig erweitert wird (vgl. I. Lehmann [100]), rechtfertigt einerseits ein bewußtes Einbeziehen von (ausgewählten) Operationen in den Unterricht und ermöglicht ande-

[1]) Man beachte, daß wir uns dabei stets in Quasigruppen befinden. An späterer Stelle (Korollar 5.2.5) werden wir zeigen, daß wir auf diese Voraussetzung verzichten können.

rerseits zugleich, einige Aspekte zur Weiterentwicklung des Mathematikunterrichts im Sinne seiner inhaltlichen Ausgestaltung aufzugreifen. Gerade die Kenntnis zueinander parastropher Identitäten bietet nun besonders interessante Möglichkeiten, bestimmte Zusammenhänge und Querverbindungen zwischen Operationen und zugehörigen Umkehroperationen[1]) bewußt zu machen.

Am Beispiel einer Gegenüberstellung der vier Grundrechenarten sei kurz skizziert, welche neuen Erkenntnisse Schüler gewinnen können, ohne daß deshalb neuer Stoff einzuführen wäre. Während wohl die meisten Schüler für Addition und Multiplikation eine Reihe von Eigenschaften aufzählen können, sind sicher nur noch sehr wenige in der Lage, in gleicher Weise bezüglich der Subtraktion bzw. bezüglich der Division Auskunft zu geben. Interessant dabei ist, daß die folgenden Eigenschaften den Schülern dem Inhalt nach bekannt sind:

$(\mathbf{Z}, +)$ $(\mathbf{R}^*, \cdot)$	$(\mathbf{Z}, -)$ $(\mathbf{R}^*, :)$	Nachweis der Gleichwertigkeit
Kommutativität	L-Systemregel	Satz 2.4.5
Assoziativität	R-Transitivität	Satz 2.4.3
Bisymmetrie	Bisymmetrie	Satz 2.4.8, Kor. 2.4.9
rechts-neutrales Element	rechts-neutrales Element	Satz 2.4.1
links-neutrales Element	Unipotenz	Satz 2.4.1

Was den Schülern ungewohnt erscheinen wird, sie vielleicht sogar überraschen dürfte, ist dabei die Tatsache, daß die hier genannten Eigenschaften der Subtraktion und Division aus bestimmten Eigenschaften der Addition bzw. Multiplikation folgen. Darüber hinaus sind einander entsprechende Eigenschaften sogar in dem Sinne gleichwertig, daß es z. B. gleichgültig ist, ob man von der Assoziativität und Kommutativität einer (Quasigruppen-) Operation $\circ$ oder aber von der Rechts-Transitivität und Links-Systemregel ihrer Umkehroperation / spricht (vgl. I. Lehmann [103]).

Daß Operation und zugehörige Umkehroperation dieselbe Eigenschaft haben können, verdient sicher besondere Aufmerksamkeit. Da Addition und Multiplikation bisymmetrisch sind, müssen wegen der Parastrophieinvarianz der Bisymmetrie auch Subtraktion und Division bisymmetrisch sein.[2]) Ein Anreiz für eine kleine Beweisführung ergibt sich aus der folgenden Feststellung. Da die Bisymmetrie einerseits aus der Assoziativität und Kommutativität folgt, andererseits diese beiden Rechengesetze (A) und (K) für die Umkehroperation / von $\circ$ die Identitäten (R-T) bzw. (L-SR) nach sich ziehen, liegt die Vermutung nahe, daß auch aus der R-Transi-

[1]) Man beachte dabei, daß im Mathematikunterricht anstelle der Links- und Rechts-Division $\backslash$ bzw. / die parastrophen Operationen / und $\backslash^*$ als Umkehroperationen von $:$ aufgefaßt werden (vgl. 1.3. und 2.1).

[2]) In diesem Zusammenhang sei auf die Aufgabe 6B der XVI. Olympiade Junger Mathematiker (3. Stufe) hingewiesen, in der es im Prinzip um die Parastrophieinvarianz der Bisymmetrie geht.

tivität und der L-Systemregel die Bisymmetrie folgt. Das ist auch tatsächlich der Fall (vgl. Satz 5.2.11).

Schließlich wirft der Vergleich der vier Grundrechenarten auch solche Fragen auf, die ganz zwanglos zu einfachen und überschaubaren Axiomensystemen für (abelsche) Gruppen führen. So drängt sich anhand obiger Tabelle z. B. einerseits die Frage auf, welche Eigenschaften — neben (BS) und (N_r) — die Addition als eine abelsche Gruppenoperation auszeichnen, andererseits stellt sich die Frage, welche weiteren Eigenschaften — neben (BS) und (N_r) — der Subtraktion zukommen, so daß sie gerade eine Umkehroperation einer abelschen Gruppenoperation ist. Mit anderen Worten: Wie läßt sich eine abelsche Gruppe $(G, \circ)$ unmittelbar, d. h. allein anhand der Operation $\circ$, aber auch mittelbar, d. h. über die Umkehroperationen $\backslash$ bzw. $/$ mit Hilfe der Eigenschaften (BS) und (N_r) charakterisieren? Eine Antwort geben wir in Kapitel 5.

2.4.12. Aufgaben

1. Es sei $(Q, \circ)$ eine rechts-transitive Quasigruppe, in der die Links-Systemregel erfüllt ist. Man zeige, daß die Rechtsdivision eine abelsche Gruppenoperation ist.

2. Man zeige, daß die vier Grundrechenarten sämtlich die Identitäten (17), (20) und (33) erfüllen.

3. Es sei $Q = \{0, 1, 2, \ldots, n-1\}$. Man zeige, daß $(Q, \circ)$ mit $x \circ y := -(x + y) \bmod n$ eine totalsymmetrische Quasigruppe ist, und überzeuge sich, daß sämtliche parastrophen Operationen zusammenfallen.

4. Obwohl weder die Identität (27) noch die Identität (28) parastrophieinvariant ist (vgl. Satz 2.4.8), besitzt die *Selbstdistributivität* diese Eigenschaft. Dabei heißt eine Operation selbstdistributiv, wenn sie sowohl links- als auch rechts-selbstdistributiv ist. (Hinweis: Um zu zeigen, daß in $(Q, \backslash)$ die R-Selbstdistributivität gilt, weise man zunächst nach, daß die Linksdivision $\backslash$ rechts-distributiv bezüglich $\circ$ ist: $(x \circ y)\backslash z = (x\backslash z) \circ (y\backslash z)$ — vgl. V. D. Belousov [21].)

3. Homomorphie

Die zwei Gruppen der Ordnung 4, die Kleinsche Vierergruppe $\mathfrak{B}_4$ und die zyklische Gruppe $\mathfrak{Z}_4$,

$\circ$	1	2	3	4
1	1	2	3	4
2	2	1	4	3
3	3	4	1	2
4	4	3	2	1

$\square$	1	2	3	4
1	1	2	3	4
2	2	3	4	1
3	3	4	1	2
4	4	1	2	3

sind beide kommutativ, daneben besitzen sie aber auch voneinander verschiedene Eigenschaften. So ist die Gruppe $\mathfrak{B}_4$ z. B. unipotent, rechts- und links-transitiv, die Gruppe $\mathfrak{Z}_4$ hat dagegen keine dieser Eigenschaften. Sie läßt sich stattdessen, d. h. im Unterschied zur Kleinschen Vierergruppe, aus einem einzigen Element erzeugen. Algebraische Strukturen, die nicht in allen Eigenschaften übereinstimmen, können nicht isomorph sein. Das heißt in unserem Fall, es läßt sich keine 1–1-Abbildung zwischen beiden Trägermengen finden, die alle Eigenschaften der Operation $\circ$ auf die Operation $\square$ (bzw. umgekehrt) überträgt. Gerade die Erhaltung aller algebraischen Eigenschaften macht aber das Wesen der Isomorphie aus.

Neben den Isomorphismen beschäftigen wir uns in diesem Kapitel mit weiteren Morphismen, vor allem mit Homomorphismen. Die Homomorphie ist eine Abschwächung der Isomorphie.

Im Homomorphieansatz für Gruppen[1]) spielt der Begriff des Normalteilers eine entscheidende Rolle. Will man eine analoge Aussage für beliebige Gruppoide gewinnen, so benötigt man den Begriff der Kongruenzrelation. Da Kongruenzrelationen spezielle Äquivalenzrelationen sind, verabreden wir im folgenden Abschnitt einige häufig benutzte Begriffe und Sprechweisen, die eng mit Äquivalenzrelationen zusammenhängen (vgl. auch MfL Bd. 1, 2.5.).

3.1. Äquivalenzrelationen

Unter einer (*binären* oder *zweistelligen*) *Relation* R in einer Menge M verstehen wir eine Korrespondenz aus M in sich; d. h., R ist eine Teilmenge der Produktmenge $M \times M$: $R \subseteq M \times M$. Ist diese Relation reflexiv, symmetrisch und transitiv (vgl.

[1]) Vgl. MfL Bd. 3, 12.5.; in diesem Kapitel werden wir diesen Satz als Spezialfall des Homomorphiesatzes für Gruppoide erhalten.

MfL Bd. 1, 2.5.), so nennt man R eine *Äquivalenzrelation* in M. Die *Äquivalenzklasse* von x mod R ist die Menge $[x]_R$ aller derjenigen $y \in M$, die in Relation zu x stehen:

3.1.1. $[x]_R := \{y: y \in M \wedge yRx\}$.

Der Index R wird mitunter auch weggelassen.[1]) Zwei Äquivalenzklassen $[x]_R$ und $[y]_R$ sind genau dann gleich, wenn xRy gilt (zum Beweis siehe MfL Bd. 1, 2.5):

3.1.2. $[x]_R = [y]_R \Leftrightarrow xRy$.

Das heißt also auch, daß aus $y \in [x]_R$ sofort $[y]_R = [x]_R$ (und umgekehrt) folgt. Die Beziehung (3.1.2) nennt man in der Literatur auch oft „Abstraktionsprinzip", da man von den Unterschieden der Elemente innerhalb einer Äquivalenzklasse absieht und die Klassen selbst als neue Elemente versteht. Mit M/R bezeichnen wir die Menge aller Äquivalenzklassen mod R:

$$M/R := \{[x]_R: x \in M\}.$$

M/R heißt die *Faktormenge*[2]) von M nach R. Der Hauptsatz über Äquivalenzrelationen (vgl. MfL Bd. 1, 2.5.) besagt, daß jede Äquivalenzrelation R in M eine eindeutig bestimmte Klasseneinteilung in M (oder Zerlegung von M) bewirkt. Umgekehrt gehört zu jeder Klasseneinteilung (Zerlegung) von M eine eindeutig bestimmte Äquivalenzrelation R in M, so daß deren Äquivalenzklassen gerade die Elemente der Zerlegung sind. Die Zerlegung wird dabei durch die Faktormenge M/R realisiert, d. h., M/R erfüllt die für eine Klasseneinteilung charakteristischen Eigenschaften.

Die Eigenschaften $D(f) = M$ bzw. $W(f) = N$ einer Abbildung f aus einer Menge M in eine Menge N bringen wir durch die Sprechweise „von" bzw. „auf" zum Ausdruck. In gleicher Weise verfährt man auch bei Korrespondenzen, die nicht notwendig eindeutig zu sein brauchen. Bevor wir uns jetzt dem Kern einer Abbildung zuwenden, zeigen wir noch, daß sich diese Eigenschaften, aber auch die Eindeutigkeit und Eineindeutigkeit einer Korrespondenz f noch anders charakterisieren lassen.

3.1.3. Satz. *Eine Korrespondenz f aus M in N ist*

3.1.4. *eine Korrespondenz von M in $N \Leftrightarrow I_M \subseteq f^{-1} \bullet f$,*

3.1.5. *eine Korrespondenz aus M auf $N \Leftrightarrow I_N \subseteq f \bullet f^{-1}$,*

3.1.6. *eindeutig, also eine Abbildung $\Leftrightarrow f \bullet f^{-1} \subseteq I_N$,*

3.1.7. *eineindeutig $\Leftrightarrow f \bullet f^{-1} \subseteq I_N$ und $f^{-1} \bullet f \subseteq I_M$,*
wobei I_M und I_N die identischen Abbildungen von M bzw. N sind.

---

[1]) $[x]_R$ ist also gerade das volle Bild von $\{x\}$ bei R, d. h., es gilt $[x]_R = R(\{x\})$, wofür man kürzer auch $R(x)$ schreibt. Man beachte, daß $R(x)$ — im Unterschied zum vollen Bild von $\{x\}$ bei einer Abbildung f — i. allg. aus mehreren Elementen besteht.

[2]) G. Asser (MfL Bd. 1, 2.5.) nennt M/R das Restsystem von M nach R. An späterer Stelle (2.7.) benutzt er daneben auch das Bestimmungswort „Faktor-". J. Flachsmeyer und L. Prohaska (MfL Bd. 3) sprechen bei Gruppen von *Faktorgruppen*, bei Ringen und Körpern von *Restklassenringen* bzw. *-körpern*. In der Literatur findet man für M/R darüber hinaus noch die Benennung *Quotientenmenge*.

Beweis. a) Es sei f eine Korrespondenz von M in N. Dann ist $D(f) = M$, d. h., für jedes $x \in M$ existiert ein $y \in N$, so daß $(x, y) \in f$ und damit $(y, x) \in f^{-1}$ gilt. Nach der Definition der Nacheinanderausführung (NAF) von Korrespondenzen ist folglich $(x, x) \in f^{-1} \bullet f$ für jedes $x \in M$, also $I_M \subseteq f^{-1} \bullet f$. Der Rückschluß vervollständigt den Beweis von (3.1.4).

b) Es sei f eine Korrespondenz aus M auf N. Dann ist $W(f) = N$, d. h., für jedes $y \in N$ existiert ein $x \in M$, so daß $(x, y) \in f$ und damit $(y, x) \in f^{-1}$ gilt. Die NAF (erst f^{-1}, dann f) liefert dann $(y, y) \in f \bullet f^{-1}$ für jedes $y \in N$, also $I_N \subseteq f \bullet f^{-1}$. Der Rückschluß vervollständigt wieder den Beweis.

c) Es sei f eine Abbildung aus M in N und $(y_1, y_2) \in f \bullet f^{-1}$. Dann gibt es ein $x \in M$, so daß $(y_1, x) \in f^{-1}$ und $(x, y_2) \in f$ gilt. Folglich ist auch $(x, y_1) \in f$. Aufgrund der Eindeutigkeit von f folgt dann $y_1 = y_2$, so daß $(y_1, y_2) \in I_N$ gilt, also ist $f \bullet f^{-1} \subseteq I_N$. Umgekehrt gelte $f \bullet f^{-1} \subseteq I_N$, und es sei $(x, y_1) \in f$ und $(x, y_2) \in f$ für beliebige $x \in M$, $y_1, y_2 \in N$. Dann ist $(y_1, x) \in f^{-1}$, so daß die NAF $(y_1, y_2) \in f \bullet f^{-1} \subseteq I_N$ liefert. Also muß $y_1 = y_2$ sein, und f ist damit eindeutig.

d) Wegen (3.1.6) genügt es zu zeigen, daß die zu f inverse Korrespondenz f^{-1} aus N in M genau dann eindeutig ist, wenn $f^{-1} \bullet f \subseteq I_M$ erfüllt ist. Das folgt jedoch unmittelbar aus $f^{-1} \bullet (f^{-1})^{-1} = f^{-1} \bullet f$ und der Vertauschung von M und N gegenüber (3.1.6).

3.1.8. Korollar. *Eine Abbildung $f: M \to N$ ist*

3.1.9. *surjektiv* $\Leftrightarrow f \bullet f^{-1} = I_N$,

3.1.10. *injektiv* $\Leftrightarrow f^{-1} \bullet f = I_M$,

3.1.11. *bijektiv* $\Leftrightarrow f \bullet f^{-1} = I_N$ *und* $f^{-1} \bullet f = I_M$.

Der Beweis folgt unmittelbar aus Satz 3.1.3, wenn man berücksichtigt, daß hier immer $D(f) = M$, also $I_M \subseteq f^{-1} \bullet f$, und die Eindeutigkeit von f, also $f \bullet f^{-1} \subseteq I_N$, vorausgesetzt worden sind.

3.1.12. Definition. Gegeben sei eine Abbildung $f: M \to N$. Dann heißt $\ker f := f^{-1} \bullet f$ der *Kern der Abbildung* f.

3.1.13. Satz. *Der Kern einer Abbildung $f: M \to N$ ist eine Äqualenzrelation*[1] *in M.*

Beweis. Wegen $f^{-1} \bullet f \subseteq M \times M$ ist $\ker f$ eine Relation in M.

$\ker f$ ist reflexiv: Nach Voraussetzung gilt $D(f) = M$. Es existiert also für jedes $x \in M$ ein Bild y in N. Mit $(x, y) \in f$ gilt zugleich $(y, x) \in f^{-1}$. Damit ist nach der Definition der Nacheinanderausführung von Korrespondenzen $(x, x) \in f^{-1} \bullet f$, d. h. für jedes $x \in M$ gilt $(x, x) \in \ker f$.

[1] G. Asser (MfL Bd. 1, 2.5.) nennt deshalb den Kern die durch f induzierte Äquivalenzrelation und bezeichnet ihn mit R_f. J. Flachsmeyer und L. Prohaska (MfL Bd. 3, 3.4., 6.4., 12.5.) sprechen bei Linearformen, bei linearen Abbildungen und Homomorphismen vom Kern und bezeichnen ihn in den beiden erstgenannten Fällen ebenfalls mit ker.

$\ker f$ ist symmetrisch: Es sei $(x, y) \in \ker f$. Dann ist $(y, x) \in (\ker f)^{-1}$. Damit ergibt sich (vgl. MfL Bd. 1, 2.3. (18), (15)) $(\ker f)^{-1} = (f^{-1} \bullet f)^{-1} = f^{-1} \bullet (f^{-1})^{-1} = f^{-1} \bullet f = \ker f$. Folglich ist $(y, x) \in \ker f$.

$\ker f$ ist transitiv: Es gelte $(x, y) \in \ker f$ und $(y, z) \in \ker f$. Dann ist (nach der Definition der NAF) $(x, z) \in \ker f \bullet \ker f$. Die Eindeutigkeit von f, die wegen (3.1.6) zu $f \bullet f^{-1} \subseteq I_N$ äquivalent ist, und die Assoziativität der NAF liefern dann $\ker f \bullet \ker f = f^{-1} \bullet f \bullet f^{-1} \bullet f \subseteq f^{-1} \bullet I_N \bullet f = f^{-1} \bullet f = \ker f$. Damit ist $(x, z) \in \ker f$.

3.1.14. Korollar. *$f: M \to N$ ist injektiv genau dann, wenn* $\ker f = I_M$.

Der Beweis folgt unmittelbar aus (3.1.10).

Wir zeigen jetzt, daß der Kern einer Abbildung $f: M \to N$ gerade aus denjenigen Paaren (x, y) mit $x, y \in M$ besteht, deren Komponenten x und y bei dieser Abbildung dasselbe Bild haben:

3.1.15. $\ker f = \{(x, y): f(x) = f(y)\}$.

Es sei $(x, y) \in \ker f = f^{-1} \bullet f$. Dann gibt es ein $z \in N$ mit $(x, z) \in f$ und $(z, y) \in f^{-1}$. Letzteres zieht $(y, z) \in f$ nach sich, so daß neben $f(x) = z$ auch $f(y) = z$ gilt. Umgekehrt folgt aus $f(x) = f(y)$, daß x und y zueinander äquivalent (mod $\ker f$) sein müssen.

Da für $f: M \to N$ der Kern dieser Abbildung f eine Äquivalenzrelation in M ist, liegt es nahe, die Faktormenge $M/\ker f$ zu betrachten. Es sei $x \in M$, beliebig. Dann ist wegen (3.1.15) klar, daß jedes Element $[x]_{\ker f}$ dieser Faktormenge gerade aus dem vollen Urbild $f^{-1}(\{y\})$ besteht, wenn wir $f(x) = y$ voraussetzen. Alle Äquivalenzklassen erhalten wir, wenn y den Wertebereich $W(f)$ durchläuft.

3.1.16. Definition. Es sei R eine Äquivalenzrelation in M. Dann heißt $\operatorname{kan} R$: $M \to M/R$ mit $x \mapsto [x]_R$ die zu R gehörige *kanonische* (oder *natürliche*) *Abbildung* von M auf M/R.

Also ordnet $\operatorname{kan} R$ jedem Element von M diejenige Äquivalenzklasse zu, in der es selbst liegt. Sowohl die Eindeutigkeit als auch die Surjektivität von $\operatorname{kan} R$ sind unmittelbar klar. Nach Satz 3.1.13 muß der Kern der Abbildung $\operatorname{kan} R: M \to M/R$ wieder eine Äquivalenzrelation in M liefern. Es gilt sogar

3.1.17. $\ker(\operatorname{kan} R) = R$.

Hierzu überlegen wir uns, daß $(x, y) \in \ker(\operatorname{kan} R)$ wegen (3.1.15) $(\operatorname{kan} R)(x) = (\operatorname{kan} R)(y)$ nach sich zieht. Hieraus folgt nach Definition 3.1.16 $[x]_R = [y]_R$, so daß infolge von (3.1.2) xRy gilt. Der Rückschluß vervollständigt den Beweis.

Für eine Abbildung $f: M \to N$ hat das Bild von $x \in M$ bei der zur Äquivalenzrelation $\ker f$ gehörigen kanonischen Abbildung $\bar{f} := \operatorname{kan}(\ker f)$ von M auf die Faktormenge $M/\ker f$ folgendes Aussehen:

$$\bar{f}(x) = \big(\operatorname{kan}(\ker f)\big)(x) = [x]_{\ker f} = \{y: y \in M \wedge (x, y) \in \ker f\},$$

d. h., es gilt

3.1.18. $\bar{f}(x) = \{y: y \in M \wedge f(x) = f(y)\}$.

Die kanonische Abbildung $\tilde{f}: M \to M/\ker f$ ordnet also jedem Element $x \in M$ die Menge aller derjenigen $y \in M$ zu, die bei der Abbildung $f: M \to N$ dasselbe Bild wie x haben.

Als Spezialfall des allgemeinen Zerlegungssatzes für Mengen (vgl. etwa H. Lugowski [108]) beweisen wir jetzt für die Abbildungen $f: M \to N$ und $\tilde{f}: M \to M/\ker f$ den folgenden Satz:

3.1.19. Satz. *Es existiert eine* (*eindeutig bestimmte*) *1-1-Abbildung* $g: M/\ker f \to N$ *mit* $g \bullet \tilde{f} = f$ *genau dann, wenn* $\ker \tilde{f} = \ker f$ *gilt* (siehe Abb. 20).[1])

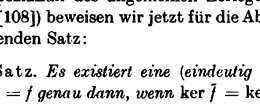


Abb. 20

Dabei ordnet die Abbildung g jedem Element $[x]_{\ker f}$ der Faktormenge $M/\ker f$ das allen $y \in [x]_{\ker f}$ gemeinsame Bild $f(x)$ $\left(= f(y)\right)$ aus N zu. Die Abbildung g ist surjektiv (und damit bijektiv) genau dann, wenn f surjektiv ist.

Beweis. 1. *Zur Existenz von g:* Setzen wir die Existenz einer 1-1-Abbildung $g: M/\ker f \to N$ mit $g \bullet \tilde{f} = f$ voraus, so gilt infolge von (3.1.17) $\ker \tilde{f} = \ker\left(\text{kan}\,(\ker f)\right) = \ker f$.

Um zu zeigen, daß umgekehrt aus $\ker \tilde{f} = \ker f$ die Existenz einer 1-1-Abbildung $g: M/\ker f \to N$ mit $g \bullet \tilde{f} = f$ folgt, definieren wir g durch $g := f \bullet \tilde{f}^{-1}$ und weisen nach, daß das so definierte g die im Satz geforderten Eigenschaften hat:

a) *g ist eine Korrespondenz aus* $M/\ker f$ *in* N. Das folgt aus der Definition der NAF von Korrespondenzen; $\tilde{f}^{-1}$ ist eine Korrespondenz (von $M/\ker f$ auf M), f ist eine eindeutige Korrespondenz (von M in N).

b) *g ist eindeutig, also eine Abbildung* (*aus* $M/\ker f$ *in* N). Für $f: M \to N$ gilt wegen (3.1.6) $f \bullet f^{-1} \subseteq I_N$. Nach Voraussetzung gilt weiterhin $\ker \tilde{f} = \ker f$, also $\tilde{f}^{-1} \bullet \tilde{f} = f^{-1} \bullet f$. Damit ergibt sich $g \bullet g^{-1} = f \bullet \tilde{f}^{-1} \bullet (f \bullet \tilde{f}^{-1})^{-1} = f \bullet \tilde{f}^{-1} \bullet \tilde{f} \bullet f^{-1} = f \bullet f^{-1} \bullet f \bullet f^{-1} \subseteq I_N \bullet I_N = I_N$, woraus wieder mit (3.1.6) die Eindeutigkeit der Korrespondenz g folgt.

c) $D(g) = M/\ker f$, d. h., *es gilt* $g: M/\ker f \to N$. Nach Voraussetzung gelten $D(f) = M$ und $W(\tilde{f}) = M/\ker f$, so daß wegen (3.1.4) und (3.1.5) neben $I_M \subseteq f^{-1} \bullet f$ auch $I_{M/\ker f} \subseteq \tilde{f} \bullet \tilde{f}^{-1}$ erfüllt ist. Das führt uns dann auf $I_{M/\ker f} \subseteq \tilde{f} \bullet \tilde{f}^{-1} = \tilde{f} \bullet I_M \bullet \tilde{f}^{-1} \subseteq \tilde{f} \bullet f^{-1} \bullet f \bullet \tilde{f}^{-1} = (f \bullet \tilde{f}^{-1})^{-1} \bullet (f \bullet \tilde{f}^{-1}) = g^{-1} \bullet g$, so daß $D(g) = M/\ker f$ gilt.

d) *g ist injektiv.* Wegen b) brauchen wir nur noch zu zeigen, daß g^{-1} eindeutig ist. Dabei stützen wir uns auf die Eindeutigkeit von $\tilde{f}$, also $\tilde{f} \bullet \tilde{f}^{-1} \subseteq I_{M/\ker f}$ und $\ker \tilde{f} = \ker f$, d. h., es gilt $g^{-1} \bullet g = (f \bullet \tilde{f}^{-1})^{-1} \bullet f \bullet \tilde{f}^{-1} = \tilde{f} \bullet f^{-1} \bullet f \bullet \tilde{f}^{-1} = \tilde{f} \bullet \tilde{f}^{-1} \bullet \tilde{f} \bullet \tilde{f}^{-1} \subseteq I_{M/\ker f} \bullet I_{M/\ker f} = I_{M/\ker f}$. Wegen (3.1.7) ist g damit eineindeutig.

[1]) Die in Abb. 20 vorgenommene Veranschaulichung nennt man ein *Diagramm*. Sind je zwei Abbildungen, die durch Verkettung längs verschiedener Wege im Diagramm zwischen demselben Paar von Endpunkten erhalten werden, gleich, so heißt das Diagramm *kommutativ*. Die Kommutativität eines Diagrammes wird durch das Gleichheitszeichen signalisiert.

e) $g \bullet \bar{f} = f$. Dazu ziehen wir neben $\ker \bar{f} = \ker f$ noch die Eindeutigkeit von f sowie $D(\bar{f}) = M$ heran, d. h., wir nutzen $\bar{f}^{-1} \bullet \bar{f} = f^{-1} \bullet f$, $f \bullet f^{-1} \subseteq I_N$ und $I_M \subseteq \bar{f}^{-1} \bullet \bar{f}$ aus. $g \bullet \bar{f} = f \bullet \bar{f}^{-1} \bullet \bar{f} = f \bullet f^{-1} \bullet f \subseteq I_N \bullet f = f = f \bullet I_M \subseteq f \bullet \bar{f}^{-1} \bullet \bar{f} = g \bullet \bar{f}$.

2. *Zur Einzigkeit von g:* Gäbe es eine weitere Abbildung $g_1 \colon M/\ker f \to N$ mit $g_1 \bullet \bar{f} = f$, so wäre $(g \bullet \bar{f})\,(x) = g\big(\bar{f}(x)\big) = g_1\big(\bar{f}(x)\big) = (g_1 \bullet \bar{f})\,(x)$ für alle $x \in M$. Da $\bar{f}$ surjektiv ist, gibt es zu jedem $y \in M/\ker f$ ein $x \in M$ mit $\bar{f}(x) = y$. Für alle Elemente $y \in M/\ker f$ gilt folglich $g(y) = g\big(\bar{f}(x)\big) = g_1\big(\bar{f}(x)\big) = g_1(y)$. Da g und g_1 wertverlaufsgleich sind, gilt $g = g_1$.

3. $g \colon [x]_{\ker f} \mapsto f(x)$. Wegen

$$\bar{f}^{-1} = \{([x]_{\ker f}, x) \colon x \in M\}, \qquad f = \big\{\big(x, f(x)\big) \colon x \in M\big\}$$

und der Definition der NAF von Korrespondenzen gilt

$$g = f \bullet \bar{f}^{-1} = \big\{\big([x]_{\ker f}, f(x)\big) \colon x \in M\big\}.$$

4. *g surjektiv $\Leftrightarrow$ f surjektiv.* Es sei f surjektiv, d. h. $W(f) = N$ bzw. $I_N \subseteq f \bullet f^{-1}$. Wegen $D(\bar{f}) = M$ gilt $I_M \subseteq \bar{f}^{-1} \bullet \bar{f}$. Damit erhalten wir

$$\begin{aligned} I_N \subseteq f \bullet f^{-1} &= f \bullet I_M \bullet f^{-1} \subseteq f \bullet \bar{f}^{-1} \bullet \bar{f} \bullet f^{-1} \\ &= f \bullet \bar{f}^{-1} \bullet (f \bullet \bar{f}^{-1})^{-1} = g \bullet g^{-1}, \end{aligned}$$

also $W(g) = N$. Umgekehrt sei g surjektiv, d. h. $I_N \subseteq g \bullet g^{-1}$. Mit $\ker \bar{f} = \ker f$ und $f \bullet f^{-1} \subseteq I_N$ folgt dann

$$\begin{aligned} I_N \subseteq g \bullet g^{-1} &= f \bullet \bar{f}^{-1} \bullet (f \bullet \bar{f}^{-1})^{-1} = f \bullet \bar{f}^{-1} \bullet \bar{f} \bullet f^{-1} \\ &= f \bullet f^{-1} \bullet f \bullet f^{-1} \subseteq I_N \bullet f \bullet f^{-1} = f \bullet f^{-1}, \end{aligned}$$

so daß auch f surjektiv ist.

Zur Veranschaulichung der vielen Begriffe und Beziehungen, die wir zum Beweis von Satz 3.1.19 benötigt haben, wählen wir abschließend ein einfaches

3.1.20. Beispiel (vgl. auch MfL Bd. 1, 2.5.)

$M = \{0, 1, 2, 3\}$, $N = \{a, b, c, d\}$, $f \colon M \to N$ sei durch $f(0) = f(2) = a$ und $f(1) = f(3) = b$ gegeben (siehe Abb. 21). Es ist

$f = \{(0, a), (1, b), (2, a), (3, b)\}$, $f^{-1} = \{(a, 0), (b, 1), (a, 2), (b, 3)\}$,
$\ker f = f^{-1} \bullet f = \{(0, 0), (1, 1), (2, 2), (3, 3), (0, 2), (2, 0), (1, 3), (3, 1)\}$,
$f \bullet f^{-1} = \{(a, a), (b, b)\}$,
$D(f) = M$, d. h. $I_M \subseteq f^{-1} \bullet f$,
f eindeutig, d. h. $f \bullet f^{-1} \subseteq I_N$,
f nicht surjektiv, denn $W(f) \subset N$ bzw. $I_N \nsubseteq f \bullet f^{-1}$,
f nicht injektiv, denn $f^{-1} \bullet f \nsubseteq I_M$,
$M/\ker f = \{[0], [1]\} = \{[0], [3]\} = \{[1], [2]\} = \{[2], [3]\}$,
$\bar{f} = \mathrm{kan}\,(\ker f) = \{(0, [0]), (1, [1]), (2, [0]), (3, [1])\}$,
$\bar{f}^{-1} = \{([0], 0), ([1], 1), ([0], 2), ([1], 3)\}$,
$\ker \bar{f} = \bar{f}^{-1} \bullet \bar{f} = \{(0, 0), (1, 1), (2, 2), (3, 3), (0, 2), (2, 0), (1, 3), (3, 1)\} = \ker f$,
$\bar{f} \bullet \bar{f}^{-1} = \{([0], [0]), ([1], [1])\}$,

$D(\tilde{f}) = M$, d. h. $I_M \subseteq \tilde{f}^{-1} \bullet \tilde{f}$,
$\tilde{f}$ eindeutig, d. h. $\tilde{f} \bullet \tilde{f}^{-1} \subseteq I_{M/\ker f}$,
$\tilde{f}$ surjektiv, d. h. $W(\tilde{f}) = M/\ker f$ bzw. $I_{M/\ker f} \subseteq \tilde{f} \bullet \tilde{f}^{-1}$
$\Rightarrow \tilde{f} \bullet \tilde{f}^{-1} = I_{M/\ker f}$,
$\tilde{f}$ nicht injektiv, denn $\tilde{f}^{-1} \bullet \tilde{f} \nsubseteq I_M$,
$g = f \bullet \tilde{f}^{-1} = \{([0], a), ([1], b)\}$, $g^{-1} = \{(a, [0]), (b, [1])\}$,
$\ker g = g^{-1} \bullet g = \{([0], [0]), ([1], [1])\}$,
$g \bullet g^{-1} = \{(a, a), (b, b)\}$,
$D(g) = M/\ker f$, d. h. $I_{M/\ker f} \subseteq g^{-1} \bullet g$,
g eindeutig, d. h. $g \bullet g^{-1} \subseteq I_N$,
g nicht surjektiv, denn $W(g) \subset N$ bzw. $I_N \nsubseteq g \bullet g^{-1}$,
g injektiv, d. h. $g^{-1} \bullet g \subseteq I_{M/\ker f}$, also $g^{-1} \bullet g = I_{M/\ker f}$,

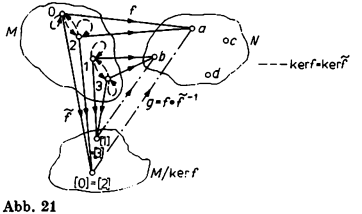


Abb. 21

3.1.21. Aufgaben

1. Es sei $\ddot{A}(M)$ die Menge aller Äquivalenzrelationen in M, $Z(M)$ die Menge aller Zerlegungen von M. Man zeige, daß der Hauptsatz über Äquivalenzrelationen auch folgendermaßen formuliert werden kann: $f\colon \ddot{A}(M) \to Z(M)$ mit $R \mapsto M/R$ ist eine bijektive Abbildung (vgl. H. Lugowski [108]).

2. Man gebe die Elemente der Faktormenge M/R an, wenn als Äquivalenzrelation R a) die identische Abbildung I_M bzw. b) die Produktmenge $M \times M$ gewählt wird.

3. Man untersuche, welche Gestalt im Fall der Injektivität von $f\colon M \to N$ die Äquivalenzklasse $\tilde{f}(x) = [x]_{\ker f}$ aus $M/\ker f$ hat.

4. Für die Abbildung $f\colon \mathbf{R}^* \to \mathbf{R}_+^*$ mit $x \mapsto |x|$ gebe man $D(f)$, $W(f)$, $\ker f$, $\tilde{f} = \mathrm{kan}\,(\ker f)$, $\mathbf{R}^*/\ker f$, $D(\tilde{f})$, $W(\tilde{f})$, $g = f \bullet \tilde{f}^{-1}$, $D(g)$, $W(g)$ und $\ker g$ an und untersuche, welche der Abbildungen f, $\tilde{f}$, g surjektiv, injektiv oder bijektiv sind.

3.2. Kongruenzrelationen und Faktorgruppoide

Ist R eine Äquivalenzrelation in der Trägermenge G eines Gruppoids $(G, \cdot)$[1]), so bleibt die Frage, ob R in irgendeiner Beziehung zur Operation $\cdot$ steht, zunächst offen. Erfüllt R nun aber neben der Reflexivität, Symmetrie und Transitivität auch noch die

[1]) Man beachte, daß das Zeichen „$\cdot$" hier nicht für die Multiplikation, sondern für eine beliebige Operation in einer Menge G steht.

Verträglichkeits- (oder Kompatibilitäts-) bedingung, so läßt sich in der Faktormenge G/R auf sehr einfache und natürliche Weise eine neue Operation definieren.

3.2.1. Definition. Es sei $\mathfrak{G} = (G, \cdot)$ ein Gruppoid. Eine Äquivalenzrelation R in G heißt *Kongruenzrelation* in $\mathfrak{G}$, wenn sie mit der Operation $\cdot$ *verträglich* (oder *kompatibel*) ist:

$$\textbf{3.2.2.} \quad \bigwedge_{a,b,c,d \in G} (aRb \wedge cRd \Rightarrow a \cdot c \; R \; b \cdot d).$$

Aus (3.2.2) folgt mit aRb und cRc (Reflexivität) $a \cdot c \; R \; b \cdot c$ und $c \cdot a \; R \; c \cdot b$. Die Forderung

$$\textbf{3.2.3.} \quad \bigwedge_{a,b,c \in G} (aRb \Rightarrow a \cdot c \; R \; b \cdot c \wedge c \cdot a \; R \; c \cdot b)$$

ist sogar äquivalent zu (3.2.2), so daß sie von manchen Autoren zur Definition der Verträglichkeit einer Äquivalenzrelation mit einer Operation herangezogen wird. Es sei aRb und cRd. Dann liefert die zweimalige Anwendung von (3.2.3) $a \cdot c \; R \; b \cdot c$ sowie $b \cdot c \; R \; b \cdot d$, so daß aufgrund der Transitivität $a \cdot c \; R \; b \cdot d$ folgt.

Die Elemente der Faktormenge G/R, d. h. die Äquivalenzklassen $[x]_R$, heißen jetzt entsprechend *Kongruenzklassen*.

Die Beziehung (3.2.2) erlaubt, in G/R eine Operation $*$ auf folgende Weise zu definieren:

3.2.4. $[x]_R * [y]_R := [x \cdot y]_R$ für alle $[x]_R, [y]_R \in G/R$.

Zur Rechtfertigung müssen wir zeigen, daß diese Definition unabhängig von der Wahl der Repräsentanten x ($\in G$) und y ($\in G$) aus den Kongruenzklassen $[x]_R$ bzw. $[y]_R$ ist. Es seien deshalb x_1 und x_2 andere Repräsentanten dieser Kongruenzklassen, d. h., es gilt $[x_1]_R = [x]_R$ und $[y_1]_R = [y]_R$. Wegen (3.1.2) erhalten wir zunächst $x_1 R x$ und $y_1 R y$, so daß $x_1 \cdot y_1 \; R \; x \cdot y$ aufgrund der Verträglichkeit von R mit $\cdot$ gilt. Daraus folgt dann (wieder wegen (3.1.2)) $[x_1 \cdot y_1]_R = [x \cdot y]_R$. Dieser Beweis läßt sich auch so interpretieren, daß aus der Gleichheit der Urbilder $([x_1]_R, [y_1]_R) = ([x]_R, [y]_R)$ die Gleichheit der Bilder $[x_1 \cdot y_1]_R = [x \cdot y]_R$ folgt. $*$ ist also eine (1-) Abbildung von $G/R \times G/R$ in G/R mit $([x]_R, [y]_R) \mapsto [x \cdot y]_R$. Folglich ist $*$ eine Operation in der Faktormenge G/R.

Das Gruppoid $\mathfrak{G}/R = (G/R, *)$ heißt das *Faktorgruppoid*[1]) von $\mathfrak{G}$ nach R.

Es ist klar, daß es zu jedem Gruppoid $\mathfrak{G} = (G, \cdot)$ mit $|\mathfrak{G}| > 1$ wenigstens zwei Faktorgruppoide gibt, da sowohl die identische Abbildung I_G als auch die Produktmenge $G \times G$ Kongruenzrelationen in $\mathfrak{G}$ sind. Diese beiden extremen Kongruenzrelationen liefern die Faktormengen $G/I_G = \{\{x\} : x \in G\}$ und $G/G \times G = \{G\}$ (vgl. Aufgabe 3.1.21.2).

3.2.5. Beispiel

In der Menge $G = \{0, 1, 2, 3\}$ ist $R = \{(0, 0), (1, 1), (2, 2), (3, 3), (0, 2), (2, 0), (1, 3), (3, 1)\}$ eine Äquivalenzrelation (vgl. Beispiel 3.1.20). Wir definieren in G durch die nachfolgenden Strük-

[1]) Der Name Faktoroid (z. B. O. Borůvka [27]) ist weniger gebräuchlich.

turtafeln zwei voneinander verschiedene Operationen:

·	0	1	2	3
0	0	1	2	3
1	1	0	1	2
2	2	1	0	1
3	3	2	1	0

∘	0	1	2	3
0	3	2	0	1
1	2	3	1	0
2	0	1	3	2
3	1	0	2	3

Während R in $(G, \cdot)$ eine Kongruenzrelation ist (Beweis!), ist R mit der Operation $\circ$ nicht verträglich. Man wähle z. B. in (3.2.2) $a = b = 0$, $c = 1$, $d = 3$. Im ersten Fall können wir also auch das zugehörige Faktorgruppoid $\mathfrak{G}/R = (G/R, *)$ konstruieren:

*	[0]	[1]
[0]	[0]	[1]
[1]	[1]	[0]

Die Bedingung (3.2.3) ist gleichwertig mit $(aRb \Rightarrow a \cdot c\, R\, b \cdot c) \wedge (aRb \Rightarrow c \cdot a\, R\, c \cdot b)$ für alle $a, b, c \in G$. Die Forderung, die wir hieraus erhalten, indem wir innerhalb beider Klammern der Konjunktion zur Umkehrung übergehen, brauchen Kongruenzrelationen i. allg. nicht zu erfüllen. Sie zeichnet aber eine spezielle Klasse unter den Kongruenzrelationen besonders aus.

3.2.6. Definition. Eine Kongruenzrelation R in einem Gruppoid $\mathfrak{G} = (G, \cdot)$ heißt *normal*, wenn sie die beiden Bedingungen

3.2.7. $\bigwedge_{a,b,c \in G} (a \cdot c\, R\, b \cdot c \Rightarrow aRb)$,

3.2.8. $\bigwedge_{a,b,c \in G} (c \cdot a\, R\, c \cdot b \Rightarrow aRb)$

erfüllt.[1])

Ist die Operation in $\mathfrak{G} = (G, \cdot)$ kommutativ, so genügt natürlich eine der beiden Bedingungen in Definition 3.2.6.

Die Zahlenkongruenz $\equiv \pmod m$, die durch $a \equiv b \pmod m :\Leftrightarrow m \mid a - b$ definiert ist ($m \in \mathbf{N}$, $m > 1$, $a, b \in \mathbf{Z}$), ist in $\mathbf{Z}$ bezüglich der Addition normal, nicht aber bezüglich der Multiplikation.

3.2.9. Satz. *Es sei $\mathfrak{G} = (G, \cdot)$ ein Gruppoid, R eine Kongruenzrelation in $\mathfrak{G}$. Die Relation R ist genau dann normal, wenn das Faktorgruppoid $\mathfrak{G}/R = (G/R, *)$ ein K-Gruppoid ist.*

Beweis. Es gelte (3.2.7). Wegen (3.1.2) heißt das, daß aus $[a \cdot c]_R = [b \cdot c]_R$ bzw. mit (3.2.4) aus $[a]_R * [c]_R = [b]_R * [c]_R$ die Beziehung $[a]_R = [b]_R$ folgt. Analog zieht (3.2.8) nach sich, daß aus $[c]_R * [a]_R = [c]_R * [b]_R$ ebenfalls $[a]_R = [b]_R$ folgt. Damit ist $*$ kürzbar. Alle Schlüsse sind umkehrbar.

Selbst für endliche Gruppoide kann der Nachweis der Verträglichkeit einer Äquivalenzrelation mit einer gegebenen Operation sehr mühselig (und langweilig) sein.

[1]) Man beachte, daß die Umkehrung von (3.2.3) nicht (3.2.7) $\wedge$ (3.2.8), sondern (3.2.7) $\vee$ (3.2.8) ist.

Im Unterschied hierzu ist der Test, ob eine Kongruenzrelation normal ist, anhand dieses Satzes sofort aus der Strukturtafel des zugehörigen Faktorgruppoids ablesbar. Wenn die Strukturtafel ein lateinisches Quadrat ist, ist die Normalität schon bewiesen. So ist also z. B. die Kongruenzrelation R in $(G, \cdot)$ aus Beispiel 3.2.5 normal. Das ist auch insofern interessant, als $(G, \cdot)$ selbst ja kein K-Gruppoid ist. Man kann also durch die Faktorisierung Eigenschaften gewinnen, die im Ausgangsgruppoid nicht erfüllt gewesen sind. Umgekehrt bleiben durch die Faktorisierung solche Eigenschaften wie Kommutativität, Assoziativität oder die Existenz eines neutralen Elementes erhalten. Ist $(G, \cdot)$ etwa kommutativ oder assoziativ, so folgt mit (3.2.4) sofort, daß auch $(G/R, *)$ kommutativ bzw. assoziativ ist. Gibt es in $(G, \cdot)$ ein neutrales Element n, so ist $[n]_R$ das neutrale Element in $(G/R, *)$ (Beweis!).

Gerade die Tatsache, daß für jede Kongruenzrelation in einer Gruppe das zugehörige Faktorgruppoid auch wieder eine Gruppe ist (Beweis!), darf uns jedoch nun nicht dazu verleiten, etwa alle algebraischen Eigenschaften eines Gruppoids ungeprüft auf das durch die jeweilige Kongruenzrelation induzierte Faktorgruppoid zu übertragen. So braucht z. B. schon das durch eine Kongruenzrelation definierte Faktorgruppoid einer Quasigruppe selbst keine Quasigruppe mehr zu sein.[1]) Es gilt aber der

3.2.10. Satz. *Ist $\mathfrak{Q} = (Q, \cdot)$ eine Quasigruppe und R eine Kongruenzrelation in $\mathfrak{Q}$, so ist das Faktorgruppoid $\mathfrak{Q}/R = (Q/R, *)$ ein D-Gruppoid.*

Beweis. Es seien $[a]_R$, $[b]_R$ beliebige Elemente aus Q/R. Die Gleichung $a \cdot x = b$ hat in $\mathfrak{Q}$ eine eindeutig bestimmte Lösung x_0, d. h. $a \cdot x_0 = b$. Also gilt $[a \cdot x_0]_R = [b]_R$ bzw. mit (3.2.4) $[a]_R * [x_0]_R = [b]_R$. Die Kongruenzklasse $[x_0]_R$ ist folglich eine Lösung der Gleichung $[a]_R * [x]_R = [b]_R$. Entsprechend zeigt man, daß die Gleichung $[y]_R * [a]_R = [b]_R$ in $\mathfrak{Q}/R$ eine Lösung hat. Damit ist $\mathfrak{Q}/R$ ein D-Gruppoid.

Wenn die Quasigruppe $\mathfrak{Q}$ endlich ist, führt die Faktorisierung wegen Satz 1.2.5 wieder auf eine Quasigruppe. Im Fall beliebiger Ordnung müssen wir zusätzlich die Normalität von R fordern, d. h., es gilt das

3.2.11. Korollar. *Ist $\mathfrak{Q}$ eine Quasigruppe und R eine normale Kongruenzrelation in $\mathfrak{Q}$, so ist das Faktorgruppoid $\mathfrak{Q}/R$ ebenfalls eine Quasigruppe.*

Der Beweis folgt unmittelbar aus den beiden vorangehenden Sätzen.

Die Frage, ob eine Kongruenzrelation normal ist, läßt sich in bestimmten Fällen auch ebenso leicht ohne die Konstruktion des jeweiligen Faktorgruppoids beantworten. In speziellen Gruppoiden ist sogar jede Kongruenzrelation normal. Zuvor beweisen wir aber noch den folgenden

3.2.12. Satz. *Es sei R eine Kongruenzrelation in einer Gruppe $(G, \cdot)$. Dann gilt für alle $a, b \in G$: $aRb \Rightarrow a^{-1} R\, b^{-1}$.*

Beweis. Es gelte aRb. Daneben gilt dann wegen der Reflexivität von R auch stets $a^{-1} R\, a^{-1}$ und $b^{-1} R\, b^{-1}$. Aufgrund der Verträglichkeit von R mit der Gruppenopera-

[1]) Wir werden im nächsten Abschnitt auf diese Problematik zurückkommen.

tion erhalten wir zunächst $b^{-1} \cdot a\, R\, b^{-1} \cdot b$, also $b^{-1} \cdot a\, R\, n$, wobei n das neutrale Element der Gruppe ist. Nutzen wir jetzt ein zweites Mal (3.2.2) aus, so ergibt sich $b^{-1} \cdot a \cdot a^{-1}\, R\, n \cdot a^{-1}$, also $b^{-1}\, R\, a^{-1}$. Die Symmetrie von R liefert damit schließlich $a^{-1}\, R\, b^{-1}$.

3.2.13. Satz. *In einer Gruppe $(G, \cdot)$ ist jede Kongruenzrelation R normal.*

Beweis. Es gelte $a \cdot c\, R\, b \cdot c$ für beliebige $a, b, c \in G$. Da R reflexiv ist, gilt auch stets $c^{-1}\, R\, c^{-1}$, so daß mit (3.2.2) $a \cdot c \cdot c^{-1}\, R\, b \cdot c \cdot c^{-1}$, also aRb folgt. Analog zeigt man (3.2.8).

Speziell für Gruppen lassen sich die Kongruenzrelationen mit Hilfe gewisser ausgezeichneter Untergruppen beschreiben bzw. konstruieren. In Anlehnung an MfL Bd. 3, 12.2., 12.3., nennen wir deshalb jede nichtleere Teilmenge K der Trägermenge G einer Gruppe $\mathfrak{G} = (G, \cdot)$ einen *Komplex* von $\mathfrak{G}$. Sind K und L Komplexe von $\mathfrak{G}$, so versteht man unter dem *Komplexprodukt* von K und L die Menge

$$KL := \{k \cdot l : k \in K \wedge l \in L\}.$$

Dabei überträgt sich die Assoziativität der Gruppenoperation auf die Komplex multiplikation. Wir schreiben zu verknüpfende Komplexe (in der entsprechenden Reihenfolge) nebeneinander, verzichten also ausnahmsweise auf ein eigenes Operationszeichen. Für $K = \{a\}$ schreiben wir statt $\{a\}\, L$ einfach aL.

$\mathfrak{U} = (U, \cdot)$ heißt *Untergruppe* von $\mathfrak{G}$, wenn U ein Komplex von $\mathfrak{G}$ und $\mathfrak{U}$ selbst eine Gruppe ist.[1]) Die Komplexprodukte aU und Ua ($a \in G$, beliebig) heißen dann *Links-* bzw. *Rechtsnebenklasse nach* $\mathfrak{U}$. Die Menge aller Linksnebenklassen liefert eine Zerlegung von G (ebenso alle Rechtsnebenklassen). Eine Untergruppe $\mathfrak{N} = (N, \cdot)$ der Gruppe $\mathfrak{G}$ heißt schließlich *Normalteiler* (oder *normale* bzw. *invariante Untergruppe*) von $\mathfrak{G}$, falls

3.2.14. $\bigwedge_{a \in G} aN = Na$

ist, d. h. alle Linksnebenklassen stimmen mit den entsprechenden Rechtsnebenklassen überein, was bekanntlich für eine beliebige Untergruppe nicht stimmt. Die Menge der Nebenklassen nach einem Normalteiler $\mathfrak{N} = (N, \cdot)$ bezeichnet man mit $G/\mathfrak{N}$ (oder G/N); sie bildet bezüglich der Komplexmultiplikation eine Gruppe (Beweis!).

Damit sind wir jetzt in der Lage, den sogenannten *Hauptsatz über Kongruenzrelationen in Gruppen* zu formulieren.

3.2.15. Satz. *Jeder Normalteiler $\mathfrak{N} = (N, \cdot)$ einer Gruppe $\mathfrak{G} = (G, \cdot)$ definiert durch*

3.2.16. $\bigwedge_{a,b \in G} aRb :\Leftrightarrow aN = bN$

[1]) Korrekt müßte man hier eigentlich $(U, \cdot \mid U \times U)$ schreiben, da die Einschränkung (vgl. MfL Bd. 1, 2.4., 2.7.) der in G definierten Operation $\cdot$ auf die Teilmenge U gemeint ist.

genau eine (normale) Kongruenzrelation R in $\mathfrak{G}$. Umgekehrt definiert jede (normale) Kongruenzrelation R in einer Gruppe $\mathfrak{G} = (G, \cdot)$ mit dem neutralen Element n durch

3.2.17. $N := [n]_R$

genau einen Normalteiler $\mathfrak{N} = (N, \cdot)$ von $\mathfrak{G}$.

3.2.18. Korollar. *Die Kongruenzklassen* mod R *sind gerade die Nebenklassen nach dem Normalteiler $\mathfrak{N}$, d. h., es gilt $[a]_R = aN$ für alle $a \in G$.*

Beweis. a) Es sei $\mathfrak{N} = (N, \cdot)$ ein Normalteiler von $\mathfrak{G} = (G, \cdot)$. Wir zeigen, daß die durch (3.2.16) definierte Relation $R := \{(x, y) : xN = yN\}$ eine Kongruenzrelation in $\mathfrak{G}$ ist. Es seien a, b, c und d im folgenden beliebige Elemente aus G.

1. *R ist eine Äquivalenzrelation:* $aN = aN$ liefert die Reflexivität. Mit $aN = bN$ gilt trivialerweise auch $bN = aN$, d. h. die Symmetrie. Schließlich gelte $aN = bN$ und $bN = cN$. Da die $=$-Relation transitiv ist, gilt folglich auch $aN = cN$, so daß die Transitivität von R erfüllt ist.

2. *R ist mit der Gruppenoperation verträglich:* Es gelte aRb und cRd, also $aN = bN$ und $cN = dN$. Aus $aNcN = bNdN$ folgt dann unter Ausnutzung von (3.2.14) und der Assoziativität der Komplexmultiplikation $acNN = bdNN$. Da $\mathfrak{N}$ Normalteiler ist, gilt insbesondere $NN = N$ (vgl. etwa H. Lugowski und H. Weinert [109]), so daß wir mit $acN = bdN$ insgesamt $a \cdot c\, R\, b \cdot d$ erhalten.

b) Es sei R eine (normale) Kongruenzrelation in $\mathfrak{G} = (G, \cdot)$. Wir zeigen, daß $\mathfrak{N} = (N, \cdot)$ mit der durch (3.2.17) definierten Menge $N := [n]_R = \{x : x \in G \wedge xRn\}$ ein Normalteiler von $\mathfrak{G}$ ist.

1. *$\mathfrak{N} = (N, \cdot)$ ist ein Gruppoid:* Wegen nRn liegt das neutrale Element n in N, d. h. $N \neq \emptyset$. Es seien $x, y \in N$. Dann gilt wegen (3.1.1) xRn und yRn. Da R nach Voraussetzung mit der Gruppenoperation verträglich ist, steht auch $x \cdot y$ in Relation zu $n \cdot n = n$. Wegen (3.1.1) bedeutet aber $x \cdot y\, R\, n$, daß $x \cdot y \in N$ gilt. Folglich ist N bezüglich $\cdot$ abgeschlossen.

2. *$\mathfrak{N} = (N, \cdot)$ ist ein Monoid:* Wegen $N \subseteq G$ ist die Assoziativität trivialerweise erfüllt, so daß $\mathfrak{N}$ Halbgruppe ist. Daß das neutrale Element n der Gruppe in N liegt, haben wir bereits gezeigt.

3. *$\mathfrak{N} = (N, \cdot)$ ist eine Untergruppe von $\mathfrak{G}$:* Es genügt zu zeigen, daß mit $x \in N$ auch das zu x inverse Element $x^{-1} (\in G)$ in N liegt. Aus xRn folgt wegen Satz 3.2.12 $x^{-1}\, R\, n^{-1}$ und wegen $n^{-1} = n$ damit $x^{-1}\, R\, n$. Also ist x^{-1} ein Element von N.

4. *$\mathfrak{N} = (N, \cdot)$ ist ein Normalteiler von $\mathfrak{G}$:* Dazu zeigen wir zunächst die Äquivalenz der folgenden Aussagen; wegen Satz 3.2.13 können wir dabei (3.2.7) und (3.2.8) benutzen:

$$x \in aN \Leftrightarrow a^{-1} \cdot x \in N \Leftrightarrow a^{-1} \cdot x\, R\, n \underset{\overset{\Leftarrow}{(3.2.7)}}{\overset{(3.2.2)}{\Rightarrow}} a \cdot a^{-1} \cdot x\, R\, a \cdot n$$
$$\Leftrightarrow xRa \underset{\Leftarrow}{\overset{\Rightarrow}{(3.1.1)}} x \in [a]_R$$

und

$$x \in Na \Leftrightarrow x \cdot a^{-1} \in N \Leftrightarrow x \cdot a^{-1}\, R\, n \underset{\overset{\Leftarrow}{(3.2.8)}}{\overset{(3.2.2)}{\Rightarrow}} x \cdot a^{-1} \cdot a\, R\, n \cdot a$$
$$\Leftrightarrow xRa \underset{\Leftarrow}{\overset{\Rightarrow}{(3.1.1)}} x \in [a]_R.$$

Insgesamt haben wir also gezeigt, daß $x \in aN \Leftrightarrow x \in Na \Leftrightarrow x \in [a]_R$ für beliebige $a, x \in G$ gilt. Damit ist (3.2.14) erfüllt. Nebenbei haben wir Korollar 3.2.18 bewiesen.

Man kann den Hauptsatz (Satz 3.2.15) auch folgendermaßen formulieren (vgl. H. Lugowski [108]): Die durch $[n]_R \mapsto N$ realisierte Abbildung f von der Menge aller Kongruenzrelationen R in einer Gruppe $\mathfrak{G} = (G, \cdot)$ auf die Menge aller Normalteiler $\mathfrak{N} = (N, \cdot)$ von $\mathfrak{G}$ ist eine Bijektion.

Bei der Faktorisierung einer Gruppe $\mathfrak{G} = (G, \cdot)$ nach einer Kongruenzrelation R hatten wir festgestellt, daß das Faktorgruppoid $\mathfrak{G}/R = (G/R, *)$ selbst wieder eine Gruppe ist. Die Elemente $[x]_R$ dieser Faktorgruppe sind nun, wie wir gesehen haben, mit den Nebenklassen xN identisch, die wir nach dem durch (3.2.16) bestimmten Normalteiler $\mathfrak{N} = (N, \cdot)$ erhalten. Man bezeichnet daher die Faktormenge G/R auch mit $G/\mathfrak{N}$ (oder G/N), was wir bereits vorweggenommen hatten. Der Komplexmultiplikation in $G/\mathfrak{N}$ entspricht gerade die Operation $*$ in G/R:

$$[x]_R * [y]_R = [x \cdot y]_R, \quad \text{also} \quad xN * yN = xyN.$$

Das neutrale Element der Faktorgruppe ist $[n]_R = N$, das zu $[x]_R = xN$ inverse Element ist $[x]_R{}^{-1} = [x^{-1}]_R = x^{-1}N$.

Die Faktorisierung einer Gruppe nach einer Kongruenzrelation R läuft zwar auf dasselbe hinaus wie die Faktorisierung nach dem entsprechenden Normalteiler $\mathfrak{N}$, auch werden sämtliche Kongruenzrelationen in einer Gruppe durch deren Zerlegungen nach den verschiedenen Normalteilern erfaßt (und umgekehrt), trotzdem dürfen wir R und N (die Trägermenge des Normalteilers $\mathfrak{N}$) deshalb nicht einfach identifizieren! R besteht aus geordneten Paaren (x, y) mit $x, y \in G$, aber N aus Elementen von G: $R \subseteq G \times G$ und $N \subseteq G$.

In Analogie zum Begriff der Untergruppe nennen wir $\mathfrak{U} = (U, \cdot)$ *Unterquasigruppe* einer Quasigruppe $\mathfrak{Q} = (Q, \cdot)$, wenn U ein Komplex von $\mathfrak{Q}$ und $\mathfrak{U}$ selbst eine Quasigruppe ist. Der Begriff des Normalteilers läßt sich nun nicht genauso einfach auf Quasigruppen übertragen.[1]) Gibt es jedoch in $\mathfrak{Q} = (Q, \cdot)$ eine normale Kongruenzrelation R, so legt Satz 3.2.15 nahe, wie eine solche Verallgemeinerung vorgenommen werden kann.[2])

3.2.19. Definition. Es sei $\mathfrak{Q} = (Q, \cdot)$ eine Quasigruppe, R eine normale Kongruenzrelation in $\mathfrak{Q}$. Eine Unterquasigruppe $\mathfrak{U} = (U, \cdot)$ der Quasigruppe $\mathfrak{Q}$ heißt *normal*[3]), falls die Trägermenge U mit einer der Kongruenzklassen mod R zusammenfällt, d. h. ein Element $a \in Q$ existiert, so daß $U = [a]_R$ gilt.

Ist $\mathfrak{Q}$ eine Gruppe, so gibt es stets ein solches Element a, nämlich das neutrale Element n. Dann geht $U = [a]_R$ in (3.2.17) über.

[1]) Für beliebige Gruppoide sind diese Schwierigkeiten noch viel größer (vgl. etwa A. G. Kuroš [95]).

[2]) (3.2.17) selbst steht uns dabei in dieser Form natürlich nicht zur Verfügung, da in einer Quasigruppe i. allg. kein neutrales Element existieren muß.

[3]) Mitunter wird auch hier der Begriff Normalteiler verwendet (etwa R. H. Bruck [32]).

Die Theorie der normalen Unterquasigruppen können wir im Rahmen dieses Buches nicht abhandeln. Weitergehende Untersuchungen und Literaturhinweise findet man z. B. bei R. H. BRUCK [32] oder V. D. BELOUSOV [21].

3.2.20 Aufgaben

1. In einem Gruppoid $\mathfrak{G} := (G, \cdot)$ sei eine der Identitäten (1) bis (33) aus 2.4. erfüllt. Es sei ferner R eine Kongruenzrelation in $\mathfrak{G}$. Man zeige, daß das Faktorgruppoid $\mathfrak{G}/R = (G/R, *)$ dieselbe Identität erfüllt wie $\mathfrak{G}$.

2. Man prüfe, ob die Äquivalenzrelation R aus Beispiel 3.2.5 mit der Operation $\square$, die durch $x \square y := x + y \pmod 4$ in $G = \{0, 1, 2, 3\}$ definiert ist, verträglich ist.

3. Warum ist die Zahlenkongruenz $\equiv \pmod m$ in $(\mathbf{Z}, \div)$ und $(\mathbf{Z}, -)$ normal, nicht aber in $(\mathbf{Z}, \cdot)$? ($\cdot$ ist hier die Multiplikation.)

4. Es sei $\mathfrak{G} = (G, \cdot)$ eine Gruppe, $\mathfrak{N} = (N, \cdot)$ ein Normalteiler von $\mathfrak{G}$. Man zeige, daß die durch $aRb :\Leftrightarrow a \cdot b^{-1} \in N$ definierte Relation eine normale Kongruenzrelation in $\mathfrak{G}$ ist.

5. Man zeige: Eine Kongruenzrelation ist in einer Quasigruppe $(Q, \cdot)$ genau dann normal, wenn sie mit den beiden Umkehroperationen von $\cdot$ verträglich ist.

3.3. Homomorphe und isomorphe Abbildungen

Es seien $\mathfrak{G} = (G, \cdot)$ und $\mathfrak{H} = (H, \circ)$ zwei beliebige Gruppoide. Liegt mit $f: G \to H$ eine Abbildung zwischen den beiden Trägermengen vor, so braucht dabei $f(x \cdot y)$, also das Bild des Produkts zweier Elemente x und y aus G, i. allg. nicht mit $f(x) \circ f(y)$, also dem Produkt der Bilder von x und y, übereinzustimmen. Es gibt jedoch Abbildungen, die die Verknüpfungen zwischen den Elementen in $\mathfrak{G}$ in dieser Weise auf die Verknüpfungen zwischen den Elementen in $\mathfrak{H}$ übertragen.

3.3.1. Definition. Es seien $\mathfrak{G} = (G, \cdot)$ und $\mathfrak{H} = (H, \circ)$ Gruppoide. Eine Abbildung $f: G \to H$ heißt ein *Homomorphismus* (oder *homomorphe Abbildung*) von $\mathfrak{G}$ in $\mathfrak{H}$, wenn f operationstreu (oder relationstreu) ist:

3.3.2. $\bigwedge_{x,y \in G} f(x \cdot y) = f(x) \circ f(y)$.

Wenn f ein Homomorphismus von $\mathfrak{G}$ in $\mathfrak{H}$ ist, schreiben wir dafür kurz $f: \mathfrak{G} \xrightarrow{\sim} \mathfrak{H}$ oder $\mathfrak{G} \xrightarrow[f]{\sim} \mathfrak{H}$. Die Operationstreue von f bedeutet gerade die Kommutativität des Diagramms in Abb. 22.

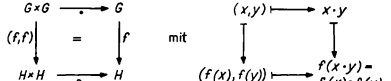


Abb. 22

Ein Homomorphismus f von einem Gruppoid $\mathfrak{G} = (G, \cdot)$ in sich heißt ein *Endomorphismus*. Die Operationstreue von f hat dann also die Gesalt $f(x \cdot y) = f(x) \cdot f(y)$ für alle $x, y \in G$.

Ist f ein Homomorphismus von $\mathfrak{G} = (G, \cdot)$ in $\mathfrak{H} = (H, \circ)$, so gilt $f(G) = \{f(x): x \in G\} = W(f) \subseteq H$ für das volle Bild von G. Sind a und b zwei beliebige Elemente aus

$f(G)$, so existieren wegen $D(f) = G$ Elemente x und y aus G mit $f(x) = a$ und $f(y) = b$. Für das Produkt von a und b, das jedenfalls in H liegt, gilt dann $a \circ b = f(x) \circ f(y) = f(x \cdot y)$, d. h., es gibt ein $z \in G$, nämlich $z = x \cdot y$, dessen Bild bei f gerade $a \circ b$ ist. Damit liegt $a \circ b$ in $f(G)$. Folglich ist $f(G)$ bezüglich $\circ$ abgeschlossen und somit $\big(f(G), \circ\big)$ ein Untergruppoid[1]) von $\mathfrak{H}$. Man nennt $f(\mathfrak{G}) = \big(f(G), \circ\big)$ das *homomorphe Bild* von $\mathfrak{G}$ (bei f). Wir haben damit folgendes bewiesen:

3.3.3. Satz. *Das homomorphe Bild $f(\mathfrak{G})$ eines Gruppoids $\mathfrak{G} = (G, \cdot)$ ist stets wieder ein Gruppoid.*

Die Homomorphie von Gruppoiden ist reflexiv (identische Abbildung I_G) und transitiv. Gilt nämlich $\mathfrak{G} \xrightarrow{f} \mathfrak{H} \xrightarrow{g} \mathfrak{K}$, so ist $g \bullet f$ ein Homomorphismus von $\mathfrak{G}$ in $\mathfrak{K}$ (Beweis der Operationstreue!). Die Symmetrie ist i. allg. nicht erfüllt (Gegenbeispiel!).

Man überzeuge sich auch, daß es mehrere Homomorphismen von einem Gruppoid $\mathfrak{G}$ in ein Gruppoid $\mathfrak{H}$ geben kann.

3.3.4. Satz. *Es sei $\mathfrak{G} = (G, \cdot)$ ein Gruppoid, in dem eine der Identitäten* (1) *bis* (33) *aus* 2.4. *gilt. Dann ist im homomorphen Bild $f(\mathfrak{G}) = \big(f(G), \circ\big)$ von $\mathfrak{G}$ (bei einem Homomorphismus von $\mathfrak{G} = (G, \cdot)$ in $\mathfrak{H} = (H, \circ)$) dieselbe Identität erfüllt wie in $\mathfrak{G}$.*

Beweis. Es gelte in $\mathfrak{G} = (G, \cdot)$ z. B. die Assoziativität, und a, b, c seien beliebige Elemente aus $f(G)$. Dann gibt es in G Elemente x, y und z mit $f(x) = a$, $f(y) = b$ sowie $f(z) = c$. Unter Ausnutzung der Operationstreue von f und der Assoziativität in $\mathfrak{G}$ erhalten wir damit

$$\begin{aligned} a \circ (b \circ c) &= f(x) \circ \big(f(y) \circ f(z)\big) = f(x) \circ f(y \cdot z) \\ &= f(x \cdot yz) = f(xy \cdot z) = f(x \cdot y) \circ f(z) \\ &= \big(f(x) \circ f(y)\big) \circ f(z) = (a \circ b) \circ c, \end{aligned}$$

so daß die Operation $\circ$ in $f(G)$ assoziativ ist. Für alle übrigen Identitäten verlaufen die Beweise analog.

Unter der Voraussetzung, daß f ein Homomorphismus von $\mathfrak{G} = (G, \cdot)$ in $\mathfrak{H} = (H, \circ)$ ist, wobei $\mathfrak{G}$ und $\mathfrak{H}$ Gruppoide sind, sind auch die folgenden Aussagen zu verstehen.

3.3.5. Satz. *Besitzt $\mathfrak{G}$ ein neutrales Element n, so ist $f(n)$ das neutrale Element in $f(\mathfrak{G})$.*

Beweis. Es sei $a \in f(G)$, beliebig. Dann existiert ein $x \in G$ mit $f(x) = a$. Wegen $x \cdot n = n \cdot x = x$ gilt dann $a \circ f(n) = f(x) \circ f(n) = f(x \cdot n) = f(x) = a = f(x) = f(n \cdot x) = f(n) \circ f(x) = f(n) \circ a$; $f(n)$ ist auch einziges neutrales Element von $f(\mathfrak{G})$ (vgl. 1.4.).

3.3.6. Satz. *Ist $\mathfrak{G}$ eine Quasigruppe, so ist $f(\mathfrak{G})$ ein D-Gruppoid.*

Beweis. Um zu zeigen, daß im Gruppoid $f(\mathfrak{G}) = (f(G), \circ)$ die Operation $\circ$ umkehrbar ist, betrachten wir die Gleichungen $a \circ x = b$ und $y \circ a = b$ für beliebige $a, b \in f(G)$. Mit $a, b \in f(G)$ gibt es in G Elemente a_1 und b_1, so daß $f(a_1) = a$ bzw.

[1]) In Analogie zu den Begriffen Untergruppe und Unterquasigruppe heißt $\mathfrak{G}_1 = (G_1, \cdot)$ *Untergruppoid* eines Gruppoids $\mathfrak{G} = (G, \cdot)$, wenn G_1 ein Komplex von $\mathfrak{G}$ und $\mathfrak{G}_1$ selbst ein Gruppoid ist; d. h. in diesem Falle also, daß G_1 bezüglich $\cdot$ abgeschlossen sein muß.

$f(b_1) = b$ gilt. Da $\mathfrak{G}$ nach Voraussetzung eine Quasigruppe ist, haben die Gleichungen $a_1 \cdot x_1 = b_1$ und $y_1 \cdot a_1 = b_1$ (eindeutig bestimmte) Lösungen x_1 bzw. y_1 in G. Folglich haben auch die Gleichungen $a \circ x = b$ und $y \circ a = b$ jeweils wenigstens eine Lösung in $f(G)$, nämlich $x = f(x_1)$ bzw. $y = f(y_1)$:

$$a \circ x = f(a_1) \circ f(x_1) = f(a_1 \cdot x_1) = f(b_1) = b$$

und

$$y \circ a = f(y_1) \circ f(a_1) = f(y_1 \cdot a_1) = f(b_1) = b.$$

Die Einzigkeit der Lösungen in $f(G)$ ist i. allg. nicht gesichert. Es gibt in der Tat homomorphe Bilder von Quasigruppen, in denen die Kürzbarkeit nicht erfüllt ist. So ist z. B. die Abbildung $f: \mathbf{C} \to \mathbf{C}^*$ mit $x \mapsto e^x$ ein Homomorphismus von der Quasigruppe $(\mathbf{C}, \square)$ mit $x \square y := 3x - 2y$ auf das D-Gruppoid $(\mathbf{C}^*, \circ)$ mit $x \circ y := x^3 \cdot y^{-2}$ (vgl. Beispiel 1.2.2.2).

Das homomorphe Bild einer endlichen Quasigruppe (Loop) ist wegen Satz 1.2.5 dagegen stets eine Quasigruppe (Loop), während das homomorphe Bild einer Gruppe sogar im Fall unendlicher Ordnung wieder eine Gruppe ist. Da nämlich die Assoziativität unter Homomorphie invariant ist, ist das homomorphe Bild einer assoziativen Quasigruppe ein assoziatives D-Gruppoid, folglich eine Gruppe. Darüber hinaus ist in einer Gruppe $\mathfrak{G} = (G, \cdot)$ das Bild des zu $x \in G$ inversen Elements $x^{-1} \in G$ mit dem Inversen des Bildes von x identisch: $f(x^{-1}) = \big(f(x)\big)^{-1}$. Da $f(\mathfrak{G})$ eine Gruppe ist, existiert zu jedem Element $a \in f(G)$ ein inverses Element $a^{-1} \in f(G)$. Wegen $a \in f(G)$ gibt es ein $x \in G$ mit $f(x) = a$. Nun gilt $a \circ a^{-1} = f(n) = f(x \cdot x^{-1}) = f(x) \circ f(x^{-1})$, also $a^{-1} = \big(f(x)\big)^{-1} = f(x^{-1})$.

Ungeachtet der Invarianz der Gruppeneigenschaften unter Homomorphie kann es indessen homomorphe Abbildungen von einem Gruppoid, das weder assoziativ noch eine Quasigruppe ist, in oder sogar auf eine Gruppe geben. Wäre die Homomorphie eine symmetrische Relation (und damit eine Äquivalenzrelation), so könnte dieser Fall nicht eintreten.

3.8.7. Beispiel

Die Abbildung $f: G \to H$ aus Beispiel 3.1.20[1]) ist ein Homomorphismus von dem nichtassoziativen Gruppoid $(G, \cdot)$ mit $G = \{0, 1, 2, 3\}$ in die Kleinsche Vierergruppe $\mathfrak{V}_4 = (H, \circ)$ mit $H = \{a, b, c, d\}$ bzw. auf deren Untergruppe $(H_1, \circ)$ mit $H_1 = \{a, b\}$. Die Operationen sind dabei durch die nachfolgenden Strukturtafeln gegeben:

$\cdot$	0	1	2	3
0	0	1	2	3
1	1	0	1	2
2	2	1	0	1
3	3	2	1	0

$f: G \to H$

$0 \mapsto a$
$1 \mapsto b$
$2 \mapsto a$
$3 \mapsto b$

$\circ$	a	b	c	d
a	a	b	c	d
b	b	a	d	c
c	c	d	a	b
d	d	c	b	a

Werden zwei Quasigruppen homomorph aufeinander abgebildet, so bleibt die Operationstreue auch zwischen den jeweiligen Umkehroperationen bestehen. Es gilt sogar der

3.3.8. Satz. *Es seien $\mathfrak{Q} = (Q, \cdot)$ und $\mathfrak{R} = (R, \circ)$ Quasigruppen. Ist f ein surjektiver*

[1]) Wir haben jetzt nur anstelle von M und N die Buchstaben G bzw. H gewählt.

Homomorphismus[1]) *von* $\mathfrak{Q}$ *auf* $\mathfrak{R}$, *so ist* f *auch ein surjektiver Homomorphismus von* $\mathfrak{Q}_i = (Q, \pi_i(\cdot))$ *auf* $\mathfrak{R}_i = (R, \pi_i(\circ))$ *mit* $i = 1, 2, \ldots, 6$.

Beweis. Der Fall $i = 1$ entspricht der Voraussetzung. Wegen Satz 2.1.10 genügt es wieder, die Behauptung für die Parastrophismen π_2 und π_6 zu beweisen. Im Fall von π_2 haben wir also zu zeigen, daß aus $(Q, \cdot) \overset{\sim}{\underset{f}{\to}} (R, \circ)$ auch $(Q, \backslash) \overset{\sim}{\underset{f}{\to}} (R, \backslash\backslash)$ folgt, wobei $\backslash = \pi_2(\cdot)$ und $\backslash\backslash = \pi_2(\circ)$ die Linksdivisionen von $\cdot$ bzw. $\circ$ sind. Für beliebige $x, y, z \in Q$ gilt $x \cdot y = z$ genau dann, wenn $x \backslash z = y$ ist, so daß $f(x \cdot y) = f(x) \circ f(y)$ auch in der Form $f(z) = f(x) \circ f(x \backslash z)$ geschrieben werden kann. Da nach Voraussetzung $R = f(Q)$ gilt und $(R, \circ)$ eine Quasigruppe ist, erhalten wir hieraus schon die gesuchte Beziehung $f(x \backslash z) = f(x) \backslash\backslash f(z)$. Mit x und y durchläuft in $x \cdot y = z$ auch z ganz Q, so daß wir ebenso x und y als unabhängige Variable wählen können. Analog zeigt man die Operationstreue von f zwischen $/ = \pi_5(\cdot)$ und $// = \pi_5(\circ)$.

Sind $\mathfrak{G} = (G, \cdot)$ und $\mathfrak{H} = (H, \circ)$ Gruppoide, für die ein Homomorphismus $f: \mathfrak{G} \overset{\sim}{\to} \mathfrak{H}$ existiert, so ist $f: \mathfrak{G} \overset{\sim}{\to} f(\mathfrak{G})$ ein surjektiver Homomorphismus. Dessen ungeachtet können $\mathfrak{G}$ und $f(\mathfrak{G})$ natürlich unterschiedlich viele Elemente haben. Es gibt nun allerdings Homomorphismen, die jede Eigenschaft eines Gruppoids (also auch die Ordnung $|\mathfrak{G}|$) auf das homomorphe Bild übertragen. Das sind gerade diejenigen Homomorphismen, die die Trägermengen von $\mathfrak{G}$ und von $f(\mathfrak{G})$ eineindeutig aufeinander abbilden.

3.3.9. Definition. Es seien $\mathfrak{G} = (G, \cdot)$ und $\mathfrak{H} = (H, \circ)$ Gruppoide. Ein Homomorphismus $f: \mathfrak{G} \overset{\sim}{\to} \mathfrak{H}$ heißt ein *Isomorphismus* (oder *isomorphe Abbildung*) von $\mathfrak{G}$ auf $\mathfrak{H}$, wenn f bijektiv ist.[2]) Wir schreiben dann hierfür $f: \mathfrak{G} \overset{\sim}{\underset{f}{\to}} \mathfrak{H}$ oder $\mathfrak{G} \overset{\sim}{\underset{f}{\leftarrow}} \mathfrak{H}$. Damit treffen die für die Homomorphismen gültigen Aussagen natürlich auch für Isomorphismen zu. Es kann durchaus mehrere Isomorphismen zwischen zwei Gruppoiden geben.

Ein Isomorphismus f von einem Gruppoid $\mathfrak{G} = (G, \cdot)$ auf sich heißt ein *Automorphismus*; d. h. mit anderen Worten, daß f ein bijektiver Endomorphismus von $\mathfrak{G}$ ist.

Im Gegensatz zur Homomorphie ist die Isomorphie in der Menge aller Gruppoide eine Äquivalenzrelation (Beweis!). Nach dem Hauptsatz über Äquivalenzrelationen (vgl. 3.1.) liefert die Isomorphie somit eine Einteilung aller Gruppoide in disjunkte Klassen. Eine solche Klasse isomorpher Gruppoide wird *abstraktes Gruppoid* genannt. Isomorphe Gruppoide unterscheiden sich in ihren Eigenschaften nicht, lediglich in der Bezeichnung ihrer Elemente.

Die Isomorphie ist insbesondere auch eine Äquivalenzrelation in der Gesamtheit aller Gruppoide, an die bestimmte Forderungen (wie z. B. die Gruppenaxiome) gestellt werden. Eine wichtige Aufgabe der Gruppentheorie ist das sogenannte *Struktur-*

[1]) Dabei heißt ein Homomorphismus $f: \mathfrak{G} \overset{\sim}{\to} \mathfrak{H}$ *surjektiv* (*injektiv*, *bijektiv*), wenn dies für die Abbildung $f: G \to H$ zutrifft.

[2]) Mitunter wird nur gefordert, daß f injektiv sein soll. Man spricht dann von einem Isomorphismus von $\mathfrak{G}$ in $\mathfrak{H}$. An dieser Stelle sei auch angemerkt, daß die (von uns hier nicht verwendeten) Begriffe Monomorphismus und Epimorphismus z. T. unterschiedlich definiert werden (vgl. etwa H. Lugowski [108], M. A. Arbib [16] und H. Kurzweil [97]).

problem, das darin besteht, jede Klasse isomorpher Gruppen so genau zu beschreiben, daß man die Rechnung in allen Gruppen der Klasse vollständig beherrscht. Für spezielle Gruppen, nämlich endliche abelsche Gruppen, ist das Strukturproblem gelöst.

Mit der folgenden Tabelle geben wir einen Überblick über die Anzahl der abstrakten Gruppen bis zur Ordnung 12:

Ordnung m	1	2	3	4	5	6	7	8	9	10	11	12
Anzahl der abstrakten Gruppen	1	1	1	2	1	2	1	5	2	2	1	5
davon abelsche	1	1	1	2	1	1	1	3	2	1	1	2

Für jede Ordnung m gibt es wenigstens eine abstrakte Gruppe, die zyklische Gruppe $\mathfrak{Z}_m$. Ist m Primzahl, so ist $\mathfrak{Z}_m$ auch die einzige. Daß es genau eine abstrakte Gruppe mit Primzahlordnung gibt, ist also gleichbedeutend mit der (oft benutzten) Formulierung, daß bis auf Isomorphie genau eine Gruppe dieser Ordnung existiert.

Beide abstrakten Gruppen der Ordnung 4, die zyklische Gruppe $\mathfrak{Z}_4$ und die Kleinsche Vierergruppe $\mathfrak{B}_4$, sind kommutativ. Die kleinste nichtkommutative Gruppe ist die symmetrische Gruppe $\mathfrak{S}_3$ der Ordnung 6.

Zur Problematik der Anzahlbestimmung der abstrakten Quasigruppen und Loops endlicher Ordnung sei auf die Monographie von J. Dénes und A. D. Keedwell [44] verwiesen (vgl. auch E. Schönhardt [157]).

3.3.10. Beispiele von Homomorphismen und Isomorphismen

1. Die Abbildung $f: \mathbf{R} \to \mathbf{R}$ mit $x \mapsto 2x$ ist
a) ein injektiver Endomorphismus sowohl von $(\mathbf{Z}, +)$ als auch von $(\mathbf{Z}, -)$. f ist nicht surjektiv. f ist kein Endomorphismus von $(\mathbf{Z}, \cdot)$[1]) bzw. $(\mathbf{Z}^*, :)$;
b) ein Isomorphismus von $(\mathbf{Z}, +)$ auf $(\mathbf{G}, +)$ mit $\mathbf{G} := \{x : x \in \mathbf{Z} \wedge 2 \mid x\}$ sowie von $(\mathbf{Z}, -)$ auf $(\mathbf{G}, -)$;
c) ein Automorphismus sowohl von $(\mathbf{R}, +)$ als auch von $(\mathbf{R}, -)$.

2. Die Abbildung $f: \mathbf{R} \to \mathbf{R}$ mit $x \mapsto x^2$ ist
a) ein Endomorphismus von $(\mathbf{R}, \cdot)$, $(\mathbf{R}^*, \cdot)$ und $(\mathbf{R}^*, :)$; f ist weder injektiv noch surjektiv. (Ein typischer Schülerfehler unterstellt, daß f auch Endomorphismus von $(\mathbf{R}, +)$ bzw. $(\mathbf{R}, -)$ ist: $(x + y)^2 = x^2 + y^2$ und $(x - y)^2 = x^2 - y^2$.);
b) ein Automorphismus von $(\mathbf{R}_+, \cdot)$, $(\mathbf{R}_+^*, \cdot)$ und $(\mathbf{R}_+^*, :)$.

3. Die Abbildung $f: \mathbf{R} \to \mathbf{R}$ mit $x \mapsto 2^x$ ist
a) ein injektiver Homomorphismus von $(\mathbf{R}, +)$ in $(\mathbf{R}, \cdot)$ sowie von $(\mathbf{R}, -)$ in $(\mathbf{R}^*, :)$. f ist nicht surjektiv;
b) ein Isomorphismus von $(\mathbf{R}, +)$ auf $(\mathbf{R}_+^*, \cdot)$ sowie von $(\mathbf{R}, -)$ auf $(\mathbf{R}_+^*, :)$.

4. Die Abbildung $f: \mathbf{R} \to \mathbf{R}$ mit $x \mapsto |x|$ ist
a) ein Endomorphismus von $(\mathbf{R}, \cdot)$, $(\mathbf{R}^*, \cdot)$ und $(\mathbf{R}^*, :)$. f ist weder injektiv noch surjektiv. f ist kein Endomorphismus von $(\mathbf{R}, +)$ bzw. $(\mathbf{R}, -)$;
b) ein surjektiver Homomorphismus von $(\mathbf{R}, \cdot)$ auf $(\mathbf{R}_+, \cdot)$, von $(\mathbf{R}^*, \cdot)$ auf $(\mathbf{R}_+^*, \cdot)$ sowie von $(\mathbf{R}^*, :)$ auf $(\mathbf{R}_+^*, :)$. f ist nicht injektiv.

[1]) In den Beispielen 3.3.10.1 bis 3.3.10.8 ist $\cdot$ die Multiplikation reeller Zahlen, in Beispiel 3.3.10.9 ist $\mathfrak{a} \cdot \mathfrak{x}$ das Skalarprodukt der Vektoren $\mathfrak{a}$ und $\mathfrak{x}$.

5. Die Abbildung $f\colon \mathbf{R} \to \mathbf{R}$ mit $x \mapsto a \cdot x$ ($a \in \mathbf{R}^*$, beliebig, fest) ist ein Automorphismus von $(\mathbf{R}, \circ)$, wobei $x \circ y := \frac{x+y}{2}$ das arithmetische Mittel der reellen Zahlen x und y ist. Der Operationstreue von f entspricht die *Jensensche Funktionalgleichung* $f\left(\frac{x+y}{2}\right) = \frac{f(x)+f(y)}{2}$, deren allgemeinste stetige Lösung gerade $f(x) = ax + b$ ($a \neq 0$) ist (vgl. J. Aczél [3]).[1])

6. Die Abbildung $f\colon \mathbf{R} \to \mathbf{R}$ mit $x \mapsto a^x$ ($a \in \mathbf{R}^*$, $a \neq 1$, beliebig, fest) ist ein Isomorphismus von $(\mathbf{R}, \circ)$ auf $(\mathbf{R}_+^*, *)$, wobei $x \circ y := \frac{x+y}{2}$ das arithmetische und $x * y := \sqrt{x \cdot y}$ das geometrische Mittel der reellen Zahlen x und y ist.

7. Die Abbildung $\sin\colon \mathbf{R} \to D$, wobei D die Menge der ebenen Drehungen (um einen festen Punkt) ist, ist ein surjektiver Homomorphismus von der additiven Gruppe $(\mathbf{R}, +)$ der reellen Zahlen auf die Gruppe $(D, \bullet)$ der ebenen Drehungen bezüglich der Nacheinanderausführung (vgl. G. Pickert [135]). $\sin$ ist nicht injektiv. Dagegen ist $\sin$ (und ebenso $\cos$) kein Endomorphismus von $(\mathbf{R}, +)$ — ein ebenfalls „beliebter" Schülerfehler: $\sin(x+y) = \sin x + \sin y$.

8. Es seien M und N beliebige Mengen. Mit $\mathfrak{P}(M) := \{X : X \subseteq M\}$ bezeichnen wir die Potenzmenge von M. Dann ist die Abbildung $f\colon \mathfrak{P}(M) \to \mathfrak{P}(M)$ mit $X \mapsto M \backslash X$ ein Isomorphismus von $(\mathfrak{P}(M), \cap)$ auf $(\mathfrak{P}(M), \cup)$. Diese Abbildung heißt *de-Morgansscher Isomorphismus*, da man zum Beweis der Operationstreue die de-Morganschen Regeln (vgl. MfL Bd. 1, 1.4.) benutzt.

9. Auch das Skalar- und Vektorprodukt lassen sich als Homomorphismen deuten (wobei ein Faktor dann festbleibt). So bildet $f\colon V \to \mathbf{R}$ mit $\mathfrak{x} \mapsto \mathfrak{a} \cdot \mathfrak{x}$ die Vektoraddition $\oplus$ operationstreu auf die Addition reeller Zahlen ab, während $g\colon V \to V$ mit $\mathfrak{x} \mapsto \mathfrak{a} \times \mathfrak{x}$ ($\mathfrak{a} \neq \mathfrak{o}$) ein Endomorphismus von $(V, \oplus)$ ist. Dabei ist V die Menge der Vektoren des dreidimensionalen euklidischen Vektorraumes.

Es fällt auf, daß eine Reihe von Aussagen, die wir im Zusammenhang mit der Faktorisierung eines Gruppoids nach einer Kongruenzrelation bewiesen haben, mitunter fast wörtlich auch auf den Übergang von einem Gruppoid auf ein homomorphes Bild zutreffen. Es ist deshalb sicherlich nicht verwunderlich, daß zwischen Kongruenzrelationen und homomorphen Abbildungen ein enger Zusammenhang besteht. Darauf gehen wir im folgenden ein.

3.3.11. Satz. *Es seien $\mathfrak{G} = (G, \cdot)$ und $\mathfrak{H} = (H, \circ)$ Gruppoide, f ein Homomorphismus von $\mathfrak{G}$ in $\mathfrak{H}$. Dann ist der Kern des Homomorphismus, also $\ker f = f^{-1} \bullet f$, eine Kongruenzrelation in $\mathfrak{G}$.*

Beweis. Nach Satz 3.1.13 ist der Kern der Abbildung $f\colon G \to H$ eine Äquivalenzrelation in G. Bleibt also zu zeigen, daß diese Äquivalenzrelation $R = \ker f = f^{-1} \bullet f = \{(x, y) : f(x) = f(y)\}$ mit der Operation $\cdot$ verträglich ist. Es gelte deshalb $a\,R\,b$ und $c\,R\,d$ für beliebige $a, b, c, d \in G$, also $f(a) = f(b)$ und $f(c) = f(d)$. Folglich ist $f(a) \circ f(c) = f(b) \circ f(d)$, so daß wir mit (3.3.2) $f(a \cdot c) = f(b \cdot d)$ oder $a \cdot c\,R\,b \cdot d$ erhalten.

In unserem Beispiel 3.3.7 ist der Kern des Homomorphismus $f\colon (G, \cdot) \overset{\sim}{\to} (H, \circ)$ die Menge $\ker f = I_G \cup \{(0, 2), (2, 0), (1, 3), (3, 1)\}$; vgl. Beispiel 3.1.20 und Abb. 21. Daß $\ker f$ eine Kongruenzrelation in $(G, \cdot)$ ist, hatten wir — nebenbei gesagt — bereits in Beispiel 3.2.5 festgestellt.

[1]) Andere bekannte Funktionalgleichungen, wie z. B. die *Cauchyschen Grundgleichungen* $f(x+y) = f(x) + f(y)$, $f(x \cdot y) = f(x) \cdot f(y)$ und $f(x+y) = f(x) \cdot f(y)$, stehen in den Beispielen 3.3.10.1 bis 3.3.10.3 für die Operationstreue von f.

3.3.12. Korollar. *$f\colon \mathfrak{G} \to \mathfrak{H}$ ist ein Isomorphismus genau dann, wenn* $\ker f = I_G$ *und* $f(\mathfrak{G}) = \mathfrak{H}$ *ist.*

Zum Beweis verweisen wir auf Korollar 3.1.14.

Anschaulich sind also zwei Gruppoide in ihrer Struktur „um so ähnlicher", „je kleiner" der Kern $\ker f$ des entsprechenden Homomorphismus f ist. Das Maß für die Vergröberung, für die Deformation der Ausgangsstruktur läßt sich also sehr gut am Kern ablesen. Im Fall eines Isomorphismus ist die Deformation gleich Null, d. h., der Kern $\ker f$ schrumpft auf die identische Abbildung zusammen.[1])

3.3.13. Korollar. *Ist $f\colon \mathfrak{G} \to \mathfrak{H}$ ein surjektiver Homomorphismus von einem Gruppoid $\mathfrak{G} = (G, \cdot)$ auf ein K-Gruppoid $\mathfrak{H} = (H, \circ)$, so ist die Kongruenzrelation $R = \ker f$ in $\mathfrak{G}$ normal.*

Beweis. Es seien a, b und c beliebige Elemente aus G; $a \cdot c\, R\, b \cdot c$ bedeutet dann $f(a \cdot c) = f(b \cdot c)$ bzw. $f(a) \circ f(c) = f(b) \circ f(c)$. Wegen der Kürzbarkeit in $\mathfrak{H}$ folgt hieraus $f(a) = f(b)$, also aRb. Analog folgt aRb aus $c \cdot a\, R\, c \cdot b$.

Ist in Korollar 3.3.13 $\mathfrak{G}$ (und damit auch $\mathfrak{H}$) eine Gruppe, so ist wegen Satz 3.2.15 $\mathfrak{N} = (N, \cdot)$ mit $N := [n_G]_{\ker f}$ ein Normalteiler von $\mathfrak{G}$. Dabei gilt $N = \{x\colon x \in G \wedge f(x) = f(n_G) = n_H\}$, wobei n_G bzw. n_H das neutrale Element in $\mathfrak{G}$ bzw. $\mathfrak{H}$ ist. Ferner stimmen die zur Kongruenzrelation $\ker f$ gehörigen Kongruenzklassen mit den Nebenklassen nach dem Normalteiler $\mathfrak{N}$ überein (Korollar 3.2.18). Aus diesem Grunde bezeichnet man die Kongruenzklasse $N = [n_G]_{\ker f}$ ebenfalls als *Kern des Homomorphismus* $f\colon \mathfrak{G} \to \mathfrak{H}$. Zur Unterscheidung von $\ker f$ schreiben wir dafür $N = [n_G]_{\mathrm{Ker}\, f}$ $= \mathrm{Ker}\, f$. Ist f ein Isomorphismus, so schrumpft $\mathrm{Ker}\, f$ auf die Einermenge $\{n_G\}$ zusammen.

Man beachte stets, daß $\ker f$ als Kongruenzrelation eine Teilmenge von $G \times G$, $\mathrm{Ker}\, f$ dagegen eine Teilmenge von G ist.

Für die Homomorphismen aus 3.3.10., die nicht injektiv sind, geben wir jetzt den jeweiligen Kern $\ker f$ an. Ist das Ausgangsgruppoid eine Gruppe, so führen wir darüber hinaus auch $N = \mathrm{Ker}\, f$ an. Ist f injektiv, d. h. ein Automorphismus, ein Isomorphismus oder ein injektiver Homomorphismus, so gilt stets $\ker f = I_G$ und im Fall einer Gruppe noch $\mathrm{Ker}\, f = \{n_G\}$ (vgl. die Tabelle).

Beispiel	$\ker f = \{(x, y)\colon f(x) = f(y)\}$	$\mathrm{Ker}\, f = \{x\colon f(x) = f(n)\}$
3.3.10.2a)	$I_{\mathbf{R}} \cup \{(x, -x)\colon x \in \mathbf{R}\}$ $I_{\mathbf{R}^*} \cup \{(x, -x)\colon x \in \mathbf{R}^*\}$ $I_{\mathbf{R}^*} \cup \{(x, -x)\colon x \in \mathbf{R}^*\}$	— $\{1, -1\}$ —
3.3.10.4a), b)	wie bei 3.3.10.2a)	wie bei 3.3.10.2a)
3.3.10.7	$\{(x, y)\colon x \equiv y \bmod 2\pi\}$	$\{x\colon x \equiv 0 \bmod 2\pi\}$
3.3.10.9	$\{(\mathfrak{x}, \mathfrak{y})\colon \mathfrak{a} \cdot \mathfrak{x} = \mathfrak{a} \cdot \mathfrak{y}\}$ $\{(\mathfrak{x}, \mathfrak{y})\colon \mathfrak{a} \times \mathfrak{x} = \mathfrak{a} \times \mathfrak{y}\}$	$\{\mathfrak{x}\colon \mathfrak{x} \perp \mathfrak{a}\}$ $\{\mathfrak{x}\colon \mathfrak{x} \parallel \mathfrak{a}\}$

[1]) Zur Möglichkeit, Homomorphismen mit Hilfe von Zahnrädern zu veranschaulichen, sei auf den Artikel [87] von A. Kirsch verwiesen.

Jedem Homomorphismus $f: \mathfrak{G} \to \mathfrak{H}$ entspricht, wie wir gesehen haben, eine Kongruenzrelation in $\mathfrak{G}$. Ist jetzt umgekehrt in einem Gruppoid eine Kongruenzrelation gegeben, so läßt sich ihr ihrerseits genau ein Homomorphismus zuordnen. Es gilt sogar der

3.3.14. Satz. *Es sei $\mathfrak{G} = (G, \cdot)$ ein Gruppoid, R eine Kongruenzrelation in $\mathfrak{G}$. Dann ist die zu R gehörige kanonische Abbildung* kan $R: G \to G/R$ *mit $x \mapsto [x]_R$ ein surjektiver Homomorphismus von $\mathfrak{G}$ auf das Faktorgruppoid $\mathfrak{G}/R$.*

kan $R: \mathfrak{G} \to \mathfrak{G}/R$ heißt der zur Kongruenzrelation R gehörige *kanonische* (oder *natürliche*) *Homomorphismus.*

Dieser kanonische Homomorphismus kan R läßt sich auch als ein Homomorphismus f von $\mathfrak{G} = (G, \cdot)$ auf ein Gruppoid $\mathfrak{H} = (H, \circ)$ interpretieren. Zu diesem Zwecke verketten wir ihn mit einem Isomorphismus $g: \mathfrak{G}/R \rightleftarrows \mathfrak{H}$, so daß $g \bullet$ kan $R = f$ gilt. Jede NAF des kanonischen Homomorphismus kan R mit einem solchen Isomorphismus g ist dann ein surjektiver Homomorphismus $f: \mathfrak{G} \to \mathfrak{H}$ (siehe Abb. 23).

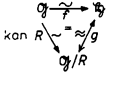


Abb. 23

Beweis. Da nach Voraussetzung R eine Kongruenzrelation in $\mathfrak{G}$ ist, ist die Existenz des zugehörigen Faktorgruppoids $\mathfrak{G}/R = (G/R, *)$ gemäß (3.2.4) stets gewährleistet. Wir haben demnach zu zeigen, daß die kanonische Abbildung kan $R: G \to G/R$ mit $x \mapsto [x]_R$ operationstreu ist. Daß sie surjektiv ist, versteht sich von selbst. Es seien x, y beliebige Elemente aus G. Dann gilt kan $R(x \cdot y) = [x \cdot y]_R = [x]_R * [y]_R$ = kan $R(x) *$ kan $R(y)$. Ferner ist die Nacheinanderausführung der Abbildungen kan $R: G \to G/R$ und $g: G/R \to H$ eine Abbildung $f: G \to H$. Mit g ist offenbar auch f surjektiv. Die Homomorphie ist eine transitive Relation in der Menge aller Gruppoide, so daß f auch operationstreu ist.

Wählt man, um bei dem Gruppoid $(G, \cdot)$ aus Beispiel 3.3.7 zu bleiben, die Relation $R = I_G \cup \{(0, 2), (2, 0), (1, 3), (3, 1)\}$ als Kongruenzrelation, so hat die zu R gehörige kanonische Abbildung kan $R: G \to G/R$ mit $x \mapsto [x]_R$ die Gestalt $0 \mapsto [0]_R$, $1 \mapsto [1]_R$, $2 \mapsto [0]_R$ und $3 \mapsto [1]_R$. Sie ist der kanonische Homomorphismus von $(G, \cdot)$ auf $(G/R, *)$. Die Strukturtafel des Faktorgruppoids ist in Beispiel 3.2.5 angegeben.

Ist das Gruppoid $\mathfrak{G} = (G, \cdot)$ in Satz 3.3.14 eine Quasigruppe und R eine normale Kongruenzrelation in $\mathfrak{G}$, dann ist kan $R: G \to G/R$ ein surjektiver Homomorphismus von $\mathfrak{G}$ auf die Faktorquasigruppe $\mathfrak{G}/R$. Zum Beweis verweisen wir auf Korollar 3.2.11. Ist $\mathfrak{G} = (G, \cdot)$ sogar eine Gruppe, so erlaubt der Hauptsatz über Kongruenzrelationen in Gruppen (Satz 3.2.15), die gegebene Kongruenzrelation R durch den entsprechenden Normalteiler $\mathfrak{N} = (N, \cdot)$ von $\mathfrak{G}$ zu ersetzen. Zu jedem Normalteiler $\mathfrak{N} = (N, \cdot)$ einer Gruppe $\mathfrak{G} = (G, \cdot)$ existiert also ein Homomorphismus f von $\mathfrak{G}$ auf die Faktorgruppe $\mathfrak{G}/\mathfrak{N} = (G/\mathfrak{N}, *)$, wobei $N = \operatorname{Ker} f$ gilt.

Als Abschluß unserer Betrachtungen zeigen wir jetzt noch, daß Satz 3.3.14 in gewissem Sinne umkehrbar ist, d. h., wir werden zeigen, daß durch die kanonischen

Homomorphismen eines Gruppoids $\mathfrak{G} = (G, \cdot)$ auf seine sämtlichen Faktorgruppoide alle homomorphen Abbildungen dieses Gruppoids erschöpft werden.

3.3.15. Homomorphiesatz für Gruppoide. *Es seien* $\mathfrak{G} = (G, \cdot)$ *und* $\mathfrak{H} = (H, \circ)$ *Gruppoide, f sei ein surjektiver Homomorphismus von* $\mathfrak{G}$ *auf* $\mathfrak{H}$ *und* $\tilde{f}: G \to G/\ker f$ *mit* $x \mapsto [x]_{\ker f}$ *der zur Kongruenzrelation* $\ker f$ *gehörige kanonische Homomorphismus* $\tilde{f} = \mathrm{kan}\,(\ker f)$ *von* $\mathfrak{G}$ *auf* $\mathfrak{G}/\ker f$. *Dann gibt es einen (eindeutig bestimmten) Isomorphismus g von* $\mathfrak{G}/\ker f$ *auf* $\mathfrak{H}$, *so daß* $g \bullet \tilde{f} = f$ *gilt* (siehe Abb. 24).[1])

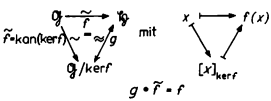


Abb. 24

Beweis. Wegen Satz 3.1.19 ist das Diagramm in Abb. 20 ($M = G$, $N = H$) kommutativ. Folglich gibt es eine (eindeutig bestimmte) 1–1-Abbildung g: $G/\ker f \to H$ mit $g \bullet \tilde{f} = f$. Da nach Voraussetzung f surjektiv ist, gilt dies auch für g. Damit ist g bijektiv. Es seien $[x]_{\ker f}$, $[y]_{\ker f}$ beliebige Elemente aus $G/\ker f$. Dann gilt

$$g([x] * [y]) = g([x \cdot y]) = f(x \cdot y) = f(x) \circ f(y) = g([x]) \circ g([y]),$$

d. h., g ist operationstreu, also ein Isomorphismus von $\mathfrak{G}/\ker f$ auf $\mathfrak{H}$.

Jeder surjektive Homomorphismus f von $\mathfrak{G}$ auf $\mathfrak{H}$ läßt sich also in den zur Kongruenzrelation $\ker f$ gehörigen kanonischen Homomorphismus des Gruppoids $\mathfrak{G}$ auf $\mathfrak{G}/\ker f$ und einen Isomorphismus von $\mathfrak{G}/\ker f$ auf $\mathfrak{H}$ (mit $[x]_{\ker f} \mapsto f(x)$) aufspalten. Das heißt mit anderen Worten: Das homomorphe Bild $\mathfrak{H} = f(\mathfrak{G})$ von $\mathfrak{G}$ bei einem surjektiven Homomorphismus f läßt sich derart isomorph auf ein bestimmtes Faktorgruppoid von $\mathfrak{G}$ abbilden, daß der Homomorphismus f in den kanonischen Homomorphismus des Gruppoids $\mathfrak{G}$ auf dieses Faktorgruppoid übergeht. Noch anders gesagt: Die homomorphen Bilder eines Gruppoids $\mathfrak{G}$ sind (bis auf Isomorphie) gerade die sämtlichen Faktorgruppoide des Gruppoids nach allen Kongruenzrelationen in $\mathfrak{G}$.

Läßt man in Satz 3.3.15 die Voraussetzung $f(\mathfrak{G}) = \mathfrak{H}$ fallen, so nimmt die Behauptung des Homomorphiesatzes folgende Gestalt an: Es gibt (genau) einen Isomorphismus g von $\mathfrak{G}/\ker f$ auf $f(\mathfrak{G})$ und (genau) einen injektiven Homomorphismus[2]) h von $f(\mathfrak{G})$ in $\mathfrak{H}$, so daß $h \bullet g \bullet \tilde{f} = f$ gilt. Wenn f surjektiv ist, d. h. $f(\mathfrak{G}) = \mathfrak{H}$ gilt, ist h gerade die identische Abbildung I_H von H, so daß wir wieder $g \bullet \tilde{f} = f$ erhalten.

Für unser Beispiel 3.3.7 stellt sich der Homomorphiesatz folgendermaßen dar: $f = \begin{pmatrix} 0 & 1 & 2 & 3 \\ a & b & a & b \end{pmatrix}$[3])

[1]) Wenn man Satz 3.3.11 in den Homomorphiesatz integriert, nimmt er die folgende Gestalt an: *Es seien* $\mathfrak{G} = (G, \cdot)$ *und* $\mathfrak{H} = (H, \circ)$ *Gruppoide, und* $f: \mathfrak{G} \to \mathfrak{H}$ *sei ein surjektiver Homomorphismus. Dann gibt es in* $\mathfrak{G}$ *eine Kongruenzrelation R, so daß das Faktorgruppoid* $\mathfrak{G}/R$ *isomorph zu* $\mathfrak{H}$ *ist.*

[2]) Man sagt auch, $f(\mathfrak{G})$ wird isomorph in $\mathfrak{H}$ eingebettet.

[3]) Diese den Permutationen entlehnte Schreibweise soll bedeuten: $0 \mapsto a$, $1 \mapsto b$, $2 \mapsto a$ und $3 \mapsto b$.

ist ein surjektiver Homomorphismus von $(G, \cdot)$ auf $(H_1, \circ)$. Der zu $\ker f = I_G \cup \{(0,2), (2,0), (1,3), (3,1)\}$ gehörige kanonische Homomorphismus $\tilde{f} = \operatorname{kan}(\ker f)$ von $(G, \cdot)$ auf $(G/\ker f, *)$ hat die Gestalt $\tilde{f} = \begin{pmatrix} 0 & 1 & 2 & 3 \\ [0] & [1] & [0] & [1] \end{pmatrix}$. Dann ist $g = \begin{pmatrix} [0] & [1] \\ a & b \end{pmatrix}$ der (eindeutig bestimmte) Isomorphismus von $(G/\ker f, *)$ auf $(H_1, \circ)$. Man überzeugt sich mühelos, daß

$$\begin{pmatrix} [0] & [1] \\ a & b \end{pmatrix} \bullet \begin{pmatrix} 0 & 1 & 2 & 3 \\ [0] & [1] & [0] & [1] \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 & 3 \\ a & b & a & b \end{pmatrix}$$

gilt. Faßt man f dagegen als Homomorphismus von $(G, \cdot)$ in die Kleinsche Vierergruppe $\mathfrak{B}_4 = (H, \circ)$ auf, so bleibt g (in der angegebenen Weise) ein Isomorphismus von $(G/\ker f, *)$ auf $(H_1, \circ)$, wobei $H_1 = f(G) \subset H$ gilt. Der injektive Homomorphismus h von $(f(G), \circ)$ in $(H, \circ)$ hat dann die Gestalt $h = \begin{pmatrix} a & b \\ a & b \end{pmatrix}$. Insgesamt erhalten wir somit

$$\begin{pmatrix} a & b \\ a & b \end{pmatrix} \bullet \begin{pmatrix} [0] & [1] \\ a & b \end{pmatrix} \bullet \begin{pmatrix} 0 & 1 & 2 & 3 \\ [0] & [1] & [0] & [1] \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 & 3 \\ a & b & a & b \end{pmatrix}.$$

Sind in Satz 3.3.15 $\mathfrak{G}$ und $\mathfrak{H}$ Gruppen, so entspricht der Kongruenzrelation $\ker f$ nach Satz 3.2.15 genau ein Normalteiler $\mathfrak{N}$ von $\mathfrak{G}$. Die durch $\mathfrak{N}$ induzierte Abbildung von G auf $G/\mathfrak{N}$ stimmt dabei mit dem kanonischen Homomorphismus $\tilde{f} = \operatorname{kan}(\ker f)$ überein. Mit der Bemerkung in der Fußnote 1 auf S. 96 ergäbe sich damit z. B. die folgende Form des Homomorphiesatzes für Gruppen: Ist $f\colon \mathfrak{G} \to \mathfrak{H}$ ein surjektiver Gruppenhomomorphismus, dann existiert in $\mathfrak{G}$ ein Normalteiler $\mathfrak{N}$, so daß die Faktorgruppe $\mathfrak{G}/\mathfrak{N}$ isomorph zu $\mathfrak{H}$ ist.

Die Theorie der homomorphen Abbildungen bricht mit dem Homomorphiesatz nicht ab. Beispielsweise lassen sich die aus der Gruppentheorie bekannten Isomorphiesätze ebenfalls verallgemeinern und hier unmittelbar anschließen. Weitere Resultate findet der interessierte Leser vor allem bei R. H. Bruck [32] und V. D. Belousov [21].

3.3.16. Aufgaben

1. Man zeige, daß das isomorphe Bild einer Quasigruppe stets wieder eine Quasigruppe ist.

2. Man zeige, daß $f\colon \mathbf{R}_+{}^* \to M$ mit $M := \{x : x \in \mathbf{R} \wedge x > 1\}$ und $x \mapsto x + 1$ ein Isomorphismus zwischen den Quasigruppen $(\mathbf{R}_+{}^*, :)$ und $(M, \circ)$ mit $x \circ y := \dfrac{x + y - 2}{y - 1}$ ist. (Werden dabei die jeweiligen rechts-neutralen Elemente aufeinander abgebildet?)

3. Es sei $\mathfrak{G} = (G, \cdot)$ eine Gruppe. Man zeige, daß $f\colon G \to G$ mit $x \mapsto x^{-1}$ genau dann ein Automorphismus von $\mathfrak{G}$ ist, wenn $\mathfrak{G}$ abelsch ist.

4. Man zeige, daß es keinen Isomorphismus zwischen den Gruppen $(\mathbf{R}, +)$ und $(\mathbf{R}^*, \cdot)$ und keinen Homomorphismus von $(\mathbf{R}, \circ)$ mit $x \circ y := \dfrac{x + y}{2}$ auf $(\mathbf{R}, +)$ gibt.

5. Jede natürliche Zahl $n > 1$ läßt sich auf genau eine Weise in der Form $n = \prod_{\lambda=1}^{l} p_\lambda{}^{\nu_\lambda}$ mit paarweise verschiedenen Primzahlen $p_1 < p_2 < \cdots < p_l$ und von Null verschiedenen natürlichen Exponenten $\nu_1, \nu_2, \ldots, \nu_l$ darstellen (Primzahlpotenzdarstellung (vgl. MfL Bd. 1, 3.7.)). Dann ist $f\colon \mathbf{N} \to \mathfrak{P}(\mathbf{N})$ mit $n \mapsto \{p_1{}^{\nu_1}, p_2{}^{\nu_2}, \ldots, p_l{}^{\nu_l}\}$ und $f(0) := \mathbf{N}$ sowie $f(1) := \emptyset$ ein injektiver Homomorphismus sowohl von $(\mathbf{N}, \sqcap)$ in $(\mathfrak{P}(\mathbf{N}), \cap)$ als auch von $(\mathbf{N}, \sqcup)$ in $(\mathfrak{P}(\mathbf{N}), \cup)$, wobei $x \sqcap y := \mathrm{ggT}(x, y)$ und $x \sqcup y := \mathrm{kgV}(x, y)$ gelten (Homomorphismus von Stone). (Beispiel:

$f(12) = \{2^2, 3^1\} = \{3, 4\}$, $f(15) = \{3^1, 5^1\} = \{3, 5\}$, $f(12 \sqcap 15) = f(3) = \{3\}$ und $f(12) \cap f(15) = \{3\}$.) Man zeige die Eineindeutigkeit und Operationstreue von f und überzeuge sich, daß die Zahlen 0 und 1 nicht ausgeklammert werden müssen, wenn man ihre Bilder in der angegebenen Weise definiert.

6. Man gebe einen surjektiven Endomorphismus an, der kein Automorphismus ist.

7. Man zeige, daß verschiedene Kongruenzrelationen isomorphe Faktorgruppoide liefern können.

4. Isotopie

Wie die Homomorphie ist auch die Isotopie eine Abschwächung der Isomorphie. Während die (Homo-, Endo-, Iso- und Auto-) Morphismen jedoch jeweils aus nur einer Abbildung bestehen, setzen sich die Topismen[1]) aus Tripeln von Abbildungen zusammen. Die Operationstreue der Morphismen wird dabei auf eine sinnvolle Weise verallgemeinert.

Speziell die Isotopie spielt in der Theorie der Quasigruppen, vor allem aber auch in der Anwendung dieser Theorie eine große Rolle (vgl. 11.2.). In der Gruppentheorie erweist sich dieser Begriff dagegen als bedeutungslos, da isotope Gruppen stets isomorph sind. Dessen ungeachtet muß nicht jedes isotope Bild einer Gruppe selbst wieder eine Gruppe sein. Darüber hinaus sind die Topismen ein sehr wirksames Instrument zur Konstruktion neuer Operationen; sie übertragen dabei i. allg. jedoch nicht alle algebraischen Eigenschaften der Ausgangsoperationen.

4.1. Isotope und Hauptisotope

4.1.1. Definition. Es seien $\mathfrak{G} = (G, \cdot)$ und $\mathfrak{H} = (H, \circ)$ Gruppoide. Ein (geordnetes) Tripel $T = (f, g, h)$ bijektiver Abbildungen f, g und h von G auf H heißt ein *Isotopismus* (oder *isotope Abbildung*) von $\mathfrak{G}$ auf $\mathfrak{H} :\Leftrightarrow$

4.1.2. $\bigwedge_{x,y \in G} h(x \cdot y) = f(x) \circ g(y)$.

Wenn $T = (f, g, h)$ ein Isotopismus von $\mathfrak{G}$ auf $\mathfrak{H}$ ist, schreiben wir dafür $T: \mathfrak{G} \overset{\sim}{\to} \mathfrak{H}$ oder $\mathfrak{G} \overset{\sim}{\underset{T}{\to}} \mathfrak{H}$. Wir sagen auch, $\mathfrak{G}$ ist *isotop* zu $\mathfrak{H}$ oder $\mathfrak{H}$ ist ein *Isotop* (*isotopes Bild*) von $\mathfrak{G}$. In Analogie zur Homomorphie schreiben wir dafür $T(\mathfrak{G}) = \mathfrak{H}$.[2]) Der Bedingung (4.1.2) entspricht — man vgl. mit Abb. 22 — die Kommutativität des Diagramms in Abb. 25.

Für die Operationstreue eines Homomorphismus $f: \mathfrak{G} \to \mathfrak{H}$ gilt bekanntlich die einprägsame „Formel":

Bild des Produktes = Produkt der Bilder.

---

[1]) Wir beschränken uns im Rahmen dieses Buches auf die Behandlung von Iso- und Autotopismen. Zur Homotopie sei z. B. auf V. D. Belousov [21] und N. I. Prodan [139] verwiesen.

[2]) Würde man in Definition 4.1.1 auf die Surjektivität der drei Abbildungen f, g und h verzichten, so ließe sich dennoch — in Analogie zu Satz 3.3.3 — erreichen, daß das isotope Bild $T(\mathfrak{G})$ von $\mathfrak{G}$ ein Untergruppoid von $\mathfrak{H}$ wäre.

Daß wir in gewissem Sinne auch bei Isotopismen bei dieser Merkregel bleiben können, zeigt die folgende Veranschaulichung. Wir müssen uns dabei natürlich stets bewußt sein, daß jetzt drei Abbildungen zu berücksichtigen sind (siehe Abb. 26).

Ist $T = (f, g, h)$ ein Isotopismus von $\mathfrak{G}$ auf $\mathfrak{H}$ und stimmen alle drei Komponenten des Tripels überein ($f = g = h$), so geht (4.1.2) in $f(x \cdot y) = f(x) \circ f(y)$ über. $T = (f, f, f)$ ist also ein Isomorphismus von $\mathfrak{G}$ auf $\mathfrak{H}$.

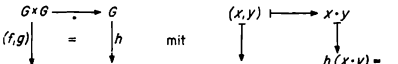


Abb. 25

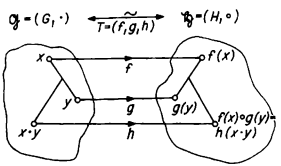


Abb. 26

Ein Isotopismus $T = (f, g, h)$ von einem Gruppoid $\mathfrak{G} = (G, \cdot)$ auf sich heißt ein *Autotopismus*. (4.1.2) nimmt in diesem Fall also die Gestalt $h(x \cdot y) = f(x) \cdot g(y)$ an. Es gilt $T(\mathfrak{G}) = \mathfrak{G}$. Sind in einem Autotopismus $T = (f, g, h)$ alle Komponenten gleich, so ist $T = (f, f, f)$ ein Automorphismus von $\mathfrak{G}$. Am Rande sei vermerkt, daß die dritte Komponente h eines Autotopismus Quasiautomorphismus genannt wird.

Ein weiterer Spezialfall, der im folgenden eine große Rolle spielen wird, liegt vor, wenn die dritte (oder Hauptkomponente) eines Isotopismus $T = (f, g, h)$ von $\mathfrak{G}$ auf $\mathfrak{H}$ die identische Abbildung I_G von G ist.

4.1.3. Definition. Es seien $\mathfrak{G}_1 = (G, \cdot)$ und $\mathfrak{G}_2 = (G, \circ)$ Gruppoide. Ein Isotopismus $T = (f, g, h)$ von $\mathfrak{G}_1$ auf $\mathfrak{G}_2$ heißt *Hauptisotopismus*, wenn $h = I_G$ gilt.

Ist also $(G, \circ)$ ein Hauptisotop von $(G, \cdot)$, so geht (4.1.2) über in

4.1.4. $\bigwedge_{x,y \in G} x \cdot y = f(x) \circ g(y)$.

Wegen $h = I_G$ müssen $\mathfrak{G}_1$ und $\mathfrak{G}_2$ notwendigerweise dieselbe Trägermenge haben.

4.1.5. Beispiele isotoper Gruppoide

1. Das Tripel $T_1 = (f_1, g_1, h_1)$ mit den bijektiven Abbildungen $f_1, g_1, h_1 \colon G \to H$ mit $f_1 = h_1 = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ b_4 & b_1 & b_2 & b_3 \end{pmatrix}$ und $g_1 = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ b_3 & b_2 & b_1 & b_4 \end{pmatrix}$ ist ein Isotopismus von $\mathfrak{G}_1 = (G, \cdot)$ auf $\mathfrak{H} = (H, \circ)$, wobei beide Gruppoide durch die folgenden Strukturtafeln gegeben sind:

·	a_1	a_2	a_3	a_4
a_1	a_1	a_2	a_3	a_4
a_2	a_2	a_1	a_2	a_3
a_3	a_3	a_2	a_1	a_2
a_4	a_4	a_3	a_2	a_1

∘	b_1	b_2	b_3	b_4
b_1	b_1	b_4	b_1	b_2
b_2	b_4	b_1	b_2	b_1
b_3	b_1	b_2	b_3	b_4
b_4	b_2	b_1	b_4	b_3

Beide Gruppoide sind kommutativ und besitzen ein neutrales Element (a_1 bzw. b_2). Sie sind weder Halbgruppen noch Quasigruppen. Sie sind nicht isomorph, da $\mathfrak{G}_1$ unipotent ist ($x \cdot x = y \cdot y$), $\mathfrak{H}$ aber nicht.

Das Tripel $T_2 = (f_2, g_2, h_2)$ mit den Permutationen $f_2, g_2, h_2 \colon G \to G$ mit $f_2 = h_2 = I_G$ und $g_2 = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ a_4 & a_3 & a_2 & a_1 \end{pmatrix}$ ist ein Hauptisotopismus von $\mathfrak{G}_1 = (G, \cdot)$ auf das Gruppoid $\mathfrak{G}_2 = (G, *)$ mit

*	a_1	a_2	a_3	a_4
a_1	a_4	a_3	a_2	a_1
a_2	a_3	a_2	a_1	a_2
a_3	a_2	a_1	a_2	a_3
a_4	a_1	a_2	a_3	a_4

In $\mathfrak{G}_2$ ist a_4 neutrales Element; $\mathfrak{G}_1$ und $\mathfrak{G}_2$ sind nicht isomorph.

2. Die nichtassoziativen Quasigruppen $\mathfrak{Q}_1 = (Q, \cdot)$ und $\mathfrak{R} = (R, \circ)$ mit den lateinischen Quadraten

·	q_1	q_2	q_3	q_4
q_1	q_4	q_3	q_1	q_2
q_2	q_3	q_1	q_2	q_4
q_3	q_1	q_2	q_4	q_3
q_4	q_2	q_4	q_3	q_1

∘	r_1	r_2	r_3	r_4
r_1	r_1	r_4	r_2	r_3
r_2	r_3	r_1	r_4	r_2
r_3	r_2	r_3	r_1	r_4
r_4	r_4	r_2	r_3	r_1

sind isotop vermöge $T_1 = (f_1, g_1, h_1)$ mit $f_1 = h_1 = \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ r_4 & r_1 & r_2 & r_3 \end{pmatrix}$ und $g_1 = \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ r_3 & r_2 & r_1 & r_4 \end{pmatrix}$ (vgl. T_1 in Beispiel 1).
$\mathfrak{Q}_1$ und $\mathfrak{R}$ sind nicht isomorph. Die Quasigruppe $\mathfrak{Q}_2 = (Q, *)$ mit dem lateinischen Quadrat

*	q_1	q_2	q_3	q_4
q_1	q_2	q_4	q_1	q_3
q_2	q_4	q_3	q_2	q_1
q_3	q_1	q_2	q_3	q_4
q_4	q_3	q_1	q_4	q_2

ist ein Hauptisotop von $\mathfrak{Q}_1$. Der Hauptisotopismus hat die Gestalt $T_2 = (f_2, g_2, h_2)$ mit

$$f_2 = \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ q_1 & q_2 & q_4 & q_3 \end{pmatrix}, \quad g_2 = \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ q_2 & q_4 & q_3 & q_1 \end{pmatrix}, \quad h_2 = I_Q.$$

$\mathfrak{Q}_2$ ist eine abelsche Gruppe (isomorph zur zyklischen Gruppe $\mathfrak{Z}_4$), ihr neutrales Element ist q_3.

3. Das Tripel $T = (f, g, h)$ mit $f, g, h \colon \mathbf{R}_+^* \to \mathbf{R}_+^*$ und $f = h = I_{\mathbf{R}_+^*}$ sowie $g = x \mapsto x^2$ ist ein Hauptisotopismus von der Quasigruppe $(\mathbf{R}_+^*, \circ)$ mit $x \circ y := x \cdot y^2$ auf die multiplikative Gruppe $(\mathbf{R}_+^*, \cdot)$ der positiven reellen Zahlen. $(\mathbf{R}_+^*, \circ)$ ist weder kommutativ noch assoziativ noch eine Loop (vgl. auch Beispiel 11.1.2).

4. $A = (f, g, h)$ mit $f, g, h\colon \mathbf{R}_+{}^* \to \mathbf{R}_+{}^*$ und $f\colon x \mapsto p \cdot x^r$, $g\colon x \mapsto q \cdot x^r$, $h\colon x \mapsto p \cdot q \cdot x^r$ (mit $p, q, r \in \mathbf{R}_+{}^*$, beliebig, fest) ist ein Autotopismus der multiplikativen Gruppe $(\mathbf{R}_+{}^*, \cdot)$ der positiven reellen Zahlen. $h(x \cdot y) = f(x) \cdot g(y)$ ist die vierte *Pexidersche Funktionalgleichung* (vgl. J. Aczél [3]).

5. Das Tripel $A = (f, g, h)$ mit den 1-1-Abbildungen $f, g, h\colon \mathbf{R} \to \mathbf{R}$ und $f\colon x \mapsto r \cdot x + p$, $g\colon x \mapsto r \cdot x + q$, $h\colon x \mapsto r \cdot x + \frac{1}{2}(p + q)$ (mit $p, q, r \in \mathbf{R}$, beliebig, fest, $r \neq 0$) ist ein Autotopismus der nichtassoziativen Quasigruppe $(\mathbf{R}, \circ)$, wobei $x \circ y := \frac{x + y}{2}$ das arithmetische Mittel zweier reeller Zahlen ist. $h(x \circ y) = f(x) \circ g(y)$ ist eine Verallgemeinerung der Jensenschen Funktionalgleichung (vgl. J. Aczél [3]).

Die Isotopie ist in der Menge aller Gruppoide eine reflexive und symmetrische Relation. Die Reflexivität ist durch $T = (I_G, I_G, I_G)$ für jedes Gruppoid $\mathfrak{G} = (G, \cdot)$ gewährleistet. Setzen wir $T\colon \mathfrak{G} \rightsquigarrow \mathfrak{H}$ mit $T = (f, g, h)$ voraus,[1]) so liefert das zu T inverse Tripel $T^{-1} := (f^{-1}, g^{-1}, h^{-1})$ einen Isotopismus von $\mathfrak{H}$ auf $\mathfrak{G}$. Mit $f(x) = u$ und $g(y) = v$ bzw. gleichbedeutend damit $f^{-1}(u) = x$ und $g^{-1}(v) = y$ erhalten wir $x \cdot y = f^{-1}(u) \cdot g^{-1}(v)$, so daß aus (4.1.2)

$$4.1.6.\quad \bigwedge_{u,v \in H} h^{-1}(u \circ v) = f^{-1}(u) \cdot g^{-1}(v)$$

folgt. Der Doppelpfeil in $\mathfrak{G} \overset{T}{\rightsquigarrow} \mathfrak{H}$ bringt die Symmetrie auch anschaulich zum Ausdruck. Es seien $\mathfrak{G} \overset{T_1}{\rightsquigarrow} \mathfrak{H}$ und $\mathfrak{H} \overset{T_2}{\rightsquigarrow} \mathfrak{K}$ Isotopismen mit $\mathfrak{G} = (G, \cdot)$, $\mathfrak{H} = (H, \circ)$ und $\mathfrak{K} = (K, *)$ sowie $T_1 = (f_1, g_1, h_1)$ und $T_2 = (f_2, g_2, h_2)$. Setzt man

$$4.1.7.\quad T_2 \bullet T_1 := (f_2 \bullet f_1, g_2 \bullet g_1, h_2 \bullet h_1),$$

so ist $T_2 \bullet T_1$ ein Isotopismus von $\mathfrak{G}$ auf $\mathfrak{K}$. Diese Nacheinanderausführung von Isotopismen wird also in natürlicher Weise durch die Nacheinanderausführung der jeweiligen Komponenten definiert. Wir überzeugen uns, daß $T_2 \bullet T_1$ die Bedingung (4.1.2) erfüllt:

$$\begin{aligned}(f_2 \bullet f_1)(x) * (g_2 \bullet g_1)(y) &= f_2\big(f_1(x)\big) * g_2\big(g_1(y)\big) = h_2\big(f_1(x) \circ g_1(y)\big)\\ &= h_2\big(h_1(x \cdot y)\big) = (h_2 \bullet h_1)(x \cdot y).\end{aligned}$$

Auf diese Weise haben wir zugleich bewiesen, daß die Isotopie transitiv und damit insgesamt eine Äquivalenzrelation ist. Dasselbe gilt für die Hauptisotopie.

Die Bedeutung der Hauptisotopie spiegelt sich in dem folgenden von A. A. Albert [12] gefundenen Satz wider:

4.1.8. Satz. *Es seien $\mathfrak{G} = (G, \cdot)$ und $\mathfrak{H} = (H, \circ)$ isotope Gruppoide, d. h., es gelte $T\colon \mathfrak{G} \rightsquigarrow \mathfrak{H}$. Dann ist $\mathfrak{H}$ zu einem Hauptisotop von $\mathfrak{G}$ isomorph.*

Beweis. Mit $T = (f, g, h)$ als Isotopismus von $\mathfrak{G}$ auf $\mathfrak{H}$ ist das Gruppoid $(G, *)$, dessen Operation $*$ durch $x \cdot y = (h^{-1} \bullet f)(x) * (h^{-1} \bullet g)(y)$ für alle $x, y \in G$ definiert

[1]) Wir verabreden, daß im weiteren das Tripel T stets die Komponenten f, g, h habe. In den hiervon abweichenden Fällen geben wir die Komponenten ohnehin an.

wird, offensichtlich ein Hauptisotop von $(G, \cdot)$. Die bijektive Abbildung $h^{-1}\colon H \to G$ ist dann ein Isomorphismus von $(H, \circ)$ auf $(G, *)$. Es gilt nämlich für alle $u, v \in H$

$$\begin{aligned} h^{-1}(u \circ v) &= f^{-1}(u) \cdot g^{-1}(v) = (h^{-1} \bullet f)\left(f^{-1}(u)\right) * (h^{-1} \bullet g)\left(g^{-1}(v)\right) \\ &= h^{-1}(u) * h^{-1}(v). \end{aligned}$$

4.1.9. Korollar. *Jeder Isotopismus $(G, \cdot) \overset{}{\underset{R}{\rightsquigarrow}} (G, *)$ und einen Isomorphismus $(G, *) \overset{}{\underset{S}{\rightsquigarrow}} (H, \circ)$ zerlegen.*

Mit $T = (f, g, h)$ lassen sich R und S in der Gestalt $R = (h^{-1} \bullet f, h^{-1} \bullet g, I_G)$ und $S = (h, h, h)$ angeben (siehe Abb. 27). (Mit h^{-1} ist auch h operationstreu, also ein Isomorphismus.)

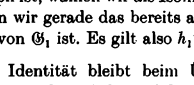


Abb. 27

Wenn wir in Beispiel 4.1.5.1 nach einem Hauptisotop von $\mathfrak{G}_1$ fahnden, das zu $\mathfrak{H}$ isomorph ist, wählen wir als Isomorphismus die Abbildung $h_1^{-1}\colon H \to G$. Im Ergebnis erhalten wir gerade das bereits angeführte Gruppoid $\mathfrak{G}_2 = (G, *)$, das ja ein Hauptisotop von $\mathfrak{G}_1$ ist. Es gilt also $h_1^{-1} \bullet f_1 = I_G = f_2$ und $h_1^{-1} \bullet g_1 = g_2$.

Eine Identität bleibt beim Übergang zu einem homomorphen Bild erhalten: Gruppen werden wieder auf Gruppen abgebildet. Einzige Ausnahme war die Quasigruppeneigenschaft (Q). Isotopismen verhalten sich in dieser Hinsicht genau entgegengesetzt. Keine der Identitäten (1) bis (33) aus 2.4. ist z. B. isotopie invariant. Ein Isotop einer Gruppe muß keine Gruppe sein, wie wir in Beispiel 4.1.5.2 bereits gesehen haben. Es gilt aber (der ebenfalls schon von A. A. Albert [12] bewiesene)

4.1.10. Satz. *Jedes Isotop einer Quasigruppe ist eine Quasigruppe.*

Beweis. Es seien $\mathfrak{Q} = (Q, \cdot)$ und $\mathfrak{R} = (R, \circ)$ isotope Gruppoide mit $\mathfrak{R} = T(\mathfrak{Q})$, $\mathfrak{Q}$ eine Quasigruppe. Ferner seien a, b beliebige Elemente aus R. Dann hat die Gleichung $a \circ x = b$ wenigstens eine Lösung, nämlich $x = g\left(f^{-1}(a) \backslash h^{-1}(b)\right)$. Mit (4.1.6) und (1.3.4) gilt in der Tat

$$\begin{aligned} a \circ x &= h\left(f^{-1}(a) \cdot g^{-1}(x)\right) = h\Big(f^{-1}(a) \cdot g^{-1}\Big(g\big(f^{-1}(a) \backslash h^{-1}(b)\big)\Big)\Big) \\ &= h\left(f^{-1}(a) \cdot \left(f^{-1}(a) \backslash h^{-1}(b)\right)\right) = h\left(h^{-1}(b)\right) = b. \end{aligned}$$

Diese Lösung ist aber auch die einzige. Ist nämlich $a \circ x_1 = a \circ x_2 (= b)$, so folgt mit (4.1.6)

$$h\left(f^{-1}(a) \cdot g^{-1}(x_1)\right) = h\left(f^{-1}(a) \cdot g^{-1}(x_2)\right).$$

Hieraus folgt nacheinander

$$\begin{aligned} &f^{-1}(a) \cdot g^{-1}(x_1) = f^{-1}(a) \cdot g^{-1}(x_2), \\ &g^{-1}(x_1) = g^{-1}(x_2), \quad x_1 = x_2. \end{aligned}$$

Entsprechend geht man vor und zeigt, daß $y = f(h^{-1}(b)/g^{-1}(a))$ die eindeutig bestimmte Lösung der Gleichung $y \circ a = b$ ist.

Aus diesem Satz folgt aufgrund der Symmetrie der Isotopie auch, daß das Ausgangsgruppoid $\mathfrak{G}$ bei einem Isotopismus $\mathfrak{G} \overset{T}{\rightarrow} \mathfrak{H}$ eine Quasigruppe sein muß, wenn das isotope Bild $\mathfrak{H}$ diese Eigenschaft hat.

Ähnlich wie im Fall der Homomorphie überträgt sich die Isotopie zwischen zwei Quasigruppen auf die jeweiligen Parastrophen. Dabei hängt allerdings das den Isotopismus realisierende Tripel T vom Parastrophismus π_i ab.

4.1.11. Satz. *Es seien* $\mathfrak{Q} = (Q, \cdot)$ *und* $\mathfrak{R} = (R, \circ)$ *Quasigruppen. Ist* $T = (f, g, h)$ *ein Isotopismus von* $\mathfrak{Q}$ *auf* $\mathfrak{R}$, *so ist* $\pi_i(T)$ *ein Isotopismus von* $\mathfrak{Q}_i = (Q, \pi_i(\cdot))$ *auf* $\mathfrak{R}_i = (R, \pi_i(\circ))$ *mit* $i = 1, 2, \ldots, 6$.
Im einzelnen gilt:

$$\pi_1(T) = (f, g, h)\colon (Q, \cdot) \rightarrow (R, \circ),$$
$$\pi_2(T) = (f, h, g)\colon (Q, \backslash) \rightarrow (R, \backslash\backslash),$$
$$\pi_6(T) = (h, g, f)\colon (Q, /) \rightarrow (R, //),$$
$$\pi_5(T) = (g, h, f)\colon (Q, /^*) \rightarrow (R, //^*),$$
$$\pi_4(T) = (h, f, g)\colon (Q, \backslash^*) \rightarrow (R, \backslash\backslash^*),$$
$$\pi_3(T) = (g, f, h)\colon (Q, \cdot^*) \rightarrow (R, \circ^*).$$

Beweis. Den Fall $i = 1$ brauchen wir nicht zu beweisen. Wegen Satz 2.1.10 genügt es, wenn wir die Fälle $i = 2$ und $i = 6$ betrachten, d. h. die Isotopie zwischen den einander entsprechenden Umkehroperationen. Infolge der Äquivalenz der Gleichungen $x \cdot y = z$, $z/y = x$ und $x\backslash z = y$ in $\mathfrak{Q}$ bzw. $u \circ v = w$, $w//v = u$ und $u\backslash\backslash w = v$ in $\mathfrak{R}$ können wir wie folgt schließen:

$$h(z) = h(x \cdot y) = f(x) \circ g(y) = f(x) \circ g(x\backslash z) \Rightarrow$$
$$f(x)\backslash\backslash h(z) = g(x\backslash z) \text{ für alle } x, z \in Q, \text{ d. h. } \pi_2(T) = (f, h, g);$$
$$h(z) = h(x \cdot y) = f(x) \circ g(y) = f(z/y) \circ g(y) \Rightarrow$$
$$h(z)//g(y) = f(z/y) \text{ für alle } z, y \in Q, \text{ d. h. } \pi_6(T) = (h, g, f).$$

Unter sämtlichen Isotopen einer Loop können sowohl Quasigruppen mit als auch ohne neutrales Element auftreten. Interessant ist nun, daß unter allen Isotopen einer beliebigen Quasigruppe sich stets auch Loops befinden (A. A. Albert [12]).

4.1.12. Satz. *Jede Quasigruppe ist zu einer Loop isotop.*

Beweis. Es seien a, b beliebige Elemente einer Quasigruppe $(Q, \cdot)$. Da in einer Quasigruppe die Links- und Rechts-Translationen bijektive Abbildungen sind (Korollar 1.7.5), definiert $T = (R_a, L_b, I_Q)$ gemäß (4.1.4) ein Hauptisotop $(Q, \circ)$ von $(Q, \cdot)$, d. h., es gilt

4.1.13. $\bigwedge_{x,y \in Q} x \cdot y = R_a(x) \circ L_b(y) = (x \cdot a) \circ (b \cdot y)$

und zugleich auch

4.1.14. $\bigwedge_{x,y \in Q} x \circ y = R_a^{-1}(x) \cdot L_b^{-1}(y) = (x/a) \cdot (b\backslash y)$.[1])

Wegen Satz 4.1.10 ist $(Q, \circ)$ eine Quasigruppe. Wir setzen $n := b \cdot a$. Dann ergibt sich mit Satz 1.3.6 für alle $x \in Q$

$$x \circ n = R_a^{-1}(x) \cdot L_b^{-1}(n) = R_a^{-1}(x) \cdot L_b^{-1}\big(L_b(a)\big) = R_a^{-1}(x) \cdot I_Q(a)$$
$$= R_a^{-1}(x) \cdot a = R_a\big(R_a^{-1}(x)\big) = I_Q(x) = x$$

und analog (zur Abwechslung mit \ und /)

$$n \circ x = (n/a) \cdot (b\backslash x) = \big((b \cdot a)/a\big) \cdot (b\backslash x) = b \cdot (b\backslash x) = x.$$

Somit ist $n = b \cdot a$ neutrales Element, $(Q, \circ)$ eine Loop.[2]) Damit haben wir gezeigt, daß es unter den Hauptisotopen einer Quasigruppe stets auch Loops gibt.

Ein Hauptisotop einer Quasigruppe, das eine Loop ist, heißt ein *Loop-Hauptisotop* dieser Quasigruppe (kurz: LH-Isotop). (4.1.13) bzw. (4.1.14) versetzt uns nun darüber hinaus in die Lage, ein solches LH-Isotop auch konstruieren zu können. Da außerdem mit a und b aus Q auch das Produkt $b \cdot a$ ganz Q durchläuft, können wir jedes Element als neutrales Element der Loop $(Q, \circ)$ vorgeben.

Interessanterweise erfaßt (4.1.13) bzw. (4.1.14) auch schon alle LH-Isotope der Quasigruppe $(Q, \cdot)$ (R. H. Bruck [32]). Um das zu zeigen, sei $(Q, \circ)$ ein beliebiges LH-Isotop von $(Q, \cdot)$, n sein neutrales Element. Dann gilt $x \cdot y = f(x) \circ g(y)$ für alle $x, y \in Q$. Da die Abbildungen $f, g: Q \to Q$ bijektiv sind, gibt es Elemente $a, b \in Q$ mit $f(b) = g(a) = n$. Daraus folgt für f und g

$$f(x) = f(x) \circ n = f(x) \circ g(a) = x \cdot a = R_a(x)$$

sowie

$$g(y) = n \circ g(y) = f(b) \circ g(y) = b \cdot y = L_b(y)$$

für alle $x \in Q$. Der Hauptisotopismus $T = (f, g, I_Q)$ geht damit in $T = (R_a, L_b, I_Q)$ über. Außerdem gilt $b \cdot a = f(b) \circ g(a) = n \circ n = n$. Wir fassen dieses Ergebnis noch einmal in dem folgenden Satz zusammen:

4.1.15. Satz. *Ist $(Q, \circ)$ ein LH-Isotop einer Quasigruppe $(Q, \cdot)$, so hat der Isotopismus die Gestalt $T = (R_a, L_b, I_Q)$ mit $a, b \in Q$, wobei $n = b \cdot a$ das neutrale Element in $(Q, \circ)$ ist.*

Die abelsche Gruppe $\mathfrak{D}_2$ aus Beispiel 4.1.5.2 ist ein Hauptisotop von $\mathfrak{D}_1$. Da $\mathfrak{D}_2$ als Gruppe eine Loop ist, muß sich das Tripel T_2 in der Gestalt (R_a, L_b, I_Q) angeben lassen. Man überzeuge sich, daß man mit $a = q_3$ und $b = q_4$ die Beziehungen $R_a = f_2$ sowie $L_b = g_2$ erhält; $b \cdot a = q_3$ ist das neutrale Element.

Es sei $\mathfrak{G} = (G, \cdot)$ ein endliches Gruppoid mit $G = \{a_1, a_2, \ldots, a_m\}$. Ist $T = (f, g, h)$

[1]) R_a, L_b und ihre Inversen sind Translationen in $(Q, \cdot)$ bzw. in $(Q, /)$ und $(Q, \backslash)$, nicht aber in $(Q, \circ)$ und deren Parastrophen.

[2]) (4.1.14) haben wir der Konstruktion in Beispiel 1.4.3.2 zugrunde gelegt.

ein Isotopismus von $\mathfrak{G}$ auf ein Gruppoid $\mathfrak{H} = (H, \circ)$, so muß $|\mathfrak{H}| = |\mathfrak{G}| = m$ gelten. Wir können also $H = \{b_1, b_2, \ldots, b_m\}$ für die Trägermenge von $\mathfrak{H}$ schreiben. (Betrachten wir der Einfachheit halber in den Abbildungen $f, g, h\colon G \to H$ nur die einander zugeordneten Indizes, so können wir diese drei Abbildungen auch als Permutationen über der Menge $M = \{1, 2, \ldots, m\}$ auffassen.) Ist die Strukturtafel von $\mathfrak{G}$ gegeben, so läßt sich die Strukturtafel von $\mathfrak{H}$ in der Weise konstruieren, daß man zunächst alle Zeilen (f) danach alle Spalten (g) und schließlich alle Elemente untereinander (h) innerhalb der Strukturtafel (bzw. Indextafel) von $\mathfrak{G}$ vertauscht. Ersetzt man anschließend noch in allen Feldern a durch b — natürlich unter Beibehaltung der Indizes —, so hat man die Strukturtafel von $\mathfrak{H}$ gewonnen. Im einzelnen erhalten wir also folgende Gruppoide (bzw. Indextafeln):

$$T_f = (f, I_M, I_M)\colon (M, \cdot) \rightleftarrows (M, *),$$
$$T_g = (I_M, g, I_M)\colon (M, *) \rightleftarrows (M, \square),$$
$$T_h = (I_M, I_M, h)\colon (M, \square) \rightleftarrows (M, \circ)$$
$$\text{mit } T_h \bullet T_g \bullet T_f = (f, g, h) = T.$$

Um einzusehen, daß hierbei nacheinander eine Zeilen-, Spalten- und elementeweise Vertauschung vorliegt, seien x, y beliebige Elemente aus M und z nacheinander $x * y$, $x \square y$ bzw. $x \circ y$:

$$T_f\colon f(x) \cdot y = z, \qquad T_g\colon x * g(y) = z, \qquad T_h\colon x \square y = h(z).$$

Wir demonstrieren die Konstruktion isotoper Strukturtafeln anhand der Indextafel der Quasigruppe $\mathfrak{Q}_1 = (Q, \cdot)$ und dem Isotopismus $T_1 = (f_1, g_1, h_1)$ aus Beispiel 4.1.5.2. Dabei fassen wir jetzt f_1, g_1 und h_1 als Permutationen von $M = \{1, 2, 3, 4\}$ auf: $f_1 = h_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix}$ und $g_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}$. Dann liefert der Isotopismus $(M, \cdot) \underset{T_{f_1}}{\rightleftarrows} (M, *) \underset{T_{g_1}}{\rightleftarrows} (M, \square) \underset{T_{h_1}}{\rightleftarrows} (M, \circ)$ die Strukturtafeln

$\cdot$	1	2	3	4
1	4	3	1	2
2	3	1	2	4
3	1	2	4	3
4	2	4	3	1

$*$	1	2	3	4
1	3	1	2	4
2	1	2	4	3
3	2	4	3	1
4	4	3	1	2

$\square$	1	2	3	4
1	2	①	3	4
2	4	2	①	3
3	3	4	2	①
4	①	3	4	2

$\circ$	1	2	3	4
1	1	4	2	3
2	3	1	4	2
3	2	3	1	4
4	4	2	3	1

$(f_1\colon Z_1^{(\cdot)} \mapsto Z_4^{(*)}, g_1\colon S_1^{(*)} \rightarrow S_3^{(\square)}, h_1\colon 1 \mapsto 4).$

Halten wir schließlich noch fest, daß Satz 4.1.12 in der Sprache der Strukturtafeln folgende Gestalt annimmt: In jedem lateinischen Quadrat kann man durch zeilen-, spalten- oder elementeweises Vertauschen erreichen, daß die Tafeleingänge mit einer Zeile und mit einer Spalte identisch sind (Existenz eines neutralen Elementes).

Die Palette der Möglichkeiten, Isotopismen zur Untersuchung spezieller Klassen von Gruppoiden oder einer bestimmten Eigenschaft eines Gruppoids einzusetzen, läßt sich im Rahmen dieses Buches nicht erfassen. So sind z. B. isotope totalsymmetrische Loops isomorph, dasselbe gilt für kommutative Moufang-Loops (R. H. Bruck [29, 32]). Es sei auch nur erwähnt, daß man eine normale Kongruenzrelation R von einer Quasigruppe $(Q, \cdot)$ auf eine Quasigruppe $(Q, \circ)$ mit Hilfe eines Isotopismus $T: (Q, \cdot) \xrightarrow{\sim} (Q, \circ)$ übertragen kann, sofern die Komponenten σ von T die Bedingung $xRy \Leftrightarrow \sigma(x)\, R\, \sigma(y)$ für alle $x, y \in Q$ erfüllen (V. D. Belousov [21]). Die Komponenten der LH-Isotopismen $T = (R_a, L_b, I_Q)$ genügen wegen (3.2.3), (3.2.7) und (3.2.8) dieser Bedingung.

4.1.16. Aufgaben

1. Zur Quasigruppe $\mathfrak{L} = (Q, \cdot)$ mit dem lateinischen Quadrat

$\cdot$	1	2	3	4	5	6
1	6	1	2	5	3	4
2	4	5	3	2	6	1
3	1	2	5	6	4	3
4	5	3	6	4	1	2
5	3	6	4	1	2	5
6	2	4	1	3	5	6

gebe man folgende Isotope von $\mathfrak{L}$ an:

a) das Isotop $T(\mathfrak{L})$ mit $T = (f, g, h)$ und $f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 6 & 1 & 3 & 4 & 5 \end{pmatrix}$, $g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 1 & 6 & 5 & 4 \end{pmatrix}$ und $h = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 2 & 3 & 5 & 1 & 6 \end{pmatrix}$;

b) die Hauptisotope $T_1(\mathfrak{L})$ bis $T_3(\mathfrak{L})$ mit $T_1 = (f, g, I_Q)$, $T_2 = (f, I_Q, I_Q)$, $T_3 = (I_Q, g, I_Q)$, wobei f und g wie in a) gegeben sind, und $T_4 = (R_3, I_Q, I_Q)$ sowie $T_5 = (I_Q, L_6, I_Q)$;

c) die LH-Isotope $T_8(\mathfrak{L})$ und $T_7(\mathfrak{L})$ mit $T_8 = (R_3, L_5, I_Q)$ sowie $T_7 = (R_5, L_3, I_Q)$.

d) Man zeige, daß nur die Paare (a, b) mit

a	1	2	3	4	5	6
b	2	4	5	6	1	3

kommutative LH-Isotope gemäß (4.1.13) bzw. (4.1.14) liefern, alle anderen LH-Isotope also nicht kommutativ sind.

2. Man zeige, daß eine kommutative und eine nichtkommutative Loop zueinander isotop sein können.

3. Man zeige, daß $A = (f, g, h)$ mit $f, g, h: \mathbb{Q} \to \mathbb{Q}$ und $f: x \mapsto 2(x - 1)$, $g: x \mapsto 2x + 3$, $h: x \mapsto 2x + 1$ ein Autotopismus von $(\mathbb{Q}, +)$ ist. Ist A auch ein Autotopismus von $(\mathbb{Q}, -)$?

4.2. Isotope von Gruppen

Eine Loop kann ein Isotop besitzen, in dem selbst kein neutrales Element existiert: ebenso kann eine Gruppe zu einer nichtassoziativen Quasigruppe isotop sein (vgl. Beispiel 4.1.5.2), d. h., weder die Existenz des neutralen Elementes noch die Assoziativität sind isotopieinvariant. Es gibt jedoch einen Satz von R. H. Bruck [32], der insbesondere auch aufgrund der folgenden Korollare von Bedeutung ist und diese beiden Eigenschaften eng miteinander verbindet.

4.2.1. Satz. *Sind ein Gruppoid* $\mathfrak{G} = (G, \cdot)$ *mit dem neutralen Element* n_G *und eine Halbgruppe* $\mathfrak{H} = (H, \circ)$ *isotop, so sind beide isomorph.*

Beweis. Es sei $\mathfrak{H} = T(\mathfrak{G})$ mit $T = (f, g, h)$. Dann gilt wegen (4.1.6) $u \circ v = h\big(f^{-1}(u) \cdot g^{-1}(v)\big)$. Damit nimmt die Assoziativität von $\circ$ wegen $h^{-1} \bullet h = I_G$ die Gestalt

a) $$\bigwedge_{u,v,w \in H} (f^{-1} \bullet h)\left(f^{-1}(u) \cdot g^{-1}(v)\right) \cdot g^{-1}(w) = f^{-1}(u) \cdot (g^{-1} \bullet h)\left(f^{-1}(v) \cdot g^{-1}(w)\right)$$

an. Da $f, g, h: G \to H$ bijektive Abbildungen sind, existieren solche $u_0, w_0 \in H$, daß $f(n_G) = u_0$ und $g(n_G) = w_0$ gilt. Für diesen Fall liefert a)

$$(f^{-1} \bullet h)\left(n_G \cdot g^{-1}(v)\right) \cdot n_G = n_G \cdot (g^{-1} \bullet h)\left(f^{-1}(v) \cdot n_G\right).$$

Daraus folgt aber sofort

b) $$\bigwedge_{v \in H} (f^{-1} \bullet h \bullet g^{-1})(v) = (g^{-1} \bullet h \bullet f^{-1})(v).$$

Setzen wir in a) nur $f^{-1}(u_0) = n_G$, so führt das unter Beachtung von b) auf

$$(g^{-1} \bullet h \bullet f^{-1})(v) \cdot g^{-1}(w) = (g^{-1} \bullet h)\left(f^{-1}(v) \cdot g^{-1}(w)\right)$$

für alle $v, w \in H$. Mit $f^{-1}(v) = y$ und $g^{-1}(w) = z$ folgt hieraus

c) $$\bigwedge_{y,z \in G} (g^{-1} \bullet h)(y) \cdot z = (g^{-1} \bullet h)(y \cdot z).$$

Setzen wir in a) dagegen nur $g^{-1}(w_0) = n_G$, so erhalten wir wieder wegen b)

$$(f^{-1} \bullet h)\left(f^{-1}(u) \cdot g^{-1}(v)\right) = f^{-1}(u) \cdot (f^{-1} \bullet h \bullet g^{-1})(v)$$

für alle $u, v \in H$, woraus mit $f^{-1}(u) = x$ und $g^{-1}(v) = y$

d) $$\bigwedge_{x,y \in G} (f^{-1} \bullet h)(x \cdot y) = x \cdot (f^{-1} \bullet h)(y)$$

folgt. Damit sind wir in der Lage zu beweisen, daß die Abbildung $\varphi := f^{-1} \bullet h \bullet g^{-1}$: $H \to G$ ein Isomorphismus von $\mathfrak{H}$ auf $\mathfrak{G}$ ist. Da f^{-1}, h und g^{-1} bijektive Abbildungen sind, ist auch φ bijektiv. φ ist auch operationstreu, denn für alle $u, v \in H$ gilt

$$\begin{aligned}
\varphi(u \circ v) &= (f^{-1} \bullet h \bullet g^{-1})(u \circ v) = (f^{-1} \bullet h \bullet g^{-1} \bullet h)\left(f^{-1}(u) \cdot g^{-1}(v)\right) \\
&\underset{c)}{=} (f^{-1} \bullet h)\left((g^{-1} \bullet h \bullet f^{-1})(u) \cdot g^{-1}(v)\right) \\
&\underset{d)}{=} (g^{-1} \bullet h \bullet f^{-1})(u) \cdot (f^{-1} \bullet h \bullet g^{-1})(v) \\
&\underset{b)}{=} (f^{-1} \bullet h \bullet g^{-1})(u) \cdot (f^{-1} \bullet h \bullet g^{-1})(v) = \varphi(u) \cdot \varphi(v).
\end{aligned}$$

Mit $\varphi: \mathfrak{H} \xrightarrow{\sim} \mathfrak{G}$ ist auch $\varphi^{-1}: \mathfrak{G} \xrightarrow{\sim} \mathfrak{H}$ ein Isomorphismus. Damit sind $\mathfrak{G}$ und $\mathfrak{H}$ isomorphe Monoide, d. h., auch das Gruppoid $\mathfrak{G}$ ist eine Halbgruppe, und die Halbgruppe $\mathfrak{H}$ besitzt ebenfalls ein neutrales Element n_H. Für dieses neutrale Element gilt $n_H = \varphi^{-1}(n_G) = (g \bullet h^{-1} \bullet f)\,(n_G)$.

4.2.2. Korollar. *Isotope Monoide sind isomorph.*

4.2.3. Korollar (Satz von ALBERT). *Sind eine Loop $(L, \cdot)$ und eine Gruppe $(G, \circ)$ isotop, so sind beide isomorph, d. h., $(L, \cdot)$ ist selbst eine Gruppe.*

Eine nichtassoziative Loop kann also niemals ein Isotop einer Gruppe sein. Daß es indessen isotope Loops gibt, die nicht isomorph sind, beweist Aufgabe 4.1.16.1c). Die Loops $T_6(\mathfrak{Q})$ und $T_7(\mathfrak{Q})$ sind isotop. Erstere ist kommutativ, die zweite nicht. Auch isotope Halbgruppen brauchen nicht isomorph zu sein (Beweis!).

4.2.4. Korollar. *Sind zwei Gruppen $\mathfrak{G} = (G, \cdot)$ und $\mathfrak{H} = (H, \circ)$ isotop, so sind sie isomorph. Dabei bestehen zwischen den Abbildungen f, g, h des Isotopismus $T: \mathfrak{G} \xrightarrow{\sim} \mathfrak{H}$ und dem Isomorphismus $\psi: \mathfrak{G} \xrightarrow{\sim} \mathfrak{H}$ die folgenden Beziehungen* (J. ACZÉL, V. D. BELOUSOV und M. HOSSZÚ [7]):

$$\bigwedge_{x \in G} f(x) = a \circ \psi(x) \wedge g(x) = \psi(x) \circ b \wedge h(x) = a \circ \psi(x) \circ b,$$

wobei $a := f(n_G)$, $b := g(n_G)$ gilt und n_G das neutrale Element von $\mathfrak{G}$ ist.

Beweis. Mit (4.1.2) erhalten wir für $y = n_G$

$$h(x) = f(x) \circ g(n_G) = f(x) \circ b, \quad \text{also} \quad f(x) = h(x) \circ b^{-1}.$$

Analog liefert $x = n_G$

$$g(x) = a^{-1} \circ h(x).$$

Dann folgt wieder mit (4.1.2)

$$h(x \cdot y) = f(x) \circ g(y) = h(x) \circ b^{-1} \circ a^{-1} \circ h(y),$$

was

$$a^{-1} \circ h(x \cdot y) \circ b^{-1} = a^{-1} \circ h(x) \circ b^{-1} \circ a^{-1} \circ h(y) \circ b^{-1}$$

nach sich zieht. $\psi: G \to H$ mit $x \mapsto a^{-1} \circ h(x) \circ b^{-1}$ ist folglich ein Isomorphismus von $\mathfrak{G}$ auf $\mathfrak{H}$. Aus der Definition von ψ erhalten wir für h sofort $h(x) = a \circ \psi(x) \circ b$, woraus sich auch die restlichen beiden Beziehungen ergeben.

Aufgrund des Korollars 4.2.4 kann eine nichtkommutative Gruppe kein Isotop einer abelschen Gruppe sein.

Ist $\mathfrak{G} = (G, \cdot)$ eine Gruppe, so gibt es i. allg. unter allen Isotopen von $\mathfrak{G}$ auch nichtassoziative Quasigruppen, wie Beispiel 4.1.5.3 beweist. Läßt man jedoch — wie in der Gruppentheorie — von vornherein nur Gruppen, also assoziative Quasigruppen zu, so wird der Begriff der Isotopie bedeutungslos.

Geht man in den Korollaren 4.2.3 und 4.2.4 gemäß Satz 4.1.11 zu den π_2- und

π_6-Parastrophen über, so ergeben sich aufgrund der Sätze und Korollar 2.4.1 bis 2.4.4 sofort zwei weitere Folgerungen:

4.2.5. Korollar. *Sind eine unipotente Links- (Rechts-) Loop* $(Q, \cdot)$ *und eine links- (rechts-) transitive Quasigruppe* $(R, \circ)$ *isotop, so sind beide isomorph.*

4.2.6. Korollar. *Sind zwei links- (rechts-) transitive Quasigruppen* $(Q, \cdot)$ *und* $(R, \circ)$ *isotop, so sind beide isomorph.*

Ist eine Quasigruppe $\mathfrak{Q}$ zu einer Gruppe $\mathfrak{G}$ isotop, so sind alle *Loop-Isotope* von $\mathfrak{Q}$ (das sind alle Isotope von $\mathfrak{Q}$, die Loops sind) wegen der Transitivität der Isotopie auch zur Gruppe $\mathfrak{G}$ isotop. Infolgedessen sind sie selbst Gruppen und sämtlich untereinander und zu $\mathfrak{G}$ isomorph. Umgekehrt muß aus der Isomorphie aller Loop-Isotope einer Quasigruppe jedoch nicht die Isotopie dieser Quasigruppe zu einer Gruppe folgen. Dennoch erweist es sich oft als vorteilhaft, die Gesamtheit aller LH-Isotope[1]) einer Quasigruppe zu betrachten. Sind nämlich z. B. alle LH-Isotope kommutativ, so sind sie auch assoziativ (J. Aczél, G. Pickert und F. Radó [11]). Daß nicht alle LH-Isotope einer Quasigruppe kommutativ sein müssen, zeigt Aufgabe 4.1.16.1, wo neben sechs kommutativen 30 nichtkommutative LH-Isotope existieren. Dieses Beispiel lehrt auch, daß aus der Kommutativität eines LH-Isotops einer Quasigruppe noch nicht seine Assoziativität folgt.

Es gibt aber auch andere Möglichkeiten, die Frage, ob eine Quasigruppe zu einer Gruppe isotop ist, zu entscheiden. So sind z. B. die Thomsen-Bedingung (T) und die Reidemeister-Bedingung (R) aus 1.6. derartige Kriterien. Wir schicken voraus, daß beide Schließungssätze — wie diese Bedingungen auch genannt werden — isotopieinvariant sind.

4.2.7. Satz. *Ist in einer Quasigruppe* $\mathfrak{Q} = (Q, \cdot)$ *die Thomsen-Bedingung* (T) *oder die Reidemeister-Bedingung* (R) *erfüllt, so ist dieselbe Bedingung auch in jeder isotopen Quasigruppe* $\mathfrak{R} = (R, \circ)$ *von* $\mathfrak{Q}$ *erfüllt.*

Beweis. Gelte $\mathfrak{R} = T(\mathfrak{Q})$ mit $T = (f, g, h)$. Mit (4.1.2) geht die Prämisse von (T) in

$$h^{-1}\big(f(x_1) \circ g(y_2)\big) = h^{-1}\big(f(x_2) \circ g(y_1)\big)$$
$$\wedge\, h^{-1}\big(f(x_1) \circ g(y_3)\big) = h^{-1}\big(f(x_3) \circ g(y_1)\big)$$

über, so daß wegen $h \bullet h^{-1} = I_R$

$$f(x_1) \circ g(y_2) = f(x_2) \circ g(y_1) \wedge f(x_1) \circ g(y_3) = f(x_3) \circ g(y_1)$$

für alle $x_i, y_i \in Q$ $(i = 1, 2, 3)$ gilt. Setzen wir $f(x_i) = u_i$ und $g(y_i) = v_i$, so erhalten wir weiter $u_1 \circ v_2 = u_2 \circ v_1 \wedge u_1 \circ v_3 = u_3 \circ v_1$. Analog liefert die Konklusion von (T) $u_2 \circ v_3 = u_3 \circ v_2$. Wenn x_i, y_i ganz Q durchlaufen, trifft dasselbe auch für u_i, v_i in bezug auf R zu. Folglich ist die Thomsen-Bedingung auch in $\mathfrak{R} = (R, \circ)$ erfüllt.

[1]) Da jedes Loop-Isotop zu einem LH-Isotop isomorph ist (Satz 4.1.8), genügt es, wenn man sich auf die LH-Isotope beschränkt.

Entsprechend zeigt man, daß die Reidemeister-Bedingung (R) unter Isotopie erhalten bleibt.[1])

4.2.8. Satz. *Eine Quasigruppe $\mathfrak{Q} = (Q, \cdot)$ ist genau dann zu einer abelschen Gruppe isotop, wenn in $\mathfrak{Q}$ die Thomsen-Bedingung* (T) *gilt.*

Beweis. a) Es sei $\mathfrak{Q} = (Q, \cdot)$ zu einer abelschen Gruppe $(R, \circ)$ isotop.[2]) Dann ist wegen Satz 1.6.3. in $(R, \circ)$ die Thomsen-Bedingung erfüllt. Aufgrund ihrer Isotopieinvarianz (Satz 4.2.7) gilt (T) folglich auch in $\mathfrak{Q} = (Q, \cdot)$.

b) Gelte in $\mathfrak{Q} = (Q, \cdot)$ die Thomsen-Bedingung. Wegen Satz 4.2.7 ist (T) dann auch in jedem *LH*-Isotop $(Q, \circ)$ von $\mathfrak{Q}$ erfüllt. $(Q, \circ)$ ist dann — wieder wegen Satz 1.6.3 — eine abelsche Gruppe.

4.2.9. Satz. *Eine Quasigruppe $\mathfrak{Q} = (Q, \cdot)$ ist genau dann zu einer Gruppe isotop, wenn in $\mathfrak{Q}$ die Reidemeister-Bedingung* (R) *gilt.*

Der Beweis folgt unmittelbar aus den Sätzen 1.6.1 und 4.2.7.

Ist $\mathfrak{Q} = (Q, \cdot)$ eine bisymmetrische Quasigruppe,[3]) d. h., gilt für alle $x, y, u, v \in Q$

(BS) $\qquad xy \cdot uv = xu \cdot yv,$

so ist jedes *LH*-Isotop $T(\mathfrak{Q})$ von $\mathfrak{Q}$ mit $T = (R_a, L_b, I_Q)$ eine abelsche Gruppe (D. G. Murdoch [121], R. H. Bruck [29]). Zum Beweis dieser Aussage stützen wir uns auf Satz 4.2.8 und das folgende Ergebnis von K. Toyoda [173]:

4.2.10. Satz. *In einer bisymmetrischen Quasigruppe $(Q, \cdot)$ ist die Thomsen-Bedingung* (T) *erfüllt.*[4])

Beweis. Wegen (BS) und der Prämisse von (T) gilt nämlich $x_2y_3 \cdot y_1x_1 = x_2y_1 \cdot y_3x_1 = x_1y_2 \cdot y_3x_1 = x_1y_3 \cdot y_2x_1 = x_3y_1 \cdot y_2x_1 = x_3y_2 \cdot y_1x_1$ für beliebige $x_i, y_i \in Q$ $(i = 1, 2, 3)$. Aufgrund der Kürzbarkeit gilt damit dann $x_2y_3 = x_3y_2$. Folglich ist (T) erfüllt.

4.2.11. Korollar. *Jede bisymmetrische Quasigruppe ist zu einer abelschen Gruppe isotop.*

Die Umkehrung gilt nicht, da es nichtbisymmetrische Isotope abelscher Gruppen gibt. Die Quasigruppe $(\mathbf{R}_+^*, \circ)$ aus Beispiel 4.1.5.3 ist nicht bisymmetrisch, aber ein Isotop einer abelschen Gruppe, nämlich $(\mathbf{R}_+^*, \cdot)$. Dieses Beispiel beweist zugleich, daß die Bisymmetrie nicht isotopieinvariant ist. Allerdings steht sie in der Frage der Isotopieinvarianz der Assoziativität näher als der Kommutativität. Sie bleibt nämlich — wie die Assoziativität — erhalten, wenn wir uns auf Loop-Isotope beschränken:

4.2.12. Korollar. *Alle Loop-Isotope einer bisymmetrischen Quasigruppe sind (isomorphe) abelsche Gruppen.*

---

[1]) Auch die Sechseckbedingung (H) aus Kapitel 11 ist isotopieinvariant (vgl. V. D. Belousov [22]). Alle diese Schließungssätze spielen in der Theorie der Gewebe eine wichtige Rolle. Da sie jedoch rein algebraischer Natur sind, können wir sie hier mit Gewinn einsetzen.

[2]) Vgl. hierzu auch Satz 11.2.3.

[3]) Beispiele für bisymmetrische Operationen sind in (5.1.10) angegeben.

[4]) Da die Sechseckbedingung (H) ein Spezialfall von (T) ist — vgl. 11.2. —, gilt in $(Q, \cdot)$ dann auch die Sechseckbedingung.

Das folgt aus den Korollaren 4.2.3 und 4.2.11 sowie der Transitivität der Isotopie.

Da aus der Thomsen-Bedingung die Reidemeister-Bedingung folgt (wie auch die übrigen Schließungssätze — vgl. z. B. J. Aczél [4]), ist in einer bisymmetrischen Quasigruppe neben (T) auch (R) erfüllt.

Bevor wir die Aussage des Korollars 4.2.11 noch weiter präzisieren, zeigen wir, daß Satz 4.2.10 ein Analogon in bezug auf den Schließungssatz (R) besitzt. Das wird unter dem Blickwinkel zueinander parastropher Identitäten nahegelegt.

4.2.13. Satz. *In einer links- (rechts-) transitiven Quasigruppe $(Q, \cdot)$ ist die Reidemeister-Bedingung erfüllt.*

Beweis. Es sei $\mathfrak{Q} := (Q, \cdot)$ eine rechts-transitive Quasigruppe. Dann ist wegen Korollar 2.4.4 die zu $\mathfrak{Q}$ parastrophe Quasigruppe $(Q, /)$ assoziativ, also eine Gruppe. Wegen Satz 1.6.1 ist in $(Q, /)$ damit die Reidemeister-Bedingung erfüllt. Die Quasigruppe $\mathfrak{Q} = (Q, \cdot)$ läßt sich nun als ein Hauptisotop der Gruppe $(Q, /)$ auffassen: $(Q, /) \underset{T}{\rightarrow} (Q, \cdot)$ mit $T = (I_Q, g, I_Q)$ und $g: x \mapsto x^{-1}$, wobei x^{-1} das Inverse von x in $(Q, /)$ ist. Satz 4.2.7 liefert dann die Behauptung. Der Beweis für die Links-Transitivität verläuft analog.

4.2.14. Korollar. *Jede links- (rechts-) transitive Quasigruppe ist zu einer Gruppe isotop.*

Mit demselben Beispiel wie im Fall von Korollar 4.2.11 folgt, daß auch hier die Umkehrung falsch ist.

4.2.15. Korollar. *Alle Loop-Isotope einer links- (rechts-) transitiven Quasigruppe sind (isomorphe) Gruppen.*

Fordert man in Satz 4.2.13, daß zusätzlich noch die Rechts- bzw. Links-Systemregel erfüllt sein soll, so kann man wieder auf die Thomsen-Bedingung schließen. Der Beweis verläuft völlig analog, nur daß neben Korollar 2.4.4 auch noch Korollar 2.4.6 sowie anstelle von Satz 1.6.1 jetzt Satz 1.6.3 herangezogen werden.[1])

4.2.16. Satz. *In einer links- (rechts-) transitiven Quasigruppe, in der die Rechts- (Links-) Systemregel gilt, ist die Thomsen-Bedingung erfüllt.*

Es ergeben sich auch die entsprechenden Folgerungen wie im Falle der Sätze 4.2.10 und 4.2.13:

4.2.17. Korollar. a) *Jede links- (rechts-) transitive Quasigruppe $(Q, \cdot)$, in der die Rechts- (Links-) Systemregel gilt, ist zu einer abelschen Gruppe isotop, und* b) *alle Loop-Isotope von $(Q, \cdot)$ sind (isomorphe) abelsche Gruppen.*

Wir kommen jetzt zu der bereits angekündigten Präzisierung von Korollar 4.2.11. Sie wird uns in die Lage versetzen, zu einer vorgegebenen abelschen Gruppe $(Q, \circ)$

[1]) Satz 4.2.16 folgt auch unmittelbar aus Satz 4.2.10, wenn wir vorgreifen und Satz 5.2.11 ausnutzen.

eine zu ihr isotope bisymmetrische Quasigruppe $(Q, \cdot)$ mit Hilfe von Automorphismen der Gruppe zu konstruieren. Wir haben uns ja bereits überlegt, daß nicht jedes Isotop einer abelschen Gruppe bisymmetrisch sein muß. Umgekehrt liefert jedes LH-Isotop einer bisymmetrischen Quasigruppe $(Q, \cdot)$ eine abelsche Gruppe $(Q, \circ)$ — Korollar 4.2.12. Darüber hinaus lassen sich dann aber auch stets Automorphismen dieser Gruppe finden, so daß $(Q, \cdot)$ linear über $(Q, \circ)$ ist, d. h., daß die folgende Beziehung (4.2.19) gilt (vgl. etwa R. H. Bruck [29]).

4.2.18. Satz von Toyoda. *Ist $(Q, \cdot)$ eine bisymmetrische Quasigruppe, so gibt es eine abelsche Gruppe $(Q, \circ)$, so daß für ein beliebiges, aber festes $c \in Q$*

$$\textbf{4.2.19.}\quad \bigwedge_{x,y\in Q} x \cdot y = f(x) \circ g(y) \circ c$$

gilt, wobei f und g vertauschbare Automorphismen von $(Q, \circ)$ sind, d. h., es gilt

$$\textbf{4.2.20.}\quad f \bullet g = g \bullet f.$$

Ist umgekehrt $(Q, \circ)$ eine abelsche Gruppe mit Automorphismen f und g, die (4.2.20) *erfüllen, so definiert* (4.2.19) *für jedes $c \in Q$ eine bisymmetrische Quasigruppe $(Q, \cdot)$.*

Beweis. 1. Es sei $\mathfrak{Q} = (Q, \cdot)$ eine bisymmetrische Quasigruppe. Dann ist wegen Korollar 4.2.12 das LH-Isotop $T(\mathfrak{Q}) = (Q, \circ)$ mit $T = (R_a, L_b, I_Q)$ eine abelsche Gruppe. Das neutrale Element von $(Q, \circ)$ ist $n = b \cdot a$. Durch $f(x) := (x \circ b) \cdot a$ und $g(x) := b \cdot (x \circ a)$ werden, da die Links- und Rechts-Translationen sowohl in $(Q, \circ)$ als auch in $(Q, \cdot)$ bijektive Abbildungen sind, zwei 1-1-Abbildungen f und g von Q auf sich definiert.

a) *f und g sind Automorphismen von $(Q, \circ)$*: Da $(Q, \cdot)$ eine Quasigruppe ist, gibt es zu vorgegebenen Elementen x und a aus Q ein eindeutig bestimmtes $r \in Q$, so daß $x = r \cdot a$ gilt. Aus demselben Grund können wir die folgenden Gleichungen ansetzen:

$$\textbf{4.2.21.}\quad \begin{cases} x = r \cdot a, \quad y = b \cdot s, \quad r \cdot s = d \cdot a, \\ a = n_l{}^a \cdot a = a \cdot n_r{}^a = b \cdot a', \\ b = n_l{}^b \cdot b = b \cdot n_r{}^b = b' \cdot a. \end{cases}$$

Dabei ist $n_l{}^x$ das individuelle links- und $n_r{}^x$ das individuelle rechts-neutrale Element von x in $(Q, \cdot)$. Dann gilt[1]) mit (4.2.21), (4.1.13) und (BS) für alle $x, y \in Q$

$$\begin{aligned} f(x) \circ f(y) &= (x \circ b) \cdot a \circ (y \circ b) \cdot a = (ra \circ bn_r{}^b) \cdot a \circ (bs \circ b'a) \cdot an_r{}^a \\ &= (ra \circ bn_r{}^b) \cdot a \circ (b'a \circ bs) \cdot an_r{}^a = rn_r{}^b \cdot a \circ b's \cdot an_r{}^a \\ &= rn_r{}^b \cdot a \circ b'a \cdot sn_r{}^a = rn_r{}^b \cdot a \circ b \cdot sn_r{}^a = rn_r{}^b \cdot sn_r{}^a \\ &= rs \cdot n_r{}^b n_r{}^a = da \cdot n_r{}^b n_r{}^a = dn_r{}^b \cdot an_r{}^a \\ &= dn_r{}^b \cdot a = (da \circ bn_r{}^b) \cdot a = (rs \circ b) \cdot a \\ &= (ra \circ bs \circ b) \cdot a = (x \circ y \circ b) \cdot a = f(x \circ y). \end{aligned}$$

[1]) Zur Klammereinsparung verabreden wir, daß die Operation $\cdot$ stärker als die Operation $\circ$ bindet. Außerdem bedeutet ab immer $a \cdot b$. Damit ist $ab \circ cd$ eine abkürzende Schreibweise für $(a \cdot b) \circ (c \cdot d)$.

f ist also eine operationstreue bijektive Abbildung, mithin ein Automorphismus von $(Q, \circ)$. Analog zeigt man mit Hilfe von (BS), (4.1.13) und (4.2.21) sowie der Kommutativität und Assoziativität von $\circ$, daß $g(x) \circ g(y) = g(x \circ y)$ für alle $x, y \in Q$ erfüllt ist.

b) *Die Automorphismen f und g sind vertauschbar:* Setzen wir (4.2.19) in (BS) ein, so erhalten wir

$$f\big(f(x) \circ g(y) \circ c\big) \circ g\big(f(u) \circ g(v) \circ c\big) \circ c$$
$$= f\big(f(x) \circ g(u) \circ c\big) \circ g\big(f(y) \circ g(v) \circ c\big) \circ c.$$

Da $(Q, \circ)$ eine abelsche Gruppe ist, f und g Automorphismen sind, läßt sich das auf

$$(f \bullet g)\,(y) \circ (g \bullet f)\,(u) = (f \bullet g)\,(u) \circ (g \bullet f)\,(y)$$

reduzieren. Für $u = n$ gilt nun $(g \bullet f)\,(n) = (f \bullet g)\,(n) = n$, so daß wir letztlich $(f \bullet g)\,(y) = (g \bullet f)\,(y)$ für alle $y \in Q$ erhalten.

2. Es sei $(Q, \circ)$ eine abelsche Gruppe mit vertauschbaren Automorphismen f und g. Dann ist das durch (4.2.19) definierte Gruppoid $(Q, \cdot)$ wegen Satz 4.1.10 eine Quasigruppe; $T = (f^{-1}, h^{-1}, I_Q)$ mit $h(x) := g(x) \circ c$ ist nämlich ein Hauptisotopismus von $(Q, \circ)$ auf $(Q, \cdot)$. Wir brauchen also nur noch zu zeigen, daß diese Quasigruppe bisymmetrisch ist. Für alle $x, y, u, v \in Q$ gilt (mit (4.2.20))

$$\begin{aligned} xy \cdot uv &= f\big(f(x) \circ g(y) \circ c\big) \circ g\big(f(u) \circ g(v) \circ c\big) \circ c \\ &= (f \bullet f)\,(x) \circ (f \bullet g)\,(y) \circ f(c) \circ (g \bullet f)\,(u) \circ (g \bullet g)\,(v) \circ g(c) \circ c \\ &= (f \bullet f)\,(x) \circ (g \bullet f)\,(u) \circ f(c) \circ (f \bullet g)\,(y) \circ (g \bullet g)\,(v) \circ g(c) \circ c \\ &= (f \bullet f)\,(x) \circ\ (f \bullet g)\,(u) \circ f(c) \circ (g \bullet f)\,(y) \circ (g \bullet g)\,(v) \circ g(c) \circ c \\ &= f\big(f(x) \circ g(u) \circ c\big) \circ g\big(f(y) \circ g(v) \circ c\big) \circ c = xu \cdot yv. \end{aligned}$$

Abschließend halten wir fest, daß wir neben den beiden Schließungssätzen von THOMSEN und REIDEMEISTER nur noch die Quasigruppeneigenschaft (Q) als Isotopieinvariante kennengelernt haben.[1]) Beschränken wir uns indessen auf Loop-Isotope, so sind auch die Assoziativität und die Bisymmetrie invariant (Korollare 4.2.3 und 4.2.12). Für die Moufang-Identität $(xy \cdot z) \cdot y = x \cdot (y \cdot zy)$, eine Abschwächung der Assoziativität, trifft dies ebenfalls zu (R. H. BRUCK [30]). Die Kommutativität hingegen ist selbst für Loop-Isotope nicht invariant. Erst bei der Beschränkung auf Gruppen (oder Moufang-Loops) liegt Invarianz vor (Korollar 4.2.4). Gilt andererseits neben der Kommutativität noch eine der Systemregeln (L-SR) oder (R-SR), also Totalsymmetrie, dann liegt wieder Invarianz für Loop-Isotope vor. Isotope totalsymmetrische Loops sind sogar isomorph (R. H. BRUCK [29]).

4.2.22. Aufgaben

1. Man zeige, daß in der Quasigruppe $(\mathbf{R}_+^*, \circ)$ mit $x \circ y := \dfrac{1}{x \cdot y}$ die Schließungssätze (T) und (R) gelten. Was fällt bei der Konstruktion der Umkehroperationen $\backslash\backslash$ und $//$ von $\circ$ auf?

[1]) Vgl. aber E. FALCONER [53].

2. Wir haben in Kapitel 1 sowohl mit Hilfe der Thomsen- als auch der Reidemeister-Bedingung die Assoziativität einer Operation getestet. Die Substration, etwa in $\mathbf{Z}$, ist — wie man leicht prüft — bisymmetrisch und rechts-transitiv, so daß infolge der Sätze 4.2.10 und 4.2.13 in der nichtassoziativen Quasigruppe $(\mathbf{Z}, -)$ beide Schließungssätze gelten. Ist das ein Widerspruch?

3. Man zeige, daß die Quasigruppen

a) $(\mathbf{R}, \circ)$ mit $x \circ y := a \cdot x + b \cdot y + c$ $(a, b, c \in \mathbf{R}, a, b \neq 0)$,
b) $(\mathbf{R}^*, \square)$ mit $x \square y := r \cdot x^s \cdot y^t$ $(r, s, t \in \mathbf{Z}, s, t$ ungerade)

bisymmetrisch sind und gebe die jeweilige abelsche Gruppe samt Automorphismen an, über der $(\mathbf{R}, \circ)$ bzw. $(\mathbf{R}^*, \square)$ aufgrund des Satzes von Toyoda linear ist ($+$ ist die Addition, $\cdot$ die Multiplikation reeller Zahlen).

5. Spezielle Gruppoide und Quasigruppen

Wollen wir uns ein möglichst genaues Bild über die Eigenschaften eines Gruppoids machen, so stehen uns mit der Parastrophie, der Homomorphie und der Isotopie schon sehr wirksame Hilfsmittel zur Verfügung. Mitunter lassen sich aber auch schon ohne derartige Werkzeuge wichtige Einsichten in die Struktur solcher Gruppoide gewinnen. Im Rahmen dieses Kapitels ist es nicht möglich, etwa alle durch die verschiedenen Identitäten definierten Klassen von Gruppoiden zu untersuchen. Deshalb haben wir uns vorrangig für eine Identität, nämlich die Bisymmetrie, entschieden. In diesem Sinne hat der folgende erste Abschnitt also auch einen stark exemplarischen Charakter. Über weitere Klassen von Gruppoiden kann sich der interessierte Leser z. B. bei R. H. BRUCK [32], V. D. BELOUSOV [21] sowie J. DÉNES und A. D. KEEDWELL [44] informieren. Im Fall assoziativer Gruppoide bzw. assoziativer Quasigruppen verweisen wir auf die einschlägige Literatur zur Theorie der Halbgruppen bzw. Gruppen.

Ein Schwerpunkt dieses Kapitels ist es, Verzahnungen zwischen Eigenschaften eines Gruppoids aufzudecken. Die hierbei gewonnenen Erkenntnisse erlauben dann u. a. (wieder unter besonderer Berücksichtigung der Bisymmetrie), abelsche Gruppen auf z. T. neuen Wegen axiomatisch zu charakterisieren. Dabei wird an vielen Stellen deutlich werden, mit welchem Gewinn die Kenntnis zueinander parastropher Identitäten ausgenutzt werden kann.

Zum Abschluß dieses Kapitels zeigen wir, daß sich mit Hilfe totalsymmetrischer Quasigruppen, die idempotent oder unipotent sind, spezielle endliche Geometrien, sogenannte Steiner-Tripel-Systeme, erzeugen lassen. Umgekehrt liefert auch jedes Steiner-Tripel-System eine derartige Quasigruppe.

5.1. Bisymmetrische Gruppoide und Quasigruppen

Wir wissen bereits, daß die Bisymmetrie[1]) unter Parastrophie und Homomorphie invariant ist (Korollar 2.4.9 und Satz 3.3.4), nicht aber unter Isotopie. Allerdings ist jede bisymmetrische Quasigruppe zu einer abelschen Gruppe isotop (Korollar 4.2.11). Da eine abelsche Gruppe bisymmetrisch ist, ist die Bisymmetrie jedoch unter Loop-

[1]) Dieser Name geht wahrscheinlich auf J. ACZÉL [1] zurück und wurde von J. ACZÉL und I. FENYÖ [9] in die deutschsprachige Literatur eingeführt. Im englischen und russischen Schrifttum heißt diese Identität vorwiegend „mediality" bzw. „medial'nost". Daneben gibt es aber auch noch andere Bezeichnungen.

Isotopie invariant. Der Satz von TOYODA stellt darüber hinaus eine noch engere Beziehung zwischen bisymmetrischen Quasigruppen und abelschen Gruppen her. Weiterhin wissen wir, daß in einer bisymmetrischen Quasigruppe die Thomsen- (und die Reidemeister-) Bedingung erfüllt ist (Satz 4.2.10).

Die Bisymmetrie, die einerseits die Vertauschung der Innenglieder erlaubt, andererseits aber auch eine bestimmte Klammerung vorschreibt, nimmt eine Zwischenstellung in bezug auf die Assoziativität und Kommutativität ein. D. C. MURDOCH [121] nannte sie deshalb auch „verallgemeinertes assoziativ-kommutatives Gesetz". Ihre Stellung zu den beiden Identitäten (A) und (K) wird durch den folgenden Satz von R. H. BRUCK [29] verdeutlicht.

5.1.1. Satz. *Ein Gruppoid $(G, \cdot)$ mit dem neutralen Element n ist genau dann bisymmetrisch, wenn es assoziativ und kommutativ ist.*

Beweis. Für alle $x, y, z, u, v \in G$ gilt wegen (N) und (BS) einerseits $x \cdot y = nx \cdot yn = ny \cdot xn = y \cdot x$ und $xy \cdot z = xy \cdot nz = xn \cdot yz = x \cdot yz$ und andererseits wegen (A) und (K) $xy \cdot uv = x \cdot (y \cdot uv) = x \cdot (yu \cdot v) = x \cdot (uy \cdot v) = x \cdot (u \cdot yv) = xu \cdot yv$.

5.1.2 Korollar. *Eine kommutative Halbgruppe ist bisymmetrisch* (S. K. STEIN [164]).

5.1.3. Korollar. *Ein D-Gruppoid mit einem neutralen Element ist genau dann bisymmetrisch, wenn es eine abelsche Gruppe ist.*

5.1.4. Korollar. *Eine Loop ist genau dann bisymmetrisch, wenn sie eine abelsche Gruppe ist* (R. H. BRUCK [29]).

5.1.5. Korollar. *Eine Gruppe ist genau dann bisymmetrisch, wenn sie abelsch ist* (S. K. STEIN [164]).

Es gibt also keine nichtkommutativen bisymmetrischen Loops (Gruppen).

Das Resultat von S. K. STEIN (Korollar 5.1.2) können wir auf folgende Weise erweitern:

5.1.6. Satz. *Ein kommutatives Gruppoid $(G, \cdot)$ ist bisymmetrisch, wenn es assoziativ oder L-transitiv oder R-transitiv ist.*[1])

Beweis. Für alle $x, y, u, v \in G$ gilt mit (L-T) und (K) $xy \cdot uv = (ux \cdot uy) \cdot (yu \cdot yv) = (yu \cdot xu) \cdot (yu \cdot yv) = xu \cdot yv$; analog für (R-T).

Daß zwei der Identitäten (A), (K) und (BS) die dritte nach sich ziehen, gilt nur für den Fall, den Korollar 5.1.2 widerspiegelt. Wenn wir zusätzlich die Kürzbarkeit fordern, können wir die Kommutativität ableiten:

5.1.7. Satz. *Ein bisymmetrisches assoziatives K-Gruppoid $(G, \cdot)$ ist kommutativ.*

Beweis. Für alle $x, y, u, v \in G$ gilt $x \cdot (yu \cdot v) = x \cdot (y \cdot uv) = xy \cdot uv = xu \cdot yv = x \cdot (u \cdot yv) = x \cdot (uy \cdot v)$, so daß die Kürzbarkeit $y \cdot u = u \cdot y$ für beliebige $u, y \in G$ liefert.

[1]) Das heißt, wenn eine parastrophe Identität der Assoziativität erfüllt ist. („L" steht für „links", „R" für „rechts").

Daß damit folglich auch eine bisymmetrische assoziative Quasigruppe stets kommutativ ist, haben wir bereits in Korollar 5.1.5 festgehalten. Für die jeweiligen parastrophen Identitäten gilt damit aber zugleich das

5.1.8. Korollar. *Eine bisymmetrische L- (R-) transitive Quasigruppe erfüllt die R- (L-)-Systemregel.*

Den Beweis liefern die Sätze 2.4.3 und 2.4.5 sowie Korollar 2.4.9.

Im Verein mit der Idempotenz kann man von der Bisymmetrie auf die Selbstdistributivität einer Operation schließen (O. Frink [56]). Ein idempotentes bisymmetrisches Gruppoid $(G, \cdot)$ erfüllt nämlich wegen $x \cdot yz = xx \cdot yz = xy \cdot xz$ und $xy \cdot z = xy \cdot zz = xz \cdot yz$ sowohl die Links- als auch die Rechts-Selbstdistributivität.

Mit $x^2 := x \cdot x$ erhalten wir $(x \cdot y)^2 = x^2 \cdot y^2$ als einen Spezialfall der Bisymmetrie. Wir nehmen diese Beziehung zum Anlaß, die m-te *Potenz* x^m eines beliebigen Elementes x eines bisymmetrischen Gruppoids $(G, \cdot)$ zu definieren. In einer (Halb-) Gruppe ist es wegen des Assoziativgesetzes gleichgültig, ob x^3 für $x^2 \cdot x$ oder für $x \cdot x^2$ stehen soll. Für m Faktoren kann schließlich x^m sowohl durch $x^{m-1} \cdot x$ als auch durch $x \cdot x^{m-1}$ erklärt werden. Das Ergebnis ist dasselbe. In einem beliebigen bisymmetrischen Gruppoid $(G, \cdot)$ liegen schon bei x^3 die Dinge anders. Man denke z. B. an die Subtraktion in $\mathbf{Z}$. Deshalb müssen wir hier bei der einmal festgelegten Reihenfolge in der induktiven Definition von x^m bleiben. Wir setzen

$$x^1 := x \quad \text{und} \quad x^{m+1} := x^m \cdot x,$$

wobei x ein beliebiges Element aus G und m aus $\mathbf{N}^*$ ist (vgl. auch MfL Bd. 1, 3.5.). Von den in abelschen Gruppen geltenden Potenzgesetzen sind zwei schon in bisymmetrischen Gruppoiden erfüllt (D. C. Murdoch [120]), nämlich $(x \cdot y)^m = x^m \cdot y^m$ und $(x^{m_1})^{m_2} = (x^{m_2})^{m_1}$, wobei x, y aus G und m, m_1 und m_2 aus $\mathbf{N}^*$ sind. Die Beweise werden durch vollständige Induktion über m bzw. m_1 geführt.

Bevor wir jetzt einige Beispiele bisymmetrischer Operationen anführen, erinnern wir uns, daß die Operation $\cdot$ eines Gruppoids $(G, \cdot)$ eine Abbildung von $G \times G$ in G ist und deshalb auch als eine zweistellige Funktion F mit $D(F) = G \times G$ und $W(F) \subseteq G$ bezeichnet werden kann. In der Theorie der Funktionalgleichungen nennt man die Identität der Bisymmetrie dann entsprechend die *Funktionalgleichung der Bisymmetrie* oder kurz *Bisymmetriegleichung*. Sie hat dann mit $F(x, y) := x \cdot y$ die Gestalt

$$F\big(F(x, y), F(u, v)\big) = F\big(F(x, u), F(y, v)\big),$$

wobei x, y, u, v beliebige Elemente aus G sind. In der Theorie der Funktionalgleichungen wird nun bewiesen (vgl. J. Aczél [3]), daß man unter den zusätzlichen Voraussetzungen der Kürzbarkeit und Stetigkeit[1]) von F sämtliche Lösungen der Bisymmetriegleichung im Bereich der reellen Zahlen angeben kann. Das sind gerade die quasilinearen Funktionen, d. h., F ist von der Gestalt

5.1.9. $F(x, y) = h^{-1}\big(c_1 h(x) + c_2 h(y) + c_3\big)$,

[1]) Man beachte, daß es sich hier um die Stetigkeit einer zweistelligen Funktion handelt.

wobei h eine reellwertige stetige und echt monotone (einstellige) Funktion ist,[1]) c_1, c_2 und c_3 beliebige, aber feste reelle Zahlen mit $c_1, c_2 \neq 0$ sind (Beweis der Bisymmetrie von F!).

Solange der Definitionsbereich $D(F)$ der Funktion F ganz $\mathbf{R} \times \mathbf{R}$ ist, können wir unbedenklich das Wort Funktion durch das Wort Operation ersetzen. Andernfalls aber, wenn $D(F) = M \times M$ eine echte Teilmenge von $\mathbf{R} \times \mathbf{R}$ ist, müssen wir immer erst prüfen, ob auch stets $F(x, y) \in M$ gilt.

5.1.10. Beispiele bisymmetrischer Operationen

1. Setzen wir in (5.1.9) $x \circ y := F(x, y)$, so ergeben sich für eine stetige und echt monotone Funktion $h\colon M \to M$ mit $M \subseteq \mathbf{R}$ und für reelle Zahlen c_1, c_2, c_3 mit $c_1, c_2 \neq 0$ z. B. folgende (stetige) Operationen $\circ$ in M:[2])

a) $x \circ y := h^{-1}(c_1 h(x) + c_2 h(y) + c_3)$;

die Operation $\circ$ heißt *quasilineare Funktion*, $(M, \circ)$ ist ein bisymmetrisches K-Gruppoid.

b) $x \circ y := h^{-1}((1 - q)\, h(x) + q h(y))$ mit $q \neq 0, q \neq 1$

($c_1 = 1 - q$, $c_2 = q$, $c_3 = 0$); $x \circ y$ heißt *quasilinearer Mittelwert* von x und y; in $(M, \circ)$ gelten (BS), (I), (L-SD), (R-SD) und (KB).

c) $x \circ y := h^{-1}\left(\dfrac{r h(x) + s h(y)}{r + s}\right)$ mit $r > 0$, $s > 0$

$\left(c_1 = \dfrac{r}{r+s}, c_2 = \dfrac{s}{r+s}, c_3 = 0\right)$; $x \circ y$ heißt *quasilineares Gewichtsmittel* von x und y, r und s heißen die *Gewichte*. In $(M, \circ)$ gelten (BS), (I), (L-SD), (R-SD) und (KB).

d) $x \circ y := h^{-1}\left(\dfrac{h(x) + h(y)}{2}\right)$

($c_1 = c_2 = \dfrac{1}{2}$, $c_3 = 0$; $q = \dfrac{1}{2}$ in b); $r = s$ in c)). $x \circ y$ heißt *quasiarithmetischer Mittelwert* von x und y; in $(M, \circ)$ gelten (BS), (I), (L-SD), (R-SD), (K) und (KB).

e) $x \circ y := h^{-1}(h(x) + h(y))$

($c_1 = c_2 = 1$, $c_3 = 0$). Für $h(t) = t$ ergibt sich die Addition in $\mathbf{R}$, für $h(t) = \ln t$ die Multiplikation in $\mathbf{R}_+{}^*$. In $(M, \circ)$ gelten (BS), (K), (A) und (KB). Wenn es ein neutrales Element n gibt, dann ist $h(n) = 0$.

f) $x \circ y := h^{-1}(h(x) - h(y))$

($c_1 = 1$, $c_2 = -1$, $c_3 = 0$). Dieselben Funktionen h wie in e) liefern jetzt die Subtraktion in $\mathbf{R}$ bzw. die Division in $\mathbf{R}_+{}^*$. In $(M, \circ)$ gelten (BS), (U), (R-T), (L-SR) und (KB). Wenn es ein rechts-neutrales Element n_r gibt, ist $h(n_r) = 0$.

g) $x \circ y := h^{-1}(-h(x) - h(y))$

($c_1 = c_2 = -1$, $c_3 = 0$). In $(M, \circ)$ gelten (BS) und (TS), also auch (K), (L-SR), (R-SR), (HS) und (Q) (vgl. 5.2.). Ein neutrales Element kann es nicht geben.[3])

[1]) Da h als echt monoton vorausgesetzt wird, ist h (und damit auch h^{-1}) injektiv.

[2]) Wie die Menge M konkret aussieht, hängt von h ab. (Im gesamten Beispiel 5.1.10.1 ist $\cdot$ die Multiplikation.)

[3]) Eine notwendige Bedingung für die Existenz eines neutralen Elementes ist nämlich $c_1 = c_2 = 1$. Deshalb gibt es auch in den Beispielen b), c) und d) kein neutrales Element.

h) $x \circ y := c_1 x + c_2 y + c_3$

($h(t) = t$; $M = \mathbf{R}$). ($\mathbf{R}$, $\circ$) ist eine bisymmetrische Quasigruppe (vgl. auch Aufgabe 4.2.22.3a)).

i) $x \circ y := c \cdot x^{c_1} \cdot y^{c_2}$ mit $c = e^{c_3}$

($h(t) = \ln t$; $M = \mathbf{R}_+^*$). ($\mathbf{R}_+^*$, $\circ$) ist eine bisymmetrische Quasigruppe (vgl. auch Aufgabe 4.2.22.3b)).

2. Für beliebige Punkte P, Q einer euklidischen Ebene ε definieren wir $P \circ Q$ als den Mittelpunkt R der Strecke $\overline{PQ}$. Für $P = Q$ setzen wir $P \circ P := P$ (vgl. Beispiel 1.3.2.3). In (ε, $\circ$) gelten

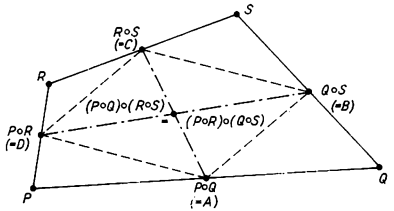


Abb. 28

(BS), (I), (L-SD), (R-SD), (K) und (Q). Kommutativität und Idempotenz liegen auf der Hand, während sich die Bisymmetrie in dem folgenden Satz widerspiegelt (Beweis!):

Die Mittelpunkte der Seiten eines Vierecks sind die Eckpunkte eines Parallelogramms (siehe Abb. 28).

Die Selbstdistributivität der Mittelpunktskonstruktion läßt sich (im Zusammenhang mit den Strahlensätzen) folgendermaßen interpretieren:

Die Verbindungsstrecken der Seitenmitten eines Dreiecks sind zu jeweils einer Dreiecksseite parallel (und halb so lang wie sie) (siehe Abb. 29).

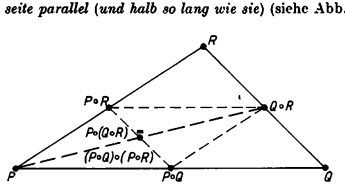


Abb. 29

Auch die Quasigruppe aus Beispiel 1.3.2.3, deren Operation man als geometrische Interpretation des arithmetischen Gewichtsmittels $x \circ y = \dfrac{bx + ay}{b + a}$ (vgl. Beispiel 5.1.10.1c)) auffassen kann, ist bisymmetrisch, idempotent und folglich auch selbstdistributiv. Kommutativität liegt jedoch nur im Fall $a : b = 1$ vor.

Das D-Gruppoid (c, $\circ$) aus Beispiel 1.2.2.3 erfüllt die Identitäten (BS), (I), (L−SD), (R−SD) und (L−SR).

Die Operationen in den Aufgaben 1.3.17.1 und 1.3.17.3 sind ebenfalls bisymmetrisch. Weitere Beispiele bisymmetrischer Operationen ergeben sich, wenn wir von einer bisymmetrischen Quasigruppe zu ihren parastrophen Quasigruppen übergehen.

5.1.11. Aufgaben

1. Man setze in Beispiel 5.10.1 d) der Reihe nach $h(t) = t$, $h(t) = \ln t$, $h(t) = \frac{1}{t}$, $h(t) = t^2$. Welche der uns bereits bekannten Mittelwerte erhält man? (Vgl. Aufgabe 1.1.8.1.)

2. Inwiefern sind die Operationen $\square$ und $*$ mit $x \square y := \frac{x \cdot y}{x + y}$ bzw. $x * y := \sqrt{x^2 + y^2}$ in $\mathbf{R}_+^*$ Spezialfälle von Beispiel 5.1.10.1 e)?

3. Man gebe die bisymmetrischen und totalsymmetrischen Quasigruppen $(M, \circ)$ an, die sich mit $h(t) = t$ bzw. $h(t) = \ln t$ in Beispiel 5.1.10.1 g) ergeben.

4. Man zeige, daß den Parametern c_1, c_2 und c_3 in den Beispielen 5.1.10.1 h) und i) die folgenden Eigenschaften entsprechen: $c_1 = 1$: (N_r); $c_2 = 1$: (N_l); $c_1 = -1$: (R-SR); $c_2 = -1$: (L-SR); $c_1 = 1$, $c_2 = -1$: (N_r), (L-SR), (R-T); $c_1 = -1$, $c_2 = 1$: (N_l), (R-SR), (L-T); $c_1 = c_2$: (K); $c_1 = c_2 = 1$: (N), (K), (A); $c_1 = c_2 = -1$: (K), (L-SR), (R-SR), (HS), (TS); $c_1 + c_2 = 1$, $c_3 = 0$: (I), (L-SD), (R-SD).

5.2. Zusammenhänge zwischen ausgewählten Identitäten

Wir haben schon mehrfach festgestellt, daß zwischen einzelnen Identitäten sehr enge Beziehungen bestehen. In diesem Abschnitt werden wir zeigen, daß es ein recht umfangreiches Beziehungsgefüge zwischen ausgewählten Eigenschaften eines Gruppoids gibt. In diesem Zusammenhang interessiert uns vor allem auch die Frage der gegenseitigen Ableitbarkeit. Gerade durch Untersuchungen über gegenseitige Ableitbarkeit und Nichtableitbarkeit von Eigenschaften binärer Operationen kann eine Bewichtung unseres Wissens über diese Operationen erreicht werden. Folgt nämlich aus der Eigenschaft E_1 einer Operation $\cdot$ die Eigenschaft E_2, nicht aber umgekehrt, so ist E_1 in diesem Sinne für die Operation $\cdot$ wichtiger als E_2.[1])

In Abb. 15 (S. 41) haben wir im übrigen bereits ein Netz von Beziehungen zwischen einigen Eigenschaften und den beiden Identitäten Kommutativität und Assoziativität zusammengestellt (vgl. auch die Sätze 5.1.1, 5.1.6 und 5.1.7 samt Folgerungen).

Die bisher in den einzelnen Kapiteln angeführten binären Operationen liefern für alle hier einbezogenen Identitäten Modelle. Eine Übersicht, in der das Erfüllt- bzw. Nichterfülltsein dieser Identitäten sowie der Eigenschaften (D), (KB), (Q) und (N) erfaßt wird, findet man in I. Lehmann [102] (vgl. auch Aufgabe 5.2.2.1.1).

Die Halbsymmetrie (HS)[2]) und die Totalsymmetrie (TS)[3]) zeigen in gewissen Fragen ein ähnliches Verhalten wie die Bisymmetrie (BS). So sind z. B. auch sie mit allen hier

[1]) Über die Notwendigkeit, die Ableitbarkeit bzw. Nichtableitbarkeit von Eigenschaften auch verstärkt in den Mathematikunterricht einzubeziehen, siehe etwa D. Ilse und I. Lehmann [78, 79, 80] sowie P. Wiechmann [179]. Insbesondere bieten die zweistelligen Relationen und Operationen sowie die (einstelligen) elementaren Funktionen viele Potenzen, die Lehrplanforderung, das Beziehungsgefüge zwischen mathematischen Sätzen sehen zu lernen, zu verwirklichen.

[2]) In 2.4. haben wir gezeigt, daß die Identitäten (L-HS) $x \cdot yx = y$ und (R-HS) $xy \cdot x = y$ zueinander äquivalent sind.

[3]) Vgl. Definition 2.4.10.

aufgeführten Eigenschaften verträglich[1]), erlauben andererseits auch eine Beschreibung spezieller abelscher Gruppen. Bevor wir uns aber letzterem widmen, schicken wir einige unmittelbare Folgerungen, die sich anhand dieser Identitäten ergeben, voraus. So sind in einem halbsymmetrischen Gruppoid (wofür wir im folgenden kürzer HS-Gruppoid schreiben) $(G, \cdot)$ die Gleichungen $a \cdot b = c$, $b \cdot c = a$ und $c \cdot a = b$ für alle $a, b, c \in G$ einander gleichwertig. Mit $a \cdot b = c$ gilt nämlich auch $b \cdot ab = bc$, so daß wegen (L-HS) $a = b \cdot c$ erfüllt ist. Analog ergeben sich daraus $ca = c \cdot bc$, also $c \cdot a = b$ und weiter $a \cdot ca = ab$, d. h. wieder $c = a \cdot b$.

5.2.1. Satz. *Ein HS-Gruppoid $(G, \cdot)$ ist eine Quasigruppe.*

Beweis (vgl. I. M. H. Etherington [51], A. Sade [154]). a) $(G, \cdot)$ ist ein D-Gruppoid: Für alle $a, b \in G$ haben die Gleichungen $a \cdot x = b$ und $y \cdot a = b$ jeweils wenigstens eine Lösung, nämlich $x = b \cdot a$ und $y = a \cdot b$:

$$a \cdot x = a \cdot ba = b, \quad y \cdot a = ab \cdot a = b.$$

b) $(G, \cdot)$ ist ein K-Gruppoid: Es sei $a \cdot x = a \cdot y$ oder $x \cdot a = y \cdot a$ für beliebige $a, x, y \in G$. Dann gilt auch $ax \cdot a = ay \cdot a$ bzw. $a \cdot xa = a \cdot ya$, so daß die Halbsymmetrie jeweils $x = y$ liefert.

5.2.2. Satz. *Eine HS-Quasigruppe $(Q, \cdot)$ ist genau dann eine Loop, wenn sie unipotent ist.*

Beweis. a) Für alle $x, y \in Q$ gilt $x \cdot x = x \cdot nx = n = y \cdot ny = y \cdot y$, wobei n das neutrale Element ist.

b) Wegen der Unipotenz ist $x \cdot x\,(= y \cdot y)$ von x unabhängig; wir setzen $n := x \cdot x$. Dann gilt $x \cdot n = x \cdot xx = x = xx \cdot x = n \cdot x$ für alle $x \in Q$.

5.2.3. Satz. *Aus der Totalsymmetrie folgt die Halbsymmetrie.*

Beweis. Für alle $x, y \in G$ eines TS-Gruppoids $(G, \cdot)$ gilt $x \cdot yx = x \cdot xy = y$.

5.2.4. Korollar. *Ein TS-Gruppoid ist eine Quasigruppe.*

Damit können wir die in 2.4. gewonnenen Äquivalenzen (TS) $\Leftrightarrow$ (K) $\wedge$ (L-SR) $\Leftrightarrow$ (K) $\wedge$ (R-SR) $\Leftrightarrow$ (L-SR) $\wedge$ (R-SR) übernehmen, auch wenn wir dort die Quasigruppeneigenschaft (Q) generell vorausgesetzt hatten. Mit anderen Worten, es gilt

5.2.5. Korollar. *Aus je zwei der Identitäten* (K), (L-SR) *und* (R-SR) *folgt die dritte.*[2])

Das bedeutet zugleich, daß in einer TS-Quasigruppe alle parastrophen Identitäten der Kommutativität erfüllt sind. Ferner sind jetzt sogar die sechs Gleichungen

[1]) Im Unterschied etwa zur Idempotenz, die mit der Unipotenz unvereinbar ist. In einer Quasigruppe schließen sich Idempotenz und Assoziativität gegenseitig aus; es gibt auch keine idempotente Loop (wenn die Ordnung jeweils größer als 1 sein soll).

[2]) Beweis ohne Bezug auf (Q):
(K) $\wedge$ (L-SR) $\Rightarrow$ (R-SR): $xy \cdot y = y \cdot xy = y \cdot yx = x$.
(K) $\wedge$ (R-SR) $\Rightarrow$ (L-SR): $x \cdot xy = xy \cdot x = yx \cdot x = y$,
(L-SR) $\wedge$ (R-SR) $\Rightarrow$ (K): $xy = (yx \cdot (yx \cdot x)) \cdot y = (yx \cdot y) \cdot y = yx$.

$ab = c$, $bc = a$, $ca = b$, $ba = c$, $cb = a$ und $ac = b$ zueinander äquivalent, was wir wegen des Zusammenfallens aller sechs parastrophen Operationen in einer TS-Quasigruppe schon in 2.4. festgestellt haben.

5.2.6. Satz. *In einer HS-Quasigruppe* $(Q, \cdot)$ *sind die drei Identitäten* (K), (L-SR) *und* (R-SR) *äquivalent.*

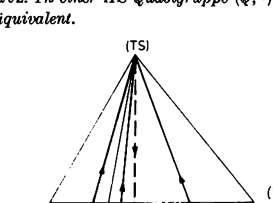


Abb. 30

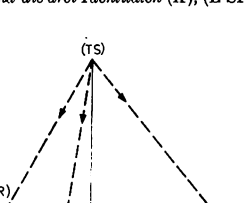


Abb. 31

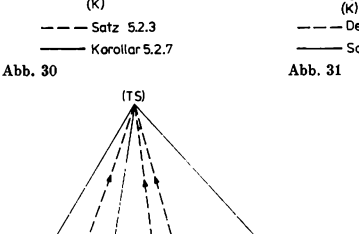


Abb. 32

Zum Beweis zeigen wir: (K) $\Rightarrow$ (L-SR) $\Rightarrow$ (R-SR) $\Rightarrow$ (K). Für alle $x, y \in Q$ gilt $x \cdot xy = x \cdot yx = y$, $xy \cdot y = xy \cdot (xy \cdot x) = x$ und $xy = x \cdot (yx \cdot x) = yx$.

5.2.7. Korollar. *Eine HS-Quasigruppe, in der eine parastrophe Identität der Kommutativität erfüllt ist, ist totalsymmetrisch.*

Die Umkehrung von Satz 5.2.3 ist dagegen nicht richtig.[1])

Die Sätze 5.2.3 und 5.2.6 lassen sich (einschließlich ihrer Folgerungen) besonders gut an einem Tetraeder veranschaulichen (siehe Abb. 30, 31, 32).

[1]) Eine HS-Quasigruppe, die nicht totalsymmetrisch ist, findet man in I. Lehmann [102].

Analog zu Satz 5.2.6 beweisen wir, daß auch die parastrophen Identitäten der Assoziativität gleichwertig sind, wenn Halb- oder Totalsymmetrie vorliegt.

5.2.8. Satz. *In einer HS-Quasigruppe* $(Q, \cdot)$ *sind die drei Identitäten* (A), (L-T) *und* (R-T) *äquivalent.*

Zum Beweis zeigen wir (A) $\Rightarrow$ (L-T) $\Rightarrow$ (R-T) $\Rightarrow$ (A). Für alle $x, y, z \in Q$ gilt $xy \cdot xz = x \cdot (y \cdot xz) = x \cdot (yx \cdot z) = (x \cdot yx) \cdot z = yz, xz \cdot yz = xz \cdot (xy \cdot xz) = xy$ und $xy \cdot z = xy \cdot (yz \cdot y) = x \cdot yz$.

Als Folgerung aus Satz 5.2.2 ergibt sich, daß auch in einer TS-Quasigruppe genau dann ein neutrales Element existiert, wenn sie unipotent ist. Beschränken wir uns auf eine der beiden Systemregeln, so ist die Unipotenz zur Existenz eines einseitig neutralen Elementes äquivalent.

5.2.9. Satz. *Ein Gruppoid* $(G, \cdot)$, *in dem die L-* (*R-*) *Systemregel gilt, ist genau dann unipotent, wenn es ein rechts-* (*links-*) *neutrales Element besitzt.*

Beweis. a) Wir setzen $n_r := x \cdot x$. Dann gilt $x \cdot n_r = x \cdot xx = x$ für alle $x \in G$, d. h., n_r ist rechts-neutrales Element.

b) Es sei n_r rechts-neutrales Element in $(G, \cdot)$. Dann gilt $x \cdot x = x \cdot xn_r = n_r$ für alle $x \in G$, d. h., $x \cdot x$ ist von x unabhängig, so daß $(G, \cdot)$ unipotent ist. Analog zeigt man (U) $\overset{\text{(R-SR)}}{\Leftrightarrow}$ (N_l). Im Überblick erhalten wir Abb. 33.

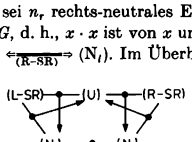


Abb. 33

Im Fall der Links- oder Rechts-Transitivität können wir, sofern die Operation umkehrbar oder kürzbar ist, sowohl auf die Unipotenz als auch die Existenz eines einseitig neutralen Elementes schließen.

5.2.10. Satz. *Ein L-* (*R-*) *transitives D- oder K-Gruppoid* $(G, \cdot)$ *ist unipotent und besitzt ein links-* (*rechts-*) *neutrales Element.*

Beweis. a) (D) $\wedge$ (L-T) $\Rightarrow$ (U) $\wedge$ (N_l). Es sei y_0 eine Lösung von $y \cdot a = b$; $a, b \in G$, beliebig. Dann gilt $b \cdot b = y_0 a \cdot y_0 a = a \cdot a$. Wir setzen $n_l := x \cdot x$. Für alle $b \in G$ gilt dann $n_l \cdot b = n_l \cdot y_0 a = xx \cdot y_0 a = y_0 y_0 \cdot y_0 a = y_0 \cdot a = b$, d. h., n_l ist links-neutrales Element in $(G, \cdot)$.

b) (KB) $\wedge$ (L-T) $\Rightarrow$ (U) $\wedge$ (N_l). Es sei $n_l := x_0 \cdot x_0$. Dann gilt $x_0 x_0 \cdot x_0 x_0 = x_0 \cdot x_0$, d. h. $n_l \cdot n_l = n_l$, also $n_l x \cdot n_l = n_l x \cdot n_l n_l = x \cdot n_l$, so daß für alle $x \in G$ wegen der Kürzbarkeit $n_l \cdot x = x$ gilt; n_l ist folglich links-neutrales Element. Die Unipotenz folgt aus $n_l \cdot xx = x \cdot x = xx \cdot xx$, wenn wir wieder die Kürzungsregel anwenden. Analog beweist man $\big((D) \vee (KB)\big) \wedge$ (R-T) $\Rightarrow$ (U) $\wedge$ (N_r) — vgl. M. Hosszú [71].

Um einmal zu illustrieren, wie wir auf den im nächsten Satz formulierten Zusammenhang gestoßen sind, rufen wir folgendes in Erinnerung: in einer assoziativen Quasigruppe gibt es ein neutrales Element, eine abelsche Gruppe ist bisymmetrisch. Also gilt unter der Voraussetzung (Q): (A) ∧ (K) ⇒ (N) ∧ (BS). Dann muß sich aber z. B. in der π_6-Parastrophen dieser abelschen Gruppe folgendes „abspielen": (R-T) ∧ (L-SR) ⇒ $\big((N_r) \wedge (U)\big)$ ∧ (BS). Es bleibt jetzt nur noch zu prüfen, ob die Quasigruppeneigenschaft (Q), auf die wir unter Parastrophie immer angewiesen sind, notwendig ist.

5.2.11. Satz. *Ein Gruppoid $(G, \cdot)$ ist genau dann L- $(R$-$)$ transitiv und erfüllt die R- $(L$-$)$Systemregel, wenn es unipotent und bisymmetrisch ist sowie ein links- (rechts-) neutrales Element n_l (n_r) besitzt.*

Beweis. 1. Für alle $x, y, z \in G$ gilt $xy \cdot xz = xx \cdot yz = n_l n_l \cdot yz = yz$ und $xy \cdot y = xy \cdot n_l y = xn_l \cdot yy = xn_l \cdot xx = xx \cdot n_l x = n_l n_l \cdot n_l x = x$.

2. Für alle $x, y \in G$ gilt

$$\begin{aligned} x \cdot x &= yx \cdot yx = \big((yx \cdot y) \cdot y\big) \cdot yx \\ &= \big((yx \cdot y) \cdot y\big) \cdot \big((yx \cdot y) \cdot (yx \cdot x)\big) \\ &= y \cdot (yx \cdot x) = y \cdot y, \end{aligned}$$

d. h. $(G, \cdot)$ ist unipotent. Wegen Satz 5.2.9 gilt dann auch (N_l). Zum Nachweis der Bisymmetrie stützen wir uns auf die beiden folgenden Identitäten

a) $xy \cdot uv = (uv \cdot y) \cdot x$,

b)[1] $xy \cdot z = zy \cdot x$,

die wegen

$$\begin{aligned} xy \cdot uv &= \big((uv \cdot x) \cdot (uv \cdot y)\big) \cdot uv \\ &= \big((uv \cdot x) \cdot (uv \cdot y)\big) \cdot \big((uv \cdot x) \cdot x\big) \\ &= (uv \cdot y) \cdot x \end{aligned}$$

und

$$xy \cdot z = xy \cdot n_l z = (n_l z \cdot y) \cdot x = zy \cdot x$$

in einem links-transitiven Gruppoid, das die Rechts-Systemregel erfüllt, gelten. Für alle $x, y, u, v \in G$ gilt folglich $xy \cdot uv = (uv \cdot y) \cdot x = (yv \cdot u) \cdot x = xu \cdot yv$. Entsprechend zeigt man: (R-T) ∧ (L-SR) ⇔ (U) ∧ (N_r) ∧ (BS). Anstelle der Identitäten a) und b) stützt man sich auf $xy \cdot uv = v \cdot (u \cdot xy)$ und die Abel-Grassmann-Identität $x \cdot yz = z \cdot yx$.

Während aus der Assoziativität und der Existenz eines neutralen Elementes nicht auf weitere Eigenschaften geschlossen werden kann, erlauben dies die Links- und Rechts-Transitivität (als parastrophe Identitäten von (A)) in beeindruckender Weise.

[1] b) ist eine parastrophe Identität der Abel-Grassmann-Identität (vgl. Satz 2.4.8).

5.2.12. Satz. *Ein L- (R-) transitives Gruppoid $(G, \cdot)$ mit einem neutralen Element n ist eine abelsche Gruppe, in der alle parastrophen Identitäten der Kommutativität und der Assoziativität erfüllt sind, folglich auch die Identitäten* (U), (BS), (HS) *und* (TS).

Beweis. Wir zeigen, daß $(G, \cdot)$ eine (unipotente) TS-Loop ist. Da dann alle parastrophen Operationen zusammenfallen, sind auch die parastrophen Identitäten der Links- bzw. Rechts-Transitivität erfüllt. Die Bisymmetrie ergibt sich dann aufgrund von Korollar 5.1.2, die Halbsymmetrie folgt aus Satz 5.2.3.

Es sei $(G, \cdot)$ also ein L-transitives Gruppoid. Dann gilt mit (N) $x \cdot xy = xn \cdot xy = ny = y$ für alle $x, y \in G$, so daß die Links-Systemregel erfüllt ist. Wegen Satz 5.2.9 ist $(G, \cdot)$ dann auch unipotent. Setzen wir in (L-T) $z = x$, so erhalten wir $xy \cdot xx = yx$. Hieraus folgt mit (U) und (N) die Kommutativität von $(G, \cdot)$. Damit ist $(G, \cdot)$ eine (unipotente) TS-Loop. Der Beweis mit (R-T) verläuft analog.

Die bis hierher gewonnenen Ergebnisse ermöglichen es, eine Reihe äquivalenter Axiomensysteme für unipotente Gruppen[1]) anzugeben (vgl. I. Lehmann [102]). Wir begnügen uns hier mit einer Möglichkeit.

5.2.13. Satz. *Ein unipotentes Gruppoid $(G, \cdot)$ ist genau dann eine Gruppe, wenn es bisymmetrisch ist und ein neutrales Element n besitzt.*

Beweis. a) In einer unipotenten Gruppe $(G, \cdot)$ gelten trivialerweise (U), (A) und (N). Da der direkte Nachweis der Bisymmetrie umständlich ist, schließen wir mit $x \cdot xy = xx \cdot y = nn \cdot y = y$ und $xy \cdot y = x \cdot yy = x \cdot nn = x$ zunächst auf die beiden Systemregeln. Das liefert mit Korollar 5.2.5 die Totalsymmetrie, folglich auch die Kommutativität. Wegen Korollar 5.1.2 ist $(G, \cdot)$ dann auch bisymmetrisch. (Darüber hinaus sind in einer unipotenten Gruppe auch (HS), (L-T) und (R-T) erfüllt.)

b) Es gelte in $(G, \cdot)$ (U), (BS) und (N). Dann ist wegen Satz 5.1.1 auch (A) erfüllt. Unter a) haben wir bewiesen, daß allein aus (U), (A) und (N) die Totalsymmetrie folgt, $(G, \cdot)$ mithin eine Quasigruppe (Gruppe) ist.

Da eine abelsche Gruppe bisymmetrisch ist, gilt dies aufgrund der Parastrophieinvarianz auch für ihre Umkehroperationen. In diesem Zusammenhang haben wir uns in 2.4 gefragt, ob sich diese — ja schon am Beispiel der vier Grundrechenoperationen leicht feststellbare — Gemeinsamkeit von Operation und Umkehroperationen nicht für elementare axiomatische Betrachtungen nutzen läßt. Auf den ersten Teil der in Kapitel 2 gestellten Frage, wann eine bisymmetrische Operation eine abelsche Gruppenoperation ist, geben bereits die Korollare 5.1.3, 5.1.4 und 5.1.5 eine Antwort (vgl. auch Satz 5.2.13). Wann eine bisymmetrische Operation nun gerade eine Umkehroperation einer abelschen Gruppenoperation ist, wird in I. Lehmann [99] ausführlich behandelt. Wir halten hier nur fest, daß z. B. schon die Eigenschaften (U) und (N_r) das Gewünschte leisten.

[1]) In der Terminologie der Gruppentheorie bedeutet Unipotenz gerade, daß jedes (vom neutralen Element n verschiedene) Gruppenelement x von der Ordnung 2 ist; zur Definition der Ordnung eines Gruppenelementes vgl. MfL Bd. 3, 12.2.

5.2.14. Satz. *Es sei $(G, /)$ ein bisymmetrisches Gruppoid. Es gibt genau dann eine abelsche Gruppe $(G, \cdot)$, so daß*

5.2.15. $x/y = x \cdot y^{-1}$

gilt, wenn $(G, /)$ unipotent ist und ein rechts-neutrales Element n_r besitzt.

Beweis. 1. Es sei $(G, \cdot)$ eine abelsche Gruppe, also mit Korollar 5.1.5 eine bisymmetrische Loop. Dann ist ihre π_6-Parastrophe $(G, /)$ wegen Satz 2.4.1 und Korollar 2.4.9 eine unipotente bisymmetrische Rechts-Loop. Ferner gilt (5.2.15) (vgl. (1.5.3)).

2. Es seien in $(G, /)$ neben (BS) auch (U) und (N_r) erfüllt. Aufgrund von Satz 5.2.11 gelten damit auch (R-T) und (L-SR). Dann ist $(G, \cdot)$ mit $x \cdot y := x/(n_r/y)$ eine abelsche Gruppe:

a) n_r ist das neutrale Element von $(G, \cdot)$:

$$x \cdot n_r = x/(n_r/n_r) = x$$

und

$$n_r \cdot x = n_r/(n_r/x) = x.$$

b) $x^{-1} := n_r/x$ ist das inverse Element von x in $(G, \cdot)$:

$$x \cdot x^{-1} = x/\big(n_r/(n_r/x)\big) = x/x = n_r/n_r = n_r$$

und

$$x^{-1} \cdot x = (n_r/x)/(n_r/x) = n_r/n_r = n_r.$$

c) $(G, \cdot)$ ist assoziativ:

$$\begin{aligned} x \cdot yz &= x/\big(n_r/\big(y/(n_r/z)\big)\big) = x/\Big(\big((n_r/z)/(n_r/z)\big)/\big(y/(n_r/z)\big)\Big) \\ &= x/\big((n_r/z)/y\big) = \big(x/(n_r/y)\big)/\Big(\big((n_r/z)/y\big)/(n_r/y)\Big) \\ &= \big(x/(n_r/y)\big)/\big((n_r/z)/n_r\big) = \big(x/(n_r/y)\big)/(n_r/z) \\ &= xy \cdot z. \end{aligned}$$

Damit ist $(G, \cdot)$ eine Gruppe.

d) Nachweis der Beziehung (5.2.15):

$$x \cdot y^{-1} = x \cdot (n_r/y) = x/\big(n_r/(n_r/y)\big) = x/y.$$

Somit ist $/$ die Rechtsdivision der Gruppenoperation $\cdot$, d. h., $(G, \cdot)$ und $(G, /)$ sind parastrophe Quasigruppen. Da $(G, /)$ nach Voraussetzung bisymmetrisch ist, zieht Korollar 2.4.9 die Bisymmetrie der Gruppe $(G, \cdot)$ nach sich. Folglich ist $(G, \cdot)$ auch kommutativ, also eine abelsche Gruppe.

Ersetzen wir in Satz 5.2.14 (N_r) durch (N_l) sowie (5.2.15) durch $x \backslash y = x^{-1} \cdot y$, so erhalten wir den entsprechenden Satz für ein bisymmetrisches Gruppoid $(G, \backslash)$. Die Gruppenoperation wird dabei durch $x \cdot y := (x \backslash n_l) \backslash y$ definiert.

Der Satz von Toyoda (vgl. 4.2.) stellt eine Beziehung zwischen einer bisymmetrischen Quasigruppe $(Q, \cdot)$ und einer abelschen Gruppe $(Q, \circ)$ her. Auf besonders ein-

fache und dabei wieder wechselseitige Weise gehen derartige Quasigruppen auseinander hervor, wenn wir in $(Q, \cdot)$ neben der Bisymmetrie auch noch Totalsymmetrie voraussetzen (I. M. H. Etherington [51]).

5.2.16. Satz. *Es sei $(Q, \cdot)$ eine bisymmetrische TS-Quasigruppe, a ein beliebiges, aber festes Element von Q. Dann ist $(Q, \circ)$ mit*

5.2.17. $$x \circ y := a \cdot xy$$

eine abelsche Gruppe, in der a das neutrale Element ist. Ist umgekehrt $(Q, \circ)$ eine abelsche Gruppe, b ein beliebiges, aber festes Element aus Q, dann ist $(Q, \cdot)$ mit

5.2.18. $$x \cdot y := b \circ (x \circ y)^{-1}$$

eine bisymmetrische TS-Quasigruppe.

Beweis. 1. Das durch (5.2.17) definierte Gruppoid $(Q, \circ)$ ist kommutativ, denn für alle $x, y \in G$ gilt $x \circ y = a \cdot xy = a \cdot yx = y \circ x$; a ist das neutrale Element: $x \circ a = a \circ x = a \cdot ax = x$. Das inverse Element von x ist $x^{-1} := aa \cdot x$, denn es gilt $x^{-1} \circ x = x \circ x^{-1} = a \cdot \big(x \cdot (aa \cdot x)\big) = a \cdot \big(x \cdot (x \cdot aa)\big) = a \cdot aa = a$. Schließlich ist $\circ$ auch assoziativ:

$$\begin{aligned} x \circ (y \circ z) &= a \cdot \big(x \cdot (a \cdot yz)\big) = a \cdot \big((z \cdot zx) \cdot (a \cdot yz)\big) \\ &= a \cdot \big(za \cdot (zx \cdot yz)\big) = a \cdot \big(az \cdot (xz \cdot yz)\big) \quad . \\ &= a \cdot \big(az \cdot (xy \cdot zz)\big) = a \cdot \big((a \cdot xy) \cdot (z \cdot zz)\big) \\ &= a \cdot \big((a \cdot xy) \cdot z\big) = (x \circ y) \circ z \end{aligned}$$

für alle $x, y, z \in Q$. Damit ist $(Q, \circ)$ eine abelsche Gruppe.

2. Aufgrund der in abelschen Gruppen geltenden Gesetze geht (5.2.18) in $x \cdot y = x^{-1} \circ y^{-1} \circ b$ über. Da die Abbildung $f: Q \to Q$ mit $x \mapsto x^{-1}$ ein Automorphismus der abelschen Gruppe $(Q, \circ)$ ist (Aufgabe 3.3.16.3), sind wegen $f = g$ die Beziehungen (4.2.19) und (4.2.20) aus dem Satz von Toyoda erfüllt. Folglich ist $(Q, \cdot)$ eine bisymmetrische Quasigruppe. Weiter erhalten wir für alle $x, y \in Q$

$$x \cdot y = b \circ (x \circ y)^{-1} = b \circ (y \circ x)^{-1} = y \cdot x$$

und

$$x \cdot xy = b \circ \Big(x \circ \big(b \circ (x \circ y)^{-1}\big)\Big)^{-1} = b \circ x^{-1} \circ b^{-1} \circ x \circ y = y,$$

so daß $(Q, \cdot)$ auch totalsymmetrisch ist.

Wenn in der Umkehrung a das neutrale Element der abelschen Gruppe $(Q, \circ)$ ist, dann erhalten wir mit (5.2.18) sofort $a \cdot a = b$. Haben wir eine bisymmetrische TS-Quasigruppe $(Q, \cdot)$ mit Hilfe dieser abelschen Gruppe $(Q, \circ)$ vermöge (5.2.18) dargestellt; dann gilt auch die Beziehung (5.2.17):

$$a \cdot xy = b \circ \Big(a \circ \big(b \circ (x \circ y)^{-1}\big)\Big)^{-1} = x \circ y.$$

Haben wir umgekehrt eine abelsche Gruppe $(Q, \circ)$ über (5.2.17) aus einer bi-

symmetrischen TS-Quasigruppe $(Q, \cdot)$ gewonnen, dann gilt unter der Voraussetzung $a \cdot a = b$ auch (5.2.18):

$$(x \circ y) \circ (x \cdot y) = a \cdot \big((a \cdot xy) \cdot xy\big) = a \cdot \big(xy \cdot (xy \cdot a)\big) = a \cdot a = b.$$

Hieraus folgt $x \cdot y = b \circ (x \circ y)^{-1}$. Das heißt mit anderen Worten: Legen wir die mit Hilfe von (5.2.17) aus einer bisymmetrischen TS-Quasigruppe $(Q, \cdot)$ erhaltene abelsche Gruppe $(Q, \circ)$ zugrunde, um mittels (5.2.18) hieraus ihrerseits eine bisymmetrische TS-Quasigruppe $(Q, *)$ zu konstruieren, so fällt $(Q, *)$ genau dann mit der Ausgangsquasigruppe $(Q, \cdot)$ zusammen, wenn $a \cdot a = b$ gilt.

H. KIESEWETTER [85] setzt in (5.2.16) anstelle der Totalsymmetrie die Halbsymmetrie und die Kommutativität voraus. Wegen Korollar 5.2.7 ist dann aber auch die Totalsymmetrie erfüllt. Darüber hinaus stellt er die Frage, ob dieser Satz auf nichtkommutative Operationen verallgemeinert werden kann. Zwar bleibt die Aussage des Satzes (in der Version von H. KIESEWETTER) gültig, wenn unter der Voraussetzung der Halbsymmetrie die Kommutativität gegen die Links- oder Rechts-Systemregel ausgetauscht wird — das ist aber wegen Korollar 5.2.7 keine Beantwortung der gestellten Frage. Wenn wir in Satz 5.2.16 auf die Kommutativität (bzw. die Links-Systemregel) verzichten, gilt der folgende

5.2.19. Satz. *Es sei $(Q, \cdot)$ eine bisymmetrische HS-Quasigruppe, c und d seien beliebige, aber feste Elemente aus Q. Dann ist $(Q, \circ)$ mit*

5.2.20. $x \circ y := cx \cdot yd$

eine abelsche Gruppe, in der $n := d \cdot c$ das neutrale Element ist.

Beweis. Aufgrund der Halbsymmetrie gilt $R_c(L_c(x)) = cx \cdot c = x$ für alle $c, x \in Q$, also $R_c \bullet L_c = I_Q$ (vgl. Aufgabe 1.7.9.1). Daraus folgt einerseits $R_c = L_c^{-1}$ und andererseits auch $L_c = R_c^{-1}$ für alle $c \in Q$. Damit nimmt der aus (5.2.20) abzulesende Hauptisotopismus $T = (L_c, R_d, I_Q)$ von $(Q, \circ)$ auf $(Q, \cdot)$ die Gestalt $T = (R_c^{-1}, L_d^{-1}, I_Q)$ an. Damit ist $(Q, \circ)$ ein LH-Isotop von $(Q, \cdot)$ und wegen Korollar 4.2.12 folglich eine abelsche Gruppe. Das neutrale Element ist $n = d \cdot c$.

Es ließe sich mittels (5.2.18) dieselbe Umkehrung wie in Satz 5.2.16 vornehmen. Die so definierte Quasigruppe wäre dann aber nicht nur bisymmetrisch und halbsymmetrisch, sondern sofort auch kommutativ (und damit totalsymmetrisch). Ist andererseits die Quasigruppe $(Q, \cdot)$ in Satz 5.2.19 auch kommutativ, so geht (5.2.20) in (5.2.17) über. Denn mit $a := c \cdot d = d \cdot c = n$ erhalten wir also $x \circ y = cx \cdot yd = cx \cdot dy = cd \cdot xy = a \cdot xy$.

5.2.21. Aufgaben

1. Man überzeuge sich von der Richtigkeit der Tabelle 6 (siehe S. 130).
Alle angegebenen Gruppoide sind Quasigruppen.

2. In der Menge $M = \{0, 1, 2, 3, 4\}$ sei $\oplus$ die Addition ganzer Zahlen modulo 5. Man gebe die Strukturtafel von $(M, \oplus)$ an und konstruiere vermöge $x \circ y := 2x + y + 1 \pmod 5$ die Strukturtafel von $(M, \circ)$. Welche Identitäten sind in den jeweiligen Gruppoiden erfüllt? (Hinweis: Satz von TOYODA.)

3. Man zeige, daß das zu x inverse Element x^{-1} in der abelschen Gruppe $(Q, \circ)$ aus Satz 5.2.19 die Gestalt $x^{-1} = d \cdot (dc \cdot cx)$ hat. Ist die bisymmetrische HS-Quasigruppe $(Q, \cdot)$ auch kommutativ, so nimmt es die Gestalt $x^{-1} = nn \cdot x$ an.

Tabelle 6

Operation	Trägermenge	N_l	N_r	U	I	BS	SD	A	L-T	R-T	K	L-SR	R-SR	HS	TS
$x \circ y := 3x + 2y$	**R**	–	–	–	–	×	–	–	–	–	–	–	–	–	–
$x \circ y := 3x + y + 1$	**R**	×	–	–	–	×	–	–	–	–	–	–	–	–	–
$x \circ y := x - 2y + 3$	**R**	–	×	–	–	×	–	–	–	–	–	–	–	–	–
$x \circ y := \frac{1}{2}(x - y)$	**R**	–	–	×	–	×	–	–	–	–	–	–	–	–	–
$x \circ y := 3x - y + 5$	**R**	–	–	–	–	×	–	–	–	–	–	×	–	–	–
$x \circ y := -x + 2y - 3$	**R**	–	–	–	–	×	–	–	–	–	–	–	×	–	–
$x \circ y := \sqrt{x \cdot y}$	$\mathbf{R}_+^*$	–	–	–	×	×	×	–	–	–	×	–	–	–	–
$x \circ y := 2x - y$	**R**	–	–	–	×	×	×	–	–	–	–	×	–	–	–
$x \circ y := \frac{y}{x^2}$	$\mathbf{R}_+^*$	–	–	–	×	×	×	–	–	–	–	–	×	–	–
$x \circ y := x + y$	**R**	×	×	–	–	×	–	×	–	–	×	–	–	–	–
$x \circ y := \frac{y}{x}$	**R***	×	–	×	–	×	–	–	×	–	–	–	×	–	–
$x \circ y := x - y$	**R**	–	×	×	–	×	–	–	–	×	–	×	–	–	–
$x \circ y := \frac{1}{x \cdot y}$	**R***	–	–	–	–	×	–	–	–	–	×	×	×	×	×
$(G_3, \circ)$ – vgl. Seite 28		–	–	–	×	×	×	–	–	–	×	×	×	×	×
$\mathfrak{B}_4$ – vgl. Seite 75		×	×	×	–	×	–	×	×	×	×	×	×	×	×

5.3. Totalsymmetrische Quasigruppen und Steiner-Tripel-Systeme

Mit den folgenden Ausführungen soll demonstriert werden, daß zwischen Quasigruppen und endlichen Geometrien sehr enge und wechselseitige Querverbindungen bestehen. Dafür müssen einige Ergebnisse zu endlichen Geometrien hier mitgeteilt werden. Die Beweise findet man in der angegebenen Literatur.

Sind $\mathfrak{P}$ und $\mathfrak{G}$ endliche disjunkte (nichtleere) Mengen, deren Elemente *Punkte* bzw. *Geraden* genannt werden, so ist mit $I \subseteq \mathfrak{P} \times \mathfrak{G}$ eine *Inzidenzrelation* gegeben. Gilt $(P, g) \in I$ bzw. PIg mit $P \in \mathfrak{P}$ und $g \in \mathfrak{G}$, so sagen wir, der Punkt P indiziert mit der Geraden g, P liegt auf g oder g geht durch P (vgl. etwa P. Dembowski [43]).[1])

Eine *endliche reguläre Ebene* ε_r^k läßt sich dann als eine Inzidenzstruktur $(\mathfrak{P}, \mathfrak{G}, I)$ charakterisieren, die folgende Bedingungen erfüllt (L. Szamkolowicz [167], N. K. Pucharev [140]):

5.3.1. Durch zwei Punkte geht genau eine Gerade;

5.3.2. Auf jeder Geraden liegen genau k Punkte und

[1]) G. Pickert [136] verlangt darüber hinaus, daß $\mathfrak{G}$ eine Menge von Teilmengen von $\mathfrak{P}$ ist. Die Inzidenz wird dann zur $\in$-Relation der Mengenlehre. (In der Tat betrachten wir im folgenden auch nur solche Inzidenzstrukturen, in denen die Geraden Punktmengen sind.)

5.3.3. Durch jeden Punkt gehen genau r Geraden[1])

$$(r \geqq k \geqq 2).$$

Die so definierte Ebene $\varepsilon_r{}^k$ besitzt $v = r(k-1)+1$ Punkte und $b = r^2 - \dfrac{r^2-r}{k}$ Geraden.[2])

Da in einer endlichen regulären Ebene $\varepsilon_r{}^k$ durch einen Punkt P, der nicht auf einer vorgegebenen Geraden g liegt, $\sigma := r - k$ Geraden gehen, die g nicht schneiden[3]), unterscheidet man für $\sigma = 0$, $\sigma = 1$ und $\sigma \geqq 2$ zwischen *projektiven*, *affinen* bzw. *Lobačevskij-Ebenen*[4]) (N. K. Pucharev [140], H. Zeitler [186]).

Eine endliche reguläre Ebene $\varepsilon_r{}^k = (\mathfrak{P}, \mathfrak{G}, I)$ mit $k = 3$ heißt ein *Steiner-Tripel-System* (kurz: STS). Diese speziellen endlichen Geometrien sind nach dem Schweizer Mathematiker Jakob Steiner (1796–1863) benannt.[5])

Wegen $k = 3$ ist Axiom (5.3.3) entbehrlich, es läßt sich aus (5.3.1) und (5.3.2) beweisen. Ein STS ist wegen $\sigma = r - 3$ für $r = 3$ eine projektive, für $r = 4$ eine affine und für $r \geqq 5$ eine Lobačevskij-Ebene.

Aufgrund der beiden Forderungen, daß

a) mit je zwei Punkten genau eine Gerade inzidiert,

b) mit jeder Geraden genau drei Punkte inzidieren,

lassen sich folgende Beziehungen ableiten (vgl. N. K. Pucharev [140], G. Pickert [136], H. Zeitler [187]):

5.3.4. $v = 2r + 1$,

5.3.5. $b = \dfrac{1}{6} v(v-1)$,

5.3.6. $r = \dfrac{3b}{v}$.

Es ist darüber hinaus sinnvoll, $v \geqq 3$ zu fordern, so daß mit (5.3.4) und (5.3.5) insgesamt die Ungleichungen

5.3.7. $v \geqq 3, r \geqq 1, b \geqq 1$

erfüllt sind.

---

[1]) Fordert man in (5.3.1), daß mit zwei Punkten stets genau λ Geraden inzidieren, so erhält man einen *Blockplan* (auch BIBD = balanced incomplete block design genannt). Läßt man die Forderung (5.3.1) ganz fallen, so liegt nur eine *taktische Konfiguration* vor (vgl. etwa H. Zeitler [188]).

[2]) Da man die Elemente von $\mathfrak{G}$ oft auch als Blöcke bezeichnet, hat sich der Buchstabe b eingebürgert; v steht in der Statistik für Varietäten.

[3]) d. h. keinen gemeinsamen Punkt mit g haben

[4]) Will man betonen, daß die Axiome der Anordnung hier keine Rolle spielen, so bezeichnet man sie auch präziser als projektive, affine bzw. Lobačevskij-Inzidenz-Ebenen. H. Zeitler [186] nennt eine endliche reguläre Ebene $\varepsilon_r{}^k$ mit $\sigma \geqq 2$ eine (nichteuklidische) σ-Inzidenzebene.

[5]) J. Steiner stellte in [165] folgende Aufgabe: „Welche Zahl n von Elementen hat die Eigenschaft, daß sich die Elemente so zu dreien ordnen lassen, daß je zwei in einer, aber nur in einer Verbindung vorkommen?" – J. Steiner war von 1834 bis zu seinem Tode Professor an der Universität Berlin.

Schon J. STEINER [165] hat 1853 gezeigt, daß

$$v \equiv 1 \pmod 6 \quad \text{oder} \quad v \equiv 3 \pmod 6$$

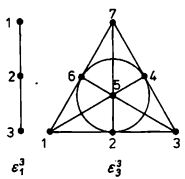
eine notwendige Bedingung für die Existenz von Steiner-Tripel-Systemen ist. M. REISS [147] bewies dann 1859, daß diese Bedingung auch hinreichend ist. T. P. KIRKMAN [86] hatte dieses Problem zwar schon 1847 gelöst, es blieb jedoch zunächst unbeachtet.[1]) Demnach existieren STS für folgende v:

$v \equiv 1 \pmod 6$		7		13		19		25	⋯
$v \equiv 3 \pmod 6$	3		9		15		21		⋯

E. NETTO [124] präzisierte diese Aussage, indem er bewies, daß für $v = 3$, 7 und 9 bis auf Isomorphie genau ein STS existiert, es für $v \geqq 13$ aber immer wenigstens zwei nichtisomorphe STS gibt. Dabei heißen zwei STS isomorph, wenn es eine bijektive Abbildung gibt, die neben den Elementen (Punkten) auch die Tripel (Geraden) ineinander überführt. Für weitere Informationen über STS sei auf die Bibliographie von J. DOYEN und A. ROSA [46] verwiesen.

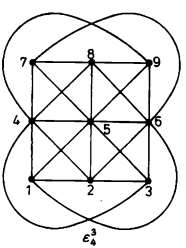
5.3.8. Beispiele für Steiner-Tripel-Systeme

1. ε_1^3 mit $k = 3$, $r = 1$, $v = 3$, $b = 1$ ($\sigma = -2$)[2]) (siehe Abb. 34).



ε_1^3 ε_3^3

Abb. 34



ε_4^3

Abb. 35

2. ε_3^3 mit $k = r = 3$, $v = b = 7$, $\sigma = 0$ (siehe Abb. 34). Dieses STS ist eine projektive Ebene. Andere (isomorphe) geometrische Modelle findet man in H. ZEITLER [187].

3. ε_4^3 mit $k = 3$, $r = 4$, $v = 9$, $b = 12$, $\sigma = 1$ (siehe Abb. 35). Dieses STS ist eine affine Ebene.

Die STS mit $v \geqq 13$ sind dann allesamt Lobačevskij-Ebenen.

Wenn wir uns mit STS beschäftigen wollen, können wir auch auf jegliche geometri-

[1]) Bekannt ist das Kirkmansche „Schulmädchen-Problem": 15 Mädchen eines Pensionates machen einen Spaziergang in 5 Reihen zu je 3 Mädchen. Welche Anordnung ist zu treffen, wenn an 7 Tagen jedes Mädchen mit jedem anderen genau einmal in einer Dreiergruppe zusammenkommt? ε_7^3 mit $k = 3$, $r = 7$, $v = 15$, $b = 35$, $\sigma = 4$ löst diese Aufgabe. Ein Zahlenmodell für dieses STS findet man z. B. in H. ZEITLER [186].

[2]) ε_1^3 wird nicht als reguläre Ebene aufgefaßt, da in diesem Fall die Ungleichungen von FISHER nicht erfüllt sind ($k \leqq r$ und $v \leqq b$).

sche Motivation verzichten (vgl. etwa R. H. Bruck [33], V. D. Belousov [21], J. Dénes und A. D. Keedwell [44]).

5.3.9. Definition. Es sei Q eine aus v Elementen bestehende Menge mit $v \geqq 3$, S eine nichtleere Menge dreielementiger Teilmengen von Q. S heißt *Steiner-Tripel-System* $:\Leftrightarrow$

Für beliebige $a, b \in Q$ mit $a \neq b$ gibt es genau ein $c \in Q$, so daß $\{a, b, c\} \in S$ gilt.

Ist $\{a, b, c\}$ eine Dreiermenge, so sind die Elemente a, b und c (nach Definition der Dreiermenge) paarweise verschieden.

Man kann nun zwischen diesen aus kombinatorischen Fragestellungen heraus entstandenen und später auch für die endliche Geometrie entdeckten STS und speziellen algebraischen Strukturen eine enge Verbindung herstellen. So liefert nämlich jedes STS eine spezielle TS-Quasigruppe. Umgekehrt lassen sich mit Hilfe derartiger Quasigruppen auch wieder STS erzeugen. Diesen wechselseitigen Zusammenhang haben zuerst A. Sade [151, 152] und R. H. Bruck [32, 33] entdeckt.[1])

5.3.10. Satz. *Es gibt eine bijektive Abbildung zwischen der Menge der idempotenten TS-Quasigruppen der Ordnung v und der Menge der STS der Ordnung v.*[2])

Beweis. 1. Es sei $(Q, \cdot)$ eine endliche idempotente TS-Quasigruppe (der Ordnung v). Da aufgrund der Totalsymmetrie die Gleichung $x \cdot y = z$ mit den Gleichungen $y \cdot x = z$, $y \cdot z = x$, $z \cdot y = x$, $x \cdot z = y$ und $z \cdot x = y$ äquivalent ist (vgl. 5.2.), definiert die Operation $\cdot$ über der Menge Q folgendermaßen ein Steiner-Tripel-System S: drei beliebige Elemente $x, y, z \in Q$ sind genau dann eine Dreiermenge $\{x, y, z\}$ aus S, wenn $x \cdot y = z$ und $x \neq y$ gilt. Die Idempotenz sichert dann nämlich, daß auch $x \neq z$ und $y \neq z$ gilt; die Totalsymmetrie, aus der ja auch (Q) folgt, garantiert, daß zwei beliebige (aber verschiedene) Elemente x und y mit genau einem dritten Element z vermöge $x \cdot y = z$ verbunden sind.

2. Es sei umgekehrt S ein STS über der Menge Q (mit $|Q| = v$). Dann läßt sich in Q wie folgt eine idempotente TS-Quasigruppe $(Q, \cdot)$ definieren:

a) Wenn x und y beliebige Elemente aus Q mit $x \neq y$ sind, ist $x \cdot y := z$ das dritte Element der Dreiermenge $\{x, y, z\}$ aus S;

b) für alle $x \in Q$ setzen wir $x \cdot x := x$.

Offensichtlich ist die so definierte Operation idempotent und kommutativ. Da ferner $z = x \cdot y$ für alle $x, y \in Q$ mit $x \neq y$ verschieden von x und y ist, gilt auch $x \cdot z = y$, also $x \cdot xy = y$. Für $x = y$ gilt $x \cdot xx = x$. Damit ist auch die Links-Systemregel erfüllt. Die Quasigruppeneigenschaft folgt aus Korollar 5.2.4. Folglich ist $(Q, \cdot)$ eine idempotente TS-Quasigruppe (der Ordnung v).

5.3.11. Korollar. *Für die Ordnung v ($v \geqq 3$) einer endlichen idempotenten TS-Quasigruppe $(Q, \cdot)$ gilt $v \equiv 1 \pmod 6$ oder $v \equiv 3 \pmod 6$.*

[1]) Für eine Verallgemeinerung dieses Resultats sei auf N. K. Pucharev [141] verwiesen.

[2]) Wegen der Forderung $v \geqq 3$ in Definition 5.3.9 klammern wir die Fälle $v = 1$ und $v = 2$ aus. Die triviale Quasigruppe der Ordnung 1 ist idempotent und totalsymmetrisch, für $v = 2$ gibt es keine idempotente TS-Quasigruppe.

5.3.12. Satz. *Es gibt eine bijektive Abbildung zwischen der Menge der TS-Loops der Ordnung $v+1$ und der Menge der STS der Ordnung v.*[1])

Beweis. 1. Es sei $(Q', \cdot)$ eine (endliche) TS-Loop[2]) der Ordnung $v+1$, d. h., es sei $Q' = \{n, q_1, q_2, \ldots, q_v\}$, wobei n das neutrale Element der Loop ist. Wegen der Sätze 5.2.2 und 5.2.3 ist $(Q', \cdot)$ auch unipotent, so daß $x \cdot x = n$ für alle $x \in Q'$ gilt. Da ferner für alle $x \in Q'$ auch $x \cdot n = n \cdot x = x$ gilt, kann das neutrale Element n in einer Gleichung wie $x \cdot y = z$ nur dann auftreten, wenn nicht alle drei Elemente paarweise verschieden sind. Deshalb definiert die Operation $\cdot$ aufgrund der Totalsymmetrie über der Menge $Q := Q' \setminus \{n\} = \{q_1, q_2, \ldots, q_v\}$ ein Steiner-Tripel-System S (vgl. Satz 5.3.10). Wegen des Ausschlusses von n sind nämlich jetzt in einer Gleichung $x \cdot y = z$ mit $x, y, z \in Q$ alle drei Elemente paarweise verschieden, so daß $\{x, y, z\}$ eine Dreiermenge ist. Die Totalsymmetrie sichert darüber hinaus, daß zwei verschiedene Elemente aus Q in genau einer Dreiermenge enthalten sind.

2. Umgekehrt sei S ein STS über der Menge $Q = \{q_1, q_2, \ldots, q_v\}$. Erweitern wir diese Menge Q durch Hinzunahme eines Elementes n zu der Menge $Q' := Q \cup \{n\}$, so läßt sich in Q' folgendermaßen eine TS-Loop $(Q', \cdot)$ definieren:

a) Wenn x und y beliebige Elemente aus Q mit $x \neq y$ sind, ist $x \cdot y := z$ das dritte Element der Dreiermenge $\{x, y, z\}$ aus S;

b) für alle $x \in Q$ setzen wir $x \cdot n = n \cdot x := x$ und $x \cdot x = n \cdot n := n$. Die so definierte Operation ist kommutativ. Wegen a) gilt $x \cdot xy = y$ für alle $x, y \in Q$ mit $x \neq y$, wegen b) auch für alle $x, y \in Q'$. Damit ist die L-Systemregel erfüllt. Folglich ist $(Q', \cdot)$ totalsymmetrisch und wegen Korollar 5.2.4 eine Quasigruppe. Da das adjungierte Element n per definitionem das neutrale Element ist, ist $(Q', \cdot)$ eine (unipotente) TS-Loop der Ordnung $v+1$.

5.3.13. Korollar. *Für die Ordnung $m := v+1$ (mit $v \geqq 3$) einer endlichen TS-Loop $(Q', \cdot)$ gilt $m \equiv 2 \pmod 6$ oder $m \equiv 4 \pmod 6$.*

Die Einschränkung $v \geqq 3$ können wir jetzt sogar auf $v \geqq 1$ reduzieren, da es auch für $m = 2$ eine TS-Loop gibt.

Während eine idempotente TS-Quasigruppe (mit $v > 1$) niemals eine Gruppe sein kann (vgl. die Fußnote 2), ist das im Fall (unipotenter) TS-Loops von ihrer Ordnung $m\,(= v+1)$ abhängig. So sind z. B. TS-Loops für $m = 2, 4, 8$ und 16 assoziativ und wegen der Sätze 5.2.3 und 5.2.8 folglich auch L- und R-transitiv. Damit ist auch die Bisymmetrie erfüllt. Für $m = 10$ und $m = 14$ ist dagegen weder eine parastrophe Identität der Assoziativität noch die Bisymmetrie erfüllt. Ist eine TS-Loop bisymmetrisch, so ist sie wegen Korollar 5.1.4 eine abelsche Gruppe. Der Assoziativität einer TS-Loop $(Q', \cdot)$ entspricht dabei in dem durch $(Q', \cdot)$ induzierten Steiner-Tripel-System S die sogenannte *Veblen-Young-Konfiguration* (vgl. H. Zeitler [187]) (siehe

[1]) Die (triviale) TS-Loop der Ordnung 1, aber auch die TS-Loop der Ordnung 2 klammern wir wegen der Voraussetzung $v \geq 3$ aus.

[2]) V. D. Belousov [21] nennt eine idempotente TS-Quasigruppe eine *Steiner-Quasigruppe* und eine TS-Loop eine *Steiner-Loop*. Da jedoch Idempotenz und Existenz eines neutralen Elementes in einer Quasigruppe miteinander unvereinbar sind, ist eine Steiner-Loop keine Steiner-Quasigruppe. Wir verzichten deshalb hier auf diese Bezeichnungen.

Abb. 36). Diese Konfiguration ist jedoch nur sinnvoll, wenn Q' mehr als drei Elemente (Punkte) enthält. Im STS der Ordnung 7 (Beispiel 5.3.8.2) ist diese Konfiguration für beliebige Punkte A, B und C erfüllt, im STS der Ordnung 9 (Beispiel 5.3.8.3) dagegen nicht.

J. W. DI PAOLA [129] hat gezeigt, daß eine TS-Loop genau dann assoziativ ist, wenn im zugehörigen STS beliebige drei nicht auf einer Geraden liegende Punkte ein STS der Ordnung 7 erzeugen.[1])

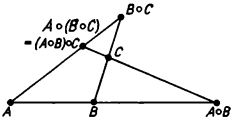


Abb. 36

Ist eine idempotente TS-Quasigruppe bisymmetrisch, so ist sie (vgl. 5.1.) auch selbstdistributiv. Zum Beispiel sind die idempotenten TS-Quasigruppen der Ordnung $v = 3$ und $v = 9$ bisymmetrisch. Für $v = 7$ ist die Bisymmetrie verletzt. Der Selbstdistributivität einer (idempotenten) TS-Quasigruppe $(Q, \cdot)$ entspricht im zugehörigen STS eine Konfiguration, die H. ZEITLER *Parallelitätsgesetz* nennt (vgl. Abb. 29). Wählt man z. B. im STS der Ordnung 7 $P = 1$, $Q = 3$ und $R = 7$, so schließt sich diese Figur nicht, wohl aber für jedes P, Q, R im STS der Ordnung 9.

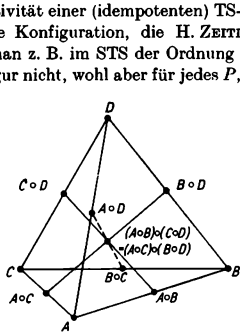


Abb. 37

Neben der Selbstdistributivität spiegelt sich auch die Bisymmetrie einer idempotenten TS-Quasigruppe $(Q, \cdot)$ im mit ihr verbundenen STS auf eine interessante Weise wider. Für ausgewählte Lagen der (sonst ja beliebigen) Punkte A, B, C und D aus Q entsteht nämlich der räumliche Eindruck eines Tetraeders. H. ZEITLER nennt diese Konfiguration deshalb auch *Tetraedergesetz*[2]) (siehe Abb. 37). Im STS der Ordnung 9 muß man zwar auf diesen räumlichen Eindruck verzichten, die „Tetraeder"-

[1]) In diesem Zusammenhang läßt sich zeigen, daß die Ordnung m einer unipotenten Gruppe eine Potenz von 2 sein muß (vgl. J. W. DI PAOLA [129]).

[2]) Dem liegt der Satz zugrunde, daß die Verbindungsgeraden der Mitten gegenüberliegender Kanten eines Tetraeders sich in einem Punkt schneiden (vgl. auch Beispiel 5.1.10.2).

Figur schließt sich aber natürlich dennoch für alle Punkte A, B, C und D (vgl. auch Abb. 28).

Die Bisymmetrie spielt, das sei hier nur erwähnt, bei der Erzeugung affiner Geometrien eine besondere Rolle (vgl. z. B. N. K. PUCHAREV [142], A. MITSCHKE und H. WERNER [119]).

Es gibt, wie wir gesehen haben, nichtbisymmetrische und bisymmetrische TS-Loops. Letztere sind (unipotente) abelsche Gruppen. Ebenso existieren unter den idempotenten TS-Quasigruppen neben bisymmetrischen auch nichtbisymmetrische. Wenn auch die Idempotenz hier die Gruppeneigenschaft generell ausschließt, erlaubt Satz 5.2.16 doch wenigstens, einen Zusammenhang zwischen bisymmetrischen idempotenten TS-Quasigruppen und abelschen Gruppen herzustellen. Vermöge $x \circ y := a \cdot xy$ definiert eine bisymmetrische (idempotente) TS-Quasigruppe $(Q, \cdot)$ eine abelsche Gruppe $(Q, \circ)$.

Aus den Sätzen 5.3.10 und 5.3.12 folgt unmittelbar, wie man direkt, d. h. ohne STS, aus einer idempotenten TS-Quasigruppe der Ordnung v eine TS-Loop der Ordnung $v + 1$ erhält (und umgekehrt). Von dieser Tatsache hat R. H. BRUCK [29] bereits Gebrauch gemacht, bevor STS und Quasigruppen miteinander in Verbindung gebracht worden sind. Man kann z. B. die beiden folgenden TS-Quasigruppen der Ordnung 7 bzw. 8 anhand des STS aus Beispiel 5.3.8.2 gewinnen, man erkennt aber auch sofort, wie sie wechselseitig auseinander hervorgehen:

·	1	2	3	4	5	6	7
1	[1]	3	2	5	4	7	6
2	3	[2]	1	6	7	4	5
3	2	1	[3]	7	6	5	4
4	5	6	7	[4]	1	2	3
5	4	7	6	1	[5]	3	2
6	7	4	5	2	3	[6]	1
7	6	5	4	3	2	1	[7]

idempotente TS-Quasigruppe

·	n	1	2	3	4	5	6	7
n	n	1	2	3	4	5	6	7
1	1	[n]	3	2	5	4	7	6
2	2	3	[n]	1	6	7	4	5
3	3	2	1	[n]	7	6	5	4
4	4	5	6	7	[n]	1	2	3
5	5	4	7	6	1	[n]	3	2
6	6	7	4	5	2	3	[n]	1
7	7	6	5	4	3	2	1	[n]

(unipotente) TS-Loop

Aus der idempotenten TS-Quasigruppe $(G_{1,3}, \circ)$ aus Aufgabe 1.3.17.1 geht auf diese Weise die Kleinsche Vierergruppe $\mathfrak{V}_4$ hervor (vgl. S. 75). Das zugehörige STS ist in Beispiel 5.3.8.1 angegeben.

Das STS $\varepsilon_4{}^3$ der Ordnung 9 (Beispiel 5.3.8.3) kann man auch als ein 4-Gewebe[1]) der Ordnung 3 auffassen. Entsprechend läßt sich die affine Ebene $\varepsilon_3{}^2$ als ein 3-Gewebe der Ordnung 2 interpretieren. Dabei heißt eine Inzidenzstruktur $(\mathfrak{P}, \mathfrak{G}, I)$, in der $\mathfrak{G}$ in r disjunkte (Parallel-) Klassen $\mathfrak{G}_1, \mathfrak{G}_2, \ldots, \mathfrak{G}_r$ zerlegt ist, ein *r-Gewebe*, wenn die

[1]) In deutschsprachigen Arbeiten wird der Begriff „Gewebe" dem Begriff „Netz" (russisch: set', englisch: net) vorgezogen — wahrscheinlich aufgrund der wichtigen Monographie von W. BLASCHKE und G. BOL [24].

folgenden beiden Bedingungen erfüllt sind (vgl. etwa P. DEMBOWSKI [43], V. D. BELOUSOV [22]):

a) Zwei Geraden aus verschiedenen Klassen inzidieren nur mit einem Punkt aus $\mathfrak{P}$;

b) jeder Punkt aus $\mathfrak{P}$ inzidiert mit genau einer Geraden aus jeder Klasse.

Zwischen den Geraden zweier verschiedener Klassen wie auch zwischen den Geraden einer Klasse und den Punkten einer beliebigen Geraden existiert eine bijektive Abbildung.

Die Anzahl der durch jeden Punkt gehenden Geraden heißt das *Geschlecht* r des Gewebes, die Anzahl der Geraden einer Klasse (Geradenschar) heißt die *Ordnung* m des r-Gewebes. (Es gilt $m = k$.) In einem endlichen r-Gewebe der Ordnung m $(= k)$ sind die Bedingungen (5.3.2) und (5.3.3) erfüllt. Ferner gelten die Beziehungen $v = k^2$, $b = r \cdot k$ und $r \leqq k + 1$. Ein r-Gewebe ist genau dann eine affine Ebene, wenn zwei beliebige Punkte kollinear sind, d. h. (5.3.1) erfüllt ist. $r = k + 1$ ist ebenfalls ein notwendiges und hinreichendes Kriterium dafür, daß ein r-Gewebe eine affine Ebene ist (vgl. P. DEMBOWSKI [43], V. D. BELOUSOV [22]).

Für die Theorie der Quasigruppen sind besonders die 3-Gewebe von Interesse (vgl. etwa J. ACZÉL, G. PICKERT und F. RADÓ [11], J. ACZÉL [4], V. D. BELOUSOV [22], J. DÉNES und A. D. KEEDWELL [44], G. PICKERT [136]). Aber auch für beliebige r $(\geqq 3)$ besteht ein Zusammenhang zwischen r-Geweben und Quasigruppen. So entspricht z. B. jedem r-Gewebe der Ordnung m ein System von $r - 2$ paarweise orthogonalen Quasigruppen der Ordnung m (und umgekehrt) (vgl. R. H. BRUCK [31]).

Schließlich seien noch die Funktionalgleichungen als ein wichtiges Anwendungsgebiet der Theorie der Quasigruppen genannt (vgl. J. ACZÉL [3] sowie die folgenden Kapitel).

5.3.14. Aufgaben

1. Man gebe eine bisymmetrische TS-Quasigruppe an, die kein STS induziert. (Hinweis: Beispiel 5.1.10.1 g), M endlich (!).)

2. Man zeige, daß die aus dem STS der Ordnung
a) $v = 9$ gewonnene idempotente TS-Quasigruppe bisymmetrisch ist und konstruiere hieraus mit Hilfe von Satz 5.2.16 eine abelsche Gruppe;
b) $v = 7$ gewonnene idempotente TS-Quasigruppe nicht bisymmetrisch ist.

3. Man zeige: Sind zwei STS isomorph, so sind auch die durch sie induzierten idempotenten TS-Quasigruppen (bzw. TS-Loops) isomorph (und umgekehrt).

4. Man zeige, daß in der aus dem STS der Ordnung 9 gewonnenen idempotenten TS-Quasigruppe jede Links- und Rechts-Translation ein Automorphismus ist.

Funktionalgleichungen

6. Über gegenseitige Ableitbarkeit bekannter Eigenschaften einiger elementarer Funktionen

Während bisher Eigenschaften zweistelliger Operationen und ihre Beziehungen zueinander im Mittelpunkt der Betrachtungen standen, wenden wir uns nun der Untersuchung von Eigenschaften vorrangig einstelliger (reeller) Funktionen zu. Dabei werden die auch im Mathematikunterricht behandelten elementaren Funktionen eine besondere Rolle spielen.

Unter den zahlreichen Eigenschaften der elementaren Funktionen befinden sich stets solche, die der betreffenden Funktion (bzw. Funktionenklasse) allein zukommen. Diese Eigenschaften kann man deshalb als typisch bezeichnen. Jede elementare Funktion hat daneben aber auch Eigenschaften, die sie mit anderen teilt, die für die betreffende Funktion in diesem Sinne also nicht typisch sind. Will man nachweisen, daß gewisse Eigenschaften für eine Funktion f nicht typisch sind, so hat man eine Funktion g mit $g \neq f$ anzugeben, die diese Eigenschaften ebenfalls besitzt. Derartige Untersuchungen könnten für eine richtige Akzentuierung bei der Behandlung der elementaren Funktionen nützlich sein. Entsteht doch beim Lernenden leicht der Eindruck, daß die meisten der im Unterricht und auch in der Anfängervorlesung behandelten Eigenschaften der elementaren Funktionen mit der jeweiligen Funktion untrennbar verbunden sind, für diese Funktionen also typisch sind, und zwar einfach deshalb, weil diese Eigenschaften als besonders wichtig hervorgehoben werden und kaum Vergleiche mit anderen Funktionen angestellt werden.

6.1. Aus den Theorien einiger elementarer Funktionen

Im Mathematik-Lehrbuch der Klasse 12 ([114], S. 94ff.) werden zur Wiederholung folgende Eigenschaften der Sinusfunktion zusammengestellt:

a) Definitionsbereich: $-\infty < x < +\infty$.

b) Wertevorrat: $-1 \leqq \sin x \leqq 1$.

c) Die Funktion ist periodisch mit der kleinsten Periode 2π.

d) Monotonieintervalle sind: $\left[\!\left[\frac{(4k-1)\,\pi}{2}, \frac{(4k+1)\,\pi}{2}\right]\!\right]$ (monoton wachsend) und $\left[\!\left[\frac{(4k+1)\,\pi}{2}, \frac{(4k+3)\,\pi}{2}\right]\!\right]$ (monoton fallend) für alle $k \in \mathbf{Z}$.

e) Die Funktion ist ungerade.

f) Die Funktion hat die Nullstellen $k \cdot \pi$ mit $k \in \mathbf{Z}$.

g) Die Funktion hat keine Pole.

Keine der hier genannten Eigenschaften ist für sich typisch für die Sinusfunktion, in dem Sinne, daß sie nur der Sinusfunktion zukommt. Auch in ihrer Gesamtheit sind die Eigenschaften a) bis g) nicht typisch, wie das folgende Beispiel zeigt. Wir definieren eine Funktion $f: \mathbf{R} \to \mathbf{R}$ durch:

$$f(x) = f(\bar{x} + 2k\pi) = g(\bar{x}) \quad \text{mit} \quad x \in \mathbf{R}, \quad \bar{x} \in [\![-\pi, \pi]\!], \quad k \in \mathbf{Z}$$

und

$$g(\bar{x}) := \begin{cases} -\frac{2}{\pi}\bar{x} - 2 & \text{für} \quad -\pi \leqq \bar{x} < -\frac{\pi}{2} \\ \frac{2}{\pi}\bar{x} & \text{für} \quad -\frac{\pi}{2} \leqq \bar{x} \leqq \frac{\pi}{2} \\ -\frac{2}{\pi}\bar{x} + 2 & \text{für} \quad \frac{\pi}{2} < \bar{x} \leqq \pi \end{cases} \qquad \text{(siehe Abb. 38).}$$

Selbst wenn man den sogenannten *trigonometrischen Pythagoras*

6.1.1. $\bigwedge_{x \in \mathbf{R}} s(x)^2 + c(x)^2 = 1$[1])

als achte Eigenschaft hinzunimmt, eine Eigenschaft also, die man nur in Verbindung mit den trigonometrischen Funktionen zu sehen gewöhnt ist, kann man doch ein

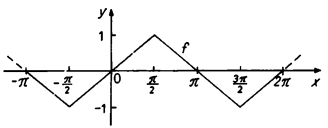


Abb. 38

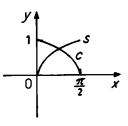


Abb. 39

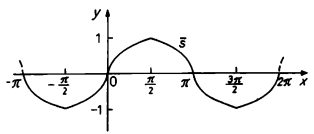


Abb. 40

von (sin, cos) verschiedenes Funktionenpaar (s, c) derart angeben, daß die Funktion s die Eigenschaften a) bis g) besitzt und zusammen mit c auch die Eigenschaft (6.1.1). Wir definieren zunächst:

$$s(x) := \sqrt{\frac{2}{\pi}x} \quad \text{und} \quad c(x) := \sqrt{1 - \frac{2}{\pi}x}$$

für $x \in \left[\!\left[0, \frac{\pi}{2}\right]\!\right]$ (siehe Abb. 39).

[1]) Zur Einsparung von Klammern wurde $(s(x))^2 = s(x)^2$ gesetzt. Die Schreibweise $s^2(x)$ wurde vermieden, um eine Verwechslung mit $s(s(x))$ auszuschließen.

In dem Intervall $\left[\!\left[0, \frac{\pi}{2} \right]\!\right]$ gilt (6.1.1), und s ist von der Sinusfunktion verschieden, denn es gilt z. B. $\sin \frac{\pi}{3} \neq s\left(\frac{\pi}{3}\right)$. Die Funktionen s und c können nun so auf die ganze Achse fortgesetzt werden, daß (6.1.1) für alle $x \in \mathbf{R}$ gilt und außerdem die Eigenschaften a) bis g) erfüllt sind. Die Funktionen $s\colon \mathbf{R} \to \mathbf{R}$ und $c\colon \mathbf{R} \to \mathbf{R}$ mit $s(x) = s(\bar{x} + 2k\pi) = \bar{s}(\bar{x})$ und $c(x) = c(\bar{x} + 2k\pi) = \bar{c}(\bar{x})$ für $x \in \mathbf{R}$, $\bar{x} \in [\![-\pi, \pi]\!]$, $k \in \mathbf{Z}$ und

$$\bar{s}(\bar{x}) := \begin{cases} \sqrt{\frac{2}{\pi}\,\bar{x}} & \text{für } \bar{x} \in \left[\!\left[0, \frac{\pi}{2} \right]\!\right], \\ \sqrt{\frac{2}{\pi}\,(\pi - \bar{x})} & \text{für } \bar{x} \in \left]\!\left] \frac{\pi}{2}, \pi \right]\!\right], \\ -\sqrt{-\frac{2}{\pi}\,\bar{x}} & \text{für } \bar{x} \in \left[\!\left[-\frac{\pi}{2}, 0 \right]\!\right[, \\ -\sqrt{\frac{2}{\pi}\,(\pi + \bar{x})} & \text{für } \bar{x} \in \left[\!\left[-\pi, -\frac{\pi}{2} \right]\!\right[, \end{cases}$$

$$\bar{c}(\bar{x}) := \begin{cases} \sqrt{1 - \frac{2}{\pi}\,\bar{x}} & \text{für } \bar{x} \in \left[\!\left[0, \frac{\pi}{2} \right]\!\right], \\ -\sqrt{1 - \frac{2}{\pi}\,(\pi - \bar{x})} & \text{für } \bar{x} \in \left]\!\left] \frac{\pi}{2}, \pi \right]\!\right], \\ \sqrt{1 + \frac{2}{\pi}\,\bar{x}} & \text{für } \bar{x} \in \left[\!\left[-\frac{\pi}{2}, 0 \right]\!\right[, \\ -\sqrt{1 - \frac{2}{\pi}\,(\pi + \bar{x})} & \text{für } \bar{x} \in \left[\!\left[-\pi, -\frac{\pi}{2} \right]\!\right[\!\left[\right. \end{cases}$$

(siehe Abb. 40) leisten das Verlangte.

Wir stellen fest, daß die Funktion s wie die Sinusfunktion stetig ist. Man kann auch differenzierbare Funktionen konstruieren, die alle genannten Eigenschaften der Sinusfunktion aufweisen und doch von ihr verschieden sind. Welche Eigenschaft kann man denn nun hinzunehmen, um nur die Sinusfunktion zu erhalten? Diese Frage werden wir im Kapitel 12 aufgreifen.

Es sei noch einmal betont: Wenn man z. B. feststellt, daß die Periodizität der Sinusfunktion für diese Funktion nicht typisch ist, dann ist dies in dem Sinne gemeint, daß es noch andere periodische Funktionen gibt. Das beeinträchtigt in keiner Weise die große Bedeutung der Sinusfunktion für die Beschreibung bestimmter periodischer Sachverhalte in der Physik.

Ein weiteres Beispiel für die hier behandelten Fragen findet man im Mathematik-Lehrbuch der Klasse 9 ([115], S. 88). Es wird dort die Erkenntnis formuliert: „Die Graphen der Funktionen $y = x^{2n}$ mit $n \in \{1; 2; 3; \ldots\}$ bilden eine Schar von Kurven, die alle die Punkte $P_1(1; 1)$, $P_2(-1; 1)$ und $P_3(0; 0)$ gemeinsam haben und die sym-

metrisch zur y-Achse sind." Auch hier sind die genannten Eigenschaften der geraden Potenzfunktionen nicht typisch für diese Funktionenklasse, denn die Funktionen $f, g\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = |x|$ und $g(x) = \sqrt[3]{|x|}$ gehören nicht zu dieser Schar, haben aber die genannten Eigenschaften.

Solche Betrachtungen, in denen nichttypische Eigenschaften der elementaren Funktionen hervorgehoben werden, können gleichzeitig als Motivation für die funktionale Charakterisierung dieser Funktionen dienen. Wenn man sich nämlich bei mehreren Eigenschaften davon überzeugt hat, daß sie für die betreffende Funktion nicht typisch sind (insbesondere bei solchen Eigenschaften, von denen man dies zunächst nicht erwartet hat), dann wird die Frage nahegelegt, ob es nun unter den Eigenschaften auch solche gibt, die die betreffende Funktion allein besitzt, die sie von allen anderen Funktionen unterscheidet. Damit wird man auf axiomatische Untersuchungen geführt. Die Menge der Eigenschaften einer (elementaren) Funktion kann als mathematische Theorie aufgefaßt werden, die man aus einer Definition durch Gleichungen, durch Potenzreihen oder auch durch geometrische Überlegungen gewonnen hat. Aus der damit gegebenen Menge von Aussagen eine Teilmenge so auszusondern, daß alle anderen Aussagen aus ihr ableitbar sind, heißt diese Theorie zu axiomatisieren. Man sagt, daß die betreffende Funktion durch diese Teilmenge von Aussagen *funktional charakterisiert* ist. Dies ist bei den elementaren Funktionen möglich, und zwar jeweils auf recht verschiedene Weise. Wir werden uns in den Kapiteln 7, 8 und 12 mit diesem Problem beschäftigen. Zuvor sollen noch einige Beispiele für Ableitbarkeitsverhältnisse innerhalb der Theorie einiger elementarer Funktionen betrachtet werden.

Es sind die Ableitbarkeitsverhältnisse zu untersuchen, in denen die *Subtraktionstheoreme der Sinus- und Kosinusfunktionen* zum „trigonometrischen Pythagoras" (6.1.1) stehen. Die Subtraktionstheoreme der Funktionen sin und cos lauten

$$\sin(x - y) = \sin x \cdot \cos y - \cos x \cdot \sin y$$

und

$$\cos(x - y) = \cos x \cdot \cos y + \sin x \cdot \sin y.$$

Wir formulieren diese Eigenschaften der Winkelfunktionen unter Verwendung von Funktionsvariablen

6.1.2. $\bigwedge_{x,y \in \mathbf{R}} s(x - y) = s(x)\, c(y) - c(x)\, s(y)$,

6.1.3. $\bigwedge_{x,y \in \mathbf{R}} c(x - y) = c(x)\, c(y) + s(x)\, s(y)$.

Dadurch gelten Schlußfolgerungen, die wir aus diesen Gleichungen ziehen, nicht nur für das Funktionenpaar (sin, cos), sondern für alle Paare (s, c) von Funktionen $s\colon \mathbf{R} \to \mathbf{R}$ und $c\colon \mathbf{R} \to \mathbf{R}$, die diese Eigenschaften besitzen. Im folgenden wird das Paar (s, c) mit $s(x) = 0$ und $c(x) = 0$ für alle $x \in \mathbf{R}$ ausgeschlossen.

Es gilt: Aus (6.1.2) und (6.1.3) folgt (6.1.1). Für jedes Paar (s, c) mit den Eigenschaften (6.1.2) und (6.1.3) folgt zunächst für $x = y = 0$ aus (6.1.2) $s(0) = s(0)\, c(0) - c(0)\, s(0) = 0$, also $s(0) = 0$, und für $y = 0$ folgt damit aus (6.1.2) $s(x) = s(x)\, c(0) - c(x)\, s(0)$, also $s(x)\,(1 - c(0)) = 0$.

Fall 1: Aus $c(0) \neq 1$ folgt $s(x) = 0$ für alle $x \in \mathbf{R}$ und damit aus (6.1.3) für $y = 0$ die Identität $c(x) = c(x)\,c(0)$, also auch $c(x) = 0$ für alle $x \in \mathbf{R}$. Diese Funktionen wurden aber ausgeschlossen, also bleibt nur der

Fall 2: $c(0) = 1$.

Setzt man jetzt in Gleichung (6.1.3) $x = y$, so erhält man $c(0) = c(x)^2 + s(x)^2 = 1$ für alle reellen Zahlen x, also (6.1.1).

Die Umkehrung der eben bewiesenen Behauptung gilt dagegen nicht. Aus (6.1.1) folgt nämlich weder (6.1.2) noch (6.1.3). Diese Behauptung beweisen wir mit Hilfe des weiter oben definierten Funktionenpaares (s, c), das zwar die Eigenschaft (6.1.1), aber nicht die Eigenschaften (6.1.2) und (6.1.3) besitzt. So gilt z. B. für $x = \frac{\pi}{3}$ und $y = \frac{\pi}{6}$ die Beziehung $s\left(\frac{\pi}{3} - \frac{\pi}{6}\right) \neq s\left(\frac{\pi}{3}\right) c\left(\frac{\pi}{6}\right) - c\left(\frac{\pi}{3}\right) s\left(\frac{\pi}{6}\right)$.

Wir wollen die Ableitbarkeitsverhältnisse zwischen den *Logarithmengesetzen* $\log_a (x \cdot y) = \log_a x + \log_a y$, $\log_a \frac{x}{y} = \log_a x - \log_a y$ und $\log_a \frac{1}{x} = -\log_a x$ für $a \in \mathbf{R}_+^*$, $a \neq 1$; $x \in \mathbf{R}_+^*$, $y \in \mathbf{R}_+^*$ untersuchen. Dazu schreiben wir diese Beziehungen wieder mit Funktionsvariablen und zwar für Funktionen $f\colon \mathbf{R}_+^* \to \mathbf{R}$

6.1.4. $\bigwedge_{x,y \in \mathbf{R}_+^*} f(x \cdot y) = f(x) + f(y)$,

6.1.5. $\bigwedge_{x,y \in \mathbf{R}_+^*} f\left(\frac{x}{y}\right) = f(x) - f(y)$,

6.1.6. $\bigwedge_{x \in \mathbf{R}_+^*} f\left(\frac{1}{x}\right) = -f(x)$.

Wir zeigen zunächst, daß (6.1.4) und (6.1.5) äquivalent sind, also, wie man auch sagen könnte, „gleichstarke" Eigenschaften der Logarithmusfunktionen sind.

Aus (6.1.4) ergibt sich (6.1.5): Aus (6.1.4) folgt für $x = y = 1$ sofort $f(1) = 0$ und damit für $y = \frac{1}{x}$: $f(1) = f(x) + f\left(\frac{1}{x}\right) = 0$, also (6.1.6). Damit erhält man (6.1.5) folgendermaßen: $f\left(x \cdot \frac{1}{y}\right) = f(x) + f\left(\frac{1}{y}\right)$, das heißt $f\left(\frac{x}{y}\right) = f(x) - f(y)$.

Aus (6.1.5) folgt (6.1.4): Für $x = y = 1$ erhält man aus (6.1.5) sofort $f(1) = 0$, damit $f\left(\frac{1}{x}\right) = f(1) - f(x) = -f(x)$, also (6.1.6) und schließlich $f(x \cdot y) = f\left(\frac{x}{\frac{1}{y}}\right)$ $= f(x) - f\left(\frac{1}{y}\right) = f(x) + f(y)$, also (6.1.4).

Die Beziehung (6.1.6) folgt also sowohl aus (6.1.4) als auch aus (6.1.5). Umgekehrt folgt aber aus (6.1.6) weder (6.1.4) noch (6.1.6). Dazu betrachten wir die folgendermaßen definierte Funktion $f\colon \mathbf{R}_+^* \to \mathbf{R}$,

$$f(x) := \begin{cases} (x-1)^2 & \text{für} \quad x \geqq 1, \\ -\left(\frac{1}{x} - 1\right)^2 & \text{für} \quad 0 < x < 1 \text{ (siehe Abb. 41).} \end{cases}$$

Wie man leicht zeigt, gilt $\bigwedge_{x \in \mathbf{R}_+^*} f\left(\frac{1}{x}\right) = -f(x)$. Da aber beispielsweise $f(6) \neq f(2) + f(3)$ und $f\left(\frac{2}{3}\right) \neq f(2) - f(3)$ ist, hat f nicht die Eigenschaften (6.1.4) und (6.1.5). Die Eigenschaft (6.1.6) ist also eine „schwächere" Eigenschaft der Logarithmusfunktionen als die Eigenschaften (6.1.4) bzw. (6.1.5). Zur Übung kann der Leser weitere Funktionen bilden, die (6.1.6) erfüllen.

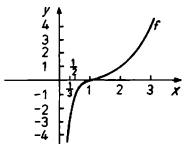


Abb. 41

Bei den Eigenschaften (6.1.1) bis (6.1.6) handelt es sich um Funktionalgleichungen. Die Variablen s, c und f sind dabei keine Zahlenvariablen, sondern Funktionsvariablen. Lösungen dieser Gleichungen sind also Funktionen über gewissen Definitionsbereichen.

Einige Funktionalgleichungen der elementaren Funktionen werden sich im Hinblick auf die funktionale Charakterisierung dieser Funktionen als besonders geeignet erweisen. Da bei unseren Betrachtungen das Anliegen der funktionalen Charakterisierung im Vordergrund steht und dafür nur ganz bestimmte Funktionalgleichungen benutzt werden, kann auf eine Definition des Begriffes *Funktionalgleichung* verzichtet werden.[1])

6.2. Aufgaben

1. Man zeige für die Eigenschaften
a) $f(x + y) = f(x) \cdot f(y)$, b) $f(x - y) = f(x) : f(y)$, c) $f(-x) = \frac{1}{f(x)}$ der Exponentialfunktionen $f: \mathbf{R} \to \mathbf{R}$:
— a) und b) sind äquivalent.
— c) folgt sowohl aus a) als auch aus b).
— Aus c) folgt weder a) noch b).
(Hinweis: Man untersuche die Konsequenz von $f(0) = 0$.)

2. Man untersuche für eine Funktion $f: \mathbf{R} \to \mathbf{R}$ die Beziehungen zwischen den Eigenschaften $f(x + y) = f(x) + f(y)$, $f(x - y) = f(x) - f(y)$ und $f(-x) = -f(x)$, die für alle reellen Zahlen x und y gelten sollen.

3. Man zeige: Eine Funktion $f: \mathbf{R} \to \mathbf{R}$ ist genau dann eine Lösung von $f(x + y) = f(x) + f(y)$, wenn sie Lösung von $f(x + y) - f(x - y) = 2f(y)$ ist.

4. Man zeige: Eine Funktion $f: \mathbf{R} \to \mathbf{R}$ ist genau dann eine Lösung von $f\left(\frac{x + y}{2}\right) = \frac{f(x) + f(y)}{2}$, wenn sie Lösung von $f(x + y) + f(x - y) = 2f(x)$ ist.

[1]) Auf die hier benutzten Funktionalgleichungen trifft die von J. Aczél in [3] gegebene Definition zu.

7. Funktionale Charakterisierungen der linearen Funktionen $f: \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x$ mit Hilfe der Funktionalgleichung $f(x + y) = f(x) + f(y)$

Bekanntlich versteht man unter Streckenmessung eine Funktion l, die jeder Strecke eine reelle Zahl zuordnet. An diese Zuordnung werden bestimmte Forderungen gestellt, z. B. die endliche *Additivität*: Liegt der Punkt B zwischen den Punkten A und C, so gilt $l(\overline{AC}) = l(\overline{AB}) + l(\overline{BC})$ (vgl. MfL Bd. 6, 1.1.4.). Entsprechend wird die Additivität bei allgemeineren Fassungen des Inhaltsproblems formuliert (vgl. MfL Bd. 5, 4.1.1.). Über der Menge der reellen Zahlen sind die schon im Mathematikunterricht recht ausführlich behandelten linearen Funktionen $f: \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x$ ($c \in \mathbf{R}$) additiv, denn es gilt wegen des Distributivgesetzes $c(x + y) = c \cdot x + c \cdot y$ für alle $x, y \in \mathbf{R}$, also $f(x + y) = f(x) + f(y)$. Man nennt diese Beziehung auch *Additionstheorem der linearen Funktionen* $f: \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x$ oder *Cauchysche Grundgleichung*. Man beachte, daß f für $c \neq 0$ damit ein Isomorphismus von der Gruppe $(\mathbf{R}, +)$ auf sich ist und das Additionstheorem die Operationstreue des Isomorphismus widerspiegelt (vgl. Kap. 3). Wir werden uns in diesem Kapitel mit der Frage nach allen derartigen Funktionen beschäftigen, die dieses Additionstheorem besitzen, und in diesem Zusammenhang bereits alle wesentlichen Probleme der funktionalen Charakterisierung der elementaren Funktionen diskutieren.

7.1. Die stetigen Lösungen der Funktionalgleichung $f(x + y) = f(x) + f(y)$

7.1.1. Satz. *Es gibt genau eine Funktion* $f: \mathbf{R} \to \mathbf{R}$ *mit folgenden Eigenschaften:*

7.1.2. $\bigwedge_{x,y \in \mathbf{R}} f(x + y) = f(x) + f(y)$,

7.1.3 *f ist stetig,*.

7.1.4. $f(1) = c$ ($c \in \mathbf{R}$).

Beweis (vgl. J. Aczél [3], A. L. Cauchy [34], E. Picard [133], T. L. Saaty [150], G. Wiswesser [182], D. Ilse [73]). Zunächst ist die Existenz der behaupteten Funktion gesichert, denn $f: \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x$ erfüllt die Bedingungen (7.1.2), (7.1.3) und (7.1.4). Für den Nachweis der Eindeutigkeit wird gezeigt: Für jede Funktion, die (7.1.2), (7.1.3) und (7.1.4) erfüllt, gilt:

a) $f(0) = 0$; denn setzt man $x = y = 0$ in (7.1.2), so erhält man $f(0) = 2f(0)$, also $f(0) = 0$.

b) Für jede reelle Zahl x und jede natürliche Zahl k gilt $f(k \cdot x) = k \cdot f(x)$. Der Beweis ergibt sich durch vollständige Induktion. Wie man ebenfalls leicht bestätigt, gilt auch $f(l \cdot x) = l \cdot f(x)$ für alle $l \in \mathbf{Z}$.

c) Für alle rationalen Zahlen r gilt $f(r) = c \cdot r$. Dazu untersuchen wir drei Fälle:

c_1) $r = 0$. Nach a) gilt $f(0) = 0$ und damit $f(0) = c \cdot 0$.

c_2) $r > 0$. Dann gibt es natürliche Zahlen p, q mit $r = \frac{p}{q}$. Wegen b) gilt $f(q \cdot r) = f\left(q \cdot \frac{p}{q}\right) = f(p \cdot 1) = c \cdot p$ und $f(q \cdot r) = q \cdot f(r)$, also $q \cdot f(r) = c \cdot p$ und damit $f(r) = c \cdot r$.

c_3) $r < 0$. Dann ist $-r > 0$, und man erhält aus (7.1.2) $f(0) = f(r) + f(-r)$ und wegen a) $f(r) + f(-r) = 0$, also $f(r) = -f(-r) = -c(-r) = c \cdot r$.

d) Für alle reellen Zahlen x gilt $f(x) = c \cdot x$. Es sei x_0 eine beliebige irrationale Zahl und (r_n) eine Folge rationaler Zahlen mit $\lim\limits_{n\to\infty} r_n = x_0$. Dann muß wegen der Stetigkeitsforderung (7.1.3) $f(x_0) := \lim\limits_{n\to\infty} f(r_n)$ gesetzt werden. Dieser Grenzwert existiert auch, denn es gilt $\lim\limits_{n\to\infty} f(r_n) = \lim\limits_{n\to\infty} c \cdot r_n = c \cdot x_0$, also $f(x) = c \cdot x$ für alle $x \in \mathbf{R}$. Damit ist Satz 7.1.1 bewiesen.

7.1.5. Bemerkung. Wir weisen darauf hin, daß die Gleichung (7.1.2) über der Menge der rationalen Zahlen nur die Lösungen $f\colon \mathbf{Q} \to \mathbf{R}$ mit $f(r) = c \cdot r$ besitzt und die Stetigkeit von f erst für die stetige Erweiterung der Lösung auf den Bereich der reellen Zahlen gebraucht wurde. Ebenfalls ohne Stetigkeitsforderung gilt bereits $f(r \cdot x) = r \cdot f(x)$ für alle rationalen Zahlen r und alle reellen Zahlen x. Dazu zeigt man zunächst, daß jede Lösung f von (7.1.2) ungerade ist. Wird in (7.1.2) y durch $-x$ ersetzt, so ergibt sich $f(x + (-x)) = f(x) + f(-x) = f(0) = 0$, also $f(x) = -f(-x)$. Es sei nun r eine beliebige positive rationale Zahl mit $r = \frac{p}{q}$ $(p, q \in \mathbf{N}^*)$. Setzt man dann $y := \frac{p}{q} x$, so gilt $f(p \cdot x) = f(q \cdot y)$, woraus mit b) $p \cdot f(x) = q \cdot f(y)$, also $r \cdot f(x) = f(y) = f(r \cdot x)$ folgt. Es sei nun $r < 0$, dann ist $-r > 0$, und es gilt $f(r \cdot x) = -f(-r \cdot x)$ wegen der Ungeradheit von f. Weiterhin ist $-f(-r \cdot x) = -(-r)\, f(x)$, da $-r > 0$ ist. Also ist für negative rationale Zahlen r auch $f(r \cdot x) = r \cdot f(x)$.

Satz 7.1.1 besagt auch, daß mit (7.1.2), (7.1.3) und (7.1.4) ein Axiomensystem (Ax) für die lineare Funktion f mit $f(x) = c \cdot x$ gegeben ist. Einerseits folgen alle Eigenschaften dieser Funktion aus (Ax), denn wir haben ja $f(x) = c \cdot x$ abgeleitet, und daraus folgen alle weiteren Eigenschaften (z. B. Monotonie, Differenzierbarkeit). Andererseits folgen aus (Ax) auch nur Eigenschaften dieser Funktion, denn sie ist ja einzige Lösung von (Ax). Somit sind die Eigenschaften (7.1.2) und (7.1.3) zusammen typisch für die Klasse der Funktionen $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x$. Da es eine Funktion f gibt, die alle drei Axiome von (Ax) erfüllt, ist das System (Ax) widerspruchsfrei. Wir wollen schließlich die Frage untersuchen, ob (Ax) unabhängig[1]) ist. Zunächst ist Axiom (7.1.4) unabhängig. Wählt man nämlich $f(1) = \bar{c} \neq c$, so ist f mit $f(x) = \bar{c} \cdot x$

[1]) Zu den Begriffen „widerspruchsfrei" und „unabhängig" siehe auch MfL Bd. 12, 4.3.

eine Funktion, die die Axiome (7.1.2) und (7.1.3) erfüllt, aber nicht das Axiom (7.1.4). Auch Axiom (7.1.2) ist unabhängig, da z. B. die Funktion $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x^3$ die Axiome (7.1.3) und (7.1.4) erfüllt, aber nicht (7.1.2). Der Beweis für die Unabhängigkeit von (7.1.3) wurde 1905 von G. HAMEL (1877–1954) [64] erbracht, indem er mit Hilfe einer Basis der reellen Zahlen unstetige Lösungen von (7.2.1) konstruierte. Bevor wir diesen Beweis führen, wollen wir uns noch ausführlicher mit der Stetigkeitsforderung (7.1.3) befassen.

7.2. Weitere funktionale Charakterisierungen der linearen Funktionen $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x$

Es hat sich gezeigt, daß auch andere (zum Teil schwächere) Bedingungen als Stetigkeit ausreichen, um die Stetigkeit von Lösungen von (7.1.2) zu sichern. Wir zeigen im folgenden die Äquivalenz einiger dieser Bedingungen. Weitere derartige Bedingungen können dem Literaturverzeichnis in J. ACZÉL [3] bzw. [5] oder auch W. SCHULZ [159] entnommen werden.

7.2.1. Satz. *Eine nichtkonstante Lösung $f\colon \mathbf{R} \to \mathbf{R}$ der Funktionalgleichung $f(x + y) = f(x) + f(y)$ mit $f(1) = c$ ist stetig genau dann, wenn gilt:*

7.2.2. *f ist in einem Punkt x_0 stetig* (vgl. M. G. DARBOUX [39], J. ACZÉL [3], T. L. SAATY [150], D. ILSE [73])

oder

7.2.3. *f ist streng monoton* (vgl. B. N. DELONE und D. A. RAJKOV [42], J. M. KOLJAGIN [90])

oder

7.2.4. *Es gibt ein Intervall $]0, \varepsilon[$ mit $\varepsilon > 0$, über dem entweder $f(x) > 0$ oder $f(x) < 0$ gilt für alle $x \in]0, \varepsilon[$* (vgl. etwa D. ILSE [73])

oder

7.2.5. *f ist auf einem endlichen (beliebig kleinen) Intervall beschränkt* (vgl. J. ACZÉL [3], E. PICARD [133], G. S. YOUNG [185], D. ILSE [73])

oder

7.2.6. *Es gibt einen Punkt der x,y-Ebene mit einer Umgebung, in der kein Punkt $(x; f(x))$ liegt* (vgl. E. HOPF [69], D. ILSE [73]).

Beweis. Die stetigen Lösungen f mit $f(x) = c \cdot x$ von (7.1.2) erfüllen die Bedingungen (7.2.2) bis (7.2.6). Es sind noch die Umkehrungen zu beweisen.

Zu (7.2.2): Es sei $\bar{x}$ eine reelle Zahl mit $\bar{x} \neq x_0$ und (x_n) eine Folge reeller Zahlen mit $\lim\limits_{n\to\infty} x_n = \bar{x}$. Dann gilt für jede in x_0 stetige Lösung von (7.1.2)

$$\begin{aligned}\lim_{n\to\infty} f(x_n) &= \lim_{n\to\infty} f\big((x_n - \bar{x} + x_0) + (\bar{x} - x_0)\big)\\ &= \lim_{n\to\infty} f(x_n - \bar{x} + x_0) + \lim_{n\to\infty} f(\bar{x} - x_0) = f(x_0) + f(\bar{x} - x_0) = f(\bar{x}).\end{aligned}$$

Das Additionstheorem bewirkt also, daß jede seiner Lösungen, die an einer Stelle stetig ist, gleich überall stetig und damit nach Satz 7.1.1 eine Funktion f mit $f(x) = c \cdot x$ ist.

Zu (7.2.3): Wir zeigen, daß aus dem streng monotonen Wachsen von f folgt, daß $f(x) = c \cdot x$ ist ($c > 0$). Gäbe es nämlich eine reelle Zahl x_0 mit $f(x_0) \neq c \cdot x_0$, etwa $f(x_0) > c \cdot x_0$, so könnte man eine rationale Zahl r finden mit $c \cdot x_0 < c \cdot r < f(x_0)$, also $c \cdot x_0 < f(r) < f(x_0)$. Wegen $c > 0$ ist $x_0 < r$ und $f(r) < f(x_0)$ im Widerspruch zum monotonen Wachsen von f. Analog beweist man den Fall, daß f streng monoton fällt und $c < 0$ ist. Weiterhin ist leicht zu sehen, daß es genügt hätte, Monotonie im weiteren Sinne zu fordern.

Zu (7.2.4): Wir betrachten den Fall, daß f in $]0, \varepsilon[$ positiv ist, und beweisen, daß dann f über $\mathbf{R}$ streng monoton wachsend ist. Nach (7.2.3) ist f dann stetig. Es seien x_1 und x_2 reelle Zahlen mit $x_1 < x_2$, und x sei diejenige positive Zahl, für die $x_1 + x_3 = x_2$ gilt.

Fall 1: Für $x_3 \in]0, \varepsilon[$ ist $f(x_2) = f(x_1 + x_3) = f(x_1) + f(x_3) > f(x_1)$.

Fall 2: Ist $x_3 \notin]0, \varepsilon[$, so gibt es eine natürliche Zahl k mit $\frac{x_3}{k} \in]0, \varepsilon[$. Damit erhält man $f(x_2) = f(x_1) + f(x_3) = f(x_1) + f\left(k \cdot \frac{x_3}{k}\right) = f(x_1) + k \cdot f\left(\frac{x_3}{k}\right) > f(x_1)$. Analog beweist man den Fall, in dem $f(x) < 0$ für alle $x \in]0, \varepsilon[$ gilt.

Zu (7.2.5): Nach Voraussetzung gibt es ein Intervall I und eine Zahl $M > 0$ mit $|f(x)| \leqq M$ für alle $x \in I$. Für die Hilfsfunktion g: $\mathbf{R} \to \mathbf{R}$ mit $g(x) := f(x) - c \cdot x$ gilt:

a) g erfüllt die Gleichung (7.1.2), denn es gilt $g(x + y) = f(x + y) - c(x + y) = \big(f(x) - c \cdot x\big) + \big(f(y) - c \cdot y\big) = g(x) + g(y)$.

b) g ist über I beschränkt, denn für alle $x \in I$ gilt $|g(x)| = |f(x) - c \cdot x| \leqq |f(x)| + |c| \cdot |x| \leqq M'$.

c) $g(r) = 0$ für alle $r \in \mathbf{Q}$, denn es ist $g(r) = f(r) - c \cdot r = c \cdot r - c \cdot r = 0$.

d) g ist über der Menge aller reeller Zahlen beschränkt; ist nämlich $x_0 \notin I$, so gibt es eine rationale Zahl r derart, daß $x_0 + r \in I$ gilt. Daraus folgt $|g(x_0)| = |g(x_0) + 0| = |g(x_0) + g(r)| = |g(x_0 + r)| \leqq M'$.

e) g ist die triviale Lösung von (7.1.2), d. h. $g(x) = 0$ für alle $x \in \mathbf{R}$. Gäbe es eine Zahl x_0, für die $g(x_0) = A \neq 0$ gilt, so könnte eine natürliche Zahl k gefunden werden mit $k \cdot |A| > M'$, und es würde $|g(k \cdot x_0)| = |k \cdot g(x_0)| > M'$ gelten im Widerspruch zu d).

Wegen e) ist $f(x) - c \cdot x = 0$, also $f(x) = c \cdot x$, und das ist die stetige Lösung von (7.1.2).

Zu (7.2.6): Wir beweisen diesen Teil des Satzes indirekt. Wenn f eine unstetige Lösung der Funktionalgleichung (7.1.2) ist, liegen in jeder Umgebung eines jeden Punktes der x,y-Ebene Punkte $\big(x; f(x)\big)$. (Dazu sagen wir auch, die Punkte $\big(x; f(x)\big)$ liegen in der Ebene dicht.) Im Beweis von Satz 7.1.1 haben wir gesehen, daß $f(r) = c \cdot r$ ($r \in \mathbf{Q}$) für jede Lösung von (7.1.2) gilt (also auch für die unstetigen Lösungen, falls es solche Lösungen überhaupt gibt). Wenn f also eine unstetige Lösung ist, muß es reelle Zahlen x_1 und x_2 derart geben, daß der Punkt $\big(x_1; f(x_1)\big)$ auf dem Gra-

phen einer Funktion f mit $f(x) = c \cdot x$ liegt und daß der Punkt $(x_2; f(x_2))$ nicht auf diesem Graphen liegt. Es sei nun $(x_0; y_0)$ irgendein Punkt der Ebene. Dann hat das Gleichungssystem

$$\begin{aligned} s \cdot x_1 + s' \cdot x_2 &= x_0, \\ s \cdot f(x_1) + s' \cdot f(x_2) &= y_0 \end{aligned}$$

genau eine Lösung (s, s'), da nach Voraussetzung für die Koeffizientendeterminante $\begin{vmatrix} x_1 & x_2 \\ f(x_1) & f(x_2) \end{vmatrix} \neq 0$ gilt, denn $f(x_1) = c \cdot x_1$ und $f(x_2) \neq c \cdot x_2$ ergibt $x_1 f(x_2) - x_2 f(x_1) \neq 0$. Sind r bzw. r' rationale Näherungswerte von s bzw. s', so ist $x^* := r \cdot x_1 + r' \cdot x_2 \approx x_0$ und $f(x^*) := f(r \cdot x_1 + r' \cdot x_2) = r \cdot f(x_1) + r' \cdot f(x_2) \approx y_0$ (vgl. das genannte Gleichungssystem). Der Abstand zwischen $f(x^*)$ und y_0 kann beliebig klein gemacht werden, falls nur r und r' genügend nahe bei s bzw. s' liegen, wie man der folgenden Abschätzung entnimmt:

$$\begin{aligned} |f(x^*) - y_0| &= |r \cdot f(x_1) + r' \cdot f(x_2) - s \cdot f(x_1) - s' \cdot f(x_2)| \\ &\leqq |r - s|\,|f(x_1)| + |r' - s'|\,|f(x_2)|. \end{aligned}$$

Damit ist (7.2.6) bewiesen.

Damit ist auch der Satz 7.2.1 vollständig bewiesen.

Wir wollen nun noch ein Beispiel zu (7.2.6) angeben. Es sei $f\colon \mathbf{R} \to \mathbf{R}$ eine an mindestens einer Stelle unstetige Lösung von $f(x + y) = f(x) + f(y)$ mit $f(1) = 1$. Dann gilt zunächst (siehe Bemerkung 7.1.5) $f(r) = r$ für alle $r \in \mathbf{Q}$. Es sei $x = \sqrt{2}$ und $f(x) = \sqrt{2}$ und $y = \sqrt{3}$ und $f(y) = \frac{4}{3}$. Als Punkt der Ebene wählen wir $P(7; 6)$ und betrachten das Gleichungssystem

$$\text{(I)} \qquad s \cdot \sqrt{2} + s' \cdot \sqrt{3} = 7,$$

$$\text{(II)} \qquad s \cdot \sqrt{2} + s' \cdot \frac{4}{3} = 6.$$

Die Koeffizientendeterminante ist wegen $\frac{4}{3} \cdot \sqrt{2} \neq \sqrt{3} \cdot \sqrt{2}$ von Null verschieden. (I) − (II) liefert $s' \left(\sqrt{3} - \frac{4}{3}\right) = 1$, also $s' \in [\![2{,}5079; 2{,}5081]\!]$. Aus (II) ergibt sich, daß $s \in [\![1{,}8779; 1{,}8782]\!]$ ist. Die Zahl $r' = 2{,}5$ ist ein rationaler Näherungswert für s', und $r = 1{,}9$ ist ein rationaler Näherungswert für s. Für x^* ergibt sich damit $x^* \in [\![6{,}9; 7{,}1]\!]$, und für $f(x^*)$ gilt $f(x^*) \in [\![5{,}9; 6{,}1]\!]$. Mit $r' = 2{,}51$ und $r = 1{,}88$ ergibt sich $x^* \in [\![6{,}99; 7{,}02]\!]$ und $f(x^*) \in [\![5{,}99; 6{,}02]\!]$.

Bekanntlich ist jede differenzierbare Funktion stetig und jede stetige Funktion integrierbar.[1]) Aus den Sätzen 7.1.1 und 7.2.1 folgt, daß für die Lösungen von (7.1.2) auch das Umgekehrte gilt: Jede (über einem abgeschlossenem Intervall) integrierbare Lösung ist stetig, und jede stetige Lösung ist differenzierbar. Ist nämlich f über

[1]) Diese Aussage gilt sowohl für die Integrierbarkeit nach B. Riemann (1826—1866) als auch nach H. L. Lebesgue (1875—1941). Beziehen wir uns ausdrücklich auf den Lebesgueschen Integralbegriff, so wählen wir im Gegensatz zum Riemannschen Integralbegriff eine besondere Bezeichnung.

einem Intervall $[\![a, b]\!]$ integrierbar, so ist f dort beschränkt, also nach Satz 7.2.1, Teil (7.2.5), stetig. Alle stetigen Lösungen sind aber die Funktionen f mit $f(x) = c \cdot x$, und diese Funktionen sind auch differenzierbar. Das muß sich natürlich auch direkt beweisen lassen.

Es sei f eine über einem abgeschlossenen Intervall $[\![a, b]\!]$ integrierbare Lösung der Cauchyschen Grundgleichung. Wir integrieren diese Funktion von a bis b über y und erhalten einerseits

$$\int_a^b f(x)\,dy + \int_a^b f(y)\,dy = f(x) \cdot (b - a) + c$$

und andererseits

$$\int_a^b f(x)\,dy + \int_a^b f(y)\,dy = \int_a^b f(x + y)\,dy.$$

Daraus ergibt sich durch die Substitution $x + y = t$

$$\int_{x+a}^{x+b} f(t)\,dt = f(x) \cdot (b - a) + c.$$

(Da f eine Lösung von (7.1.2) ist, existiert auch das Integral $\int_{x+a}^{x+b} f(t)\,dt$ für jedes $x \in [\![a, b]\!]$).

Da f integrierbar ist, ist das Integral eine stetige Funktion seiner Grenzen, woraus sich die Stetigkeit von f ergibt. Wenn aber f stetig ist, dann ist das Integral sogar eine differenzierbare Funktion seiner Grenzen. Also ist auch f selbst differenzierbar (vgl. J. Aczél [3], 4.2.2.). Damit können alle stetigen Lösungen der Gleichung (7.1.2) — übrigens sehr einfach — bestimmt werden. Es sei nämlich f eine stetige, also auch differenzierbare Funktion, für die (7.1.2) gilt. Dann folgt durch Differentiation nach x die Gleichung $f'(x + y) = f'(x)$ und damit $f'(x) = c$ für alle reellen Zahlen x, wobei c eine feste reelle Zahl ist. Dann ist zunächst $f(x) = c \cdot x + b$ und wegen $f(0) = 0$ (das folgte ja allein aus (7.1.2)) auch $b = 0$, insgesamt also $f(x) = c \cdot x$.

Wir wollen zum Abschluß dieser Betrachtungen noch zeigen, wie man aus der Meßbarkeit[1]) auf die Stetigkeit (Differenzierbarkeit) der Lösungen von (7.1.2) schließen kann.

7.2.7. Satz. *Es sei* $f: \mathbf{R} \to \mathbf{R}$ *eine Lösung der Funktionalgleichung* $f(x + y) = f(x) + f(y)$. *Dann gilt:* f *ist stetig genau dann, wenn* f *meßbar ist.*

Beweis[2]). Da jede stetige Funktion meßbar ist, ist nur noch eine Richtung zu beweisen. Es sei f eine meßbare Lösung von (7.1.2) und mit $x \in \mathbf{R}^*$ eine Zahl gegeben. Wir betrachten die mit f meßbaren Hilfsfunktionen $g: \mathbf{R} \to \mathbf{R}$ und $h: \mathbf{R} \to \mathbf{R}$ mit

[1]) Zu den Begriffen Meßbarkeit von Funktionen und Lebesgue-Integral, auf die wir hier nicht ausführlich eingehen können, vgl. H. v. Mangoldt und K. Knopp [112], Band 4.

[2]) Für diesen Satz gibt es mehrere Beweise. Wir geben hier einen Beweis von A. Alexiewicz und W. Orlicz [14] an.

$g(t) := f(t) - \frac{f(x)}{x} \cdot t$ und $h(t) := \frac{1}{1 + |g(t)|}$. Die Funktion g erfüllt (7.1.2), denn es gilt

$$g(t_1 + t_2) = f(t_1 + t_2) - \frac{f(x)}{x}(t_1 + t_2)$$

$$= f(t_1) + f(t_2) - \frac{f(x)}{x} t_1 - \frac{f(x)}{x} t_2$$

$$= f(t_1) - \frac{f(x)}{x} t_1 + f(t_2) - \frac{f(x)}{x} t_2 = g(t_1) + g(t_2).$$

Wegen $g(x) = 0$ ist g periodisch mit der Periode x, denn es ist $g(t + x) = g(t) - g(x) = g(t)$ für alle $t \in \mathbf{R}$. Die Funktion h ist ebenfalls periodisch und außerdem beschränkt. Daher existiert $^{(L)}\int_0^x h(t)\,dt$. Wir betrachten nun $^{(L)}\int_0^x h(2t)\,dt$. Die Substitution $u = 2t$ ergibt $\frac{1}{2} \cdot {}^{(L)}\int_0^{2x} h(u)\,du$. Da h die Periode x besitzt, ist

$$^{(L)}\int_0^x h(2t)\,dt = {}^{(L)}\int_0^x h(t)\,dt.$$

Also gilt

$$^{(L)}\int_0^x \frac{dt}{1 + |g(t)|} = {}^{(L)}\int_0^x h(t)\,dt = {}^{(L)}\int_0^x h(2t)\,dt = {}^{(L)}\int_0^x \frac{dt}{1 - |g(2t)|}.$$

Mit diesem Ergebnis folgt durch Partialbruchzerlegung

$$^{(L)}\int_0^x \frac{|g(t)|\,dt}{(1 + 2|g(t)|)(1 + |g(t)|)} = -{}^{(L)}\int_0^x \frac{dt}{1 + 2|g(t)|} + {}^{(L)}\int_0^x \frac{dt}{1 + |g(t)|} = 0.$$

Das bedeutet aber, daß g fast überall[1]) Null ist, denn der Integrand ist nicht negativ. Damit gilt $f(t) = \frac{f(x)}{x} t$ für fast alle t bei beliebigem, aber festem $x \in \mathbf{R}^*$. Wählt man $x = 1$, so gilt für fast alle t die Beziehung $f(t) = f(1) \cdot t$. Daraus folgt, daß es für jedes $x \neq 0$ ein $t_0 \neq 0$ gibt mit $f(t_0) = \frac{f(x)}{x} \cdot t_0$ und $f(t_0) = f(1) \cdot t_0$. Also gilt $f(x) = f(1) \cdot x$ für jedes $x \neq 0$. Das gilt aber auch für $x = 0$, da ja allein aus (7.1.2) $f(0) = 0$ folgt. Damit ist Satz 7.2.7 bewiesen.

Auf weitere, mit der Stetigkeit in diesem Zusammenhang äquivalente Bedingungen sei noch hingewiesen: „f hat eine meßbare Majorante" (vgl. W. Sierpiński [163]) und „f ist auf einer Menge von positivem Maß einseitig beschränkt" (vgl. A. Ostrowski [128], H. Kestelman [84]).

[1]) Das heißt „für alle x bis auf eine Menge vom Maß 0".

7.3. Die unstetigen Lösungen der Funktionalgleichung $f(x + y) = f(x) + f(y)$

Die Sätze 7.2.1 und 7.2.7 legen die Vermutung nahe, daß man mit den Funktionen $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x$ schon alle Lösungen von (7.1.2) gefunden hat, daß also nur diese Funktionen additiv sind. So besagt doch Teil (7.2.6) von Satz 7.2.1., daß die Graphen von unstetigen Lösungen in der x,y-Ebene dicht liegen. Im Jahre 1905 konnte G. HAMEL [64] (vgl. auch E. KAMKE [83]) die Existenz unstetiger Lösungen nachweisen. Er konstruierte dazu eine Basis der reellen Zahlen, wozu er den Wohlordnungssatz[1]) benutzte, den E. ZERMELO (1871—1953) im Jahre 1904 bewiesen hatte (vgl. E. ZERMELO [189, 190] und E. KAMKE [83]).

7.3.1. Definition[2]). Eine Teilmenge $\mathbf{R}_B$ der reellen Zahlen heißt *Basis* (*Hamelbasis*) *von* $\mathbf{R}$ *über* $\mathbf{Q} :\Leftrightarrow$

a) Für jede endliche Teilmenge $\{b_1, b_2, \ldots, b_m\} \subseteq \mathbf{R}_B$ gilt

$$\text{7.3.2.} \quad \sum_{\nu=1}^{m} r_\nu b_\nu = 0$$

nur, wenn $r_1 = r_2 = \cdots = r_m = 0$ ist mit $r_\nu \in \mathbf{Q}$ für alle ν. (Die b_ν sind also linear unabhängig über $\mathbf{Q}$.)

b) Jede reelle Zahl x ist eine Linearkombination endlich vieler Zahlen $b_1, b_2, \ldots, b_m$ (Basiselemente) aus $\mathbf{R}_B$ mit rationalen Koeffizienten:

$$\text{7.3.3.} \quad x = \sum_{\nu=1}^{m} r_\nu b_\nu.$$

Der Nachweis der Existenz einer solchen Basis $\mathbf{R}_B$ wird durch transfinite Induktion geführt (vgl. E. KAMKE [83]). Gegeben sei eine Wohlordnung der reellen Zahlen, und b sei die erste Zahl bezüglich dieser Wohlordnung, die von Null verschieden ist; dann soll b Basiselement sein. Es sei x eine beliebige reelle Zahl und R_x der zu x gehörende Abschnitt[3]) in der Wohlordnung, und für alle reellen Zahlen $y \in R_x$ sei schon entschieden, ob sie Basiselemente sind oder nicht. Dann wird die Zahl x genau dann in die Basis $\mathbf{R}_B$ aufgenommen, wenn x mit je endlich vielen der schon festgelegten Basiselemente linear unabhängig über $\mathbf{Q}$ ist.

Die so definierte Menge $\mathbf{R}_B$ ist als Teilmenge einer wohlgeordneten Menge selbst wohlgeordnet. Wir haben noch zu zeigen, daß $\mathbf{R}_B$ wirklich eine Basis der reellen Zahlen ist.

Zu (7.3.1a)): Angenommen, es würde $\sum_{\nu=1}^{m} r_\nu b_\nu = 0$ für irgendwelche endlich vielen Elemente von $\mathbf{R}_B$ gelten, ohne daß alle Koeffizienten gleich Null sind. Dann gäbe es

[1]) Der Wohlordnungssatz besagt, daß jede Menge wohlgeordnet werden kann. Eine Menge M heißt *wohlgeordnet*, wenn jede nichtleere Teilmenge von M ein kleinstes Element enthält.

[2]) Der Begriff der Hamelbasis kann allgemeiner für beliebige Vektorräume definiert werden (vgl. H. v. MANGOLDT und K. KNOPP [112], Nr. 110).

[3]) Das ist die Teilmenge der reellen Zahlen, die der Zahl x in der gegebenen Wohlordnung vorangehen.

unter diesen Zahlen eine nach der ursprünglichen Wohlordnung letzte mit von Null verschiedenem Koeffizienten. Diese Zahl ließe sich dann also aus vorhergehenden Basiselementen linear kombinieren. Das steht im Widerspruch zur Konstruktion der Menge $\mathbf{R}_B$.

Zu (7.3.1b)): Jede reelle Zahl x ist nun entweder Basiselement oder nicht.

Fall 1: $x \in \mathbf{R}_B$. Dann ist $x = 1 \cdot x$ eine Darstellung der Form (7.3.3).

Fall 2: $x \notin \mathbf{R}_B$. Dann besteht nach Konstruktion der Menge $\mathbf{R}_B$ zwischen x und endlich vielen in der gegebenen Wohlordnung vor x liegenden Basiselementen eine Gleichung der Form (7.3.2), wobei x einen von Null verschiedenen Koeffizienten hat. Löst man diese Gleichung nach x auf, so erhält man eine Darstellung der Form (7.3.3).

Die Darstellung $x = \sum_{v=1}^{m} r_v b_v$ ist für jedes $x \in \mathbf{R}$ eindeutig bestimmt.[1]) Gäbe es nämlich zwei verschiedene Darstellungen, so erhielte man durch Subtraktion eine nichttriviale Linearkombination der Null, im Widerspruch zur Konstruktion der Menge $\mathbf{R}_B$.

Jetzt können wir den folgenden Satz beweisen:

7.3.4. Satz. *Es gibt unstetige Lösungen der Funktionalgleichung* $f(x + y) = f(x) + f(y)$

Beweis. a) Es sei $\mathbf{R}_B$ eine Basis der reellen Zahlen. Dann wird f zunächst für die Basiselemente definiert, und zwar in völlig beliebiger Weise. Ist x eine beliebige reelle Zahl und $x = \sum_{v=1}^{m} r_v b_v$ ihre Linearkombination durch Basiselemente, so wird

$$f(x) := \sum_{v=1}^{m} r_v f(b_v)$$

gesetzt.

b) Die in a) definierte Funktion $f: \mathbf{R} \to \mathbf{R}$ ist Lösung der Funktionalgleichung (7.1.2). Sind nämlich x und y reelle Zahlen mit $x = \sum_{v=1}^{m} r_v b_v$ und $y = \sum_{v=1}^{m} r_v' b_v$[2]), so gilt

$$x + y = \sum_{v=1}^{m} (r_v + r_v')\, b_v$$

und

$$f(x + y) = \sum_{v=1}^{m} (r_v + r_v')\, f(b_v) = \sum_{v=1}^{m} r_v f(b_v) + \sum_{v=1}^{m} r_v' f(b_v) = f(x) + f(y).$$

[1]) In dieser Summendarstellung wird bei der Numerierung der Einfachheit halber stets mit Eins begonnen. Diese Numerierung stimmt dann i. allg. mit der von $\mathbf{R}_B$ nicht überein. Um in der Summendarstellung auch die Reihenfolge der b_v ($1 \leqq v \leqq m$) eindeutig zu machen, ordnen wir diese Basiselemente so, daß b_v auch in der Wohlordnung von $\mathbf{R}_B$ vor b_{v+1} liegt, dort aber i. allg. mit anderem Index.

[2]) Wir können o. B. d. A. annehmen, daß in beiden Linearkombinationen dieselben Basiselemente auftreten. Das kann durch Einfügen von Summanden mit dem Koeffizienten Null erreicht werden.

c) Die in a) definierte Funktion $f\colon \mathbf{R} \to \mathbf{R}$ ist stetig, wenn $f(b)$ für alle $b \in \mathbf{R}_B$ so festgelegt wird, daß $f(b) = c \cdot b$ gilt bei vorgegebenem $c \in \mathbf{R}$; denn es ist

$$f(x) = \sum_{v=1}^{m} r_v f(b_v) = \sum_{v=1}^{m} r_v \cdot c \cdot b_v = c \cdot \sum_{v=1}^{m} r_v b_v = c \cdot x.$$

Wird $f(b) \neq c \cdot b$ für wenigstens ein $b \in \mathbf{R}_B$ gewählt, so ist f unstetig. Der Graph liegt dann in der x,y-Ebene dicht (vgl. Beweis von Teil (7.2.6) des Satzes 7.2.1). Damit ist Satz 7.3.4 bewiesen.

Man erhält auf die angegebene Weise auch alle Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ von (7.1.2), wie die folgende Überlegung zeigt: Es sei wieder $\mathbf{R}_B$ eine Basis der reellen Zahlen und $f\colon \mathbf{R} \to \mathbf{R}$ eine beliebige Lösung von (7.1.2). Es ist zu zeigen, daß man f auf die im Beweis gegebene Weise gewinnen kann. Dazu definieren wir folgendermaßen eine Funktion $g\colon \mathbf{R} \to \mathbf{R}$. Zunächst gelte $g(b) := f(b)$ für alle $b \in \mathbf{R}_B$. Damit stimmt g mit f in allen Basiselementen überein. Für $x \notin \mathbf{R}_B$ mit $x = \sum_{v=1}^{m} r_v b_v$ wird $g(x) := \sum_{v=1}^{m} r_v g(b_v)$ gesetzt. Die gegebene Funktion f, die als Lösung von (7.1.2) additiv ist und für die $f(r_v b_v) = r_v f(b_v)$ gilt (vgl. Bemerkung 7.1.5), hat dann an der Stelle x den Wert $f(x) = f\left(\sum_{v=1}^{m} r_v b_v\right) = \sum_{v=1}^{m} f(r_v b_v) = \sum_{v=1}^{m} r_v f(b_v)$, also ist $g = f$, und zwar unabhängig davon, ob f stetig oder unstetig ist.

7.4. Zu Lösungen von $f(x + y) = f(x) + f(y)$ über eingeschränktem Definitionsbereich

Bisher haben wir die Lösungen der Funktionalgleichung $f(x + y) = f(x) + f(y)$ über der Menge der reellen Zahlen betrachtet. Wir wollen jetzt auf einige Probleme hinweisen, die bei der Untersuchung der Lösungen über eingeschränkten Definitionsbereichen entstehen. Dabei beschränken wir uns darauf, die Menge aller Lösungen $f\colon \mathbf{R}_+ \to \mathbf{R}$ von $f(x + y) = f(x) + f(y)$ zu bestimmen. und verweisen auf weiterführende Betrachtungen bei Z. Daróczy und L. Losonczi [41].

Zunächst folgt aus dem Beweis von Satz 7.1.1 sofort, daß alle stetigen (monotonen usw.) Lösungen die Funktionen $f\colon \mathbf{R}_+ \to \mathbf{R}$ mit $f(x) = c \cdot x$ ($c \in \mathbf{R}$) sind. Es liegt nun nahe, zur Bestimmung aller Lösungen die Konstruktion einer Hamelbasis für die nichtnegativen reellen Zahlen über $\mathbf{Q}_+$ zu versuchen. J. Aczél und P. Erdös [8] bewiesen aber 1965, daß es keine Hamelbasis der Menge $\mathbf{R}_+$ über $\mathbf{Q}_+$ gibt, daß es also keine Teilmenge $B' \subseteq \mathbf{R}_+$ derart gibt, daß sich jede nichtnegative reelle Zahl eindeutig als Linearkombination einer endlichen Menge von Zahlen aus B' mit nichtnegativen rationalen Koeffizienten darstellen läßt. Zum Beweis dieser Behauptung werde angenommen, es gäbe doch eine solche Basis B'. Wegen des Archimedischen Axioms (vgl. MfL Bd. 2, 6.2.1.) für reelle Zahlen gibt es zu beliebigen Zahlen $b_1, b_2 \in B'$ eine natürliche Zahl k mit $kb_1 > b_2$. Dann betrachten wir die reelle Zahl $y = kb_1 - b_2$. Diese Zahl ist größer als Null und hätte daher nach der Annahme eine

eindeutige Basisdarstellung $y = \sum_{\nu=1}^{m} r_\nu b_\nu$ mit $r_\nu \in \mathbf{Q}_+$ und $b_\nu \in B'$. Daraus folgt

$$kb_1 - b_2 = r_1 b_1 + r_2 b_2 + \cdots + r_m b_m$$

und daraus

$$(r_1 - k)\, b_1 + (r_2 + 1)\, b_2 + \cdots + r_m b_m = 0,$$

wobei $r_2 + 1 > 0$ gilt. Für die Gültigkeit dieser Gleichung wäre $r_1 - k < 0$ notwendig. Damit hätte die positive Zahl

$$x = -(r_1 - k)\, b_1 = (r_2 + 1)\, b_2 + \cdots + r_m b_m$$

aber zwei Basisdarstellungen.

Trotz dieses negativen Ergebnisses kann doch die Menge aller Lösungen von (7.1.2) über $\mathbf{R}_+$ recht einfach bestimmt werden, falls man alle Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ schon kennt. Zunächst ist klar, daß die Einschränkung einer Lösung über $\mathbf{R}$ auf $\mathbf{R}_+$ eine Lösung über $\mathbf{R}_+$ ist. Es gilt aber auch umgekehrt, daß jede Lösung über $\mathbf{R}_+$ Einschränkung einer Lösung über $\mathbf{R}$ ist (vgl. J. Aczél und P. Erdös [8]). Also kennt man mit den Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ von (7.1.2) auch alle Lösungen $f\colon \mathbf{R}_+ \to \mathbf{R}$ von (7.1.2). Wir beweisen den folgenden Satz:

7.4.1. Satz. *Jede Lösung $f\colon \mathbf{R}_+ \to \mathbf{R}$ der Cauchyschen Funktionalgleichung $f(x + y) = f(x) + f(y)$ ist die Einschränkung einer Lösung $f\colon \mathbf{R} \to \mathbf{R}$ dieser Gleichung.*

Beweis (nach J. Aczél und P. Erdös [8]). Es sei $f\colon \mathbf{R}_+ \to \mathbf{R}$ eine Lösung. Durch

$$\bigwedge_{x \in \mathbf{R}_+} f(-x) := -f(x) \tag{7.4.2}$$

erhalten wir eine über $\mathbf{R}$ definierte Funktion, von der wir zeigen, daß sie Lösung von (7.1.2) ist. Dazu werden sechs Fälle unterschieden:

Fall 1: $x \geqq 0$; $y \geqq 0$. Hier gilt (7.1.2) nach Voraussetzung.

Fall 2: $x \geqq 0$; $y < 0$; $x + y \geqq 0$. In diesem Fall folgt aus (7.4.2) und der Gültigkeit von (7.1.2) für nichtnegative Zahlen:

$$f(x) = f\big((x + y) + (-y)\big) = f(x + y) + f(-y) = f(x + y) - f(y),$$

also $f(x + y) = f(x) + f(y)$.

Fall 3: $x < 0$; $y \geqq 0$; $x + y \geqq 0$. Der Beweis verläuft wie im zweiten Fall.

Fall 4: $x \geqq 0$; $y < 0$; $x + y < 0$. Aus der Gültigkeit von (7.1.2) für $x \in \mathbf{R}_+$ und (7.4.2) folgt

$$\begin{aligned} f(y) &= -f(-y) = -f\big(-(x + y) + x\big) = -\big(f\big(-(x + y)\big) + f(x)\big) \\ &= f(x + y) - f(x), \end{aligned}$$

also wieder $f(x + y) = f(x) + f(y)$.

Fall 5: $x < 0$; $y \geqq 0$; $x + y < 0$. Der Beweis verläuft wie im Fall 4.

Fall 6: $x < 0$; $y < 0$. Aus den Voraussetzungen (7.1.2) und (7.4.2) für nichtnegative Argumente schließt man diesmal:

$$f(x+y) = -f\big(-(x+y)\big) = -f\big((-x)+(-y)\big)$$
$$= -\big(f(-x)+f(-y)\big) = f(x)+f(y).$$

Damit ist alles bewiesen, denn da sich jede Lösung über $\mathbf{R}_+$ zu einer Lösung über $\mathbf{R}$ fortsetzen läßt, gibt es keine Lösung über $\mathbf{R}_+$, die nicht Einschränkung einer Lösung über $\mathbf{R}$ ist. Es sei noch vermerkt, daß man die Gültigkeit von (7.1.2) auch nur für positive reelle Zahlen fordern kann. Man hat dann lediglich noch $f(0) := 0$ zu definieren.

7.4.3. Bemerkung. Gelegentlich werden wir die stetigen Lösungen der Funktionalgleichung (7.1.2) über einem Intervall $I \coloneqq [\![0, a]\!]$ mit $a \in \mathbf{R}_+$ benötigen. Durch einfache Abänderungen der Überlegungen in 7.1. erkennt man, daß dies die entsprechend eingeschränkten Funktionen $f\colon I \to \mathbf{R}$ mit $f(x) = c \cdot x$ sind. Dabei kann das Intervall I auch einseitig oder beidseitig offen sein.

7.5. Bemerkungen zur direkten Proportionalität

Bekanntlich heißen Größen zueinander *proportional*, wenn jeder Maßzahl x der einen Größe genau eine Maßzahl y der anderen Größe zugeordnet ist und wenn es eine reelle Zahl c derart gibt, daß je zwei zugeordnete Maßzahlen in der Beziehung $y = c \cdot x$ stehen. Nun findet man mitunter, auch in der Schulbuchliteratur, folgende Charakterisierung der *direkten Proportionalität*:

7.5.1. Werden die Werte der einen Größe verdoppelt, verdreifacht, ..., so werden auch die entsprechenden Werte der anderen Größe verdoppelt, verdreifacht, ...

Bezeichnet man die Zuordnung mit $y = f(x)$, so kann die Formulierung (7.5.1) durch

7.5.2. $\bigwedge_{m \in \mathbf{N}} \bigwedge_{x \in \mathbf{R}_+} f(m \cdot x) = m \cdot f(x)$

ausgedrückt werden. Wir wollen zeigen (vgl. D. Ilse [73]), daß (7.5.1) nicht ausreicht, daß also die Funktion $f\colon \mathbf{R}_+ \to \mathbf{R}$ mit $f(x) = c \cdot x$ durch (7.5.2) nur mit zusätzlichen Forderungen an f charakterisiert werden kann. Eine solche Forderung ist auch wieder die Stetigkeit.[1])

Zunächst ist klar, daß f mit $f(x) = c \cdot x$ die Eigenschaft (7.5.2) besitzt. Es bleibt nachzuweisen, daß diese Funktion auch die einzige derartige stetige Funktion ist: Aus (7.5.2) folgt $f(0) = 0$, wenn $m = 0$ gesetzt wird, und $f(m) = c \cdot m$, wenn $x = 1$ gesetzt wird, wobei $f(1) = c$ ist. Weiterhin erhält man für alle $r \in \mathbf{Q}$ mit $x = r = \frac{p}{q}$

$$f\left(q \cdot \frac{p}{q}\right) = f(p) = c \cdot p = q \cdot f\left(\frac{p}{q}\right),$$

also $f(r) = c \cdot r$. Ist schließlich $x \in \mathbf{R}_+$ und irrational, (r_n) eine Folge mit $r_n \in \mathbf{Q}$ für

[1]) Will man f mit $f(x) = c \cdot x$ über $\mathbf{R}$ charakterisieren, so braucht man in den folgenden Betrachtungen nur $m \in \mathbf{Z}$ zu fordern.

alle n und $\lim_{n\to\infty} r_n = x$, so gilt notwendigerweise $f(x) = \lim_{n\to\infty} f(r_n) = \lim_{n\to\infty} c \cdot r_n = c \cdot x$, da f stetig sein sollte.

Statt Stetigkeit genügt es auch, Monotonie zu fordern. Das beweist man mit denselben Überlegungen, mit denen Teil (7.2.3) von Satz 7.2.1 bewiesen wurde. Dagegen genügt es z. B. nicht mehr, die Positivität von f zu fordern (vgl. Teil (7.2.4) von Satz 7.2.1), denn für die Funktion f mit

$$f(x) := \begin{cases} c \cdot 2 \cdot r \cdot \sqrt{2}, \text{ falls es eine Zahl } r \in \mathbf{Q} \text{ gibt mit } x = r \cdot \sqrt{2}, \\ c \cdot x \quad \text{sonst} \end{cases}$$

gilt zunächst für $c > 0$, daß $f(x) > 0$ ist für alle $x > 0$. Sie hat auch die Eigenschaft (7.5.2); denn ist das Argument x ein rationales Vielfaches von $\sqrt{2}$, also $x = r \cdot \sqrt{2}$, so gilt $f(m \cdot x) = f\left(m \cdot r \cdot \sqrt{2}\right) = c \cdot 2 \cdot m \cdot r \cdot \sqrt{2} = m\left(c \cdot 2 \cdot r \cdot \sqrt{2}\right) = m \cdot f\left(r \cdot \sqrt{2}\right) = m \cdot f(x)$, weil mit r auch $r \cdot m$ rational ist. Außerdem ist f eineindeutig. Ist x nicht rationales Vielfaches von $\sqrt{2}$, so erhält man $f(m \cdot x) = c(m \cdot x) = m(c \cdot x) = m \cdot f(x)$. Die Funktion f ist unstetig, z. B. an der Stelle $x_0 = \sqrt{2}$. Denn für eine monoton wachsende Folge (r_n) mit $r_n \in \mathbf{Q}$ für alle n und $\lim_{n\to\infty} r_n = x_0$ gilt $\lim_{n\to\infty} f(r_n) = \lim_{n\to\infty} c \cdot r_n = c \cdot \sqrt{2}$, aber $f\left(\sqrt{2}\right) = c \cdot 2 \cdot \sqrt{2}$. Die eben betrachtete Funktion f zeigt, daß die Formulierung (7.5.1) nicht die Monotonie von f impliziert, wie man vielleicht auf den ersten Blick denken könnte. Die direkte Proportionalität wird durch (7.5.2) erst dann charakterisiert, wenn man noch eine geeignete Regularitätsforderung (z. B. Stetigkeit oder Monotonie) stellt. Dagegen folgt $f(r) = c \cdot r$ allein aus (7.5.2) für $r \in \mathbf{Q}$. Solange man also die reellen Zahlen nicht zur Verfügung hat, z. B. in Klasse 6, macht man keinen Fehler, wenn man aus (7.5.1) die direkte Proportionalität folgert. Wenn in der Experimentalphysik bei der Herleitung physikalischer Gesetze (z. B. des Hookeschen Gesetzes) auch im Reellen so geschlossen wird, so ist das deshalb berechtigt, weil (stillschweigend) die Monotonie vorausgesetzt wird (z. B. hat jede Vergrößerung der Last auch eine Vergrößerung der Dehnung zur Folge).

Unsere Betrachtungen zeigen, daß die Funktionalgleichungen

$$f(x + y) = f(x) + f(y) \quad \text{und} \quad f(m \cdot x) = m \cdot f(x)$$

über $\mathbf{R}_+$ (und damit auch über $\mathbf{R}$) nicht äquivalent sind. Zwar ist jede Lösung von (7.1.2) auch eine Lösung von (7.5.2), wie man durch vollständige Induktion beweist, aber nicht jede Lösung von (7.5.2) ist auch eine Lösung von (7.1.2), wie die eben definierte Funktion f zeigt, denn für diese ist z. B. $f(1) = c$ und $f\left(\sqrt{2}\right) = c \cdot 2 \cdot \sqrt{2}$, aber $f\left(1 + \sqrt{2}\right) = c\left(1 + \sqrt{2}\right)$. (Für den Fall, daß man f über $\mathbf{R}$ betrachten will, definiert man $f(x) := -f(-x)$ für negative x.)

Im Zusammenhang mit der direkten Proportionalität betrachten wir noch eine weitere Abschwächung von (7.1.2) und (7.5.2), und zwar

7.5.3 $\bigwedge_{x \in \mathbf{R}_+} f(2 \cdot x) = 2 \cdot f(x)$[1]).

[1]) Hier handelt es sich um eine Funktionalgleichung mit einer einzigen Argumentvariablen. Für solche Funktionalgleichungen werden wesentlich andere Lösungsmethoden benutzt (vgl. M. Kuczma [93]).

Durch (7.5.3) ist die direkte Proportionalität auch mit der Forderung nach Stetigkeit nicht charakterisiert, wie das folgende Beispiel zeigt (vgl. D. Ilse [73]): Es sei $f\colon \mathbf{R}_+ \to \mathbf{R}$ definiert durch

$$f(x) := \begin{cases} 2^k \cdot c + (c' - 2c)\,(x - 2^k) & \text{für } 2^k \leqq x < 2^{k-1} \cdot 3, \\ 2^{k-1} \cdot c' + (4c - c')\,(x - 2^{k-1} \cdot 3) & \text{für } 2^{k-1} \cdot 3 \leqq x < 2^{k-1}, \\ 0 \text{ für } x = 0, \end{cases}$$

wobei $c' \neq 3c$, $c > 0$, $c' > 0$ und $k \in \mathbf{Z}$ ist (siehe Abb. 42). Diese Funktion erfüllt die Gleichung (7.5.3). Dazu betrachten wir eine Zahl x, für die es eine ganze Zahl k gibt mit $2^{k-1} \cdot 3 \leqq x < 2^{k+1}$. Dann gilt für die Zahl $2x$ die Abschätzung $2^k \cdot 3 \leqq 2x < 2^{k-2}$, also

$$\begin{aligned} f(2x) &= 2^k \cdot c' + (4c - c')\,(2x - 2^k \cdot 3) \\ &= 2\big(2^{k-1} \cdot c' + (4c - c')\,(x - 2^{k-1} \cdot 3)\big) = 2 \cdot f(x). \end{aligned}$$

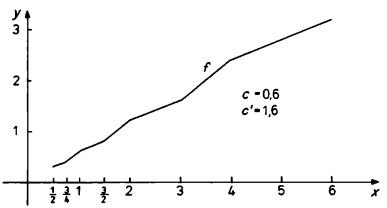


Abb. 42

Es sei nun x eine Zahl, für die es eine ganze Zahl k gibt mit $2^k \leqq x < 2^{k-1} \cdot 3$. Dann gilt für die Zahl $2x$ die Abschätzung $2^{k+1} \leqq 2x < 2^k \cdot 3$, also

$$\begin{aligned} f(2x) &= 2^{k+1} \cdot c + (c' - 2c)\,(2x - 2^{k+1}) \\ &= 2\big(2^k \cdot c + (c' - 2c)\,(x - 2^k)\big) = 2 \cdot f(x). \end{aligned}$$

Weiterhin ist die Funktion f stückweise linear, also zwischen den „Knickstellen" stetig. Sie ist aber auch an diesen Stellen stetig, wie man leicht nachprüft. Schließlich ist sie verschieden von der Funktion $f\colon \mathbf{R}_+ \to \mathbf{R}$ mit $f(x) = c \cdot x$.

Trotzdem kann man auch mit der Gleichung (7.5.3) die direkte Proportionalität charakterisieren, wenn man die Stetigkeitsforderung hinreichend verstärkt. Wir fordern, daß f an der Stelle 0 rechtsseitig stetig differenzierbar sein soll (vgl. J. Aczél und H. Kiesewetter [10]). Diese Forderung wird durch Abb. 42 nahegelegt, denn durch die rechtsseitige Tangente werden die sich gegen Null häufenden Zacken „geglättet".

Es sei f also eine solche Funktion. Dann erhält man durch Differentiation aus (7.5.3) für eine rechtsseitige Umgebung von Null

7.5.4. $f'(2x) = f'(x)$.

Nun gilt: f' ist genau dann konstant, wenn f' in $x = 0$ stetig ist. Wenn f' konstant ist, dann ist f' in $x = 0$ trivialerweise stetig. Es sei nun f' in $x = 0$ stetig und nicht konstant, dann gibt es Argumente x_1 und x_2 mit $x_1 \neq x_2$ und $f'(x_1) = c_1$ und $f'(x_2) = c_2$, wobei $c_1 \neq c_2$ ist. Aus (7.5.4) folgt $f'\left(\frac{x_1}{2^k}\right) = c_1$ und $f'\left(\frac{x_2}{2^k}\right) = c_2$ für alle $k \in \mathbf{N}$ und damit $\lim\limits_{k\to\infty} f'\left(\frac{x_1}{2^k}\right) = c_2$ und $\lim\limits_{k\to\infty} f'\left(\frac{x_2}{2^k}\right) = c_2$. Das ist jedoch ein Widerspruch zur Stetigkeit von f' in $x = 0$, denn $\left(\frac{x_1}{2^k}\right)$ und $\left(\frac{x_2}{2^k}\right)$ sind Nullfolgen. Man erhält $f'(x) = c$, also $f(x) = c \cdot x + b$ und wegen $f(0) = 0$ schließlich $f(x) = c \cdot x$.

7.6. Anwendungen

Wir wollen jetzt noch einige Anwendungen der Funktionalgleichung (7.1.2) bzw. des Satzes 7.1.1 behandeln.

Bei der Ableitung des *Hookeschen Gesetzes* wird zunächst experimentell ermittelt, daß die Verlängerung Δl eines Drahtes der Belastung P und (bei gleicher Belastung) der Länge l proportional ist, und dann daraus geschlossen, daß Δl auch dem Produkt $l \cdot P$ proportional ist (vgl. E. Grimsehl [62], § 47). In analoger Weise wird im Physikunterricht z. B. bei der Herleitung der Grundgleichung der Wärmelehre, bei der Zustandsgleichung idealer Gase usw. vorgegangen. Dabei taucht (bei Lehrern und Schülern) hin und wieder die Frage auf, inwiefern diese oft praktizierte Schlußweise gerechtfertigt ist, inwiefern man also aus der Tatsache, daß eine Größe z zu Größen x und y proportional ist, auf die Proportionalität von z zum Produkt dieser Größen x und y schließen darf. Die Größe z wird also in dem gerade betrachteten Zusammenhang als zweistellige Funktion aufgefaßt, die zu jedem ihrer Argumente proportional ist, und zwar jeweils bei konstant gehaltenem anderen Argument. Wir betrachten also zweistellige Funktionen $F\colon \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ mit folgenden Eigenschaften:

a) Es gibt Funktionen $c_1\colon \mathbf{R} \to \mathbf{R}$ und $c_2\colon \mathbf{R} \to \mathbf{R}$, so daß für alle reellen Zahlen x und y folgendes gilt:

a_1) $F(x, y) = c_1(y) \cdot x$,

a_2) $F(x, y) = c_2(x) \cdot y$,

b) $F(1, 1) = c\ (c \neq 0)$.

Dann gilt:

7.6.1. Satz. *Die einzige Funktion* $F\colon \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ *mit den Eigenschaften* a) *und* b) *ist die Funktion* F *mit* $F(x, y) = c \cdot x \cdot y$.[1])

Beweis (vgl. D. Ilse [77]). Aus a_1) und a_2) folgt

7.6.2. $\bigwedge\limits_{x,y\in\mathbf{R}} c_1(y) \cdot x = c_2(x) \cdot y$.

[1]) Man beachte, daß die Stetigkeit der Funktion F nicht gefordert zu werden braucht. Die Stetigkeit von F folgt aus den Voraussetzungen.

Setzt man hierin $y = 1$, so erhält man $c_1(1) \cdot x_1 = c_2(x_1) \cdot 1$ und $c_1(1) \cdot x_2 = c_2(x_2) \cdot 1$ für beliebige reelle Zahlen x_1 und x_2 und daraus durch Addition $c_1(1) \cdot (x_1 + x_2) = c_2(x_1) + c_2(x_2)$. Andererseits folgt aus (7.6.2) für $y = 1$ und $x = x_1 + x_2$ die Beziehung $c_1(1) \cdot (x_1 + x_2) = c_2(x_1 + x_2) \cdot 1$, so daß sich

$$\bigwedge_{x_1, x_2 \in \mathbf{R}} c_2(x_1 + x_2) = c_2(x_1) + c_2(x_2)$$

ergibt. Die Funktion F ist wegen a_1) für jedes feste y eine stetige Funktion von x und nach a_2) für jedes feste x eine stetige Funktion von y. Damit folgt wieder aus a_1) und a_2), daß c_1 und c_2 stetige Funktionen sind, also ist c_2 eine stetige Lösung der Cauchyschen Grundgleichung, so daß nach Satz 7.1.1 $c_2(x) = c_2(1) \cdot x$ gilt. Damit ergibt sich aus a_2) $F(x, y) = c_2(1) \cdot x \cdot y$. Wegen b) ist $c_2(1) = c$. Also ergibt sich $F(x, y) = c \cdot x \cdot y$. Diese Funktion genügt auch den Forderungen a) und b). Der Beweis wird noch kürzer, wenn man in (7.6.2) zunächst $x = 1$ und dann $y = 1$ setzt. Dadurch erhält man $c_1(y) = c_2(1) \cdot y$ bzw. $c_2(x) = c_1(1) \cdot x$, und daraus wieder $F(x, y) = c \cdot x \cdot y$.

Nutzen wir die in diesem Kapitel erzielten Ergebnisse für eine axiomatische Begründung der Streckenmessung, so können wir feststellen, daß es nicht genügt, nur die (endliche) Additivität zu fordern. Man erhielte dann auch unstetige Abbildungen der Menge aller Strecken der euklidischen Ebene in die Menge der reellen Zahlen. Diese Unstetigkeit wäre z. B. so beschaffen (vgl. Teil (7.2.5) von Satz 7.2.1), daß es zu Strecken mit den „üblichen" Längen a und b (etwa $a < b$) Strecken gäbe, deren „übliche" Längen zwischen a und b liegen, denen bei der unstetigen Längenmessung aber Längen zugeordnet würden, die jede noch so große reelle Zahl übertreffen! Will man also die geläufige Streckenmessung definieren, so hat man neben der Additivität etwa noch zu fordern (vgl. Teil (7.2.4) von Satz 7.2.1), daß jeder Strecke eine positive Zahl als Länge zugeordnet werden soll. So wird z. B. in MfL Bd. 6, 1.1.4., verfahren, wobei „positiv" noch zu „nichtnegativ" abgeschwächt wird. Das ist möglich, da durch die Forderung, daß es eine Strecke mit der Länge 1 geben soll, die triviale Lösung, die jeder Strecke die Zahl 0 zuordnet, ausgeschlossen wird (vgl. Aufgabe 7.7.1).

Ein Rechteck, dessen Seitenlängen x und y sind, hat den Flächeninhalt $F(x, y) = x \cdot y$. Die Flächeninhaltsfunktion F: $\mathbf{R}_+^* \times \mathbf{R}_+^* \to \mathbf{R}$ hat u. a. folgende Eigenschaften:

a) $\bigwedge_{x, y \in \mathbf{R}_+^*} F(x, y) > 0,$

b) $\bigwedge_{x_1, x_2, y \in \mathbf{R}_+^*} F(x_1 + x_2, y) = F(x_1, y) + F(x_2, y),$

c) $\bigwedge_{x, y_1, y_2 \in \mathbf{R}_+^*} F(x, y_1 + y_2) = F(x, y_1) + F(x, y_2),$

d) $F(1, 1) = 1.$

Obwohl es sich hier bei b) und c) um ein System von Funktionalgleichungen handelt, kann man doch leicht durch Zurückführung auf die Gleichung (7.1.2) zeigen, daß F: $\mathbf{R}_+^* \times \mathbf{R}_+^* \to \mathbf{R}$ mit $F(x, y) = x \cdot y$ die einzige Funktion mit den Eigenschaften a) bis d) ist(vgl. J. Aczél [3], 2.3.2): Faßt man y in b) als festen Parameter auf, so er-

hält man eine Funktion $f_y\colon \mathbf{R}_+^* \to \mathbf{R}$ mit $f_y(x) := F(x, y)$. Für diese Funktion gilt $f_y(x_1 + x_2) = f_y(x_1) + f_y(x_2)$. Da mit F auch f_y für alle positiven Zahlen x_1, x_2 positiv ist, folgt daraus nach Teil (7.2.4) von Satz 7.2.1 $f_y(x) = c(y) \cdot x$ oder $F(x, y) = c(y) \cdot x$. Mit diesem Ergebnis ergibt sich aus c)

$$F(x, y_1 + y_2) = c(y_1 + y_2) \cdot x = c(y_1) \cdot x + c(y_2) \cdot x,$$

also

$$\bigwedge_{y_1, y_2 \in \mathbf{R}_+^*} c(y_1 + y_2) = c(y_1) + c(y_2).$$

Mit F ist auch die Funktion c positiv, also gilt $c(y) = a \cdot y$ und damit $F(x, y) = a \cdot x \cdot y$; wegen d) ist $a = 1$, so daß man schließlich $F(x, y) = x \cdot y$ erhält.

Die in Mathematik, Lehrbuch für Klasse 8, [116], S. 19, gegebene Strahlensatzfigur (siehe Abb. 43) läßt folgende Überlegungen zu (vgl. A. Kirsch [88]). Die

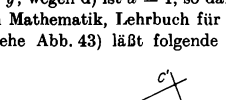


Abb. 43

Strecken $\overline{SA}$, $\overline{SB}$ und $\overline{SC}$ mögen die Längen x, y bzw. $x + y$ haben. Dann betrachten wir die Strecken $\overline{SA'}$, $\overline{SB'}$ und $\overline{SC'}$ in Abhängigkeit von x, y bzw. $x + y$, also $f(x)$, $f(y)$ bzw. $f(x + y)$. Dabei seien die Geraden durch A und A', durch B und B' bzw. durch C und C' jeweils parallel zueinander. Die Strecken $\overline{AA'}$, $\overline{BB'}$ bzw. $\overline{CC'}$ betrachten wir ebenfalls in Abhängigkeit von x, y bzw. $x + y$ und bezeichnen die Längen mit $g(x)$, $g(y)$ bzw. $g(x + y)$. Durch Verschiebung des Dreiecks $\triangle SAA'$ längs des Strahls SC' um die Strecke $\overline{SB'}$ ergibt sich $f(x + y) = f(x) + f(y)$ und $g(x + y) = g(x) + g(y)$. Beide Funktionen sind positiv, also gibt es positive reelle Zahlen c und c' mit $f(x) = c \cdot x$ und $g(x) = c' \cdot x$. Damit gilt $\frac{|\overline{SB'}|}{|\overline{SA'}|} = \frac{f(y)}{f(x)} = \frac{y}{x} = \frac{|\overline{SB}|}{|\overline{SA}|} = \frac{g(y)}{g(x)} = \frac{|\overline{BB'}|}{|\overline{AA'}|}$, also die Behauptung des ersten und zweiten Teils des Strahlensatzes[1]); vgl. auch MfL Bd. 7, 3.3.2.

7.7. Aufgaben

1. Man bestimme alle Funktionen $f\colon \mathbf{R} \to \mathbf{R}$, für die folgendes gilt: a) $\bigwedge_{x,y \in \mathbf{R}} f(x + y) = f(x) + f(y)$.

b) Wenn $x > 0$, so $f(x) \geqq 0$. (Hinweis: Man zeige zunächst, daß jede in Frage kommende Funktion monoton ist.)

[1]) In [116] wird der Beweis des ersten Teils des Strahlensatzes mit Hilfe der nicht bewiesenen Flächeninhaltsformel für Dreiecke geführt. Im Gegensatz dazu wird hier deutlich, daß für den Beweis des Strahlensatzes im Rationalen die Additivität von f und g genügt, für den Beweis im Reellen aber eine zusätzliche Forderung notwendig ist.

2. Man bestimme alle streng monotonen Funktionen $f: \mathbf{R} \to \mathbf{R}$, die $f(x+y) + f(x-y) = 2f(x)$ für alle reellen Zahlen x, y erfüllen. (Hinweis: Durch Substitutionen für x und y wird man (fast) auf die Cauchysche Grundgleichung geführt.)

3. a) Man gebe eine stetige Funktion $f: \mathbf{R} \to \mathbf{R}$ mit $f(1) = c > 0$ an, die für alle natürlichen Zahlen $n \geqq 1$ und alle reellen Zahlen x die Gleichung $f(n \cdot x) = n \cdot f(x)$ erfüllt und nicht von der Form $f(x) = c \cdot x$ ist. b) Man zeige, daß $f: \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x$ die einzige stetige Funktion mit $\bigwedge_{n \in \mathbf{N}} \bigwedge_{x \in \mathbf{R}} f(n \cdot x) = n \cdot f(x)$ und $f(1) = c$ sowie $f(-1) = -c$ ist.

4. Man zeige, daß die auf S. 160 definierte Funktion f für $3c > c' > 2c > 0$ streng monoton wachsend ist.

5. Man zeige, daß die Funktionen $f: \mathbf{R} \to \mathbf{R}$ mit $f(x) = c \cdot x^2$ $(c > 0)$ die einzigen für $x \geqq 0$ streng monoton wachsenden Funktionen sind, die für alle reellen Zahlen $f(x+y) + f(x-y) = 2f(x) + 2f(y)$ erfüllen.

6. Man bestimme alle stetigen Funktionen $f: \mathbf{R}_+^* \to \mathbf{R}$, die $f(x+y) = \dfrac{f(x) \cdot f(y)}{f(x) + f(y)}$ für alle positiven reellen Zahlen x und y erfüllen. (Hinweis: Man folge den Beweisschritten von Satz 7.1.1.)

7. Man bestimme alle stetigen Funktionen $f: \mathbf{R} \to \mathbf{R}$, die $f(x+y) = f(x) + f(y) + k \cdot x \cdot y$ für festes $k \in \mathbf{R}$ und alle reellen Zahlen x, y erfüllen. (Hinweis: Man folge den Beweisschritten von Satz 7.1.1.)

8. Funktionale Charakterisierungen der Exponential- und Logarithmusfunktionen mit Hilfe von Funktionalgleichungen

8.1. Funktionale Charakterisierungen der Exponentialfunktionen mit Hilfe der Funktionalgleichung $f(x + y) = f(x) \cdot f(y)$

Wir greifen jetzt die Problemstellung von Kapitel 7 für das Beispiel der Exponentialfunktionen auf. Im Mathematikunterricht der Klasse 9 wird der Potenzbegriff für Potenzen mit natürlichen Exponenten schrittweise auf Potenzen mit ganzzahligen bzw. rationalen Exponenten unter Beachtung der Identität

8.1.1. $a^{p+q} = a^p \cdot a^q$

erweitert. Auf diese Entwicklung stützt sich die Definition der Exponentialfunktionen für alle reellen Zahlen. Dann gilt für alle reellen Zahlen x, y

8.1.2. $a^{x+y} = a^x \cdot a^y \; (a > 0)$.

Das Logarithmieren als eine Umkehrung des Potenzierens liefert dann die Grundlage für die Definition der Logarithmusfunktionen.[1]) Die Exponentialfunktionen $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = a^x$ $(a > 0)$ erfüllen also (8.1.2). Nun ergibt sich z. B. die Frage: Sind die Exponentialfunktionen die einzigen Funktionen, die die Eigenschaft (8.1.2) besitzen? Das ist die Frage, ob die Eigenschaft (8.1.2) typisch für die Exponentialfunktionen ist. Darüber geben die Lösungen der Funktionalgleichung

8.1.3. $\bigwedge_{x,y \in \mathbf{R}} f(x + y) = f(x) \cdot f(y)$

Auskunft.

Die Funktionalgleichung (8.1.3) ergibt sich auch bei dem Versuch, Wachstums- bzw. Zerfallsprozesse mathematisch zu modellieren (vgl. A. Engel [48]).

Es sei w eine Funktion, die einen solchen Prozeß in Abhängigkeit von der Zeit t beschreibt. $w(t)$ ist die Anzahl der Individuen bzw. die Masse der betrachteten Substanz zum Zeitpunkt t. Die Beobachtung des Prozesses hat folgende Gesetzmäßigkeit ergeben: w nimmt in gleichen Zeitintervallen um den gleichen Prozentsatz zu (bzw. ab). Daraus ergibt sich für die gleichlangen Zeitintervalle $[\![0, t]\!]$ und $[\![h, t + h]\!]$

8.1.4. $$\frac{w(t)}{w(0)} = \frac{w(t + h)}{w(h)}$$

oder auch

8.1.5. $$\frac{w(t + h)}{w(0)} = \frac{w(t)}{w(0)} \cdot \frac{w(h)}{w(0)}.$$

[1]) Das heißt nicht, daß das Logarithmieren eine Umkehroperation des Potenzierens im Sinne von 1.3. ist.

Die Funktion $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(t) := \dfrac{w(t)}{w(0)}$ erfüllt dann (8.1.3). Der Versuch der mathematischen Modellierung eines solchen Prozesses führt zu der Notwendigkeit, die Existenz einer streng monotonen Lösung von (8.1.3) zu sichern.

Der folgende Satz, der ebenfalls auf A. L. Cauchy [34] zurückgeht, ist so formuliert, daß die Kenntnis der Exponentialfunktionen nicht vorausgesetzt werden muß.

8.1.6. Satz. *Es gibt genau eine Funktion* $f\colon \mathbf{R} \to \mathbf{R}$, *die der Bedingung* (8.1.3) *und den folgenden Bedingungen genügt:*

8.1.7. *f ist stetig,*

8.1.8. $f(1) = a$ $(a > 0)$.

Beweis (vgl. etwa D. Ilse [74]). Wir betrachten zunächst den Fall $a \neq 1$ und setzen den Potenzbegriff für rationale Exponenten und die in Aufgabe 8.5.1 genannten Eigenschaften als bekannt voraus[1]) und beginnen mit dem Nachweis, daß es höchstens eine Funktion mit den genannten Eigenschaften gibt. Dabei ergibt sich eine Möglichkeit zur Definition dieser Funktion, woraus schließlich auch eine Definition der Potenz mit reellen Exponenten folgt.

Es sei $f\colon \mathbf{R} \to \mathbf{R}$ eine Funktion, die (8.1.3), (8.1.7) und (8.1.8) erfüllt. Dann ergibt sich:

a) *Für alle Argumente* $x \in \mathbf{R}$ *gilt* $f(x) > 0$. Dafür zeigen wir zunächst, daß es keine Zahl x_0 gibt mit $f(x_0) = 0$. Gäbe es eine solche Zahl x_0, so würde für jedes $x \in \mathbf{R}$

$$f(x) = f\big(x_0 + (x - x_0)\big) = f(x_0) \cdot f(x - x_0) = 0$$

gelten im Widerspruch zu (8.1.8). Weiter gilt für jedes $x \in \mathbf{R}$

$$f(x) = f\left(\frac{x}{2} + \frac{x}{2}\right) = f\left(\frac{x}{2}\right) \cdot f\left(\frac{x}{2}\right) = f\left(\frac{x}{2}\right)^2,$$

d. h. $f(x) \geqq 0$. Da f aber keine Nullstelle besitzt, gilt die Behauptung.

b) *Für jede reelle Zahl x und jede natürliche Zahl $k \geqq 1$ gilt* $f(k \cdot x) = f(x)^k$. Der Beweis mittels vollständiger Induktion sei dem Leser überlassen. Für $x = 1$ ergibt sich somit $f(k) = f(1)^k = a^k$.

c) *Für jede reelle Zahl x und jede rationale Zahl r gilt* $f(r \cdot x) = f(x)^r$. Dazu untersuchen wir drei Fälle:

c_1) $r = 0$. Für $x = y = 0$ ergibt sich aus (8.1.3) $f(0) = f(0)^2$, also $f(0) = 1$, da $f(x) > 0$ für alle x gilt, also $f(0) = f(0 \cdot x) = f(x)^0$ in Übereinstimmung mit der Definition $c^0 := 1$ für $c \neq 0$.

c_2) $r > 0$. Es sei $r = \dfrac{p}{q}$ $(p, q \in \mathbf{N}^*)$. Für $y = r \cdot x$ ergibt sich nun $p \cdot x = q \cdot y$ und daher auch $f(p \cdot x) = f(q \cdot y)$, was wegen b) $f(x)^p = f(y)^q$ nach sich zieht. Also ist $f(y) = f(x)^{p/q}$ oder auch $f(r \cdot x) = f(x)^r$.

c_3) $r < 0$. Dann ist $-r > 0$ und $1 = f(0) = f\big(r \cdot x + (-r)\, x\big) = f(r \cdot x) \cdot f(-r \cdot x)$, also $f(r \cdot x) = \dfrac{1}{f(-r \cdot x)} = \dfrac{1}{f(x)^{-r}} = f(x)^r$. Für $x = 1$ ergibt sich $f(r) = f(1)^r = a^r$.

[1]) Das ist auch die Situation, in der die Schüler der Klasse 9 die Exponentialfunktion kennenlernen.

d) *Die Funktion* $f\colon \mathbf{Q} \to \mathbf{R}$ *mit* $f(r) = a^r$ *erfüllt das Additionstheorem* (8.1.3) *und ist streng monoton wachsend für* $a > 1$ *und streng monoton fallend für* $0 < a < 1$.

Diese Behauptungen sind bekannte Gesetze für das Rechnen mit Potenzen mit rationalen Exponenten (siehe Aufgabe 8.5.1).

Die bisherigen Untersuchungen haben gezeigt: Wenn es überhaupt eine für alle rationalen Zahlen definierte Funktion mit den Eigenschaften (8.1.3), (8.1.7) und (8.1.8) gibt, dann ist dies die Funktion $f\colon \mathbf{Q} \to \mathbf{R}$ mit $f(r) = a^r$. Zum Nachweis, daß diese Funktion das Gewünschte leistet, müßte noch ihre Stetigkeit nachgewiesen werden. Das folgt aber aus dem später ohnehin zu führenden Beweis dafür, daß die auf alle reellen Zahlen erweiterte Funktion f mit $f(x) = a^x$ stetig ist. Diese Erweiterung soll nun vorgenommen werden.

e) *Definition der Funktion* $f\colon \mathbf{R} \to \mathbf{R}$ *mit* $f(x) = a^x$.

e_1) Wenn (h_n) eine beliebige Folge rationaler Zahlen ist mit $h_n > 0$ und $\lim\limits_{n\to\infty} h_n = 0$, so ist $\lim\limits_{n\to\infty} f(h_n) = \lim\limits_{n\to\infty} a^{h_n} = 1$.

Fall 1: $h_n = \dfrac{1}{n}$ (siehe MfL Bd. 4, 2.1.3., Beispiel 4).

Fall 2: Es sei jetzt (h_n) eine beliebige Nullfolge rationaler Zahlen mit $h_n > 0$, wobei o. B. d. A. $h_n \leqq 1$ vorausgesetzt werden kann. Wird nun zu jedem n die größte natürliche Zahl k so bestimmt, daß $h_n \leqq \dfrac{1}{k}$ gilt, dann gilt wegen der strengen Monotonie von a^{h_n} (siehe d)) für $a > 1$ auch $a^{h_n} \leqq a^{1/k} = \sqrt[k]{a}$. Da a^{h_n} stets größer als 1 ist, erhält man mit $\lim\limits_{k\to\infty} \sqrt[k]{a} = 1$ auch $\lim\limits_{n\to\infty} a^{h_n} = 1$. Für $0 < a < 1$ kann analog geschlossen werden.

e_2) Nach diesen Vorbereitungen können die Funktionswerte von f an den irrationalen Stellen festgelegt werden. Im Hinblick darauf, daß die erweiterte Funktion stetig sein soll, muß für den Funktionswert $f(x_0)$ an einer solchen Stelle x_0 sicherlich folgendes gelten: Sind (r_n) und (r_n') monoton wachsende bzw. fallende Folgen rationaler Zahlen mit $\lim\limits_{n\to\infty} r_n = \lim\limits_{n\to\infty} r_n' = x_0$, so gilt $\lim\limits_{n\to\infty} f(r_n) = f(x_0)$ und $\lim\limits_{n\to\infty} f(r_n') = f(x_0)$.

Eine solche Definition von $f(x_0)$ ist nun tatsächlich möglich. Dazu seien (r_n) und (r_n') streng monotone Folgen rationaler Zahlen mit $\lim\limits_{n\to\infty} r_n = x_0$ ($r_n < x_0$ für alle n) und $\lim\limits_{n\to\infty} r_n' = x_0$ ($r_n' > x_0$ für alle n), wobei x_0 eine beliebige, aber fest gewählte reelle Zahl ist. Da die zugehörigen Folgen der Funktionswerte $\big(f(r_n)\big)$ und $\big(f(r_n')\big)$ streng monoton (nach d)) und beschränkt ($f(r_1) \leqq f(r_n) < f(r_1')$ bzw. $f(r_1) < f(r_n') \leqq f(r_1')$ für alle n) sind, gibt es Zahlen c und c' mit $\lim\limits_{n\to\infty} f(r_n) = c$ und $\lim\limits_{n\to\infty} f(r_n') = c'$ und $c \leqq c'$ für $a > 1$ bzw. $c \geqq c'$ für $0 < a < 1$. Es ist sogar $c = c'$, denn es gilt

$$c' = \lim_{n\to\infty} f(r_n') = \lim_{n\to\infty} f\big((r_n' - r_n) + r_n\big) = \lim_{n\to\infty} f(r_n' - r_n) \cdot \lim_{n\to\infty} f(r_n) = 1 \cdot c = c$$

wegen (8.1.3), des entsprechenden Grenzwertsatzes und e_1).

Damit kann definiert werden: $f(x_0) = a^{x_0} := c = c'$.

Somit ist gezeigt, daß es höchstens eine Funktion $f\colon \mathbf{R} \to \mathbf{R}$ gibt, die den Bedingungen (8.1.3), (8.1.7) und (8.1.8) genügt. Wir zeigen jetzt, daß die so definierte Funktion f wirklich die Eigenschaften (8.1.3), (8.1.7) und (8.1.8) besitzt.

f) $f\colon \mathbf{R} \to \mathbf{R}$ *mit* $f(x) = a^x$ *erfüllt die Funktionalgleichung* (8.1.3). Es seien x und y beliebige reelle Zahlen sowie (r_n) und (r_n') monoton wachsende Folgen rationaler Zahlen mit $\lim\limits_{n\to\infty} r_n = x$ und $\lim\limits_{n\to\infty} r_n' = y$. Wegen d) gilt dann $a^{r_n + r_n'} = a^{r_n} \cdot a^{r_n'}$ für jedes n. Für $n = 1, 2, \ldots$ erhält man auf der linken und der rechten Seite identische konvergente Folgen, die demnach denselben Grenzwert besitzen:

$$\lim_{n\to\infty} a^{r_n + r_n'} = \lim_{n\to\infty} a^{r_n} \cdot \lim_{n\to\infty} a^{r_n'}, \quad \text{also} \quad a^{x+y} = a^x \cdot a^y.$$

g) $f\colon \mathbf{R} \to \mathbf{R}$ *mit* $f(x) = a^x$ *ist stetig.*

g_1) Wir zeigen zunächst, daß f für $a > 1$ streng monoton wachsend ist. Es seien x und y beliebige reelle Zahlen mit $x < y$ sowie (r_n) und (r_n') Folgen rationaler Zahlen mit $\lim\limits_{n\to\infty} r_n = x$ ($r_n > x$ für alle n) und $\lim\limits_{n\to\infty} r_n' = y$ ($r_n' < y$ für alle n). Wegen $x < y$ gibt es dann eine natürliche Zahl N_0 derart, daß $x < r_n < r_n' < y$ für alle $n > N_0$ gilt. Nun ist offensichtlich $a^x \leqq a^{r_n}$ und $a^{r_n'} \leqq a^y$ für alle natürlichen n und $a^{r_n} < a^{r_n'}$ für $n > N_0$ (wegen d)). Also gilt $a^x \leqq a^{r_n} < a^{r_n'} \leqq a^y$ für $n > N_0$, woraus sich $a^x < a^y$ ergibt. Für $0 < a < 1$ folgt entsprechend, daß f streng monoton fallend ist.

g_2) Es sei x_0 eine beliebige reelle Zahl und (x_n) zunächst eine beliebige Folge reeller Zahlen mit $\lim\limits_{n\to\infty} x_n = x_0$ und $x_n < x_0$ für alle n. Nach f) gilt für beliebige reelle Zahlen die Funktionalgleichung (8.1.3), also gilt $a^{x_0} = a^{x_0 - x_n} \cdot a^{x_n}$. Es genügt daher zu beweisen, daß $\lim a^{x_0 - x_n} = 1$ ist. Dazu kann $x_0 - x_n < 1$ vorausgesetzt und für jedes n das größte natürliche k gewählt werden mit $x_0 - x_n \leqq \frac{1}{k}$. Daraus folgt wegen $1 \leqq a^{x_0 - x_n} \leqq a^{1/k}$ die Behauptung für den genannten Fall, wobei g_1) benutzt wird.

Die Folge (x_n') konvergiere jetzt von rechts gegen das Argument x_0, d. h. $\lim\limits_{n\to\infty} x_n' = x_0$ mit $x_n' > x_0$ für alle n. In diesem Fall gilt ebenfalls $\lim\limits_{n\to\infty} a^{x_n'} = a^{x_0}$, denn aus $a^{-x_0} = a^{x_n' - x_0} \cdot a^{-x_n'}$ folgt entsprechend $\lim\limits_{n\to\infty} a^{-x_n'} = a^{-x_0}$. Das bedeutet aber $\lim\limits_{n\to\infty} \frac{1}{a^{x_n'}} = \frac{1}{a^{x_0}}$, also auch $\lim\limits_{n\to\infty} a^{x_n'} = a^{x_0}$.

h) *Für* $f\colon \mathbf{R} \to \mathbf{R}$ *mit* $f(x) = a^x$ *gilt* $f(1) = a$. Das ergibt sich aus der Definition von a^1.

Man überzeugt sich leicht, daß auch für $a = 1$ die Funktion $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = a^x$ die einzige Funktion ist, die (8.1.3), (8.1.7) und (8.1.8) erfüllt. Damit ist der Beweis von Satz 8.1.6 beendet. Die konstruierte Funktion f wird ***Exponentialfunktion*** mit der Basis a genannt.

Wie der Beweis von Satz 8.1.6 zeigt, hat (8.1.3) im Rationalen nur die Lösungen f mit $f(r) = a^r$ und damit keine unstetigen Lösungen. Weiterhin folgt $f(r \cdot x) = f(x)^r$ für alle rationalen Zahlen r und alle reellen Zahlen x allein aus (8.1.3) und (8.1.8).

Die in Satz 8.1.6 konstruierten stetigen Lösungen von (8.1.3) enthalten die kon-

stante Lösung f mit $f(x) = 1$ für alle reellen Zahlen. Auch $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = 0$ ist eine konstante Lösung von (8.1.3). Damit hat man aber auch alle konstanten Lösungen von (8.1.3) gefunden; denn ist $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = c$ eine konstante Lösung von (8.1.3), dann gilt $c = f(x + y) = f(x) \cdot f(y) = c \cdot c = c^2$, woraus $c = 0$ oder $c = 1$ folgt.

Die in Satz 8.1.6 formulierte Forderung (8.1.7) kann wie in Satz 7.2.1 durch andere Forderungen ersetzt werden. Wir geben einige Beispiele an:

8.1.9. f ist an einer Stelle x_0 stetig (vgl. etwa W. F. Osgood [127]).

8.1.10. f ist streng monoton (vgl. z. B. MfL Bd. 4, 1.4.).

8.1.11. Für alle $x \in \mathbf{R}$ gilt: Wenn $x > 0$, so $f(x) > 1$ (vgl. etwa D. Ilse [74]). (Die Forderung der Positivität an f für positive Argumente recht hier wegen a) im Beweis von Satz 8.1.6 nicht aus.)

8.1.12. f ist über einem Intervall beschränkt (vgl. etwa S. Minetti [118]).

8.1.13. f ist differenzierbar (vgl. etwa P. Rossier [149]).

8.1.14. f ist integrierbar.

Wegen der ausführlichen Betrachtung der entsprechenden Fragestellung in Kapitel 7 und der Möglichkeit der Übertragung der Beweisgedanken auf Satz 8.1.6 verzichten wir auf die Beweise.

Unsere Betrachtungen zeigen also, daß die zu Beginn des Kapitels erwähnten Wachstums- bzw. Zerfallsprozesse durch Exponentialfunktionen beschrieben werden.

Wir geben noch eine Charakterisierung der Exponentialfunktionen mit Hilfe einer Ungleichung anstelle der Stetigkeitsforderung an. Die dabei benutzte Ungleichung bringt den Sachverhalt zum Ausdruck, daß die Exponentialfunktionen stets oberhalb ihrer Tangente im Punkte $(0; 1)$ liegen. Von der Tatsache, daß die Gerade Tangente für die Exponentialfunktionen ist, wird im folgenden jedoch kein Gebrauch gemacht. Es sei c eine beliebige, aber feste reelle Zahl und $f\colon \mathbf{R} \to \mathbf{R}$ eine Funktion, die (8.1.3) und

8.1.15. $\bigwedge_{x \in \mathbf{R}} f(x) \geqq 1 + c \cdot x$

erfüllt. Wir benutzen einige aus Satz 8.1.6 bekannte Eigenschaften, die alle nichtkonstanten Lösungen von (8.1.3) besitzen.

a) $f(x) > 0$ für alle $x \in \mathbf{R}$,

b) $f(0) = 1$,

c) $1 = f(x) \cdot f(-x)$ für alle $x \in \mathbf{R}$,

d) $f(r \cdot x) = f(x)^r$ für alle rationalen Zahlen r und alle reellen Zahlen x, also auch $f\left(\frac{x}{k}\right) = f(x)^{1/k}$ für $k \in \mathbf{N}^*$.

Aus (8.1.15) ergibt sich für $-x$: $f(-x) \geqq 1 - c \cdot x$, also mit c) $f(x) \leqq \dfrac{1}{1 - c \cdot x}$ für alle x mit $c \cdot x < 1$. Insgesamt gilt daher für diese x

8.1.16. Wenn $c \cdot x < 1$, so $1 + c \cdot x \leqq f(x) \leqq \dfrac{1}{1 - c \cdot x}$.

Aus (8.1.16) ergibt sich die Stetigkeit jeder Lösung f an der Stelle Null; denn ist (x_n) eine Folge mit $\lim\limits_{n\to\infty} x_n = 0$, so gilt $\lim\limits_{n\to\infty} (1 + c \cdot x_n) = 1$ und $\lim\limits_{n\to\infty} \dfrac{1}{1 - c \cdot x_n} = 1$. Daher gilt auch $\lim\limits_{n\to\infty} f(x_n) = 1$, also $\lim\limits_{n\to\infty} f(x_n) = f(0)$.

Die Stetigkeit von f für jedes x ergibt sich wie bei (8.1.9).

Aus (8.1.16) ergibt sich nun wie folgt eine Definition für eine stetige Lösung $f: \mathbf{R} \to \mathbf{R}$ von (8.1.3). Wir ersetzen in (8.1.16) x durch $\dfrac{x}{k}$ mit $k \in \mathbf{N}^*$. Für diese Argumente gilt (8.1.16) ebenfalls. Es ergibt sich $1 + \dfrac{c \cdot x}{k} \leqq f\left(\dfrac{x}{k}\right) \leqq \dfrac{1}{1 - \dfrac{c \cdot x}{k}}$ oder auch wegen d) $\left(1 + \dfrac{c \cdot x}{k}\right)^k \leqq f(x) \leqq \dfrac{1}{\left(1 - \dfrac{c \cdot x}{k}\right)^k}$. Für $k \to \infty$ konvergiert $\left(1 + \dfrac{c \cdot x}{k}\right)^k$ bekanntlich gegen $e^{c \cdot x}$ (vgl. MfL Bd. 4, 2.1.5., Satz 2) und $\dfrac{1}{\left(1 - \dfrac{c \cdot x}{k}\right)^k} = \dfrac{1}{\left(1 + \dfrac{-c \cdot x}{k}\right)^k}$ gegen $\dfrac{1}{e^{-c \cdot x}} = e^{c \cdot x}$. Daher ist $f(x) = e^{c \cdot x}$.

Für $a := e^c$ ist $\ln a = c$. Für $c > 0$ ist $a > 1$, und für $c < 0$ ist $0 < a < 1$, während für $c = 0$ dann $a = 1$ ist. Damit ist gezeigt, daß die Funktion $f: \mathbf{R} \to \mathbf{R}$ mit $f(x) = e^{c \cdot x}$ die einzige Funktion ist, die den Bedingungen (8.1.3) und (8.1.15) genügt.

Setzt man die Kenntnis der Exponential- und Logarithmusfunktionen voraus, so ergibt sich ein Beweis von Satz 8.1.6 durch Zurückführung auf Satz 7.1.1. Da jede nichtkonstante Funktion $f: \mathbf{R} \to \mathbf{R}$, für die (8.1.3) gilt, nur positive Funktionswerte aufweist, kann bezüglich einer Basis a ($a \in \mathbf{R}_+^*$, $a \neq 1$) logarithmiert werden, und es ergibt sich

$$\log_a f(x + y) = \log_a \big(f(x) \cdot f(y)\big) = \log_a f(x) + \log_a f(y).$$

Für $g: \mathbf{R} \to \mathbf{R}$ mit $g(x) := \log_a f(x)$ gilt somit $g(x + y) = g(x) + g(y)$. Mit f ist auch g stetig, und es gilt $g(1) = \log_a f(1) = \log_a a = 1$. Also ist $g(x) = x$, d. h. $x = \log_a f(x)$ oder $f(x) = a^x$.

Wegen $x^v := a^{v \cdot \log_a x}$ ($v \in \mathbf{R}$, $x \in \mathbf{R}_+^*$, $a \in \mathbf{R}_+^*$, $a \neq 1$) kann aus Satz 8.1.6 eine Charakterisierung der Potenzfunktionen $f: \mathbf{R}_+^* \to \mathbf{R}$, die $\bigwedge\limits_{x,y \in \mathbf{R}^*} f(x \cdot y) = f(x) \cdot f(y)$ genügen, erhalten werden, die dann auf Potenzfunktionen mit $D(f) = \mathbf{R}$ bzw. $D(f) = \mathbf{R}^*$ ausgedehnt werden können (vgl. etwa D. Ilse [74]).

8.2. Die unstetigen Lösungen der Funktionalgleichung $f(x+y)=f(x)\cdot f(y)$

8.2.1. Satz. *Die Funktionalgleichung $f(x+y)=f(x)\cdot f(y)$ hat unstetige Lösungen.*

Beweis. Es sei $\mathbf{R}_B$ eine Hamelbasis der reellen Zahlen. Dann wird f für die Basiselemente durch beliebige positive reelle Zahlen definiert. Ist x eine beliebige reelle Zahl und $x=\sum_{\nu=1}^{m} r_\nu b_\nu$ ihre Darstellung durch Basiselemente, so wird $f(x):=\prod_{\nu=1}^{m} f(b_\nu)^{r_\nu}$ gesetzt. Die so definierte Funktion $f\colon \mathbf{R}\to\mathbf{R}$ ist Lösung von (8.1.3). Sind nämlich x und y reelle Zahlen mit $x=\sum_{\nu=1}^{m} r_\nu b_\nu$ und $y=\sum_{\nu=1}^{m} r_\nu' b_\nu$, so gilt

$$f(x+y)=\prod_{\nu=1}^{m} f(b_\nu)^{r_\nu+r_\nu'}=\prod_{\nu=1}^{m} f(b_\nu)^{r_\nu}\cdot\prod_{\nu=1}^{m} f(b_\nu)^{r_\nu'}=f(x)\cdot f(y).$$

f ist genau dann unstetig, wenn für mindestens ein Basiselement b_ν der Funktionswert $f(b_\nu)\neq f(1)^{b_\nu}$ definiert wird.

Auf die angegebene Weise erhält man alle Lösungen von (8.1.3), wie man sich durch entsprechende Überlegungen wie nach Satz 7.3.4 verdeutlicht.

Der folgende Satz, der zuerst von E. B. Wilson 1899 [180] bewiesen wurde, beschreibt eine Eigenschaft unstetiger Lösungen von (8.1.3).

8.2.2. Satz. *Wenn $f\colon \mathbf{R}\to\mathbf{R}$ eine unstetige Lösung von $f(x+y)=f(x)\cdot f(y)$ ist, dann liegt in jeder Umgebung eines jeden Punktes der oberen Halbebene ein Punkt des Graphen von f.*

Beweis. Es sei $f\colon \mathbf{R}\to\mathbf{R}$ eine unstetige Lösung der Gleichung (8.1.3). Die Funktion $g\colon \mathbf{R}\to\mathbf{R}$ mit $g(x):=\dfrac{f(x)}{f(1)^x}$ hat folgende Eigenschaften:

a) g ist Lösung von (8.1.3), wie man sofort nachrechnet.

b) Wenn $x\in\mathbf{Q}$, dann $g(x)=1$, denn für $r\in\mathbf{Q}$ gilt $f(r)=f(1)^r$.

c) Wenn $a, b\in\mathbf{Q}$, so $g(a\cdot x+b)=g(x)^a$ für beliebige $x\in\mathbf{R}$.

Aus $g(a\cdot x+b)=g(a\cdot x)\cdot g(b)=g(x)^a\cdot g(b)$ ergibt sich wegen b) die Behauptung.

d) Es gibt ein $x'\in\mathbf{R}$ mit $g(x')\neq 1$; anderenfalls wäre $f(x)=f(1)^x$ für alle $x\in\mathbf{R}$ und f damit stetig.

Es sei nun $(x_0; y_0)$ ein beliebiger Punkt der oberen Halbebene. Dann kann wegen $y_0>0$ und $g(x')\neq 1$ zu jedem $\varepsilon>0$ eine rationale Zahl a so gewählt werden, daß $|g(x')^a-y_0|<\varepsilon$ gilt. Weiterhin kann eine rationale Zahl b so gewählt werden, daß $|(a\cdot x'+b)-x_0|<\varepsilon$ ist. Damit liegt die Zahl $a\cdot x'+b$ in der ε-Umgebung von x_0, und $g(a\cdot x'+b)$ liegt wegen c) in der ε-Umgebung von y_0. Damit kommt der Graph von g jedem Punkt der oberen Halbebene beliebig nahe. Wegen $f(x)=g(x)\cdot f(1)^x$ überträgt sich diese Eigenschaft von g auf f, denn $f(1)^x$ ist stetig und immer von Null verschieden.

8.3. Funktionale Charakterisierungen der Logarithmusfunktionen mit Hilfe der Funktionalgleichung $f(x \cdot y) = f(x) + f(y)$

Da die Logarithmusfunktionen Umkehrfunktionen entsprechender Exponentialfunktionen sind, ergibt sich aus (8.1.3) für die Logarithmusfunktionen

8.3.1. $\bigwedge_{x,y \in \mathbf{R}_+^*} f(x \cdot y) = f(x) + f(y)$.

Weiterhin gilt:

8.3.2. Satz. *Es gibt genau eine Funktion* $f: \mathbf{R}_+^* \to \mathbf{R}$, *die* (8.3.1) *sowie den folgenden Bedingungen genügt:*

8.3.3. f *ist stetig,*

8.3.4. $f(a) = 1$ $(a \in \mathbf{R}_+^*, a \neq 1)$.

Beweis (vgl. etwa D. Ilse [74]). Für jede Lösung $f: \mathbf{R}_+^* \to \mathbf{R}$ von (8.3.1), (8.3.3) und (8.3.4) gilt:

a) $f(1) = 0$. Für $x = y = 1$ in (8.3.1) ergibt sich die Behauptung.

b) *Für jede positive reelle Zahl* x *und für jede rationale Zahl* r *gilt* $f(x^r) = r \cdot f(x)$, insbesondere gilt $f(a^r) = r$.

Es sei $x > 0$ eine reelle Zahl. Dann gilt

b_1) $f(x^m) = m \cdot f(x)$ *für alle natürlichen Zahlen* m. Der Beweis wird mit vollständiger Induktion geführt.

b_2) $f(x^k) = k \cdot f(x)$ *für alle ganzen Zahlen* k. Es sei $k < 0$, dann ist $-k > 0$. Nun gilt $0 = f(1) = f(x^k \cdot x^{-k}) = f(x^k) + f(x^{-k})$, also $f(x^k) = -f(x^{-k}) = -(-k) \cdot f(x) = k \cdot f(x)$.

b_3) $f\left(\sqrt[m]{x}\right) = \frac{1}{m} \cdot f(x)$ *für alle natürlichen Zahlen* $m \geqq 2$. Man setzt in b_1) $\sqrt[m]{x}$ für x ein und erhält $f\left(\left(\sqrt[m]{x}\right)^m\right) = f(x) = m \cdot f\left(\sqrt[m]{x}\right)$, also $f\left(\sqrt[m]{x}\right) = \frac{1}{m} \cdot f(x)$.

b_4) $f(x^r) = r \cdot f(x)$ *für alle rationalen Zahlen* r. Es sei $r = \frac{p}{q}$ mit p, q ganzzahlig und $q \geqq 2$. Dann gilt $f(x^r) = f(x^{p/q}) = f\left((x^{1/q})^p\right) = p \cdot f(x^{1/q}) = \frac{p}{q} \cdot f(x) = r \cdot f(x)$. Insbesondere ergibt sich für $x = a$ somit $f(a^r) = r$.

c) *Für jede reelle Zahl* x *gilt* $f(a^x) = x$. Da die Funktion $g: \mathbf{R} \to \mathbf{R}$ mit $g(x) = a^x$ $(a \in \mathbf{R}_+^*, a \neq 1)$ stetig und streng monoton ist, gilt für jede Folge (x_n) mit $\lim_{n \to \infty} x_n = x$ auch $\lim_{n \to \infty} a^{x_n} = a^x$, und umgekehrt folgt aus $\lim_{n \to \infty} a^{x_n} = a^x$ auch $\lim_{n \to \infty} x_n = x$. Also ist f mit $f(a^x) = x$ $(x \in \mathbf{R})$ die einzige stetige Fortsetzung von f mit $f(a^r) = r$ $(r \in \mathbf{Q})$, so daß nur diese Funktion f die Bedingungen (8.3.1), (8.3.3) und (8.3.4) erfüllen kann. Daß f diese Eigenschaften auch wirklich besitzt, erkennt man wie folgt: Die Funktion f ist die inverse Funktion oder Umkehrfunktion der Funktion g, so daß für die Funktion f mit $a^x = y$ die Beziehung $a^{f(y)} = a^x = y$ gilt. Es handelt sich also um die Funktion $f: \mathbf{R}_+^* \to \mathbf{R}$ mit $f(x) = \log_a x$. Da f Umkehrfunktion von g ist, ergibt sich für f die Gültigkeit von (8.3.1), (8.3.3) und (8.3.4) aus der Gültigkeit von (8.1.3), (8.1.7) und (8.1.8) für g.

Eine Möglichkeit der Charakterisierung der Logarithmusfunktionen ohne Bezugnahme auf die Exponentialfunktionen liefert

8.3.5. Satz. *Zu jeder reellen Zahl a mit $a > 1$ gibt es genau eine Funktion $f\colon \mathbf{R}_+^* \to \mathbf{R}$, für die neben* (8.3.1) *die folgenden Bedingungen gelten:*

8.3.6. *f ist streng monoton wachsend,*

8.3.7. $f(a) = 1$.

Beweis (vgl. J. DIEUDONNÉ [45]). Wir zeigen zunächst, daß es höchstens eine Funktion $f\colon \mathbf{R}_+^* \to \mathbf{R}$ gibt, die den Bedingungen (8.3.1), (8.3.6) und (8.3.7) genügt. Im Verlaufe dieses Beweises ergibt sich eine Konstruktion für eine Funktion f, mit der dann die Existenz der behaupteten Funktion nachgewiesen wird. Für jede Funktion f, die (8.3.1), (8.3.6) und (8.3.7) erfüllt, gilt:

a) $f(1) = 0$ (siehe a) im Beweis von Satz 8.3.2).

b) *Für jede positive reelle Zahl x und jede ganze Zahl k gilt* $f(x^k) = k \cdot f(x)$ (siehe b_1), b_2) im Beweis von Satz 8.3.2).

c) *Für jede positive reelle Zahl x läßt sich eine Menge von Intervallen* $\left[\!\left[\frac{k_m}{m}, \frac{k_m}{m} + \frac{1}{m}\right]\!\right]$ *konstruieren, deren Durchschnitt höchstens eine Zahl enthält.* Da $a > 1$ ist, ist die Folge (a^k) für $k \to \infty$ unbeschränkt wachsend und die Folge (a^{-k}) ist eine Nullfolge. Daher gibt es zu jeder Zahl x^m ($x \in \mathbf{R}_+^*$, $m \in \mathbf{N}^*$) eine ganze Zahl k_m mit $a^{k_m} \leqq x^m \leqq a^{k_m+1}$. Wegen (8.3.6) ergibt sich daraus $f(a^{k_m}) \leqq f(x^m) \leqq f(a^{k_m+1})$. Wegen b) und (8.3.7) gilt dann auch $k_m \leqq m \cdot f(x) \leqq k_m + 1$ und schließlich $\frac{k_m}{m} \leqq f(x) \leqq \frac{k_m + 1}{m}$. Falls es also überhaupt eine derartige Funktion f mit den oben genannten Eigenschaften gibt, so gilt $f(x) \in \left[\!\left[\frac{k_m}{m}, \frac{k_m}{m} + \frac{1}{m}\right]\!\right]$ für jedes positive x und für jede natürliche Zahl $m > 0$. Da die Intervallänge $\frac{1}{m}$ beträgt, ist durch den Durchschnitt $\bigcap_{m \in \mathbf{N}^*} \left[\!\left[\frac{k_m}{m}, \frac{k_m}{m} + \frac{1}{m}\right]\!\right]$ höchstens eine Zahl erfaßt, wodurch die Eindeutigkeit von f gezeigt ist. Mit Hilfe der Intervalle $\left[\!\left[\frac{k_m}{m}, \frac{k_m}{m} + \frac{1}{m}\right]\!\right]$ wird nun eine Intervallschachtelung (vgl. MfL Bd. 4, 2.1.4.) definiert, die zur Definition der Funktion f führt.

d) *Definition einer Funktion $f\colon \mathbf{R}_+^* \to \mathbf{R}$.* Da die Folge $\left(\frac{k_m}{m}\right)$ nicht monoton wachsend sein muß, bildet die Menge der Intervalle $\left[\!\left[\frac{k_m}{m}, \frac{k_m}{m} + \frac{1}{m}\right]\!\right]$ nicht immer eine Intervallschachtelung. Wir definieren daher

$$(r_m \mid r_m') := \begin{cases} [\![r_1, r_1']\!] = [\![k_1, k_1 + 1]\!], \\ [\![r_m, r_m']\!] = [\![r_{m-1}, r_{m-1}']\!] \cap \left[\!\left[\frac{k_m}{m}, \frac{k_m}{m} + \frac{1}{m}\right]\!\right]. \end{cases}$$

Da $r_m \leqq r_{m+1}, r_m' \geqq r_{m+1}', r_m \leqq r_m'$ und $r_m' - r_m \leqq \frac{1}{m}$ für alle $m \in \mathbf{N}^*$ gilt, ist die Folge der Intervalle $[\![r_m, r_m']\!]$ eine Intervallschachtelung. Daher läßt sich bei fest gewähltem $a > 1$ jeder positiven Zahl x durch $(r_m \mid r_m')$ eine Zahl zuordnen, die $f(x)$ genannt wird[1]). Dabei gilt

8.3.8. $$\frac{k_m}{m} \leqq r_m \leqq f(x) \leqq r_m' \leqq \frac{k_m}{m} + \frac{1}{m}.$$

Wir zeigen nun, daß die so definierte Funktion f die geforderten Bedingungen erfüllt.

e) *f erfüllt die Funktionalgleichung* (8.3.1). Es seien x und y positive reelle Zahlen. Dann existiert für jede natürliche Zahl m eine ganze Zahl k_m und eine ganze Zahl p_m mit $a^{k_m} \leqq x^m \leqq a^{k_m+1}$ bzw. $a^{p_m} \leqq y^m \leqq a^{p_m+1}$. Durch Multiplikation ergibt sich daraus $a^{k_m+p_m} \leqq (x \cdot y)^m \leqq a^{k_m+p_m+2}$. Wegen der Definition von f und (8.3.8) gilt dann auch

$$\frac{k_m}{m} \leqq f(x) \leqq \frac{k_m}{m} + \frac{1}{m}, \quad \frac{p_m}{m} \leqq f(y) \leqq \frac{p_m}{m} + \frac{1}{m}$$

und

$$\frac{k_m + p_m}{m} \leqq f(x \cdot y) \leqq \frac{k_m + p_m + 2}{m}.$$

Zusammengefaßt ergibt sich schließlich $-\frac{2}{m} \leqq f(x) + f(y) - f(x \cdot y) \leqq \frac{2}{m}$ für jedes $m \in \mathbf{N}^*$, woraus sich (8.3.1) ergibt.

f) *f erfüllt die Bedingung* (8.3.6). Zunächst zeigen wir: Wenn $x > 1$, so $f(x) > 0$. Zu jedem $x > 1$ gibt es eine natürliche Zahl m, so daß $a < x^m$ gilt. Nun sei k diejenige natürliche Zahl, für die $a^k \leqq x^m \leqq a^{k+1}$ gilt. Nach Definition von f gilt dann $\frac{k}{m} \leqq f(x)$, und $f(x)$ ist positiv.

Es seien nun mit x und y zwei reelle Zahlen gegeben, für die $0 < x < y$ gilt. Dann existiert eine Zahl $z > 1$ mit $x \cdot z = y$. Dann gilt wegen (8.3.1) $f(y) = f(x \cdot z) = f(x) + f(z) > f(x)$, denn $f(z) > 0$.

g) *f erfüllt die Bedingung* (8.3.7). Aus der Konstruktion von f ergibt sich für $f(a)$ die Ungleichung $\frac{k_m}{m} \leqq f(a) \leqq \frac{k_m}{m} + \frac{1}{m}$ für alle $m \in \mathbf{N}^*$. Wegen $x = a$ ergibt sich $k_m = m$ und somit $1 \leqq f(a) \leqq 1 + \frac{1}{m}$ für alle $m \in \mathbf{N}^*$, woraus $f(a) = 1$ folgt.

Damit ist Satz 8.3.5 bewiesen, und die konstruierte Funktion $f: \mathbf{R}_+^* \to \mathbf{R}$ heißt *Logarithmusfunktion* zur Basis a. Für die streng monoton fallenden Lösungen von (8.3.1) sei auf die Aufgabe 8.5.3 verwiesen.

[1]) Die hier gegebene Definition von f ermöglicht auch das näherungsweise Berechnen von Funktionswerten der vorliegenden Logarithmusfunktion (siehe Aufgabe 8.5.8).

Einen Zugang zur Funktion ln liefert

8.3.9. Satz. *Es gibt genau eine Funktion* $f\colon \mathbf{R}_+^* \to \mathbf{R}$, *die die Bedingungen* (8.3.1) *und*

$$\textbf{8.3.10.}\quad \bigwedge_{x\in\mathbf{R}_+^*} f(x) \leqq x - 1$$

erfüllt.

Beweis (vgl. M. BARNER [19]). Wir nehmen zunächst an, daß es eine solche Funktion f gibt. Für jede Funktion f, die (8.3.1) erfüllt, gilt:

a) $f(1) = 0$ (vgl. a) im Beweis von Satz 8.3.2).

b) $f\left(\frac{1}{x}\right) = -f(x)$ *für jedes positive* x, wegen $1 = x \cdot \frac{1}{x}$ und a).

c) *Für jede positive reelle Zahl* x *und jede natürliche Zahl* m *gilt* $f(x^m) = m \cdot f(x)$ (vgl. b_1) im Beweis von Satz 8.3.2).

d) *Für jede positive reelle Zahl* x *und jede natürliche Zahl* $m \geqq 2$ *gilt die Beziehung* $f(x^{1/m}) = \frac{1}{m} \cdot f(x)$ (vgl. b_3) im Beweis von Satz 8.3.2).

e) *Für jede Funktion* f, *die* (8.3.1) *und* (8.3.10) *erfüllt, gilt*

$$\textbf{8.3.11.}\quad \bigwedge_{x\in\mathbf{R}_+^*} 1 - \frac{1}{x} \leqq f(x) \leqq x - 1.$$

Dazu ersetzen wir in (8.3.10) x durch $\frac{1}{x}$ und benutzen b).

f) *Definition einer Funktion* $f\colon \mathbf{R}_+^* \to \mathbf{R}$. Da (8.3.11) für alle positiven reellen Zahlen gilt, gilt (8.3.11) auch für $\sqrt[2^m]{x}$ ($m \in \mathbf{N}^*$), also

$$1 - \frac{1}{\sqrt[2^m]{x}} \leqq f\left(\sqrt[2^m]{x}\right) \leqq \sqrt[2^m]{x} - 1.$$

Wegen d) gilt $f\left(\sqrt[2^m]{x}\right) = \frac{1}{2^m} \cdot f(x)$, also

$$1 - \frac{1}{\sqrt[2^m]{x}} \leqq \frac{1}{2^m} \cdot f(x) \leqq \sqrt[2^m]{x} - 1.$$

Multiplikation mit 2^m liefert

$$2^m\left(1 - \frac{1}{\sqrt[2^m]{x}}\right) \leqq f(x) \leqq 2^m\left(\sqrt[2^m]{x} - 1\right).$$

Wir betrachten nun die Folgen (a_m) und (b_m) mit $a_m = 2^m\left(1 - \frac{1}{\sqrt[2^m]{x}}\right)$ und $b_m = 2^m\left(\sqrt[2^m]{x} - 1\right)$. Wir setzen $\sqrt[2^m]{x} =: z$ und deshalb $\sqrt[2^{m-1}]{x} = z^2$ und erhalten

$$a_m - a_{m-1} = 2^m\left(1 - \frac{1}{z}\right) - 2^{m-1}\left(1 - \frac{1}{z^2}\right) = 2^{m-1}\left(2 - \frac{2}{z} - 1 + \frac{1}{z^2}\right)$$

$$= 2^{m-1}\left(1 - \frac{2}{z} + \frac{1}{z^2}\right) = 2^{m-1}\left(\frac{1}{z} - 1\right)^2 \geqq 0.$$

Daher ist (a_m) monoton wachsend. Entsprechend ergibt sich

$$b_{m-1} - b_m = 2^{m-1}(z^2 - 1) - 2^m(z - 1) = 2^{m-1}(z^2 - 1 - 2z + 2)$$

$$= 2^{m-1}(z^2 - 2z + 1) = 2^{m-1}(z - 1)^2 \geqq 0.$$

Daher ist (b_m) monoton fallend. Wegen

$$b_m - a_m = 2^m(z - 1) - 2^m\left(1 - \frac{1}{z}\right) = 2^m\left(z - 1 - 1 + \frac{1}{z}\right)$$

$$= 2^m\left(z + \frac{1}{z} - 2\right) = 2^m\left(\frac{z^2 + 1 - 2z}{z}\right) = 2^m \cdot \frac{1}{z} \cdot (z - 1)^2 \geqq 0$$

ist weiterhin $a_m \leqq b_m$ für alle $m \in \mathbf{N}^*$. Daher ist b_1 eine obere Schranke von (a_m) und a_1 eine untere Schranke von (b_m), folglich sind die Folgen (a_m) und (b_m) konvergent. Es ergibt sich $\lim\limits_{m\to\infty} a_m = d$ und $\lim\limits_{m\to\infty} b_m = d'$, und es gilt $d \leqq d'$. Nun ist $b_m = a_m \cdot \sqrt[2^m]{x}$ und $\lim\limits_{m\to\infty} \sqrt[2^m]{x} = 1$ (vgl. MfL Bd. 4, 2.1.3., Beispiel 4). Also ergibt sich $d' = \lim\limits_{m\to\infty} b_m = \lim\limits_{n\to\infty}\left(a_m \cdot \sqrt[2^m]{x}\right) = d \cdot 1 = d$. Daher können wir definieren $f(x) := d = d'$.

Damit ist bewiesen, daß es höchstens eine Funktion $f\colon \mathbf{R}_+^* \to \mathbf{R}$ gibt, die (8.3.1) und (8.3.10) erfüllt. Nun beweisen wir

g) *Die in* f) *definierte Funktion erfüllt die Bedingung* (8.3.1). Wegen

$$2^m\left((x \cdot y)^{1/2^m} - 1\right) = 2^m(x^{1/2^m} \cdot y^{1/2^m} - 1) = 2^m(x^{1/2^m} - 1)\, y^{1/2^m} + 2^m(y^{1/2^m} - 1),$$

$$\lim_{m\to\infty} 2^m\left((x \cdot y)^{1/2^m} - 1\right) = f(x \cdot y), \quad \lim_{m\to\infty} 2^m(y^{1/2^m} - 1) = f(y)$$

und

$$\lim_{m\to\infty} 2^m(x^{1/2^m} - 1)\, y^{1/2^m} = f(x) \cdot 1$$

ergibt sich $f(x \cdot y) = f(x) + f(y)$.

h) *Die in* f) *definierte Funktion erfüllt die Bedingung* (8.3.10). Wegen

$$f(x) \leqq b_m \leqq b_1 = 2\left(\sqrt{x} - 1\right) = 2\left(\sqrt{x} - 1\right) + (x - 1) - (x - 1)$$

$$= x - 1 - \left(x - 2\sqrt{x} + 1\right) = (x - 1) - \left(\sqrt{x} - 1\right)^2 \leqq x - 1$$

gilt (8.3.10).

Damit ist der Satz 8.3.9 bewiesen.

Man kann weiterhin zeigen, daß die in Satz 8.3.9 konstruierte Funktion f stetig ist und $f(e) = 1$ gilt. Damit ist nach Satz 8.3.2 $f = \log_e = \ln$.

Wird die Bedingung (8.3.10) entsprechend (8.1.15) formuliert, so ergibt sich eine Charakterisierung für alle Logarithmusfunktionen (siehe Aufgabe 8.5.7).

Auf weitere Charakterisierungen der Logarithmusfunktionen gehen wir hier nicht ein und verweisen auf J. RÄTZ [146].

8.3.12. Bemerkung. Die stetigen Lösungen $f: \mathbf{R}_+^* \to \mathbf{R}$ von $f(x \cdot y) = f(x) + f(y)$ sind die Logarithmusfunktionen mit $f(x) = \log_a x$ ($a \in \mathbf{R}_+^*$, $a \neq 1$), die bekanntlich auch differenzierbar sind und die Ableitung f' mit $f'(x) = \frac{k}{x}$ haben. Dieser Sachverhalt gibt die Möglichkeit, die Logarithmusfunktionen als bestimmtes Integral mit variabler oberer Grenze zu definieren.

8.4. Zusammenfassung

Wir können feststellen, daß die stetigen Lösungen der Funktionalgleichungen (8.1.3) und (8.3.1) mit den differenzierbaren Lösungen dieser Gleichungen übereinstimmen.

Jede nichtkonstante stetige Lösung von (8.1.3) ist streng monoton, daher vermittelt jede Exponentialfunktion einen Isomorphismus der Gruppe $(\mathbf{R}, +)$ auf die Gruppe $(\mathbf{R}_+^*, \cdot)$ (vgl. MfL Bd. 3, 12.3.). Jede nichtkonstante stetige Lösung von (8.3.1), also jede Logarithmusfunktion, vermittelt einen Isomorphismus von $(\mathbf{R}_+^*, \cdot)$ auf $(\mathbf{R}, +)$.[1])

Die zuletzt genannte Tatsache ermöglicht die Konstruktion eines Rechenstabes für die Multiplikation.

Für die Einführung der Exponential- und Logarithmusfunktionen ergeben sich aus den bisherigen Ausführungen zwei prinzipielle Wege, die auch in der Literatur diskutiert worden sind:

a) ein algebraischer Weg, bei dem die schrittweise Erweiterung des Potenzbegriffes die tragende Rolle bei der Definition der Exponentialfunktionen spielt und die Logarithmusfunktionen als deren Umkehrfunktionen gewonnen werden (vgl. Satz 8.1.6 und Satz 8.3.2).

b) ein analytischer Weg, bei dem die Logarithmusfunktionen als Flächeninhaltsfunktionen definiert werden und sich die Exponentialfunktionen als deren Umkehrfunktionen ergeben (vgl. Bemerkung 8.3.12).

Beide Wege sind mathematisch gleichwertig, da sie zu denselben Funktionen führen. Sie sind auch auf ihre Realisierbarkeit im Mathematikunterricht vielfältig untersucht worden. Im folgenden wollen wir einen mathematischen Gesichtspunkt der Gegenüberstellung der beiden Wege weiter verfolgen. Dabei wird sich in Kapitel 9 zeigen, daß bei entsprechender Verallgemeinerung der Problemstellung diese Gleichwertigkeit verlorengeht.

[1]) Vgl. auch 3.3., speziell Beispiel 3.3.10.3, aber auch Aufgabe 3.3.16.4.

8.5. Aufgaben

1. Man zeige
a) $a^k \cdot b^k = (a \cdot b)^k$ für $a, b \in \mathbf{R}^*$ und $k \in \mathbf{N}$.
b) $(a^k)^l = a^{k \cdot l}$ für $a \in \mathbf{R}^*$ und $k \in \mathbf{N}$, $l \in \mathbf{Z}$.
c) $a^{\frac{k}{l}} = a^{\frac{k \cdot m}{l \cdot m}}$ für $a \in \mathbf{R}_+^*$ und $k \in \mathbf{Z}$, $l, m \in \mathbf{N}^*$.
d) $a^k \cdot a^l = a^{k+l}$ für $a \in \mathbf{R}^*$ und $k, l \in \mathbf{Z}$ sowie für $a \in \mathbf{R}_+^*$ und $k, l \in \mathbf{Q}$.
e) Für beliebige rationale Zahlen k, l mit $k < l$ gilt $a^k < a^l$ für $a > 1$ und $a^k > a^l$ für $0 < a < 1$.

2. Man zeige durch Rückführung auf die Funktionalgleichung

$$\bigwedge_{x,y \in \mathbf{R}} f(x + y) = f(x) + f(y),$$

daß die einzigen stetigen Lösungen der Funktionalgleichung

$$\bigwedge_{x,y \in \mathbf{R}_+^*} f(x \cdot y) = f(x) + f(y)$$

mit $f(a) = 1$ ($a \in \mathbf{R}_+^*$, $a \neq 1$) die Funktionen $f: \mathbf{R}_+^* \to \mathbf{R}$ mit $f(x) = \log_a x$ sind. (Hinweis: Man setze $x = a^v$ und $y = a^w$.)

3. Man zeige: Zu jeder reellen Zahl a mit $0 < a < 1$ gibt es genau eine Funktion $f: \mathbf{R}_+^* \to \mathbf{R}$ mit folgenden Eigenschaften:
a) $\bigwedge_{x,y \in \mathbf{R}_+^*} f(x \cdot y) = f(x) + f(y)$, b) f ist streng monoton fallend, c) $f(a) = 1$. (Hinweis: Vgl. Satz 8.3.5.)

4. Man zeige, daß der Körper der reellen Zahlen nur den identischen Automorphismus besitzt. (Hinweis: Man zeige, daß jede Lösung $f: \mathbf{R}_+^* \to \mathbf{R}$ von $f(x \cdot y) = f(x) \cdot f(y)$ nur nichtnegative Funktionswerte besitzt.)

5. a) Man zeige, daß (1) $\bigwedge_{m \in \mathbf{N}} \bigwedge_{x \in \mathbf{R}} f(mx) = f(x)^m$ eine Abschwächung von (0) $\bigwedge_{x,y \in \mathbf{R}} f(x + y) = f(x) \cdot f(y)$ ist.
b) Bekanntlich (vgl. Aufgabe 6.2.1) ist (2) $\bigwedge_{x \in \mathbf{R}} f(-x) = \frac{1}{f(x)}$ ebenfalls eine Abschwächung von (0). Man zeige, daß (1) und (2) unabhängig sind.
c) Man zeige: Die einzige differenzierbare Lösung $f: \mathbf{R} \to \mathbf{R}$ von (1) und (2) mit $f(1) = a$ ($a > 0$, $a \neq 1$) ist $f(x) = a^x$.

6. Man zeige: Jede über $\mathbf{R}$ bzw. $\mathbf{R}_+^*$ definierte und über dem Intervall $[\![1, 2]\!]$ integrierbare Lösung von $f(x + y) = f(x) \cdot f(y)$ bzw. $f(x \cdot y) = f(x) + f(y)$ ist differenzierbar. (Hinweis: Man vergleiche mit der entsprechenden Fragestellung in Kapitel 7.)

7. Man zeige: Zu jedem $c > 0$ gibt es genau eine Funktion $f: \mathbf{R}_+^* \to \mathbf{R}$, die die Bedingungen
a) $\bigwedge_{x,y \in \mathbf{R}_+^*} f(x \cdot y) = f(x) + f(y)$, b) $\bigwedge_{x \in \mathbf{R}_+^*} f(x) \leqq c(x - 1)$ erfüllt. (Hinweis: Vgl. Satz 8.3.9.)

8. Man berechne mit Hilfe der Intervallschachtelung aus Satz 8.3.5 $f(5)$ für $a = 10$, also $\log_{10} 5 = \lg 5$ und $f(3)$ für $a = 2$, also $\log_2 3$ auf zwei Stellen genau.

9. Man zeige, daß die Gruppen $(\mathbf{Q}_+^*, \cdot)$ und $(\mathbf{Q}, +)$ nicht isomorph sind.

9. Verallgemeinerung der Exponential- und Logarithmusfunktionen unter Beibehaltung der Stetigkeitsforderung

Eine Analyse des Beweises von Satz 8.1.6 zeigt, daß zur Konstruktion der Exponentialfunktionen und ihrer Umkehrfunktionen nur benutzt wurde, daß die Multiplikation positiver reeller Zahlen eine stetige kommutative Gruppenoperation ist. Dies legt die Vermutung nahe, daß sich aus jeder zweistelligen Funktion F, die in einer gewissen Menge A eine stetige kommutative Gruppenoperation ist, einstellige Funktionen $f\colon \mathbf{R} \to A$ konstruieren lassen, die einen Isomorphismus zwischen der additiven Gruppe der reellen Zahlen und der Menge A mit der Operation F vermitteln, für die also

$$\bigwedge_{x,y\in\mathbf{R}} f(x+y) = F\big(f(x), f(y)\big)$$

gilt. Solche Funktionen würden auch die Konstruktion eines Rechenstabes zulassen (vgl. 11.1.).

Wir werden sehen, daß für Funktionen $F\colon A\times A \to A$ mit den genannten Eigenschaften Funktionen $f\colon \mathbf{R} \to A$ existieren und daß diese in völliger Analogie zur Konstruktion der Exponentialfunktionen (d. h. durch schrittweise Erweiterung des Potenzbegriffes) definiert werden können. Daher werden wir von *verallgemeinerten Exponentialfunktionen* sprechen.

Aus der Funktionalgleichung $f(x\cdot y) = f(x) + f(y)$ läßt sich durch eine entsprechende Verallgemeinerung die Funktionalgleichung

$$\bigwedge_{x,y\in A} f\big(F(x,y)\big) = f(x) + f(y)$$

gewinnen. Diese Funktionalgleichung werden wir im Zusammenhang mit der Verallgemeinerung des analytischen Weges behandeln. Die Lösungen sind dann *verallgemeinerte Logarithmusfunktionen.*

9.1. Verallgemeinerung der Exponential- und Logarithmusfunktionen auf algebraischem Weg mit Hilfe der Funktionalgleichung $f(x+y) = F\big(f(x), f(y)\big)$

Wir untersuchen nun, unter welchen Bedingungen für F die Funktionalgleichung

9.1.1. $\bigwedge_{x,y\in\mathbf{R}} f(x+y) = F\big(f(x), f(y)\big)$

eine nichtkonstante stetige Lösung f hat. Ist $f\colon \mathbf{R} \to \mathbf{R}$ eine nichtkonstante stetige

Funktion, dann ist $W(f)$ ein (nicht notwendig endliches) Intervall $I \subseteq \mathbf{R}$ (vgl. MfL Bd. 4, 2.4.2., Satz 2). Ist f auch eine Lösung von (9.1.1), so muß F (wegen der rechten Seite von (9.1.1)) über $I \times I$ definiert sein. Dann ist F (wegen der linken Seite von (9.1.1)) eine Funktion von $I \times I$ in I, also eine Operation in I (vgl. 1.1.).

Mit $I \subseteq \mathbf{R}$ sei ein offenes Intervall und mit $F: I \times I \to I$ eine stetige Funktion[1]) gegeben. Dann gilt

9.1.2. Satz. *Die Funktionalgleichung* (9.1.1) *hat genau dann eine nichtkonstante stetige Lösung f mit $D(f) = \mathbf{R}$ und $W(f) = I$ (I offenes Intervall), wenn F in I eine stetige kommutative Gruppenoperation ist.*

Beweis (vgl. J. Aczél [3]). Es sei $f: \mathbf{R} \to \mathbf{R}$ eine nichtkonstante stetige Lösung von (9.1.1) mit dem Wertebereich I. Dann gilt: F ist stetig über $I \times I$, was sich sofort aus (9.1.1) ergibt. Wie wir schon gesehen haben, ist F eine Operation in I. F ist in I auch assoziativ, denn für alle $f(x), f(y), f(z) \in I$ ist

$$\begin{aligned} F\big(F(f(x), f(y)), f(z)\big) &= F\big(f(x \dotplus y), f(z)\big) = f\big((x \dotplus y) \dotplus z\big) \\ &= f\big(x \dotplus (y \dotplus z)\big) = F\big(f(x), f(y \dotplus z)\big) \\ &= F\big(f(x), F\big(f(y), f(z)\big)\big). \end{aligned}$$

F ist kommutativ in I, denn für alle $f(x), f(y) \in I$ ist

$$F\big(f(x), f(y)\big) = f(x \dotplus y) = f(y \dotplus x) = F\big(f(y), f(x)\big).$$

Da für alle $f(x) \in I$ wegen (9.1.1) $F\big(f(0), f(x)\big) = f(x)$ ist, ist $f(0)$ neutrales Element von F in I. $f(-x)$ ist inverses Element von $f(x)$ in I, da für alle $f(x) \in I$ wegen (9.1.1) $F\big(f(-x), f(x)\big) = f(0)$ ist.

Damit ist gezeigt, daß die Existenz eines Intervalls, in dem F eine kommutative stetige Gruppenoperation ist, notwendig für die Existenz einer nichtkonstanten stetigen Lösung $f: \mathbf{R} \to I$ von (9.1.1) ist. Je nach Zweckmäßigkeit schreiben wir $F(u, v)$ oder $u \circ v$ für $u, v \in I$, so daß (9.1.1) auch als

9.1.3. $\bigwedge_{x,y \in \mathbf{R}} f(x \dotplus y) = f(x) \circ f(y)$

geschrieben werden kann.

Es sei nun durch F über einem Intervall $I =]\!]a, b[\![$ (es sei auch $I =]\!]a, \infty[\![$ bzw. $I =]\!]-\infty, b[\![$ bzw. $I = \mathbf{R}$ zugelassen[2]) eine stetige kommutative Gruppe $(I, \circ)$ mit dem neutralen Element n gegeben. Dann gilt zunächst

9.1.4. *Für alle $v \in I$ und alle $k, l \in \mathbf{Z}$ gilt $v^{k+l} = v^k \circ v^l$*[3]).

Fall 1: $k, l \in \mathbf{N}$. Das ergibt sich mittels vollständiger Induktion über k unter Benutzung der Definition von v^m, der Assoziativität und der Kommutativität von $\circ$.

[1]) Zur Stetigkeit zweistelliger Funktionen vgl. MfL Bd. 4, 2.3.1.

[2]) Die Beweise können auch für diese Fälle in gleicher Weise geführt werden.

[3]) $v^0 := n$; $v^{m+1} := v^m \circ v$ ($m \in \mathbf{N}$); $v^m := (v^{-1})^{-m}$ ($m \in \mathbf{Z} \setminus \mathbf{N}$) (vgl. 5.1.). Es handelt sich also um Potenzbildungen bezüglich der Operation $\circ$. Nur wenn Mißverständnisse möglich erscheinen, schreiben wir $v^{\circ m}$.

Fall 2: $k, l \in \mathbf{Z} \setminus \mathbf{N}$. Der Beweis wird unter Benutzung der Definition der Potenz auf **Fall 1** zurückgeführt: $v^k \circ v^l = (v^{-1})^{-k} \circ (v^{-1})^{-l} = (v^{-1})^{-(k+l)} = v^{k+l}$.

Fall 3: $k \in \mathbf{N}$, $l \in \mathbf{Z} \setminus \mathbf{N}$. Das ergibt sich mittels vollständiger Induktion über k unter Benutzung der Assoziativität und der Kommutativität von $\circ$.

Fall 4: $k \in \mathbf{Z} \setminus \mathbf{N}$, $l \in \mathbf{N}$. Das ergibt sich wegen der Kommutativität von $\circ$ aus Fall 3.

Weiter gilt der Reihe nach:

A) *F ist streng monoton wachsend in beiden Argumenten.* Es genügt wegen der Kommutativität von $\circ$ den Beweis für ein Argument (wir wählen das zweite) zu führen. Es sei $v, w \in I$ mit $v < w$. Zu zeigen ist, daß $u \circ v < u \circ w$ für jedes $u \in I$ ist. Zunächst gilt $n \circ v < n \circ w$. Gäbe es ein u mit $u \circ v \geqq u \circ w$, so gäbe es wegen der Stetigkeit von $\circ$ ein t mit $t \circ v = t \circ w$ (Zwischenwertsatz). Das gilt genau dann, wenn $t^{-1} \circ t \circ v = t^{-1} \circ t \circ w$ ist (t^{-1} sei das zu t inverse Element), was $v = w$ nach sich zieht im Widerspruch zur Voraussetzung.

B) *Die Funktion $p_m : I \to I$ mit $p_m(v) := v^m$ ($m \in \mathbf{N}$, $m > 1$) ist streng monoton wachsend und stetig.* Das ergibt sich unmittelbar aus A) und der Stetigkeit von F.

C) *Für jedes $v \in I$ ist die Folge (v^m) mit $m \in \mathbf{N}$ monoton. Ist $v > n$, so ist (v^m) streng monoton wachsend. Ist $v < n$, so ist (v^m) streng monoton fallend. Ist $v = n$, so ist (v^m) eine konstante Folge.* Das ergibt sich wegen (9.1.4) aus $v^{m-1} = v^m \circ v \overset{\S}{=} v^m \circ n = v^m$.

D) *Es sei $v \in I =]a, b[$. Ist $v > n$, so ist $\lim_{m \to \infty} v^m = b$. Ist $v < n$, so ist $\lim_{m \to \infty} v^m = a$.* Jede streng monotone Folge ist konvergent oder bestimmt divergent (vgl. MfL Bd. 4, 2.1.4.). Im Fall der bestimmten Divergenz ist nichts mehr zu beweisen. Es sei nun $b \in \mathbf{R}$. Angenommen, es existiert ein t mit $t \in]a, b[$ und $\lim_{m \to \infty} v^m = t$. Dann gilt $t = t \circ n < t \circ v = \lim_{m \to \infty} v^m \circ v = \lim_{m \to \infty} (v^m \circ v) = \lim_{m \to \infty} v^{m-1} = t$, was ein Widerspruch ist. Damit ist die Behauptung für $v > n$ bewiesen. Entsprechend kann der Beweis für $v < n$ geführt werden.

E) *Die Gleichung $v^m = u$ hat für jedes $u \in I$ und jedes $m \in \mathbf{N}^*$ eine eindeutig bestimmte Lösung $v \in I$.* Für festes m ist p_m stetig und streng monoton (siehe B)). Weiterhin gilt $\lim_{v \to b} v^m = b$, denn es ist $b = \lim_{v \to b} v \leqq \lim_{v \to b} v^m \leqq b$. Entsprechend gilt $\lim_{v \to a} v^m = a$. Also ist der Wertebereich von p_m gleich I, und jeder Funktionswert wird genau einmal angenommen.

Für die somit eindeutig bestimmte Zahl $v \in I$ ist dann die Schreibweise $v := u^{1/m}$ gerechtfertigt, ebenso wie die Definition $u^{k/m} := (u^{1/m})^k$ für $m \in \mathbf{N}^*$ und $k \in \mathbf{Z}$. Es gilt dann $(u^{1/m})^m = u$ für $m \in \mathbf{N}^*$ und $u \in I$. Wie bei der Potenzbildung bezüglich der Multiplikation ist $(u^{1/m})^k = \left(u^{\frac{1}{m \cdot l}}\right)^{k \cdot l}$ für beliebige $l, m \in \mathbf{N}^*$ und $k \in \mathbf{Z}$. Zum Beweis dieser Tatsache benötigt man, daß $(a^k)^l = a^{k \cdot l}$ für $a \in I$, $k \in \mathbf{N}^*$, $l \in \mathbf{Z}$ ist (Beweis durch vollständige Induktion bei Benutzung der Definition der Potenz und (9.1.4)) und $u^{\frac{1}{m \cdot l}} = (u^{1/m})^{1/l}$ für $u \in I$; $m, l \in \mathbf{N}^*$. (Es sei $z := (u^{1/m})^{1/l}$, dann ist $z^l = u^{1/m}$ und

$(z^l)^m = u = z^{l \cdot m}$, also $z = u^{\frac{1}{l \cdot m}}$.) Damit ergibt sich $\left(u^{\frac{1}{m \cdot l}}\right)^{k \cdot l} = \left(\left((u^{1/m})^{1/l}\right)^l\right)^k = (u^{1/m})^k$. Man vergleiche die entsprechenden Beweise für $(\mathbf{R}_+^*, \cdot)$ in Aufgabe 8.5.1.

Nach diesen Vorbereitungen können wir nun eine Funktion $f\colon \mathbf{R} \to I$ konstruieren, die (9.1.1) erfüllt.

a) Es sei $c \in I$ mit $c \neq n$. Wir behandeln im folgenden den Fall $c > n$. *Es wird* $f(1) := c$ *gesetzt.*

b) $f(m) := c^m$ *für* $m \in \mathbf{N}^*$ (vgl. mit b) im Beweis von Satz 8.1.6).

c) $f\left(\frac{1}{q}\right) := c^{1/q}$, $f\left(\frac{p}{q}\right) := c^{p/q}$ *für* $p \in \mathbf{Z}$, $q \in \mathbf{N}^*$ *mit* $f(0) := n\,(= c^0)$. Da wegen E) $f\left(\frac{k}{m}\right) = f\left(\frac{k \cdot l}{m \cdot l}\right)$ für $m, l \in \mathbf{N}^*$, $k \in \mathbf{Z}$ ist, ist f auf $\mathbf{Q}$ korrekt definiert (vgl. mit c) im Beweis von Satz 8.1.6).

d) *Die Funktion* $f\colon \mathbf{Q} \to I$ *mit* $f(r) = c^r$ *erfüllt*

d_1) *das Additionstheorem* (9.1.1) *und*

d_2) *ist streng monoton wachsend für* $c > n$ *und streng monoton fallend für* $c < n$.

Zu d_1). Es seien $r_1 = \frac{m_1}{k}$ und $r_2 = \frac{m_2}{k}$ mit $m_1, m_2 \in \mathbf{Z}$ und $k \in \mathbf{N}^*$ gegeben. Dann ist $f(r_1 + r_2) = f\left(\frac{m_1 + m_2}{k}\right) = c^{\frac{m_1+m_2}{k}} = (c^{1/k})^{m_1+m_2}$. Wegen (9.1.4) und c) gilt $(c^{1/k})^{(m_1+m_2)} = (c^{1/k})^{m_1} \circ (c^{1/k})^{m_2} = f(r_1) \circ f(r_2) = F\big(f(r_1), f(r_2)\big)$, also (9.1.1).

Zu d_2). Ist $c > n$, so ist $c^{1/k} > n$ für $k \in \mathbf{N}^*$. Angenommen, es gäbe ein k mit $c^{1/k} \leqq n$. Dann wäre $c = (c^{1/k})^k \leqq n^k = n$ nach B) im Widerspruch zu $c > n$. Damit erhält man $f\left(\frac{m+1}{k}\right) = f\left(\frac{m}{k}\right) \circ f\left(\frac{1}{k}\right) > f\left(\frac{m}{k}\right) \circ n = f\left(\frac{m}{k}\right)$, woraus folgt, daß f über $\mathbf{Q}$ streng monoton wachsend für $c > n$ ist. Für $c < n$ wird der Beweis entsprechend geführt. Damit ist d) bewiesen (vgl. mit d) im Beweis von Satz 8.1.6

e) *Definition einer Funktion* $f\colon \mathbf{R} \to I$.

e_1) Ist (h_k) eine beliebige Folge rationaler Zahlen mit $h_k > 0$ und $\lim\limits_{k\to\infty} h_k = 0$, so ist $\lim\limits_{k\to\infty} f(h_k) = \lim\limits_{k\to\infty} c^{h_k} = n$.

Fall 1: $h_k = \frac{1}{k}$. Es gilt $\lim\limits_{k\to\infty} f\left(\frac{1}{k}\right) = \lim\limits_{k\to\infty} c^{1/k} = n$. Da $(c^{1/k})$ monoton fallend und durch n nach unten beschränkt ist, existiert $\lim\limits_{k\to\infty} c^{1/k}$. Aus der Annahme $\lim\limits_{k\to\infty} f\left(\frac{1}{k}\right) = d > n$ ergibt sich

$$d = \lim_{k\to\infty} f\left(\frac{1}{k}\right) = \lim_{k\to\infty} f\left(\frac{1}{2k} + \frac{1}{2k}\right) = \lim_{k\to\infty} f\left(\frac{1}{2k}\right) \circ \lim_{k\to\infty} f\left(\frac{1}{2k}\right)$$
$$= \left(\lim_{k\to\infty} f\left(\frac{1}{2k}\right)\right)^2 = d^2 > d,$$

was einen Widerspruch darstellt. Also gilt die Behauptung.

Fall 2: Es sei jetzt (h_k) eine beliebige Nullfolge rationaler Zahlen mit $h_k > 0$, wobei o. B. d. A. $h_k \leqq 1$ vorausgesetzt werden kann. Wird nun zu jedem k die größte natürliche Zahl v so bestimmt, daß $h_k \leqq \frac{1}{v}$ gilt, so gilt wegen der strengen Monotonie von c^{h_k} (siehe d)) für $c > n$ auch $c^{h_k} \leqq c^{1/v}$. Da die linke Seite dieser Ungleichung stets größer als n ist, erhält man mit $\lim\limits_{v\to\infty} c^{1/v} = n$ (siehe Fall 1) auch $\lim\limits_{k\to\infty} c^{h_k} = n$. Für $c < n$ kann entsprechend geschlossen werden.

e_2) Gegeben seien $x_0 \in \mathbf{R}$, eine monoton wachsende Folge rationaler Zahlen (r_k) und eine monoton fallende Folge rationaler Zahlen (r_k') mit $\lim\limits_{k\to\infty} r_k = x_0 = \lim\limits_{k\to\infty} r_k'$. Wegen der Monotonie und Beschränktheit von $\big(f(r_k)\big)$ und $\big(f(r_k')\big)$ existieren $g := \lim\limits_{k\to\infty} f(r_k)$ und $g' := \lim\limits_{k\to\infty} f(r_k')$, wobei $g \leqq g'$ ist. Es gilt sogar $g = g'$ wegen

$$\begin{aligned} g' &= \lim_{k\to\infty} f(r_k') = \lim_{k\to\infty} f\big((r_k' - r_k) + r_k\big) \\ &= \lim_{k\to\infty} f(r_k' - r_k) \circ \lim_{k\to\infty} f(r_k) = n \circ g = g. \end{aligned}$$

Daher ist $f(x_0) = c^{x_0} := g = g'$ eindeutig bestimmt, und f ist über $\mathbf{R}$ definiert (vgl. e) im Beweis von Satz 8.1.6). Wir zeigen nun

f) *Die Funktion* $f\colon \mathbf{R} \to I$ *mit* $f(x) = c^x$ *erfüllt* (9.1.1). Es seien $x, y \in \mathbf{R}$ und (r_k), (r_k') monoton wachsende Folgen rationaler Zahlen mit $\lim\limits_{k\to\infty} r_k = x$ und $\lim\limits_{k\to\infty} r_k' = y$. Es gilt zunächst für jedes k nach d_1) $f(r_k + r_k') = f(r_k) \circ f(r_k')$. Daher gilt auch

$$\lim_{k\to\infty} f(r_k + r_k') = \lim_{k\to\infty} \big(f(r_k) \circ f(r_k')\big) = \lim_{k\to\infty} f(r_k) \circ \lim_{k\to\infty} f(r_k'),$$

also

$$f(x + y) = f(x) \circ f(y) = F\big(f(x), f(y)\big)$$

(vgl. f) im Beweis von Satz 8.1.6).

g) $f\colon \mathbf{R} \to I$ *mit* $f(x) = c^x$ *ist stetig.*

g_1) Wir zeigen zunächst, daß f für $c > n$ streng monoton wachsend ist. Es seien x und y beliebige reelle Zahlen mit $x < y$ sowie (r_k) und (r_k') Folgen rationaler Zahlen mit $\lim\limits_{k\to\infty} r_k = x$ ($r_k > x$ für alle k) und $\lim\limits_{k\to\infty} r_k' = y$ ($r_k' < y$ für alle k). Wegen $x < y$ gibt es dann eine natürliche Zahl N_0 derart, daß $x < r_k < r_k' < y$ für alle $k > N_0$ gilt. Nun ist offensichtlich $c^x \leqq c^{r_k}$ und $c^{r_k'} \leqq c^y$ für alle natürlichen Zahlen k und $c^{r_k} < c^{r_k'}$ für $k > N_0$ (wegen d)). Also gilt $c^x \leqq c^{r_k} < c^{r_k'} \leqq c^y$ für $k > N_0$, woraus sich $f(x) < f(y)$ ergibt.

Für $c < n$ folgt entsprechend, daß f streng monoton fallend ist.

g_2) Es sei x_0 eine beliebige reelle Zahl und (x_k) zunächst eine Folge reeller Zahlen mit $\lim\limits_{k\to\infty} x_k = x_0$ und $x_k < x_0$ für alle k. Nach f) gilt für reelle Zahlen die Funktionalgleichung (9.1.1), also gilt $c^{x_0} = c^{x_0 - x_k} \circ c^{x_k}$. Es genügt daher zu zeigen, daß $\lim\limits_{k\to\infty} c^{x_0 - x_k} = n$ ist. Dazu kann $x_0 - x_k < 1$ vorausgesetzt werden und für jedes k das größte natür-

liche v gewählt werden mit $x_0 - x_k \leqq \frac{1}{v}$. Daraus folgt wegen $n \leqq c^{x_0 - x_k} \leqq c^{1/v}$ die Behauptung für den genannten Fall, wobei g_1) benutzt wird.

Die Folge (x_k') konvergiere jetzt von rechts gegen das Argument x_0, d. h. $\lim_{k\to\infty} x_k' = x_0$ mit $x_k' > x_0$ für alle k. In diesem Fall gilt ebenfalls $\lim_{k\to\infty} c^{x_k'} = c^{x_0}$, denn aus $c^{-x_0} = c^{x_k' - x_0} \circ c^{-x_k'}$ folgt wie eben $\lim_{k\to\infty} c^{-x_k'} = c^{-x_0}$. Das bedeutet aber $\lim_{k\to\infty} (c^{x_k'})^{-1} = (c^{x_0})^{-1}$, also $\lim_{k\to\infty} c^{x_k'} = c^{x_0}$.

Insgesamt ergibt sich daher die Stetigkeit von f an der Stelle x_0 (vgl. g) im Beweis von Satz 8.1.6). Damit ist Satz 9.1.2 bewiesen.

9.1.5. Bemerkung. Jede nichtkonstante stetige Lösung $f\colon \mathbf{R} \to I$ von (9.1.1) hat sich als streng monoton erwiesen. Daher existiert zu ihr eine Umkehrfunktion f^{-1}, die eine verallgemeinerte Logarithmusfunktion ist, denn wenn $f\colon \mathbf{R} \to I$ eine stetige und streng monotone Funktion ist, für die $f(x + y) = F\big(f(x), f(y)\big)$ gilt, so ergibt sich für $f(x) = u, f(y) = v$ bzw. $f^{-1}(u) = x$ und $f^{-1}(v) = y$ aus (9.1.1) $f^{-1}\big(f(x + y)\big) = f^{-1}\Big(F\big(f(x), f(y)\big)\Big)$, also $x + y = f^{-1}\big(F(u, v)\big)$ oder $f^{-1}(u) + f^{-1}(v) = f^{-1}\big(F(u, v)\big)$. Damit ist $f^{-1}\colon I \to \mathbf{R}$ eine stetige und streng monotone Lösung von

$$\textbf{9.1.6.} \qquad \bigwedge_{x,y\in A} f\big(F(x, y)\big) = f(x) + f(y).$$

Satz 9.1.2 besagt auch, daß für jede stetige und streng monotone Funktion $f\colon \mathbf{R} \to I$ eine stetige Gruppenoperation $\circ$ existiert, so daß (9.1.3) bzw .(9.1.1) gilt. Zum Beispiel ergibt sich für $f\colon \mathbf{R} \to I$ mit $f(x) = \arctan x$ und $u := f(x)$ bzw. $v := f(y)$ die Beziehung $F(u, v) = \arctan(\tan u + \tan v)$. Man bestimme I, n, u^{-1}.

Wir wollen uns nun einen Überblick über alle nichtkonstanten, für alle reellen Zahlen definierten Lösungen, von (9.1.1) verschaffen. Es sei $f\colon \mathbf{R} \to I$ eine nichtkonstante stetige Lösung von (9.1.1), und $g\colon \mathbf{R} \to I$ sei eine beliebige nichtkonstante Lösung von (9.1.1). Wegen der Voraussetzungen existiert f^{-1}, und mit $f(x) = u$ bzw. $f^{-1}(u) = x$ und $f(y) = v$ bzw. $f^{-1}(v) = y$ ergibt sich $f\big(f^{-1}(u) + f^{-1}(v)\big) = F(u, v)$ oder auch

$$g(x + y) = F\big(g(x), g(y)\big) = f\Big(f^{-1}\big(g(x)\big) + f^{-1}\big(g(y)\big)\Big).$$

Wir definieren nun durch

$$\textbf{9.1.7.} \qquad \bigwedge_{x\in\mathbf{R}} h(x) := f^{-1}\big(g(x)\big)$$

eine Funktion $h\colon \mathbf{R} \to \mathbf{R}$.[1]) Für h gilt

$$\textbf{9.1.8.} \qquad \bigwedge_{x,y\in\mathbf{R}} h(x + y) = h(x) + h(y).$$

Fall 1: Wenn g stetig ist (oder g ist in einem Punkt stetig, oder g ist auf einem endlichen Intervall durch Zahlen aus I nach oben bzw. nach unten beschränkt, oder

[1]) Man überzeuge sich davon, daß h korrekt definiert ist.

es gibt einen Punkt (x, y) mit $x \in \mathbf{R}$, $y \in I$ und eine Umgebung dieses Punktes, die keinen Punkt des Graphen von g enthält, oder g ist meßbar), dann überträgt sich wegen der Stetigkeit von f^{-1} diese Eigenschaft auf h, und h mit $h(x) = c \cdot x$ ($c \in \mathbf{R}^*$) ist die einzige Lösung von (9.1.8) (vgl. 7.1. und 7.2.). Also ist $g(x) = f(c \cdot x)$ wegen (9.1.7). Damit haben wir auch alle stetigen nichtkonstanten Lösungen von (9.1.1) gefunden.

Fall 2: Es sei $h: \mathbf{R} \to \mathbf{R}$ eine unstetige Lösung von (9.1.8) und sei f eine nichtkonstante stetige Lösung von (9.1.1). Dann ist g mit $g(x) = f(h(x))$ eine unstetige Lösung von (9.1.1), wie man sich sofort überzeugt. Also läßt die Funktionalgleichung (9.1.1) auch unstetige Lösungen zu. Für unstetige Lösungen gilt

9.1.9. Satz. *Jede unstetige Lösung* $f: \mathbf{R} \to I$ *von* (9.1.1) *liegt dicht in* $\mathbf{R} \times I$.

Der Beweis (vgl. V. Drobot [47]) verläuft wie für Satz 8.2.2 und wird dem Leser überlassen.

9.1.10. Beispiel. Es sei $I =]-1, \infty[$ und $F(u, v) = u \circ v = u + v + u \cdot v$ für $u, v \in I$. Dann hat (9.1.1) die Form $f(x + y) = f(x) + f(y) + f(x) \cdot f(y)$. $(I, \circ)$ ist eine kommutative stetige Gruppe mit $n = 0$ und $u^{-1} = \dfrac{-u}{1 + u}$ (vgl. auch Aufgabe 1.5.6.3). In dieser Gruppe ist $u^{\circ m} = (u + 1)^m - 1$ ($m \in \mathbf{N}^*$), woraus sich schließlich $f: \mathbf{R} \to I$ mit $f(x) = (a + 1)^x - 1$ mit $a \in I$, $a \neq 0$ ergibt.

9.2. Verallgemeinerung der Exponential- und Logarithmusfunktionen auf analytischem Weg mit Hilfe der Funktionalgleichung $f(F(x, y)) = f(x) + f(y)$

Die in 9.1. gewonnenen verallgemeinerten Exponential- und Logarithmusfunktionen wurden mit Hilfe der Potenzbildung, also auf algebraischem Weg, konstruiert. Es ergibt sich die Frage, ob diese Funktionen auch auf analytischem Weg, d. h. in diesem Fall mit Hilfe des Integralbegriffs, gewonnen werden können. Die Beantwortung dieser Frage hängt sicher von Eigenschaften der Operation $\circ$ ab. Wenn $\circ$ die Multiplikation $\cdot$ reeller Zahlen ist, sind die Wege gleichwertig. Der folgende Satz zeigt, daß dies auch für andere Operationen $F = \circ$ gilt:

9.2.1. Satz. *Es sei* $I \subseteq \mathbf{R}$ *ein Intervall und* $F: I \times I \to I$ *eine Funktion, für die gilt:*

a) F *ist Gruppenoperation in* I *und* n *das neutrale Element.*

b) F *hat stetige partielle Ableitungen erster Ordnung* $\partial_1 F$ *und* $\partial_2 F$, *die immer von Null verschieden sind.*[1])

Dann ist die Funktion $f: I \to \mathbf{R}$ *mit*

$$\textbf{9.2.2.}\quad f(x) := k \int_n^x \frac{dt}{\partial_2 F(t, n)} \qquad (k \in \mathbf{R}^*)$$

eine über I *differenzierbare, streng monotone Lösung von* (9.1.6).

[1]) Zur Differentiation zweistelliger Funktionen vgl. MfL Bd. 5, 3.2.

Beweis (vgl. H. RÅDSTRÖM [145]). Wir zeigen zunächst, daß jede differenzierbare Lösung $f: I \to \mathbf{R}$ von (9.1.6) von der Form (9.2.2) ist. Es sei n das neutrale Element der durch F gegebenen Gruppenoperation, d. h., es gilt $F(x, n) = F(n, x) = x$ für alle $x \in I$. Es ist $f(n) = 0$, denn für $y = n$ folgt aus (9.1.6) $f\big(F(x, n)\big) = f(x) + f(n)$, also $f(x) = f(x) + f(n)$. Für jede differenzierbare Lösung $f: I \to \mathbf{R}$ von (9.1.6) ergibt sich durch Differenzieren nach x bzw. y

$$f'\big(F(x, y)\big) \cdot \partial_1 F(x, y) = f'(x) \quad \text{bzw.} \quad f'\big(F(x, y)\big) \cdot \partial_2 F(x, y) = f'(y).$$

Unter Beachtung von b) ergibt sich

$$\textbf{9.2.3.}\quad f'(x) = \frac{\partial_1 F(x, y)}{\partial_2 F(x, y)} \cdot f'(y).$$

Indem wir in (9.2.3) y durch n, $f'(n)$ durch k ersetzen und berücksichtigen, daß $F(x, n) = x$, also $\partial_1 F(x, n) = 1$ ist, ergibt sich $f'(x) = k \cdot \dfrac{1}{\partial_2 F(x, n)}$, was wegen $f(n) = 0$ zu (9.2.2) führt. Die durch (9.2.2) definierte Funktion $f: I \to \mathbf{R}$ ist als Integral einer stetigen Funktion differenzierbar und wegen b) streng monoton.

Es bleibt zu zeigen, daß die durch (9.2.2) definierte Funktion Lösung von (9.1.6) ist. Da F eine Gruppenoperation ist, gilt $F\big(x, F(y, z)\big) = F\big(F(x, y), z\big)$ für alle $x, y, z \in I$. Wir differenzieren die Funktion G mit $G(x, y, z) := F\big(x, F(y, z)\big) = F\big(F(x, y), z\big)$ nach z und erhalten

$$\partial_3 G(x, y, z) = \partial_2 F\big(x, F(y, z)\big) \cdot \partial_2 F(y, z) = \partial_2 F\big(F(x, y), z\big).$$

Für $z = n$ und $y = s$ ergibt sich

$$\partial_2 F\big(x, F(s, n)\big) \cdot \partial_2 F(s, n) = \partial_2 F\big(F(x, s), n\big),$$

also $\partial_2 F(x, s) \cdot \partial_2 F(s, n) = \partial_2 F\big(F(x, s), n\big)$. Daher gilt auch

$$\frac{\partial_2 F(x, s) \cdot \partial_2 F(s, n) - \partial_2 F\big(F(x, s), n\big)}{\partial_2 F\big(F(x, s), n\big) \cdot \partial_2 F(s, n)} = 0$$

und

$$\textbf{9.2.4.}\quad k \cdot \int_n^y \left(\frac{\partial_2 F(x, s)}{\partial_2 F\big(F(x, s), n\big)} - \frac{1}{\partial_2 F(s, n)} \right) ds = 0.$$

$k \displaystyle\int_n^y \frac{\partial_2 F(x, s)\, ds}{\partial_2 F\big(F(x, s), n\big)}$ geht durch die Substitution $t = F(x, s)$ mit $\dfrac{dt}{ds} = \partial_2 F(x, s)$ über in $k \displaystyle\int_{F(x, y)}^{x} \frac{dt}{\partial_2 F(t, n)}$. Dafür können wir aufgrund der Additivität des bestimmten Integrals $k \displaystyle\int_n^{F(x, y)} \frac{dt}{\partial_2 F(t, n)} - k \int_n^x \frac{dt}{\partial_2 F(t, n)}$ schreiben. Das berücksichtigend, ergibt

sich aus (9.2.4) unter Verwendung der Definition von f die Beziehung $f(F(x, y)) - f(x) - f(y) = 0$ für alle $x, y \in I$, also (9.1.6)[1].

Die durch Satz 9.2.1 erhaltenen Funktionen $f: I \to \mathbf{R}$ sind verallgemeinerte Logarithmusfunktionen, die, da sie streng monoton sind, Umkehrfunktionen $f^{-1}: \mathbf{R} \to I$ besitzen, die dann verallgemeinerte Exponentialfunktionen sind. Die Umkehrfunktionen der durch Satz 9.2.1 ermittelten Funktionen sind in der Menge der in 9.1. gefundenen Funktionen enthalten, denn dort wurden u. a. alle stetigen nichtkonstanten Lösungen von (9.1.1) bestimmt.

9.2.5. Beispiele

1. Es sei $I = \mathbf{R}_+^*$ und F die Multiplikation der positiven reellen Zahlen. $(\mathbf{R}_+^*, \cdot)$ ist eine Gruppe, und die partiellen Ableitungen $\partial_1 F$ und $\partial_2 F$ sind stetig und immer von Null verschieden. Daher sind die Funktionen $f: \mathbf{R}_+^* \to \mathbf{R}$ mit $f(x) = k \int_1^x \frac{dt}{t}$ für jedes $k \in \mathbf{R}^*$ streng monotone und differenzierbare Lösungen von $f(x \cdot y) = f(x) + f(y)$ (vgl. Bemerkung 8.3.12).

2. Gegeben seien $I =]-\infty, 1[$ und F mit $F(x, y) = x + y - x \cdot y$. Man zeigt leicht, daß F in I eine Gruppenoperation ist und daß $\partial_1 F$ und $\partial_2 F$ stetig und immer von Null verschieden sind. Daher sind die Funktionen $f:]-\infty, 1[\to \mathbf{R}$ mit $f(x) = k \int_1^x \frac{dt}{1-t}$ für jedes $k \in \mathbf{R}_+^*$ streng monotone und differenzierbare Lösungen von $f(x + y - x \cdot y) = f(x) + f(y)$.

Man könnte auch folgendermaßen versuchen, solche stetigen Gruppenoperationen $F = \circ$ auszusondern, für die die zugehörigen verallgemeinerten Exponential- und Logarithmusfunktionen f bzw. f^{-1} differenzierbar sind. (Die Differenzierbarkeit von f war ja notwendig für den analytischen Weg!) Wir bilden den Differenzenquotienten einer stetigen Lösung $f: \mathbf{R} \to I$ von (9.1.1) an der Stelle x und erhalten $\frac{f(x+h) - f(x)}{h} = \frac{c^{\circ(x+h)} - c^{\circ x}}{h} = \frac{c^{\circ x} \circ c^{\circ h} - c^{\circ x}}{h}$. Könnte man jetzt wie bei der Differentiation der Exponentialfunktionen $c^{\circ x}$ ausklammern, so würde für alle diese stetigen Gruppenoperationen $\circ$ die Differenzierbarkeit von f an jeder Stelle aus der Differenzierbarkeit an der Stelle Null folgen. Solche Operationen gibt es aber außer der Multiplikation der reellen Zahlen nicht, wie folgender Satz zeigt.

9.2.6. Satz. *Die einzigen stetigen und zugleich kommutativen zweistelligen Funktionen* $F: \mathbf{R} \times \mathbf{R} \to \mathbf{R}$, *die*

$$\text{9.2.7.} \quad \bigwedge_{x,y,z \in \mathbf{R}} F(x + y, z) = F(x, z) + F(y, z)$$

genügen, die also bezüglich der Addition distributiv sind, sind die Funktionen F *mit* $F(x, z) = c \cdot x \cdot z$ $(c \in \mathbf{R})$.

[1]) Bei dem Beweis sind also gar nicht alle geforderten Gruppeneigenschaften von F in I notwendig. Der Satz wurde in dieser Weise formuliert, um den angestrebten Vergleich mit den in 9.1. betrachteten Funktionen anzustellen.

Beweis (vgl. J. Aczél [2]). Für festes z definieren wir eine Funktion $f_z\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) := F(x, z)$, die mit F stetig ist. Aus (9.2.7) ergibt sich für f_z

$$f_z(x + y) = f_z(x) + f_z(y)$$

mit den einzigen stetigen Lösungen f_z mit $f_z(x) = c \cdot x$ (vgl. 7.1.), wobei c eine reelle Zahl ist, die von z abhängen kann. Also ist $F(x, z) = x \cdot c(z)$. Wegen der Kommutativität von F gilt nun $c(z) \cdot x = F(z, x) = F(x, z) = c(x) \cdot z$, also $\dfrac{c(z)}{z} = \dfrac{c(x)}{x} = c$ für alle von Null verschiedenen Zahlen x und z. Daraus ergibt sich $c(z) = c \cdot z$ und $F(x, z) = c \cdot x \cdot z$, und nur die Festlegungen $F(0, z) := F(x, 0) := F(0, 0) = 0$ widersprechen nicht der Stetigkeit von F.

9.2.8. Bemerkung. Wir haben in Satz 9.1.2 gesehen, daß auf dem algebraischen Weg für jede stetige kommutative Gruppenoperation F eine streng monotone stetige Lösung f von (9.1.1) konstruiert werden kann. Die Umkehrfunktion f^{-1} ist ebenfalls streng monoton und stetig. Da es streng monotone stetige Funktionen gibt, die keine Differenzierbarkeitsintervalle aufweisen (siehe etwa K. Kursawe [96]), gibt es stetige Gruppenoperationen F (vgl. Bemerkung 9.1.5), deren zugehörige stetige Lösungen f von (9.1.1) über keinem Intervall differenzierbar sind. Wenn man den Riemannschen Integralbegriff zugrunde legt, sind die auf dem analytischen Weg erhaltenen verallgemeinerten Exponentialfunktionen also eine echte Teilmenge der auf dem algebraischen Weg erhaltenen verallgemeinerten Exponentialfunktionen. Auch bei Berücksichtigung des Lebesgueschen Integralbegriffs (vgl. H. v. Mangoldt und K. Knopp, Band IV, [112]) wird dieses Ergebnis bestätigt. Um dies zu zeigen, benutzen wir eine auf G. Cantor (1845—1918) zurückgehende Funktion $c\colon [\![0, 1]\!] \to [\![0, 1]\!]$, die folgendermaßen schrittweise definiert werden kann. Im ersten Schritt wird c auf einem Teilintervall definiert:

$$c(x) := \frac{1}{2} \quad \text{für} \quad x \in \left]\!\left]\frac{1}{3}, \frac{2}{3}\right[\!\right[.$$

Im zweiten Schritt wird c auf zwei Teilintervallen definiert:

$$c(x) := \frac{1}{4} \quad \text{für} \quad x \in \left]\!\left]\frac{1}{9}, \frac{2}{9}\right[\!\right[\quad \text{und} \quad c(x) := \frac{3}{4} \quad \text{für} \quad x \in \left]\!\left]\frac{7}{9}, \frac{8}{9}\right[\!\right[.$$

Im k-ten Schritt wird c auf 2^{k-1} Teilinterwallen definiert:

$$\text{Von } c(x) := \frac{1}{2^k} \quad \text{für} \quad x \in \left]\!\left]\frac{1}{3^k}, \frac{2}{3^k}\right[\!\right[\quad \text{über} \quad c(x) := \frac{3}{2^k} \quad \text{für} \quad x \in \left]\!\left]\frac{7}{3^k}, \frac{8}{3^k}\right[\!\right[$$

und

$$c(x) := \frac{5}{2^k} \quad \text{für} \quad x \in \left]\!\left]\frac{19}{3^k}, \frac{20}{3^k}\right[\!\right[\quad \text{bis} \quad c(x) := \frac{2^k - 1}{2^k}$$

$$\text{für} \quad x \in \left]\!\left]\frac{3^k - 2}{3^k}, \frac{3^k - 1}{3^k}\right[\!\right[.$$

Für von diesem Prozeß nicht erfaßte Zahlen x_0 aus dem Intervall $]0, 1[$ definieren wir $c(x_0) := \sup\{c(x): x < x_0\}$. Schließlich sei $c(0) := 0$ und $c(1) := 1$. Die Funktion c ist stetig und monoton wachsend. Die Ableitung von c ist fast überall Null,[1]) und durch $c(x + 1) = c(x) + 1$ kann c zu einer Funktion $d: \mathbf{R} \to \mathbf{R}$ fortgesetzt werden, die ebenfalls die genannten Eigenschaften besitzt. Die von S. Saks ([155], Kap. III, § 13) konstruierte Funktion $h: [0, 1] \to [0, 2]$ mit $h(x) := \sum_{k=1}^{\infty} \frac{d(kx)}{2^k}$ ist dann streng monoton wachsend und stetig, und ihre Ableitung ist fast überall Null. Schließlich ist die Funktion $g:]0, 2[\to \mathbf{R}$ mit $g(x) = \frac{x - 1}{x(2 - x)}$ streng monoton wachsend und differenzierbar. Daher hat die Funktion $f:]0, 1[\to \mathbf{R}$ mit $f(x) = g(h(x))$ folgende Eigenschaften:

a) f ist streng monoton wachsend und stetig.

b) Die Ableitung von f ist fast überall Null.

c) f und ihre Umkehrfunktion f^{-1} definieren durch $F(x, y) := f^{-1}(f(x) + f(y))$ eine Funktion $F:]0, 1[\times]0, 1[\to]0, 1[$.

f ist für die in c) definierte Operation F eine stetige und streng monotone Lösung von (9.1.6), die sich wegen b) nicht als Lebesguesches Integral mit variabler oberer Grenze darstellen läßt. Damit ist ein Beispiel für eine verallgemeinerte Logarithmusfunktion gefunden, die auf dem analytischen Weg nicht gewonnen werden kann. Hiermit ist die am Anfang von 9.2. gestellte Frage negativ beantwortet. Der algebraische Weg ist verallgemeinerungsfähiger als der analytische Weg.

[1]) Eine ausführliche Behandlung der Eigenschaften dieser Funktion findet man z. B. bei I. P. Natanson [122], Kap. VIII, § 2.

10. Verallgemeinerung der Exponentialfunktionen ohne Stetigkeitsforderung

Da die Stetigkeit von Funktionen im Mathematikunterricht bis Klasse 10 nicht behandelt wird, wollen wir eine weitere Verallgemeinerung vornehmen und untersuchen, welche Aussagen über Exponentialfunktionen gelten, wenn auf Stetigkeitsforderungen verzichtet wird. Dabei werden dividierbare Gruppen[1]) und ihre Eigenschaften eine wesentliche Rolle spielen. Wir geben hier nur einige Definitionen und Sätze an, die für die folgenden Ausführungen von Bedeutung sind.

10.1. Dividierbare Gruppen

10.1.1. Definition. Eine Gruppe $(B, \circ)$ heißt *dividierbar* $\Leftrightarrow$ Für alle $u \in B$ und für alle ganzen Zahlen $k \neq 0$ existiert ein $v \in B$, so daß $v^{\circ k} = u$ gilt.

10.1.2. Beispiele dividierbarer Gruppen

1. Die Gruppe $(\mathbb{R}_+^*, \cdot)$ ist eine dividierbare Gruppe, denn die Gleichung $v^k = u$ hat für jedes $u \in \mathbb{R}_+^*$ und jede ganze Zahl $k \neq 0$ eine Lösung $v \in \mathbb{R}_+^*$, die für $k > 0$ durch $\sqrt[k]{u}$ und für $k < 0$ durch $\frac{1}{\sqrt[-k]{u}}$ gegeben ist.

2. Die Gruppe $(\mathbb{Q}, +)$ ist eine dividierbare Gruppe, denn die Gleichung $k \cdot v = u$ hat für jedes $u \in \mathbb{Q}$ und jede ganze Zahl $k \neq 0$ eine Lösung $v = \frac{u}{k}$, die eine rationale Zahl ist.

3. Es sei p eine feste Primzahl. Wir betrachten die Menge

$$P_p = \left\{ \frac{k}{p^i} : i = 0, 1, 2, \ldots; k = 0, 1, 2, \ldots, (p^i - 1) \right\}$$

und erklären eine Operation $\circ$ in P_p durch

$$a \circ b := \begin{cases} a + b & \text{für} \quad a + b < 1, \\ a + b - 1 & \text{für} \quad a + b \geqq 1. \end{cases}$$

Für $p = 2$ ergibt sich z. B.

$$P_2 = \left\{ 0, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \frac{1}{16}, \ldots \right\}$$

und

$$\frac{1}{2} \circ \frac{1}{2} = 0, \qquad \frac{1}{2} \circ \frac{1}{4} = \frac{3}{4}, \qquad \frac{1}{2} \circ \frac{3}{4} = \frac{1}{4}, \qquad \frac{3}{8} \circ \frac{7}{8} = \frac{1}{4}.$$

---

[1]) Dividierbare Gruppen werden ausführlich z. B. bei L. Fuchs [58] erörtert.

Man erkennt leicht, daß $(P_p, \circ)$ isomorph zu einer Untergruppe der Faktorgruppe $\mathfrak{Q}/\mathfrak{Z}$ der Gruppe $\mathfrak{Q} = (\mathbb{Q}, +)$ nach (dem Normalteiler) $\mathfrak{Z} = (\mathbb{Z}, +)$ ist (vgl. 3.2.). $(P_p, \circ)$ ist eine dividierbare Gruppe. Die zu ihr isomorphe Untergruppe von $\mathfrak{Q}/\mathfrak{Z}$ bezeichnen wir mit $Z(p^\infty)$. Ein Beweis für die Dividierbarkeit von $Z(p^\infty)$ wird in 10.2. gegeben.

4. Die direkte Summe[1]) Z der Gruppen $Z(2^\infty) \oplus Z(3^\infty) \oplus \cdots \oplus Z(p^\infty) \oplus \cdots$ ist ebenfalls eine dividierbare abelsche Gruppe.

10.1.3. Definition. Eine Gruppe $(B, \circ)$ heißt *torsionsfrei* $\Leftrightarrow$ Für jedes $a \in B$ und jede ganze Zahl $k \neq 0$ gilt: Wenn $a^{\circ k} = n$, so $a = n$, wobei n das neutrale Element von $(B, \circ)$ ist.

10.1.4. Beispiele

1. Die Gruppe $(\mathbb{R}_+^*, \cdot)$ ist torsionsfrei, denn für jede positive Zahl a und jede ganze Zahl $k \neq 0$ gilt $a^k = 1$ genau dann, wenn $a = 1$ ist.

2. Die Gruppe $(\mathbb{R}^*, \cdot)$ ist nicht torsionsfrei, denn es gilt $(-1)^2 = 1$. Die Zahl -1 ist ein Torsionselement der Gruppe $(\mathbb{R}^*, \cdot)$.

10.1.5. Definition. Eine Gruppe $(T, \circ)$ heißt *Torsionsgruppe* $\Leftrightarrow$ Jedes Element der Gruppe ist von endlicher Ordnung.

10.1.6. Beispiel. Jede Gruppe mit endlich vielen Elementen ist eine Torsionsgruppe.

10.1.7. Satz. *Jede dividierbare Gruppe $(B, \circ)$ ist direkte Summe endlich oder unendlich vieler Gruppen der Form $Z(p^\infty)$ und einer torsionsfreien dividierbaren Gruppe $(D, \circ)$.*

Für einen Beweis verweisen wir auf L. Fuchs [58].

10.2. Allgemeine Lösung der Funktionalgleichung $f(x+y) = F(f(x), f(y))$

Es sei A eine beliebige Menge (die Elemente von A brauchen keine Zahlen zu sein) und $F: A \times A \to A$ eine gegebene Funktion. Wir wollen Lösungen $f: \mathbb{R} \to A$ des Additionstheorems

10.2.1. $\bigwedge_{x,y \in \mathbb{R}} f(x+y) = F(f(x), f(y))$

bestimmen, ohne eine Regularitätsforderung (wie z. B. Stetigkeit, Meßbarkeit) an f zu stellen. Als notwendige Bedingung für die Existenz von Lösungen f von (10.2.1) erweist sich

10.2.2. Lemma. *Ist $f: \mathbb{R} \to A$ eine Lösung von* (10.2.1), *so existiert eine Menge B mit $B \subseteq A$ derart, daß $(B, \circ)$ eine dividierbare abelsche Gruppe ist, wobei $\circ$ die gegebene Operation F ist.*

[1]) Der Begriff der direkten Summe bezüglich der Addition entspricht dem Begriff des direkten Produktes bezüglich der Multiplikation. Zum Begriff des direkten Produktes siehe MfL Bd. 3, 12.7. Im Fall unendlich vieler Summanden hat man die diskrete und die vollständige direkte Summe zu unterscheiden (siehe L. Fuchs [58]). Ist im folgenden von der direkten Summe unendlich vieler Gruppen die Rede, so ist immer die diskrete direkte Summe gemeint.

Beweis (vgl. M. Neagu [123]). Es sei $f: \mathbf{R} \to A$ eine Lösung von (10.2.1) und $B := W(f)$. Das Additionstheorem (10.2.1) läßt sich auch als

10.2.3. $\bigwedge_{x,y \in \mathbf{R}} f(x+y) = f(x) \circ f(y)$

schreiben. Die Funktion f ist ein Homomorphismus von der abelschen Gruppe $(\mathbf{R}, +)$ auf $(\mathbf{B}, \circ)$, und somit ist $(B, \circ)$ eine abelsche Gruppe (vgl. 3.3.). Es seien nun $u = f(x) \in B$ und eine ganze Zahl $k \neq 0$ beliebig gegeben. Dann gilt für $v = f\left(\frac{x}{k}\right) \in B$ die Gleichung $v^{\circ k} = u$, denn es ist

$$v^{\circ k} = f\left(\frac{x}{k}\right)^{\circ k} = f\left(\frac{x}{k} + \frac{x}{k} + \cdots + \frac{x}{k}\right) = f(x)$$

wegen (10.2.3). Also ist $(B, \circ)$ eine dividierbare Gruppe. $f(0)$ ist das neutrale Element von $(B, \circ)$. (Hat die Gleichung (10.2.3) für $x = y = 0$ mehrere Lösungen, so wählen wir eine als neutrales Element von $(B, \circ)$.) Das inverse Element von $f(x)$, also $f(x)^{-1}$, ist durch $f(-x)$ definiert.

Hieraus folgt sofort

10.2.4. Korollar. *Ist jede Menge $B \subseteqq A$, für die $(B, \circ)$ Gruppe ist, endlich, dann ist $f: \mathbf{R} \to A$ mit $f(x) = f(0)$ einzige Lösung von* (10.2.1).

Beweis. $f: \mathbf{R} \to A$ mit $f(x) = f(0)$ ist Lösung, wie man sich durch Einsetzen in (10.2.1) überzeugt. Es sei $k > 1$ die Ordnung von $(B, \circ)$. (Für $k = 1$ gilt trivialerweise nur $f(x) = f(0)$.) Nach dem Fermatschen Satz der Gruppentheorie (vgl. MfL Bd. 3, 12.2., Folgerung 2) gilt: Für alle $a \in B$ ist $a^k = f(0)$. Es sei $b \neq f(0)$ ein Element von $(B, \circ)$. Für jedes $a \in B$ gilt $a^k \neq b$. Also ist $(B, \circ)$ keine dividierbare Gruppe.

Es wird sich zeigen, daß die in Lemma 10.2.2 formulierte notwendige Bedingung für die Existenz von Lösungen $f: \mathbf{R} \to A$ von (10.2.1) auch hinreichend ist. Zunächst gehen wir auf die Frage ein, welche Rolle Torsionselemente bei unserem Problem spielen. Es sei $(B, \circ)$ eine dividierbare Gruppe. Dann hat die Gleichung $v^k = u$ für jede ganze Zahl $k \neq 0$ und jedes $u \in B$ eine (nicht notwendig eindeutige) Lösung $v \in B$. Wir bezeichnen solche Lösungen mit $v = u^{1/k}$. Nun gilt

10.2.5. Lemma. *Es sei $(B, \circ)$ eine dividierbare torsionsfreie Gruppe. Dann hat die Gleichung $v^k = u$ für jede ganze Zahl $k \neq 0$ und jedes $u \in B$ genau eine Lösung $v = u^{1/k} \in B$.*

Beweis. Es seien $u \in B$ und $k \neq 0$ ganzzahlig gegeben. Dann existiert wegen der Dividierbarkeit von $(B, \circ)$ ein $v \in B$ mit $v^k = u$. Es sei nun $w \in B$ mit $w \neq v$ und $w^k = u$. Dann gilt $(w^k)^{-1} = u^{-1}$ und $v^k \circ (w^k)^{-1} = u \circ u^{-1} = n$, wobei n das neutrale Element von $(B, \circ)$ ist. Da $(B, \circ)$ eine Gruppe ist, gilt $(w^k)^{-1} = (w^{-1})^k$. Also ergibt sich $n = u \circ u^{-1} = v^k \circ (w^k)^{-1} = v^k \circ (w^{-1})^k = (w^{-1} \circ v)^k$. Da $(B, \circ)$ torsionsfrei ist, gilt $w^{-1} \circ v = n$ und somit $v = w$, was ein Widerspruch ist.

Nun können wir den folgenden Satz (vgl. W. Schulz [159]) beweisen.

10.2.6. Satz. *Wenn A eine Menge B derart enthält, daß $(B, \circ)$ eine dividierbare torsionsfreie abelsche Gruppe ist, dann existieren Lösungen $f: \mathbf{R} \to A$ von $f(x+y) = F(f(x), f(y))$, die mit Hilfe einer Basis der reellen Zahlen gebildet werden können.*

Beweis. Es sei $c \in B$, dann existiert $c^p \in B$ für jede ganze Zahl p, da $(B, \circ)$ eine Gruppe ist. Wegen der Dividierbarkeit von $(B, \circ)$ existiert $c^{1/q}$ für jede natürliche Zahl $q \neq 0$. Wegen der Torsionsfreiheit von $(B, \circ)$ ist $c^{1/q}$ eindeutig bestimmt (siehe Lemma 10.2.5). Folglich gibt es für jedes $c \in B$ und jede rationale Zahl $r = \frac{p}{q}$ in B ein eindeutig bestimmtes Element c^r ($c^0 := n$). Es sei $\mathbf{R}_B$ eine Basis von $\mathbf{R}$ über $\mathbf{Q}$ und $\bar{f}: \mathbf{R}_B \to B$ eine Funktion. Ferner sei $x \in \mathbf{R}$, und x habe die Darstellung $x = \sum_{i=1}^{m} r_i b_i$ mit $r_i \in \mathbf{Q}$ und $b_i \in \mathbf{R}_B$ für alle i. Für $a_1 \circ a_2 \circ \cdots \circ a_m$ schreiben wir auch $\prod_{i=1}^{m} a_i$[1]). Dann ist die Funktion $f: \mathbf{R} \to B$ mit

$$\textbf{10.2.7.} \quad f(x) := \prod_{i=1}^{m} \left(\bar{f}(b_i)\right)^{r_i}$$

eine Lösung von (10.2.1). Denn ist $y \in \mathbf{R}$ und hat y die Darstellung $y = \sum_{i=1}^{m} r_i' b_i$ (vgl. die Fußnote 1 auf S. 155), dann ist $x + y = \sum_{i=1}^{m} (r_i + r_i') b_i$, und es gilt wegen der Definition von f und der Tatsache, daß $\circ$ eine Gruppenoperation ist,

$$f(x+y) = \prod_{i=1}^{m} \left(\bar{f}(b_i)\right)^{r_i + r_i'} = \prod_{i=1}^{m} \left(\bar{f}(b_i)\right)^{r_i} \circ \prod_{i=1}^{m} \left(\bar{f}(b_i)\right)^{r_i'} = f(x) \circ f(y),$$

also (10.2.3) und somit auch (10.2.1).

10.2.8. Bemerkung. Bei jeder Wahl von $\bar{f}$ ergibt sich bei fester Wahl der Basis $\mathbf{R}_B$ von $\mathbf{R}$ über $\mathbf{Q}$ eine Lösung f. Wie in Kapitel 7 für die Funktionalgleichung $f(x+y) = f(x) + f(y)$ gibt es entsprechend keine Lösungen von $f(x+y) = F(f(x), f(y))$, die nicht von der Form (10.2.7) sind.

Wir betrachten jetzt den Fall, daß $(B, \circ)$ Torsionselemente enthält und geben zunächst ein Beispiel an.

10.2.9. Beispiel. Es sei $(\bar{B}, \circ)$ die Faktorgruppe der Gruppe $\mathfrak{R} = (\mathbf{R}, +)$ nach der Gruppe $\mathfrak{Z} = (\mathbf{Z}, +)$, $\circ$ also die übliche (Komplex- oder) Nebenklassenaddition. Der kanonische Homomorphismus f von $\mathfrak{R}$ auf $(\mathbf{R}/\mathfrak{Z}, \circ)$[2]) ist dann eine Lösung von (10.2.3). $(\mathbf{R}/\mathfrak{Z}, \circ)$ besitzt Torsionselemente. Die Menge der Torsionselemente von $\mathbf{R}/\mathfrak{Z}$ stimmt mit der Menge $\mathbf{Q}/\mathfrak{Z}$ überein.

Es sei nun $(B, \circ)$ eine dividierbare abelsche Gruppe mit Torsionselementen. Wie das Beispiel 10.2.9 zeigt, können auch in diesem Fall Lösungen $f: \mathbf{R} \to A$ von (10.2.1) existieren. Die Konstruktion aus Satz 10.2.6 versagt, denn es existieren zwar Ele-

[1]) Da Mißverständnisse ausgeschlossen sind, verzichten wir auf ein besonderes Zeichen für die Produktbildung bezüglich der Operation $\circ$.

[2]) Vgl. 3.3.

mente $c^r \in B$, aber diese müssen nicht mehr eindeutig bestimmt sein, was für die Konstruktion von Lösungen f mit Hilfe einer Basis der reellen Zahlen wesentlich ist.

Die Menge T der Torsionselemente von B bildet eine Untergruppe $(T, \circ)$ von $(B, \circ)$, wie man sich leicht überzeugt. Diese Gruppe ist ebenfalls dividierbar, denn für jede ganze Zahl $m \neq 0$ und jedes $y \in B$ existiert ein $x \in B$ mit $x^m = y$. Ist $y \in T$, dann gibt es eine ganze Zahl $k \neq 0$ mit $y^k = n$. Daraus ergibt sich $(x^m)^k = x^{m \cdot k} = n$, also $x \in T$.

Wegen Satz 10.1.7 ist $(B, \circ) = (T, \circ) \oplus (D, \circ)$, wobei $(D, \circ)$ eine torsionsfreie dividierbare abelsche Gruppe und $(T, \circ)$ eine direkte Summe von Gruppen der Form $Z(p^\infty)$ ist. Die Lösungen $f\colon \mathbf{R} \to A$ von (10.2.3) stimmen mit den Homomorphismen von $(\mathbf{R}, +)$ in $(B, \circ)$ überein. Daher bezeichnen wir die Menge der Lösungen $f\colon \mathbf{R} \to A$ mit $\mathrm{Hom}\big((\mathbf{R}, +), (B, \circ)\big)$, die mit der üblichen Addition von Homomorphismen abelscher Gruppen eine Gruppe bildet. Nun gilt auch (siehe L. Fuchs [58]), daß diese Gruppe isomorph zur direkten Summe der Gruppen $\mathrm{Hom}\big((\mathbf{R}, +), (T, \circ)\big)$ und $\mathrm{Hom}\big((\mathbf{R}, +), (D, \circ)\big)$ ist. Für den Fall $(D, \circ)$ können wir nach Satz 10.2.6 und Bemerkung 10.2.8 Lösungen angeben. Daher beschäftigen wir uns noch mit der Konstruktion von Lösungen $f\colon \mathbf{R} \to T$. Hierzu betrachten wir den Fall $(T, \circ) = Z(p^\infty)$ für eine feste Primzahl p und untersuchen $Z(p^\infty)$ auf Dividierbarkeit.

Dafür muß geprüft werden, ob für jedes $\alpha = \left[\frac{k}{p^i}\right] \in Z(p^\infty)$ $(i \in \mathbf{N},\ k \in \mathbf{N})$ und jede ganze Zahl $m \neq 0$ ein $\xi = \left[\frac{r}{p^s}\right] \in Z(p^\infty)$ $(s \in \mathbf{N},\ r \in \mathbf{N})$ existiert, so daß $\xi^m = \alpha$ ist. Wegen $\xi^{-m} = (\xi^{-1})^m$ können wir uns auf positive ganze Zahlen m beschränken. Wegen der Definition der Gruppenoperation in $Z(p^\infty)$ ist $\xi^m = m\xi$. Zu $m \in \mathbf{N}^*$ gibt es genau eine zu p teilerfremde Zahl $m' \in \mathbf{N}^*$ und genau ein $j \in \mathbf{N}$ mit $m = m'p^j$.

Wenn die Klasse ξ der Bedingung $m\xi = \alpha$ genügt, gilt $\left[\frac{mr}{p^s}\right] = \left[\frac{k}{p^i}\right]$, und folglich ist $l := \frac{mr}{p^s} - \frac{k}{p^i} \in \mathbf{Z}$. Also ist auch $lp^i + k = \frac{mrp^i}{p^s} = \frac{m'rp^{i-j}}{p^s} \in \mathbf{Z}$. Da $m' \sqcap p = ggT(m', p) = 1$ ist, muß $x := \frac{rp^{i-j}}{p^s} \in \mathbf{Z}$ sein. Jede Lösung $\xi = \left[\frac{r}{p^s}\right]$ der Gleichung $m\xi = \alpha$ kann also in der Form $\xi = \left[\frac{x}{p^{i-j}}\right]$ dargestellt werden, wobei die ganze Zahl x wegen $m'x = k + lp^i$ eine Lösung der Kongruenz $m'x \equiv k \bmod p^i$ ist. Diese Kongruenz besitzt wegen $m' \sqcap p = 1$ stets eine Lösung x_0, und jede Lösung kann in der Form $x_t = x_0 + tp^i$ mit $t \in \mathbf{Z}$ dargestellt werden. Für die Klassen $\xi_t := \left[\frac{x_t}{p^{i-j}}\right] = \left[\frac{x_0}{p^{i-j}}\right] + \left[\frac{t}{p^j}\right]$ gilt nun wegen $m'x_t \equiv k \bmod p^i$ tatsächlich

$$m\xi_t = \left[\frac{mx_t}{p^{i-j}}\right] = \left[\frac{m'x_t}{p^i}\right] = \left[\frac{k}{p^i}\right] = \alpha.$$

Da es genau p^j Klassen der Form $\left[\frac{t}{p^j}\right]$ mit $t \in \mathbf{Z}$ gibt, existieren somit genau p^j Lösungen ξ_t der Gleichung $m\xi = \alpha$.

Insgesamt ist nun gezeigt, daß $Z(p^\infty)$ dividierbar ist und die Gleichung $m\xi = \alpha$

mehr als eine Lösung haben kann. So hat z. B. in $Z(2^\infty)$ die Gleichung $6\xi = \left[\frac{1}{4}\right]$ die Lösungen $\xi_1 = \left[\frac{3}{8}\right]$ und $\xi_2 = \left[\frac{7}{8}\right]$, während die Gleichung $7\xi = \left[\frac{3}{4}\right]$ die Lösung $\xi = \left[\frac{1}{4}\right]$ hat.

Nachdem wir nun einen Überblick über die Anzahl der Lösungen der Gleichung $m\xi = \alpha$ haben, definieren wir für jede natürliche Zahl m mit $m > 1$ eine Funktion $g_m\colon Z(p^\infty) \to Z(p^\infty)$, die aus den Lösungen von $m\xi = \alpha$ eine Lösung auswählt. Dabei können wir erreichen, daß für jedes $x \in Z(p^\infty)$ und alle natürlichen Zahlen u, v, s folgende Beziehungen gelten:

10.2.10. $\big(g_m(x)\big)^m = x$

und

10.2.11. $\big(g_{u\cdot v}(x)\big)^{u\cdot s} = \big(g_v(x)\big)^s$.

Es sei p die Primzahl, die $Z(p^\infty)$ erzeugt. Für alle natürlichen Zahlen a mit $a \sqcap p = 1$ ist g_a eindeutig bestimmt, da $a\xi = \alpha$ genau eine Lösung hat. Zur Festlegung von $g_p(x)$ bestehen p Möglichkeiten. Wir entscheiden uns für $g_p(x) := \frac{x}{p}$. Es sei $m = l \cdot p$. Von den p Möglichkeiten wählen wir die eine, für die $\big(g_{l\cdot p}(x)\big)^l := g_p(x)$ ist. Es sei nun $m = p^w$, dann bestehen für $g_{p^w}(x)$ gerade p^w Möglichkeiten. Wir entscheiden uns für $g_{p^w}(x) := \frac{x}{p^w}$. Für $m = l \cdot p^w$ mit $l = 2, 3, \ldots, p - 1$ verfahren wir entsprechend wie bei $m = l \cdot p$.

Auf diese Weise sind die Funktionen g_m unter Berücksichtigung der Eigenschaften (10.2.10) und (10.2.11) definiert. Nun können wir Lösungen f von (10.2.1) für den Fall $A = Z(p^\infty)$ mit einer Basis $\mathbf{R}_B$ von $\mathbf{R}$ über $\mathbf{Q}$ definieren. Es sei $\mathbf{R}_B$ eine Basis der reellen Zahlen und $\bar{f}\colon \mathbf{R}_B \to Z(p^\infty)$ eine Funktion. Es sei $x \in \mathbf{R}$, und x habe die Darstellung $x = \sum_{i=1}^{m} \frac{s_i}{t_i} b_i$ mit $\frac{s_i}{t_i} \in \mathbf{Q}$ und $s_i \sqcap t_i = 1$, falls $s_i \neq 0$ und $b_i \in \mathbf{R}_B$ für alle i ist. Wir definieren nun eine Funktion $f\colon \mathbf{R} \to Z(p^\infty)$ mit

$$f(x) := \prod_{i=1}^{m} \big(g_{t_i}(\bar{f}(b_i))\big)^{s_i},$$

die eine Lösung von (10.2.1) ist. Denn ist $y \in \mathbf{R}$ und hat y die Darstellung $y = \sum_{i=1}^{m} \frac{s_i'}{t_i'} b_i$ mit $\frac{s_i'}{t_i'} \in \mathbf{Q}$ und $s_i' \sqcap t_i' = 1$, falls $s_i' \neq 0$ und $b_i \in \mathbf{R}_B$ für alle i ist, dann ist

$$x + y = \sum_{i=1}^{m} \frac{s_i t_i' + s_i' t_i}{t_i t_i'} b_i.$$

Wegen der Definition von f und unter Berücksichtigung der Tatsache, daß $\circ$ eine Gruppenoperation ist, und wegen der Eigenschaft (10.2.11) der Funktionen g_m gilt

$$\begin{aligned} f(x+y) &= \prod_{i=1}^{m} \left(g_{t_i t_i'}(\bar{f}(b_i))\right)^{s_i t_i' + s_i' t_i} \\ &= \prod_{i=1}^{m} \left(g_{t_i t_i'}(\bar{f}(b_i))\right)^{s_i t_i'} \circ \prod_{i=1}^{m} \left(g_{t_i t_i'}(\bar{f}(b_i))\right)^{s_i' t_i} \\ &= \prod_{i=1}^{m} \left(g_{t_i}(\bar{f}(b_i))\right)^{s_i} \circ \prod_{i=1}^{m} \left(g_{t_i'}(\bar{f}(b_i))\right)^{s_i'} = f(x) \circ f(y). \end{aligned}$$

f ist also bei jeder Wahl von $\bar{f}$ eine Lösung von (10.2.3) und damit auch von (10.2.1). Weitere Lösungen können sich ergeben, wenn die Funktionen g_m unter Beachtung von (10.2.10) und (10.2.11) auf andere Weise definiert werden.

Nach Satz 10.1.7 besitzt jede dividierbare Gruppe $(B, \circ)$, sofern sie nicht torsionsfrei ist, eine Untergruppe $(U, \circ)$, die isomorph ist zu einer Gruppe $Z(p^\infty)$. Eine Lösung $f\colon \mathbf{R} \to Z(p^\infty)$ können wir daher auch als eine Lösung $f\colon \mathbf{R} \to B$ von (10.2.1) ansehen.

Also gilt (vgl. W. Schulz [160])

10.2.12. Satz. *Wenn A eine Menge B enthält, die mit der Operation $\circ = F$ eine dividierbare abelsche Gruppe bildet, dann existieren Lösungen $f\colon \mathbf{R} \to A$ der Gleichung $f(x+y) = F\big(f(x), f(y)\big)$, die mit einer Basis der reellen Zahlen angegeben werden können.*

Satz 10.2.2 und Satz 10.2.12 zusammengefaßt, ergeben schließlich den

10.2.13. Satz. *Es sei A eine gegebene Menge und $F\colon A \times A \to A$ eine gegebene Funktion. Die Funktionalgleichung $f(x+y) = F\big(f(x), f(y)\big)$ hat genau dann eine Lösung $f\colon \mathbf{R} \to A$, wenn eine Menge $B \subseteq A$ derart existiert, daß $(B, \circ)$ eine dividierbare abelsche Gruppe ist, wobei $\circ$ die gegebene Operation F ist.*

Zur Illustration von Satz 10.2.13. betrachten wir das

10.2.14. Beispiel. Für B wählen wir die Menge $\mathbf{R} \cup \{\infty\}$ und als F die Operation $\circ$, die wie folgt definiert ist: Für alle $u, v \in \mathbf{R}$ ist

$$u \circ v := \begin{cases} \dfrac{u+v}{1-u\cdot v} & \text{für} \quad u \cdot v \neq 1, \\ \infty & \text{für} \quad u \cdot v = 1. \end{cases}$$

Außerdem wird folgendes vereinbart: $\infty \circ \infty := 0$; $\infty \circ u := u \circ \infty := -\dfrac{1}{u}$, falls $u \in \mathbf{R}^*$; $\infty \circ 0 := 0 \circ \infty := \infty$. (10.2.1) hat dann die Form

$$\underset{x,y\in\mathbf{R}}{\wedge} f(x+y) = \frac{f(x)+f(y)}{1-f(x)\cdot f(y)}.$$

Die Funktionen $f\colon \mathbf{R} \to B$ mit

$$f(x) = \begin{cases} \tan \lambda x & \text{für} \quad \lambda x \neq \dfrac{2k+1}{2}\pi, \quad k \in \mathbf{Z}, \quad \lambda \in \mathbf{R}, \\ \infty & \text{für} \quad \lambda x = \dfrac{2k+1}{2}\pi, \quad k \in \mathbf{Z}, \quad \lambda \in \mathbf{R}, \end{cases}$$

gehören zur Lösungsmenge von (10.2.15).

Wir überzeugen uns davon, daß $(B, \circ)$ eine dividierbare abelsche Gruppe mit Torsionselementen ist.
$K = \{e^{i\varphi}: 0 \leqq \varphi < 2\pi\}$ ist die Menge der komplexen Zahlen mit dem Betrag 1. In K wird die Operation $\cdot$ wie folgt definiert:

$$e^{i\varphi_1} \cdot e^{i\varphi_2} := \begin{cases} e^{i(\varphi_1+\varphi_2)} & \text{für } \varphi_1 + \varphi_2 < 2\pi, \\ e^{i(\varphi_1+\varphi_2-2\pi)} & \text{für } \varphi_1 + \varphi_2 \geqq 2\pi. \end{cases}$$

$(K, \cdot)$ ist eine abelsche Gruppe. $(K, \cdot)$ ist auch dividierbar, denn zu gegebenem $u = e^{i\varphi}$ und gegebenem natürlichem $m \neq 0$ ist $v = e^{i\frac{\varphi}{m}}$ eine Lösung der Gleichung $v^m = u$. Die Menge $K_T = \left\{e^{i\varphi}: \varphi = \frac{2\pi r}{m},\ r \in \mathbf{N},\ m \in \mathbf{N}^*,\ r < m\right\}$ der m-ten Einheitswurzeln der Elemente aus K ist die Menge der Torsionselemente von K. Die Abbildung $g: K \to B$ mit

$$g(e^{i\varphi}) := \begin{cases} \tan\frac{\varphi}{2} & \text{für } \varphi \neq \pi, \\ \infty & \text{für } \varphi = \pi \end{cases}$$

ist ein Isomorphismus und überträgt die genannten Eigenschaften von $(K, \cdot)$ auf $(B, \circ)$.

10.2.16. Bemerkung. Aus Satz 10.2.13 ergibt sich, daß für die Existenz von Homomorphismen von $(\mathbf{R}, +)$ auf $(B, \circ)$ letzten Endes das „Radizieren" bezüglich der Operation $\circ$ von entscheidender Bedeutung ist und nicht die Stetigkeit oder die Differenzierbarkeit von $\circ$. Die Stetigkeitsforderung in Satz 9.1.2 sichert nur, daß das „Radizieren" eindeutig ist.

10.2.17. Bemerkung. Unter den Lösungen $f: \mathbf{R} \to A$ von $\bigwedge_{x,y \in \mathbf{R}} f(x+y) = F(f(x), f(y))$ gibt es auch Funktionen, die keine Umkehrfunktionen besitzen, die also nicht eineindeutig sind. Solche Lösungen ergeben sich, wenn für zwei Basiselemente der gleiche Funktionswert gewählt wird. Die eineindeutigen Lösungen f von (10.2.1) besitzen Umkehrfunktionen, die wir als verallgemeinerte Logarithmusfunktionen ansehen.

10.3. Aufgaben

1. Man zeige, daß die Menge der p^m-ten komplexen Einheitswurzeln mit der Multiplikation eine zu $Z(p^\infty)$ isomorphe Gruppe bildet, wobei p eine Primzahl und $m \in \mathbf{N}^*$ ist.

2. Es sei $B = (\mathbf{R} \setminus \{-1\}) \cup \{\infty\}$ und $\circ$ durch

$$u \circ v := \begin{cases} \frac{2 \cdot u \cdot v + u + v}{1 - u \cdot v} & \text{für } u \cdot v \neq 1, \\ \infty & \text{für } u \cdot v = 1 \end{cases}$$

für alle $u, v \in \mathbf{R} \setminus \{-1\}$ und $\infty \circ \infty := -2$; $0 \circ \infty := \infty \circ 0 := \infty$; $\left(-\frac{1}{2}\right) \circ \infty := \infty \circ \left(-\frac{1}{2}\right) := 0$; $\infty \circ u := u \circ \infty := \frac{2u+1}{-u}$, falls $u \in \mathbf{R}^* \setminus \left\{-1, -\frac{1}{2}\right\}$ ist, definiert.

a) Man zeige, daß $(B, \circ)$ eine torsionsfreie dividierbare abelsche Gruppe ist.
b) Welche Form hat $f(x+y) = f(x) \circ f(y)$?
c) Man gebe eine Lösung $f: \mathbf{R} \to B$ von $f(x+y) = f(x) \circ f(y)$ an. (Hinweis: Man bilde $u^{\circ\frac{p}{q}}$ für $p \in \mathbf{Z}$ und $q \in \mathbf{N}^*$.)

11. Die Funktionalgleichung $h(F(x, y)) = f(x) + g(y)$

11.1. Rechenstäbe, Netztafeln, Fluchtlinientafeln und die Schließungsbedingung von Brianchon[1])

In der Klasse 9 (und schon vorher in Klasse 7) wird das Rechnen mit dem Rechenstab behandelt und nicht nur im Mathematikunterricht genutzt. Aufgrund der Tatsache, daß die Logarithmusfunktion (zur Basis 10) ein (stetiger und monotoner) Isomorphismus von der multiplikativen Gruppe $(\mathbf{R}_+^*, \cdot)$ der positiven reellen Zahlen auf die additive Gruppe $(\mathbf{R}, +)$ der reellen Zahlen ist, kann für die zweistellige Operation Multiplikation ein *Additionsstab* gebaut werden. Man trägt auf dem unteren Teil des Stabkörpers und auf der Zunge die logarithmische Skala (zur Basis 10) auf, beschriftet diese Skalen aber nicht mit den Funktionswerten $\log_{10} x = \lg x$, sondern mit den Argumenten x. Zur Berechnung des Produktes $x \cdot y$ stellt man die 1 der Zunge über die Zahl x des Stabkörpers und liest unter der Zahl y der Zunge auf dem Stabkörper das Produkt $x \cdot y$ ab (siehe Abb. 44). Es gilt $c + d = \lg x + \lg y = \lg (x \cdot y)$.

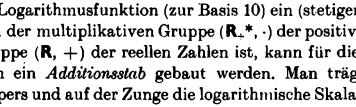


Abb. 44

Da es nun für jede über einem offenen Intervall I reeller Zahlen stetige Gruppenoperation F eine echt monotone Funktion $f\colon \mathbf{R} \to I$ gibt mit $f(x + y) = F(f(x), f(y))$ (vgl. Kap. 9), könnte man auch für jede derartige Operation $F = \circ$ einen solchen Stab konstruieren. Die beiden Skalen sind durch die Funktion f^{-1} gegeben. Für beliebig gewählte Zahlen $u, v \in I$ gibt es nämlich eindeutig bestimmte reelle Zahlen x und y mit $x = f^{-1}(u)$ und $y = f^{-1}(v)$, und aus Bemerkung 9.1.5 folgt damit $f^{-1}(u \circ v) = f^{-1}(u) + f^{-1}(v)$. Dabei würde man allerdings den Definitionsbereich von f^{-1} in geeigneter Weise bzw. für den jeweiligen Zweck benötigter Weise einschränken müssen. Bei den logarithmischen Rechenstäben kann man sich wegen $\lg 10^m = m$ $(m \in \mathbf{Z})$ bekanntlich auf Skalen für die Zahlen x mit $1 \leqq x \leqq 10$ beschränken. So etwas ist bei Rechenstäben für beliebige stetige Gruppenoperationen über offenen Intervallen wegen $f^{-1}(c^{\circ m}) = m$ (vgl. Kap. 9) im Prinzip auch möglich, wird im

[1]) Die Nomographie ist durch die moderne Rechentechnik weitgehend verdrängt worden. Der enge Zusammenhang aber von Grundbegriffen dieses interessanten Gebietes der klassischen Mathematik mit den hier behandelten Problemen, insbesondere mit der Multiplikation und den Logarithmusfunktionen, rechtfertigt wohl die folgenden Ausführungen als zusätzliche Motivation für die in 11.2. durchzuführende Charakterisierung.

allgemeinen aber komplizierter sein als das einfache „Abspalten" von Zehnerpotenzen.

Für die Existenz eines Rechenstabes zu einer Funktion F ist es nun keineswegs notwendig, daß F eine Gruppenoperation ist. Gibt es nämlich stetige und streng monotone Funktionen f, g und h mit

11.1.1. $\bigwedge_{(x,y)\in D(F)} h\big(F(x,y)\big) = f(x) + g(y)$,

so trägt man auf dem Stabkörper und auf der Zunge drei durch die Funktionen f, g und h gegebene Skalen auf (siehe Abb. 45). Es gilt $c + d = f(x) + g(y) = h\big(F(x,y)\big)$.

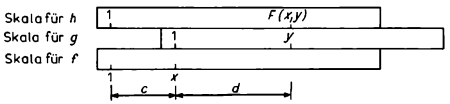


Abb. 45

Diese Funktionen $F = \circ$ sind im allgemeinen weder assoziativ noch kommutativ, wie man aus (11.1.1) sofort erkennt. Wir können (11.1.1) auch als *verallgemeinertes Logarithmengesetz* bezeichnen. Für $F = \cdot$ (Multiplikation positiver reeller Zahlen) und für $h = f = g = \log_a$ erhält man ja aus (11.1.1) ein bekanntes und im Mathematikunterricht behandeltes Logarithmengesetz.[1])

In 11.2. werden wir Klassen zweistelliger Funktionen F charakterisieren, für die es eine Darstellung (11.1.1) gibt. Dabei werden die Funktionen h, f und g wie in 8.1. $\big(f(x) = a^x\big)$ und in 9.1. $\big(f(x) = a^{\circ x}\big)$ durch schrittweise Erweiterung des Potenzbegriffs konstruiert.

Wir wollen die nomographische Bedeutung der Gleichung (11.1.1) noch auf andere Weise untersuchen.

Nomogramme (z. B. von zweistelligen Funktionen F mit $z = F(x,y)$) sind graphische Darstellungen, aus denen man eine der Zahlen x, y oder z ablesen kann, wenn die beiden anderen gegeben sind. (Wir werden stets voraussetzen, daß F (wie die Multiplikation) stetig und in beiden Argumenten streng monoton ist.) Additionsstäbe sind solche Nomogramme. Verbreitete Nomogramme, auf die wir hier etwas näher eingehen wollen, sind Netztafeln und Fluchtlinientafeln.

Zunächst sei kurz skizziert, wie man zu einer Funktion F mit $z = F(x,y)$ eine *Netztafel* konstruiert. Bekanntlich (vgl. MfL Bd. 5) kann man reellwertige Funktionen F, die in einem Gebiet[2]) definiert sind, unter gewissen Voraussetzungen durch eine Fläche im dreidimensionalen Raum veranschaulichen. Diese Fläche ist die Menge aller Punkte $P\big(x,y,F(x,y)\big)$ mit $(x,y) \in D(F)$. Zur graphischen Darstellung von F in der x,y-Ebene denke man sich die Fläche im dreidimensionalen Raum durch Ebenen $z = \text{const}$ geschnitten und die so entstandenen Schnittkurven (Höhen-,

[1]) Aufgrund der Voraussetzungen sind die Funktionen f, g und h 1-1-Abbildungen. Deshalb ist das Tripel $T = (f,g,h)$ ein Isotopismus vom Gruppoid $(I, \circ)$ auf die abelsche Gruppe $(\mathbf{R}, +)$, wobei $I \times I$ der Definitionsbereich von F ist. Vgl. 4.1.

[2]) Die Funktionen F sind im folgenden stets reellwertig und über einem offenen Rechteck des Raumes $\mathbf{R}^2$ definiert.

Schicht- oder *Niveaulinien*), in denen F jeweils denselben Funktionswert hat, in die x,y-Ebene projiziert.

Wir werden im folgenden stets voraussetzen, daß F stetig und in beiden Argumenten streng monoton ist. In diesem Fall sind die Niveaulinien Graphen stetiger und streng monotoner einstelliger Funktionen. Auf den Beweis dieses anschaulich klaren Sachverhalts verzichten wir. Da Niveaulinien zusammenfallen oder keinen Punkt gemeinsam haben, bilden sie eine Kurvenschar, die zusammen mit den Geraden $x = \text{const}$ und $y = \text{const}$ eine Netztafel[1]) bilden (siehe Abb. 46).

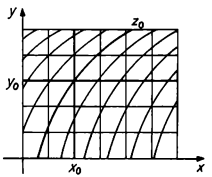


Abb. 46

Ein Tripel (x_0, y_0, z_0) mit $z_0 = F(x_0, y_0)$ ist als Schnittpunkt von drei Kurven $x = x_0$, $y = y_0$ und $z = z_0$ gegeben.

11.1.2. Beispiel (vgl. Beispiel 4.1.5.3). Die Funktion mit $z = x \cdot y^2$ $(x, y \in \mathbb{R}_+^*)$ ist durch eine Netztafel darzustellen. Die Kurven $z = \text{const}$ sind die Hyperbeläste $y = \frac{\sqrt{z_0}}{\sqrt{x}}$ (vgl. Abb. 47).

Aus der Netztafel kann man z bzw. y bzw. x bestimmen, wenn x und y bzw. x und z bzw. y und z gegeben sind. Natürlich können nicht alle Kurven $z = \text{const}$ eingezeichnet werden, so daß man i. allg. interpolieren muß. Solche Interpolationen sind besonders einfach, wenn alle Kurven Geraden sind. Für die Funktion mit $z = x \cdot y^2$ läßt sich dies durch Logarithmieren

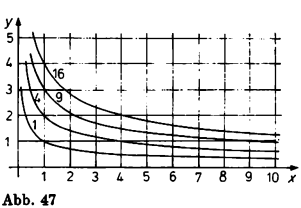


Abb. 47

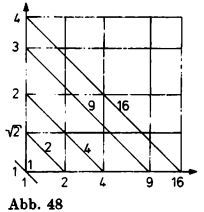


Abb. 48

leicht erreichen. Man erhält $\lg z = \lg x + 2 \cdot \lg y$ und trägt auf die x-Achse und die y-Achse entsprechende logarithmische Skalen auf, beschriftet aber diese Skalen wieder mit den Argumenten (vgl. Abb. 48). Es ist eine geradlinige Netztafel entstanden. Durch Streckung der Skala auf der y-Achse im Verhältnis 2:1 schneiden die Geraden der Schar $z = \text{const}$ die Geraden der beiden anderen Scharen unter einem Winkel von 135°.

[1]) Man sagt auch, daß diese drei Kurvenscharen ein „3-Gewebe", „Kurvengewebe" oder einfach „Gewebe" bilden (vgl. etwa W. Blaschke und G. Bol [24], J. Aczél [4]). Für den endlichen Fall vergleiche man auch 5.3.

Im allgemeinen Fall besteht die Netztafel einer Funktion F mit $z = F(x, y)$ aus drei Scharen von Kurven. Eine Schar ist Träger der in Frage kommenden x-Werte, eine zweite Träger der y-Werte und eine dritte Träger der z-Werte (vgl. Abb. 49). Ein wichtiges Problem der *Nomographie* besteht nun darin, diejenigen Funktionen auszusondern, deren Nomogramme durch geeignete Transformationen in drei Scharen von Geraden übergeführt werden können. Man sagt, daß diese Funktionen eine *Anamorphose* zulassen (vgl. Abb. 50). Läßt sich das Nomogramm sogar in drei

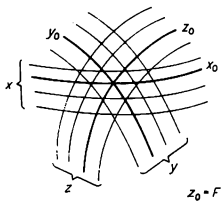


Abb. 49

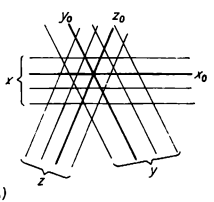


Abb. 50

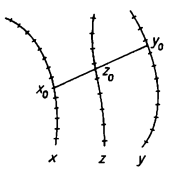


Abb. 51

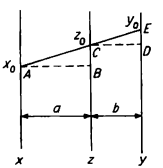


Abb. 52

Scharen paralleler Geraden transformieren, so spricht man von *paralleler Anamorphose*. In diesem Fall kann man erreichen, daß die Geraden zweier Scharen jeweils zueinander senkrecht stehen und außerdem die Geraden der dritten Schar die Geraden der beiden anderen Scharen unter einem Winkel von 135° schneiden. Bezeichnet man die Transformationsfunktionen der ursprünglichen Skalen x, y und z mit $u = f(x)$, $v = g(y)$ und $w = h(z)$, so gilt für $w = \text{const}$ auf den Geraden der dritten Schar $w = u + v$, also $h(z) = h\big(F(x, y)\big) = f(x) + g(y)$, also wieder (11.1.1). Umgekehrt besitzt jede Funktion F, für die es eine Darstellung (11.1.1) gibt, ein solches Nomogramm.

Es sei schließlich noch angedeutet, wie man von einer gegebenen *Fluchtlinientafel* auf die Gleichung (11.1.1) kommt. Ein derartiges Nomogramm für die Funktion F mit $z = F(x, y)$ besteht aus drei Kurven (Funktionsleitern), auf die die Skalen für x, y und z aufgetragen sind. Werte x_0, y_0 und z_0 mit $z_0 = F(x_0, y_0)$ liegen dann auf einer Geraden, der *Fluchtgeraden* (die drei Punkte liegen in einer „Flucht") (vgl. Abb. 51). Die Form einer Fluchtlinientafel hängt natürlich von der Funktion F ab, von der ein Nomogramm hergestellt werden soll. In bestimmten Fällen ist es möglich, eine Fluchtlinientafel mit drei parallelen Leitern zu konstruieren (vgl. Abb. 52).

Auf den Leitern, deren Teilung durch die Funktionen f, g und h bestimmt wird, sind wieder nur die Argumente x, y und z vermerkt. Aus der Ähnlichkeit der Dreiecke $\triangle ABC$ und $\triangle CDE$ folgt $\frac{\overline{BC}}{a} = \frac{\overline{ED}}{b}$, also $\frac{h(z) - f(x)}{a} = \frac{g(y) - h(z)}{b}$, und daraus

$$h(z) = \frac{b}{a+b} f(x) + \frac{a}{a+b} g(y)$$

oder

$$h(z) = h\big(F(x, y)\big) = \bar{f}(x) + \bar{g}(y).$$

Also gibt es für F auch in diesem Fall einstellige Funktionen h, f und g, so daß (11.1.1) gilt. Umgekehrt besitzt jede Funktion F, für die es eine Darstellung (11.1.1) gibt, eine Fluchtlinientafel mit drei parallelen Leitern.

Es gibt also für zweistellige stetige und in beiden Argumenten streng monotone Funktionen F genau dann einstellige und streng monotone Funktionen h, f und g mit (11.1.1), wenn für F eine Netztafel mit drei Scharen paralleler Geraden oder eine Fluchtlinientafel mit drei parallelen Leitern existiert.

Zwischen den Netztafeln und den Fluchtlinientafeln von Funktionen, die eine Anamorphose zulassen, kann auch ein geometrischer Zusammenhang hergestellt werden. Dazu benötigt man den in der projektiven Geometrie so wichtigen Begriff der Dualität (vgl. auch MfL Bd. 7). Durch diese Dualität wird jedem Punkt der projektiven Ebene eine Gerade dieser Ebene umkehrbar eindeutig zugeordnet (Punkte und Geraden sind zueinander duale Gebilde). Jeder Punktreihe (Menge aller Punkte auf einer Geraden) entspricht als duales Bild ein Geradenbüschel (Menge aller Geraden durch einen Punkt). Der Verbindungsgeraden zweier Punkte P_1 und P_2 mit $P_1 \neq P_2$ ist der Schnittpunkt der den Punkten P_1 und P_2 zugeordneten Geraden g_1 und g_2 zugeordnet. Dabei ist zu beachten, daß zwei verschiedene Geraden der projektiven Ebene stets genau einen Schnittpunkt haben.

Es sei nun F mit $z = F(x, y)$ eine Funktion, die eine parallele Anamorphose zuläßt, die also eine Netztafel besitzt, die aus drei Scharen von Parallelen besteht. Jede dieser Parallelenscharen hat als Bild eine Punktreihe, also eine Punktschar, die auf einer Geraden liegt, da alle Geraden der Parallelenschar durch einen Punkt gehen. Die Geraden x_0, y_0 und z_0 mit $z_0 = F(x_0, y_0)$ haben in der Netztafel einen gemeinsamen Schnittpunkt, also liegen die diesen Geraden zugeordneten Punkte X_0, Y_0 und Z_0 auf einer Geraden, nämlich auf der Fluchtgeraden. Da sich schließlich erreichen läßt, daß die den drei Parallelenscharen zugeordneten Punktscharen auf drei parallelen Geraden liegen, sind also die Fluchtlinientafeln mit parallelen Leitern die dualen Bilder der Netztafeln mit drei Parallelenscharen.

Gehen die Geraden einer Schar nicht durch einen gemeinsamen Punkt, so liegt die duale Punktschar nicht auf einer Geraden. Man erhält unter bestimmten Voraussetzungen für $z = F(x, y)$ als duales Bild einer solchen Schar eine gekrümmte Skala. Die allgemeinen Fluchtlinientafeln (vgl. Abb. 51) sind also die projektiven Bilder von Netztafeln, die aus drei beliebigen Scharen von Geraden bestehen.[1])

[1]) Weitergehende Ausführungen über diese Zusammenhänge findet man in W. Meyer zur Capellen [117] und M. W. Pentkowski [130].

Schließlich sei noch darauf hingewiesen, daß einfache Aufgaben der Nomographie auch im Mathematikunterricht vorkommen bzw. in ihm angelegt sind. Zum Beispiel werden die Graphen der elementaren Funktionen gezeichnet, aber an diesen Graphen werden auch numerische Bestimmungen vorgenommen. So werden etwa quadratische Gleichungen dadurch gelöst, daß der Graph der zugehörigen Parabel mit einer Schablone gezeichnet wird und aus dem Bild die Nullstellen abgelesen werden.

In den unteren Klassen wird mit Diagrammen gearbeitet, die „verstümmelte" Nomogramme zweistelliger Funktionen darstellen. In Klasse 3 (vgl. Mathematik-Lehrbuch für Klasse 3, [113], S. 72) arbeiten die Schüler z. B. mit dem in Abb. 53 dargestellten Diagramm. Es handelt sich hier um die graphische Darstellung einiger Wertepaare der einstelligen Funktion f mit $y = f(x) = x - 200$. Fassen wir das Diagramm so auf, daß den eingetragenen Punkten jeweils die Summe der betreffenden Koordinaten zugeordnet ist, und zeichnen alle anderen Punkte im ersten Quadran-

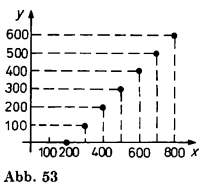


Abb. 53

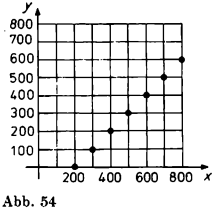


Abb. 54

ten ein, denen jeweils dieselbe Summe zugeordnet ist, so erhalten wir eine Netztafel für die Funktion F mit $z = F(x, y) = x + y$ (vgl. Abb. 54). Für festes $z = c$ erhält man aus dieser Netztafel eine ganze Reihe von Nomogrammen, in der das der Abb. 53 enthalten ist.

Die bereits angekündigte funktionale Charakterisierung von Klassen zweistelliger Funktionen F, die eine Darstellung (11.1.1) besitzen[1]), erfordert bekanntlich, notwendige und hinreichende Bedingungen anzugeben, denen solche Funktionen genügen müssen. Das sind also Bedingungen, die einerseits aus der Darstellung (11.1.1) von F folgen und die andererseits die Existenz von Funktionen h, f und g mit (11.1.1) zur Folge haben. Auf der Suche nach solchen Bedingungen für F wird man folgendermaßen auf den bereits in 4.2. behandelten Schließungssatz von REIDEMEISTER geführt. An dieser Stelle interessiert uns jedoch eine spezielle Reidemeisterfigur, die sogenannte *Sechseckbedingung* (von CH. J. BRIANCHON (1783—1864)).

Wir betrachten eine Funktion F mit einer Darstellung (11.1.1), die also eine parallele Anamorphose zuläßt und damit eine Netztafel besitzt, in der die Geraden $z = \text{const}$ die Geraden der beiden anderen Scharen unter einem Winkel von 135°

[1]) (11.1.1) ist eine Funktionalgleichung, in der die Funktion F bekannt und die Funktionen h, f und g unbekannt sind. Es geht dann darum, Klassen von Funktionen F anzugeben, für die (11.1.1) lösbar ist, d. h. (in der Sprache von 4.2.), es gilt isotope Quasigruppen der Gruppe $(\mathbf{R}, -)$ zu finden. Es sind also Existenzsätze zu beweisen.

schneiden. Es sei P ein beliebiger Punkt dieser Netztafel (vgl. Abb. 55). Auf der Parallelen durch P zur x-Achse wählen wir, ebenfalls beliebig, einen Punkt B. Die Parallele zur y-Achse durch B schneidet die z-Linie durch P in C. Die Parallelen zur x-Achse durch C und zur y-Achse durch P schneiden sich in D, und die z-Linie durch D schneidet die Parallele zur x-Achse durch P in E. Die Parallele zur y-Achse durch E schneidet die z-Linie durch P in F, und die Parallelen zur x-Achse durch F und zur y-Achse durch P schneiden sich in G. Aufgrund der vorliegenden geometrischen Verhältnisse liegen B und G auf derselben z-Linie. Das heißt mit anderen Worten: Die durchgeführte Sechseckkonstruktion mit dem Anfangspunkt B hat diesen Punkt auch als Endpunkt, das Sechseck schließt sich. Wie man sich leicht klar macht, kann man die schrittweise Konstruktion der Punkte B, C, D, E, F und G in beliebigen

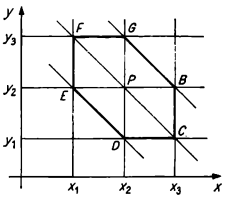


Abb. 55

Geweben nachvollziehen. Allerdings wird man i. allg. nicht erwarten können, daß G und B auf derselben z-Linie liegen, daß sich das Sechseck schließt. Das hängt von der Funktion F ab, um deren Netztafel es sich handelt. Es gibt Funktionen, bei denen sich diese Sechsecke schließen, und solche, bei denen das nicht der Fall ist (bei denen man auf diese Weise also gar kein Sechseck erhält).

Wir wollen die Tatsache, daß sich das Sechseck in Abb. 55 schließt, daß also G und B auf derselben z-Linie liegen, wenn dies auch für D und E bzw. für C, P und F der Fall ist, mit Hilfe der Koordinaten dieser Punkte ausdrücken. Da F auf jeder z-Linie (Niveaulinie) jeweils denselben Wert hat, gilt also für alle (x_i, y_j) $(i, j = 1, 2, 3)$ aus $D(F)$: Wenn $F(x_1, y_2) = F(x_2, y_1)$ und $F(x_1, y_3) = F(x_2, y_2) = F(x_3, y_1)$, so $F(x_2, y_3) = F(x_3, y_2)$. Das ist gerade die *Sechseckbedingung*[1]). Wir werden zeigen, daß diese Bedingung von jeder Funktion F, die eine Darstellung der Form (11.1.1) besitzt, trivialerweise erfüllt ist, daß diese Bedingung also für (11.1.1) notwendig ist. Daß sie auch hinreichend ist, wird dann für bestimmte Funktionenklassen in 11.2. bewiesen bzw. für noch umfassendere Klassen mitgeteilt.

Es gibt weitere derartige Schließungsbedingungen. Neben den bereits genannten gibt es z. B. noch die drei Bolschen Bedingungen (vgl. G. Bol [26], G. Thomsen [170], W. Blaschke und G. Bol [24], J. Aczél [4]). Über ihre Rolle im Zusammenhang mit algebraischen Strukturen findet man über das bereits in 1.6. und 4.2. Gesagte hinaus viele weitere Beziehungen sowohl zwischen diesen Schließungssätzen als

[1]) Man erhält sie als Spezialfall der Schließungsbedingung von Reidemeister (vgl. (R) in 1.6. und (R) in 4.2.), indem man $x_3 = x_2$ und $y_3 = y_2$ setzt.

auch zu anderen algebraischen Strukturen bei J. Aczél [4], S. Gołąb [60], V. D. Belousov [22]. Wir wollen hier auf diese (sehr interessanten) Zusammenhänge und Beziehungen jedoch nicht weiter eingehen. Uns kommt es hier in erster Linie darauf an, Konstruktionsmethoden für die Funktionen h, f und g aus (11.1.1) vorzuführen und dabei insbesondere herauszuarbeiten, daß unter bestimmten Voraussetzungen über F (vgl. Satz 11.2.3) auch hierfür die von der Definition der Funktion f mit $f(x) = a^x$ (vgl. Kap. 8) her bekannte schrittweise Erweiterung des Potenzbegriffes eingesetzt werden kann. Damit haben wir ein weiteres Beispiel für die starke Verallgemeinerungsfähigkeit der Potenzbildung. Für diesen Zweck benötigen wir lediglich die Sechseckbedingung. Sie sei hier noch einmal formuliert, und zwar gleich für zweistellige monotone Funktionen F mit dem Definitionsbereich $D(F)$:

11.1.3. Definition. Die zweistellige reellwertige und in beiden Argumenten streng monotone Funktion F erfüllt in $D(F)$ die *Schließungsbedingung* (H)[1] (von Brianchon): $\Leftrightarrow$

$$\bigwedge_{\substack{x_i,y_i\in\mathbb{R}\\ i,j=1,2,3}} \Big(\big((x_1,y_2),(x_2,y_1),(x_1,y_3),(x_2,y_2),(x_3,y_1),(x_2,y_3),(x_3,y_2)\in D(F)$$
$$\wedge\, F(x_1,y_2)=F(x_2,y_1)\wedge F(x_1,y_3)=F(x_2,y_2)=F(x_3,y_1)\big)$$
$$\Rightarrow F(x_2,y_3)=F(x_3,y_2)\Big)$$

(vgl. Abb. 56 und Abb. 55).

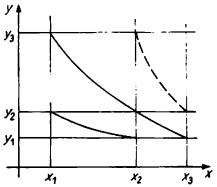


Abb. 56

In dieser Definition wird für die x_i und die y_i ($i = 1, 2, 3$) keine bestimmte Reihenfolge gefordert, so daß nicht notwendig $x_1 < x_2 < x_3$ und $y_1 < y_2 < y_3$ gelten muß (wie in Abb. 56). Wegen der strengen Monotonie von F sind andererseits auch nicht beliebige Reihenfolgen möglich.[2]) Für $x_3 < x_2 < x_1$ und $y_3 < y_2 < y_1$ vertauschen die beiden äußeren Niveaulinien ihre Rolle, so daß man (geometrisch ausgedrückt) auch von den beiden oberen Niveaulinien auf die untere schließen kann. Dagegen kann man nicht von den beiden äußeren Linien auf die mittlere schließen. Bei der in 11.2. durchzuführenden Charakterisierung wird es notwendig sein, in der Sechseck-

[1]) Abkürzung für Hexagon (Sechseck). Die Bedingung (H) wird auch Sechseckbedingung genannt.

[2]) Der Leser veranschauliche sich diese Verhältnisse an Skizzen nach dem Muster der Abb. 56.

Figur von den beiden äußeren Niveaulinien und einem Abschnitt der mittleren Niveaulinie auf den zweiten Abschnitt dieser Linie schließen zu können. Dazu definieren wir:

11.1.4. Definition. Die zweistellige reellwertige und in beiden Argumenten streng monotone Funktion F erfüllt in $D(F)$ die *Schließungsbedingung* (H'):

$$\bigwedge_{\substack{x_i, y_i \in \mathbb{R} \\ i,j=1,2,3}} \Big(\big((x_1, y_2), (x_2, y_1), (x_1, y_3), (x_2, y_2), (x_3, y_1), (x_2, y_3), (x_3, y_2) \in D(F)$$
$$\wedge\, F(x_1, y_2) = F(x_2, y_1) \wedge F(x_2, y_3) = F(x_3, y_2) \wedge F(x_1, y_3) = F(x_2, y_2)\big)$$
$$\Rightarrow F(x_2, y_2) = F(x_3, y_1)\Big)$$

(vgl. Abb. 57).

Dann gilt der

11.1.5. Hilfssatz. *Die Bedingungen* (H) *und* (H') *sind äquivalent.*

Beweis (vgl. J. Aczél [4]). a) *Aus* (H) *folgt* (H'). Es sei F eine Funktion, für die (H) gilt, und es seien weiterhin (x_1, y_2), (x_2, y_1), (x_1, y_3), (x_2, y_2), (x_2, y_3) und (x_3, y_2)

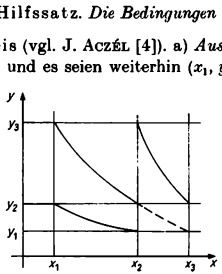


Abb. 57

beliebige Zahlenpaare aus $D(F)$, für die auch $(x_3, y_1) \in D(F)$ ist und die Voraussetzungen von (H') gelten. Dann ist zu zeigen, daß $F(x_2, y_2) = F(x_3, y_1)$ gilt. Dazu definieren wir eine Zahl $\bar{x}_3$ durch

11.1.6. $F(x_2, y_2) = F(\bar{x}_3, y_1)$.

Wegen der vorausgesetzten Stetigkeit und der strengen Monotonie von F gibt es höchstens eine solche Zahl. Daß auch $(\bar{x}_3, y_1) \in D(F)$ gilt, daß also die Niveaulinie durch (x_1, y_3) und (x_2, y_2) die Parallele zur x-Achse durch $(0, y_1)$ noch in $D(F)$ schneidet, ist zunächst nicht sicher. Es wird sich aber zeigen, daß $\bar{x}_3 = x_3$ ist, daß es also insgesamt genau eine Zahl $\bar{x}_3$ mit (11.1.6) gibt. Nun gelten für x_1, x_2, $\bar{x}_3$, y_1, y_2 und y_3 die Voraussetzungen von (H):

$$F(x_1, y_2) = F(x_2, y_1) \quad \text{und} \quad F(x_1, y_3) = F(x_2, y_2) = F(\bar{x}_3, y_1),$$

so daß $F(x_2, y_3) = F(\bar{x}_3, y_2)$ ist. Mit der zweiten Voraussetzung von (H') folgt daraus $F(x_3, y_2) = F(\bar{x}_3, y_2)$, also ist aufgrund der Kürzbarkeit $\bar{x}_3 = x_3$. Damit geht (11.1.6) über in $F(x_2, y_2) = F(x_3, y_1)$, und das ist die Behauptung von (H').

b) *Aus* (H') *folgt* (H). Der Beweis dieser Behauptung wird ganz analog geführt.

Schließlich kann eine Schließungsbedingung (H'') formuliert werden, in der auf den „oberen Abschnitt" der mittleren Niveaulinie geschlossen wird. Die Formulierung dieser Bedingung und der Beweis der Äquivalenz bleibt ebenfalls dem Leser überlassen. Die Äquivalenz folgt auch unmittelbar daraus, daß (H'') und (H') gleichbedeutend sind. Man braucht in (H') nur x_1 mit x_3 und y_1 mit y_3 zu vertauschen.

Der folgende Satz stellt einen ersten Zusammenhang her zwischen zweistelligen Funktionen, die eine parallele Anamorphose zulassen, und der Schließungsbedingung (H).

11.1.7. Satz. *Jede zweistellige reellwertige Funktion F, für die es einstellige, stetige und streng monotone Funktionen h, f und g gibt mit*

$$\textbf{11.1.1.}' \quad \bigwedge_{(x,y)\in D(F)} F(x, y) = h^{-1}\big(f(x) + g(y)\big),$$

erfüllt in $D(F)$ die Bedingung (H).

Beweis (vgl. F. Radó [144], J. Aczél [4] sowie Satz 4.2.8). Wir setzen voraus, daß alle auftretenden Punkte im Definitionsbereich von F liegen. Mit (11.1.1') lauten die Voraussetzungen von (H):

$$h^{-1}\big(f(x_1) + g(y_2)\big) = h^{-1}\big(f(x_2) + g(y_1)\big)$$

und

$$h^{-1}\big(f(x_1) + g(y_3)\big) = h^{-1}\big(f(x_2) + g(y_2)\big) = h^{-1}\big(f(x_3) + g(y_1)\big).$$

Wegen der strengen Monotonie von h^{-1} sind auch jeweils die Argumente von h^{-1} gleich. Man erhält durch Subtraktion $f(x_2) - f(x_3) = g(y_2) - g(y_3)$ und daraus $h^{-1}\big(f(x_2) + g(y_3)\big) = h^{-1}\big(f(x_3) + g(y_2)\big)$. Das ist die Behauptung von (H): $F(x_2, y_3) = F(x_3, y_2)$.[1])

Wir wollen anschließend in 11.2. für gewisse Klassen von Funktionen F auch umgekehrt beweisen: Wenn sich alle Sechseck-Figuren schließen, die im Definitionsbereich von F liegen, dann gilt (11.1.1). Mit Satz 11.1.7 sind dann diese zweistelligen reellwertigen stetigen Funktionen mit paralleler Anamorphose durch die Sechseckbedingung (H) charakterisiert.

11.1.8. Aufgaben

1. Man zeichne je eine Netztafel für die Funktionen F mit
a) $z = F(x, y) = x^2 \cdot y$ $(x, y \in \mathbf{R}_+^*)$,
b) $z = F(x, y) = x^2 \cdot y^3$ $(x, y \in \mathbf{R}_+^*)$,
c) $z = F(x, y) = \frac{1}{2} x \cdot y$ $(x, y \in \mathbf{R}_+^*)$.

2. Man gebe für die drei Funktionen aus 11.1.8.1 jeweils Funktionen h, f und g mit $h(F(x, y)) = f(x) + g(y)$ an.

3. Wie müssen die Skalen auf der x- bzw. y-Achse in Aufgabe 11.1.8.2 jeweils verändert werden, damit die z-Geraden die x-Achse unter einem Winkel von 135° schneiden?

[1]) Dazu wurden nicht alle Voraussetzungen von (H) gebraucht. Es wurden nur die Voraussetzungen der Schließungsbedingung von Thomsen (vgl. 1.6. und 4.2.) benutzt.

11.2. Charakterisierung einer Klasse zweistelliger reellwertiger Funktionen F mit $h(F(x, y)) = f(x) + g(y)$

Im folgenden betrachten wir stetige Funktionen $F: I \times I \to I$ mit $z = F(x, y)$, wobei I ein offenes (nicht notwendig endliches) Intervall $]a, b[$ ist. Der Definitionsbereich ist also das (quadratische) Gebiet $G = \{(x, y): a < x < b \wedge a < y < b\}$. Solche Funktionen können wir auch wie die Addition und Multiplikation reeller Zahlen als Operationen bezeichnen (vgl. MfL Bd. 1 und Kap. 1). Wir fordern außerdem, daß diese Operationen (ebenfalls wie die Addition in $\mathbf{R}$ und die Multiplikation in $\mathbf{R}_+$*) umkehrbar und kürzbar sein sollen (vgl. 1.2.). Setzen wir noch $\circ := F$, so ist also die Struktur $(I, \circ)$ eine stetige Quasigruppe (vgl. 1.3.).

Die Forderung nach Umkehrbarkeit stellt eine recht starke Einschränkung dar. Wir werden am Schluß dieses Kapitels noch einen Satz erwähnen, in dem diese Forderung fallengelassen wird und (neben der Stetigkeit) für F nur strenge Monotonie in beiden Argumenten vorausgesetzt wird. Daß es sich dabei wirklich um eine Erweiterung handelt, zeigen die folgenden Überlegungen.

Jede in beiden Argumenten streng monotone Funktion F mit $z = F(x, y)$ ist kürzbar. Aus $F(x, y_1) = F(x, y_2)$ folgt dann ja $y_1 = y_2$. Eine solche Funktion muß aber nicht umkehrbar sein. So ist z. B. die Funktion F mit $F(x, y) = x \cdot y$ und $x, y \in]4, \infty[$ ($\cdot$ ist die Multiplikation reeller Zahlen) streng monoton und damit kürzbar, aber nicht umkehrbar, denn z. B. hat die Gleichung $5 \cdot x = 10$ keine Lösung.

Umgekehrt ist jede über einem Intervall $I =]a, b[$ stetige und kürzbare Operation F (und damit erst recht jede stetige Quasigruppenoperation) in beiden Argumenten streng monoton, so daß diese Funktionen Spezialfälle der in 11.1. betrachteten Funktionen sind. Zum Beweis dieser Behauptung sei $x \in I$ fest gewählt, und es gelte $F(x, y_1) < F(x, y_2)$ für zwei Zahlen $y_1, y_2 \in I$ mit $y_1 < y_2$. Dann ist F für dieses x in y streng monoton wachsend. Gäbe es nämlich Zahlen $\overline{y}_1, \overline{y}_2 \in I$ mit $\overline{y}_1 < \overline{y}_2$ und $F(x, \overline{y}_1) \geqq F(x, \overline{y}_2)$, so ergäbe zunächst die Gleichheit sofort einen Widerspruch zur Kürzbarkeit von F. Im anderen Fall gäbe es drei Zahlen t_1, t_2 und t_3 mit $t_1 < t_2 < t_3$ derart, daß $F(x, t_1) - F(x, t_2)$ und $F(x, t_2) - F(x, t_3)$ verschiedene Vorzeichen hätten. Es gelte etwa $F(x, t_1) < F(x, t_2)$ und $F(x, t_2) > F(x, t_3)$. Aufgrund des Zwischenwertsatzes gäbe es dann für jede Zahl c mit $\operatorname{Max}\{F(x, t_1), F(x, t_3)\} < c < F(x, t_2)$ Zahlen η_1 und η_2 mit $\eta_1 \in]t_1, t_2[$, $\eta_2 \in]t_2, t_3[$ und $F(x, \eta_1) = F(x, \eta_2) = c$. Daraus würde wegen der Kürzbarkeit von F aber $\eta_1 = \eta_2$ folgen. Damit ist das streng monotone Wachsen in y für ein festes x bewiesen. Darüber hinaus gilt sogar, daß F dann für jedes $x \in I$ in y monoton wächst. Gäbe es nämlich Zahlen x_0 und x_1, so daß für alle $y_1, y_2 \in I$ mit $y_1 < y_2$ und $F(x_0, y_1) < F(x_0, y_2)$ und $F(x_1, y_1) > F(x_1, y_2)$ gelten würde, so gäbe es wieder wegen des Zwischenwertsatzes ein $x_2 \in I$ mit $F(x_2, y_1) = F(x_2, y_2)$. Damit wäre $y_1 = y_2$ wegen der Kürzbarkeit von F.

Im weiteren Verlauf des Beweises werden wir von der gegebenen Struktur $(I, \circ)$ durch folgende Definition zu einer Struktur $(I, \cdot)$ übergehen:

11.2.1. Definition. Für beliebige, aber fest gewählte Zahlen $u, v \in I$ sei $H(x, y) = x \cdot y := x' \circ y' = F(x', y')$ für alle $x, y \in I$, wobei x' die Lösung der Gleichung $x = x' \circ v$ und y' die Lösung der Gleichung $y = u \circ y'$ ist.

Wie wir aus den Sätzen 4.1.12 und 4.1.15 wissen, ist $(I, \cdot)$ das durch die Zahlen u und v eindeutig bestimmte Loop-Hauptisotop von $(I, \circ)$. Wir wollen das für den vorliegenden Fall noch einmal beweisen.

Zunächst ist $(I, \cdot)$ eine stetige Quasigruppe, die ein neutrales Element besitzt, d. h. eine stetige Loop.

Die Funktion $H = \cdot$ ist für alle $x, y \in I$ definiert, denn die Gleichungen $x = x' \circ v$ und $y = u \circ y'$ haben wegen der Umkehrbarkeit und wegen der Kürzbarkeit von $\circ$ eindeutig bestimmte Lösungen $x', y' \in I$. Der Funktionswert $x \cdot y = x' \circ y'$ liegt nach Definition von $\circ$ wieder in I, also ist $H = \cdot$ eine Operation in I. Diese Operation ist auch umkehrbar und kürzbar, d. h., für alle $c, d \in I$ haben die Gleichungen $c \cdot x = d$ bzw. $y \cdot c = d$ eindeutig bestimmte Lösungen $x \in I$ bzw. $y \in I$. Wir weisen die Richtigkeit der ersten Behauptung nach. Die Gleichung $c' \circ x' = d$, wobei c' durch $c = c' \circ v$ eindeutig bestimmt ist, hat eine eindeutig bestimmte Lösung x'. Dann ist die Zahl x mit $x = u \circ x'$ eindeutig bestimmte Lösung von $c \cdot x = d$, denn es gilt

$$d = c' \circ x' = (c' \circ v) \cdot (u \circ x') = c \cdot x.$$

Die Zahl $n := u \circ v$ ist neutrales Element von $(I, \cdot)$, denn sind $x, y \in I$ beliebig gewählt, so gilt

$$x \cdot n = x \cdot (u \circ v) = (x' \circ v) \cdot (u \circ v) = x' \circ v = x$$

und

$$n \cdot y = (u \circ v) \cdot y = (u \circ v) \cdot (u \circ y') = u \circ y' = y.$$

Mit $F = \circ$ ist auch $H = \cdot$ stetig.

Die Funktionen $f\colon I \to I$ und $g\colon I \to I$ mit $f(x) := x \circ v = R_v(x)$ und $g(y) := u \circ y = L_u(y)$ sind 1-1-Abbildungen von I auf sich, also ist $(I, \cdot)$ ein Loop-Hauptisotop von $(I, \circ)$ wegen $x \circ y = (x \circ v) \cdot (u \circ y) = f(x) \cdot g(y)$. Von diesem Loop-Hauptisotop werden wir zeigen, daß es isomorph ist zur additiven Gruppe der reellen Zahlen, woraus dann die Darstellung (11.1.1) für F folgt (vgl. den Beweis zum folgenden Satz 11.2.3).

Dazu wird in $(I, \cdot)$ zunächst der Potenzbegriff definiert und dann das Rechnen mit diesen Potenzen untersucht. Liegt als zweistellige Operation die Multiplikation der reellen Zahlen zugrunde, so wird x^m bekanntlich (vgl. MfL Bd. 1 und MfL Bd. 4) für jedes $x \in \mathbf{R}$ und jedes $m \in \mathbf{N}$ induktiv definiert durch $x^0 := 1$ $(x \neq 0)$, $x^{m+1} := x^m \cdot x$. Dann wird $x^{-m} := \dfrac{1}{x^m}$ festgelegt für alle $x \in \mathbf{R}^*$ und alle $m \in \mathbf{N}$. Durch vollständige Induktion über p beweist man anschließend leicht das Potenzgesetz $x^p \cdot x^q = x^{p+q}$. Insgesamt wird dabei wesentlich die Tatsache benutzt, daß die Multiplikation eine kommutative Gruppenoperation ist. In Kapitel 9 haben wir diese Betrachtungen für beliebige kommutative Gruppenoperationen durchgeführt.

In einer beliebigen Loop dagegen, die ja weder assoziativ noch kommutativ zu sein braucht, sind i. allg. schon die Produkte $x \cdot (x \cdot x)$ und $(x \cdot x) \cdot x$ nicht gleich. Man muß sich also bei der Potenzdefinition zunächst auf eine bestimmte Klammerung festlegen. Weiterhin gibt es wegen der eindeutigen Umkehrbarkeit der Loopoperation zwar für jedes $x \in I$ genau ein Linksinverses $x_l = {}^{-1}x$ mit $n = {}^{-1}x \cdot x$ und genau

ein Rechtsinverses $x_r = x^{-1}$ mit $n = x \cdot x^{-1}$, aber die Zahlen ${}^{-1}x$ und x^{-1} brauchen nicht gleich zu sein (vgl. 1.4.). Man kann sogar beweisen, daß es in jeder nichtkommutativen Loop, die mindestens die Ordnung 5 hat, Elemente gibt, deren Rechts- und Linksinverse nicht identisch sind (vgl. Aufgabe 11.2.7.2).

Die folgende Definition einer m-ten *Potenz* kann in jeder Loop vorgenommen werden (vgl. aber auch 5.1., S. 118).

11.2.2. Definition. Es sei $(M, \cdot)$ eine Loop. Dann wird durch folgende Festsetzungen in $(M, \cdot)$ eine Potenzfunktion $p_m\colon M \to M$ mit $p_m(x) = x^m$ $(x \in M, m \in \mathbf{Z})$ definiert:[1])

a) $x^0 := n$,

b) $x^{m+1} := x^m \cdot x$ $(m \in \mathbf{N})$,

c) $x^m := (x^{-1})^{-m}$ $(m \in \mathbf{Z} \setminus \mathbf{N})$, wobei x^{-1} das eindeutig bestimmte Rechtsinverse von x ist.

Jetzt können wir folgenden Satz beweisen:

11.2.3. Satz von Aczél. *Für jede zweistellige reellwertige Funktion F, die*

a) *über einem (beschränkten oder unbeschränkten) offenen Intervall reeller Zahlen $I =]a, b[$ definiert und dort eine stetige, umkehrbare und kürzbare Operation ist und*

b) *in $I \times I$ die Sechseckbedingung* (H) *erfüllt,*

gibt es einstellige stetige und streng monotone Funktionen $h\colon I \to \mathbf{R}$, $f\colon I \to \mathbf{R}$ und $g\colon I \to \mathbf{R}$ mit (11.1.1) *für $D(F) = I \times I$.*

Beweis (vgl. J. Aczél [4]). Der Beweis dieses Satzes beruht auf folgendem Grundgedanken. Gegeben ist die Struktur $(I, F) = (I, \circ)$, wobei für $F = \circ$ die Schließungsbedingung (H) gilt. Wir gehen nach Definition 11.2.1 zu $(I, \mathrm{H}) = (I, \cdot)$ über und zeigen, daß es für diese Struktur eine stetige und streng monotone Funktion $h\colon I \to \mathbf{R}$ gibt mit $h(x \cdot y) = h(x) + h(y)$, die surjektiv ist, so daß also $(I, \cdot)$ isomorph ist zur additiven Gruppe der reellen Zahlen. Dann gilt für die gegebene Funktion $F = \circ$

$$\begin{aligned} h\big(F(x, y)\big) &= h(x \circ y) = h\big((x \circ v) \cdot (u \circ y)\big) \\ &= h(x \circ v) + h(u \circ y) = f(x) + g(y) \end{aligned}$$

für alle $x, y \in I$ mit $f(x) := h(x \circ v)$ und $g(y) := h(u \circ y)$. Die Funktionen f und g sind mit h stetig und streng monoton.

Es genügt also zu zeigen, daß $(I, \cdot)$ und $(\mathbf{R}, +)$ zueinander isomorph sind. Damit ist dann übrigens $(I, \circ)$ isotop zu $(\mathbf{R}, +)$ (vgl. Satz 4.2.8). Dazu gliedern wir die Überlegungen in zwei Abschnitte 11.2.4 und 11.2.5.

11.2.4. In der Loop $(I, \cdot)$ gilt für den in Definition 11.2.2 definierten Potenzbegriff das Potenzgesetz $x^p \cdot x^q = x^{p+q}$ für alle $x \in I$ und $p, q \in \mathbf{Z}$. (Beweis durch vollständige Induktion über p und q.)

---

[1]) Man beachte im folgenden stets, daß es sich bei diesen Potenzbildungen i. allg. nicht um die Multiplikation handelt. Dies könnte (siehe auch Kap. 9) durch Bezeichnungen wie „$x^{\cdot m}$" zum Ausdruck gebracht werden, worauf wir aber zu Gunsten einer einfachen Schreibweise verzichten wollen.

11.2.5. Die Struktur $(I, \cdot)$ ist isomorph zu $(\mathbf{R}, +)$ und damit selbst eine kommutative Gruppe. (Beweis durch schrittweise Erweiterung des Potenzbegriffes.)

Zu (11.2.4) (vgl. auch J. Aczél, G. Pickert und F. Radó [11], G. Pickert [134]):

1. *In der Loop $(I, \cdot)$ sind Links- und Rechtsinverse jeder Zahl $x \in I$ identisch.*

Zum Beweis wird die Gültigkeit von (H) in $(I, \circ)$ ausgenutzt. Es seien $u, v \in I$ beliebig, aber für den ganzen folgenden Beweis fest gewählt. Wir setzen $x_2 := u$ und $y_2 := v$ und betrachten eine beliebige Zahl $x \in I$. Dann sind durch

$$\alpha) \qquad x = x_1 \circ v = u \circ y_1$$

die Zahlen x_1, $y_1 \in I$ und durch

$$\beta) \qquad u \circ v = x_1 \circ y_3 = x_3 \circ y_1$$

die Zahlen x_3, $y_3 \in I$ jeweils eindeutig bestimmt.

Nun sind $\alpha)$ und $\beta)$ gerade die Voraussetzungen von (H), also gilt $u \circ y_3 = x_3 \circ v$.

Wir setzen weiterhin $\bar{x} := x_3 \circ v = u \circ y_3$ und beachten, daß $u \circ v$ das neutrale Element von $(I, \cdot)$ ist. Für diese Zahl $\bar{x}$ gilt

$$\bar{x} \cdot x = (x_3 \circ v) \cdot (u \circ y_1) = x_3 \circ y_1 = u \circ v = n$$

und

$$x \cdot \bar{x} = (x_1 \circ v) \cdot (u \circ y_3) = x_1 \circ y_3 = u \circ v = n$$

(vgl. Abb. 58). Damit ist $\bar{x}$ linksinverses als auch rechtsinverses Element von x. Insbesondere gilt also $x \cdot x^{-1} = n$ und $(x^{-1})^{-1} \cdot x^{-1} = n$. Daraus folgt wegen der Kürzbarkeit von $\cdot$, daß in $(I, \cdot)$ für jedes x gilt: $x = (x^{-1})^{-1}$ (vgl. Aufgabe 1.4.7.5).

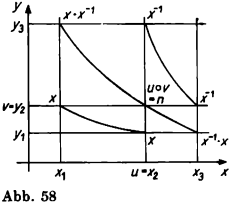


Abb. 58

2. *In $(I, \cdot)$ gilt für alle $x \in I$ und alle $p, q \in \mathbf{Z}$ das Potenzgesetz $x^p \cdot x^q = x^{p+q}$.*

Wir betonen noch einmal, daß uns für den Beweis dieses Gesetzes weder die Kommutativität noch die Assoziativität zur Verfügung steht. Es wird sich herausstellen, daß $(I, \cdot)$ eine kommutative Gruppe ist, die dazu noch isomorph ist zur additiven Gruppe der reellen Zahlen. Aber das gilt es ja erst zu beweisen. Zunächst ist von der in I definierten Operation $\cdot$ bekannt, daß sie umkehrbar und kürzbar ist, daß ein neutrales Element existiert und daß die Links- und Rechtsinversen eines Elementes jeweils identisch sind.

Deshalb wird der Beweis durch vollständige Induktion über p und q geführt, wobei

vier Fälle zu unterscheiden sind. Die Zahl $x \in I$ sei wieder beliebig, aber fest gewählt.

Fall 1: Wir zeigen, daß $x^p \cdot x^q = x^{p+q}$ für alle $p, q \in \mathbf{N}$ gilt.

Wegen $x^0 = n$ gilt bereits $x^0 \cdot x^q = x^q = x^{0+q}$, $x^p \cdot x^0 = x^p = x^{p+0}$ für beliebige $p, q \in \mathbf{N}$ und nach Definition 11.2.2 auch $x^p \cdot x^1 = x^{p+1}$. Dagegen gilt $x^p \cdot x^2 = x^{p-2}$ zunächst nur für $p = 0$. Man hat die Induktion jetzt so zu führen, daß aus den bereits gültigen Fällen der Reihe nach $x^1 \cdot x^2 = x^{1+2}, x^2 \cdot x^2 = x^{2+2}, \ldots, x^p \cdot x^2 = x^{p+2}, \ldots$ folgt. Daraus müßte dann weiter $x^1 \cdot x^3 = x^{1+3}, x^2 \cdot x^3 = x^{2+3}, \ldots, x^p \cdot x^3 = x^{p+3}, \ldots$ folgen. Das heißt, wir brauchten ein Induktionsschema, durch das für jedes q jeweils vollständige Induktion über p geführt wird. Oder: Wenn P und Q feste natürliche Zahlen sind und die Behauptung sowohl für beliebige p und alle $q \leqq Q$ als auch für $P - 1$ und $Q + 1$ gilt, so gilt sie auch für P und $Q + 1$. Wie der nachfolgende Rechtfertigungssatz zeigen wird, genügt weniger. Sind nämlich P und Q beliebige, aber fest gewählte natürliche Zahlen, dann gilt in $(I, \cdot)$:

$$\text{Wenn } x^P \cdot x^{Q-1} = x^{P+Q-1} = x^{P-1} \cdot x^Q \quad \text{und}$$

$$x^{P+1} \cdot x^{Q-1} = x^{P+Q} = x^P \cdot x^Q = x^{P-1} \cdot x^{Q+1} \quad \text{und}$$

$$x^{P+1} \cdot x^Q = x^{P+Q+1}, \quad \text{so} \quad x^P \cdot x^{Q+1} = x^{P+Q+1}.$$

Zum Beweis dieser Behauptung definieren wir folgendermaßen Zahlen $x_1, x_2, x_3, y_1, y_2, y_3 \in I$:

$$x^{P-1} = x_1 \circ v, \quad x^P = x_2 \circ v, \quad x^{P+1} = x_3 \circ v,$$
$$x^{Q-1} = u \circ y_1, \quad x^Q = u \circ y_2, \quad x^{Q+1} = u \circ y_3.$$

Damit gilt (vgl. Definition 11.2.1)

$$\begin{aligned} x_1 \circ y_2 &= (x_1 \circ v) \cdot (u \circ y_2) = x^{P-1} \cdot x^Q = x^P \cdot x^{Q-1} \\ &= (x_2 \circ v) \cdot (u \circ y_1) = x_2 \circ y_1 \end{aligned}$$

(erste Induktionsvoraussetzung) und

$$\begin{aligned} x_1 \circ y_3 &= (x_1 \circ v) \cdot (u \circ y_3) = x^{P-1} \cdot x^{Q+1} = x^P \cdot x^Q \\ &= (x_2 \circ v) \cdot (u \circ y_2) = x_2 \circ y_2 = x^{P+1} \cdot x^{Q-1} \\ &= (x_3 \circ v) \cdot (u \circ y_1) = x_3 \circ y_1 \end{aligned}$$

(zweite Induktionsvoraussetzung). Wegen der Gültigkeit von (H) in $(I, \circ)$ folgt daraus $x_2 \circ y_3 = x_3 \circ y_2$. Damit erhält man bei Benutzung der dritten Induktionsvoraussetzung:

$$\begin{aligned} x^P \cdot x^{Q+1} &= (x_2 \circ v) \cdot (u \circ y_3) = x_2 \circ y_3 = x_3 \circ y_2 \\ &= (x_3 \circ v) \cdot (u \circ y_2) = x^{P+1} \cdot x^Q = x^{P+Q+1}; \end{aligned}$$

das ist die Induktionsbehauptung. Die Abb. 59 spiegelt den Zusammenhang wider zwischen dem Induktionsschluß in $(I, \cdot)$ und der Gültigkeit von (H) in $(I, \circ)$.

Zur Rechtfertigung des angegebenen Induktionsschlusses haben wir zu zeigen, daß durch ihn die Gültigkeit von $x^p \cdot x^q = x^{p+q}$ für alle $p, q \in \mathbf{N}$ bewiesen ist. Angenommen, das wäre nicht der Fall. Dann betrachten wir die Menge aller Paare (p, q) natürlicher Zahlen, für die das Potenzgesetz nicht gilt. In dieser Menge gibt es nach

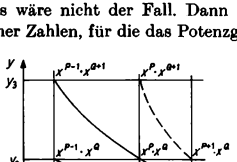


Abb. 59

dem Prinzip der kleinsten Zahl genau ein Paar $(\overline{p}, \overline{q})$, so daß $\overline{p}$ zunächst die kleinste erste Komponente und $\overline{q}$ die kleinste zu diesem $\overline{p}$ gehörende zweite Komponente ist. Dann würden für $P = \overline{p}$ und $Q = \overline{q} - 1$ die Induktionsvoraussetzungen

$$x^{\overline{p}} \cdot x^{\overline{q}-2} = x^{\overline{p}+\overline{q}-2} = x^{\overline{p}-1} \cdot x^{\overline{q}-1},$$

$$x^{\overline{p}+1} \cdot x^{\overline{q}-2} = x^{\overline{p}+\overline{q}-1} = x^{\overline{p}} \cdot x^{\overline{q}-1} = x^{\overline{p}-1} \cdot x^{\overline{q}} \quad \text{und} \quad x^{\overline{p}+1} \cdot x^{\overline{q}-1} = x^{\overline{p}+\overline{q}}$$

gelten. Dies hätte die Gültigkeit der Induktionsbehauptung $x^{\overline{p}} \cdot x^{\overline{q}} = x^{\overline{p}+\overline{q}}$ zur Folge. Das ist ein Widerspruch, denn $(\overline{p}, \overline{q})$ gehörte zu den Paaren, für die das Potenzgesetz nicht gelten sollte.

Fall 2: Das Potenzgesetz $x^p \cdot x^q = x^{p+q}$ gilt in $(I, \cdot)$ auch für alle $p, q \in \mathbf{Z} \setminus \mathbf{N}$.

Der Beweis dieser Behauptung kann mit Definition 11.2.2 auf den ersten Fall zurückgeführt werden:

$$x^p \cdot x^q = (x^{-1})^{-p} \cdot (x^{-1})^{-q} = (x^{-1})^{-(p+q)} = x^{p+q}.$$

Fall 3: Es gilt $x^p \cdot x^q = x^{p+q}$ für alle $p \in \mathbf{N}$ und alle $q \in \mathbf{Z} \setminus \mathbf{N}$.

Wir schreiben $x^p \cdot x^{-r} = x^{p+(-r)}$ $(q = -r,\ r > 0)$ und führen den Beweis durch vollständige Induktion über p und r. Es gilt schon $x^p \cdot x^1 = x^{p+1}$ und $x^p \cdot x^0 = x^{p+0}$ für beliebige $p \in \mathbf{N}$ sowie $x^0 \cdot x^{-1} = n \cdot x^{-1} = x^{0+(-1)}$ und $x \cdot x^{-1} = n = x^0 = x^{1+(-1)}$. Daraus muß auf $x^2 \cdot x^{-1} = x^{2+(-1)}$, $x^3 \cdot x^{-1} = x^{3+(-1)}, \ldots, x^p \cdot x^{-1} = x^{p+(-1)}, \ldots$ geschlossen werden und dann so weiter für $q = -2$, $q = -3, \ldots$ Sind P und R feste natürliche Zahlen, so muß aus der Gültigkeit des Potenzgesetzes für beliebige p und alle q mit $-R \leqq q \leqq 1$ und für $p = P$ und $q = -R - 1$ auf die Gültigkeit für $p = P + 1$ und $q = -R - 1$ geschlossen werden. Das gelingt durch folgende Induktion:

Wenn $x^{P-1} \cdot x^{-R+1} = x^{P+(-R)} = x^P \cdot x^{-R}$ und

$x^P \cdot x^{-R+1} = x^{P+(-R)+1} = x^{P+1} \cdot x^{-R}$ und

$x^{P-1} \cdot x^{-R} = x^{P-1+(-R)} = x^P \cdot x^{-R-1}$, so $x^{P+1} \cdot x^{-R-1} = x^{P+(-R)}$.

Wir definieren wieder Zahlen $x_1, x_2, x_3, y_1, y_2, y_3 \in I$ durch folgende Gleichungen

$$x^{P-1} = x_1 \circ v, \quad x^P = x_2 \circ v, \quad x^{P+1} = x_3 \circ v, \quad x^{-R-1} = u \circ y_1,$$
$$x^{-R} = u \circ y_2, \quad x^{-R+1} = u \circ y_3.$$

Damit gilt in $(I, \circ)$

$$x_1 \circ y_2 = (x_1 \circ v) \cdot (u \circ y_2) = x^{P-1} \cdot x^{-R} = x^{P-1+(-R)} = x^P \cdot x^{-R-1}$$
$$= (x_2 \circ v) \cdot (u \circ y_1) = x_2 \circ y_1$$

(dritte Induktionsvoraussetzung),

$$x_2 \circ y_3 = (x_2 \circ v) \cdot (u \circ y_3) = x^P \cdot x^{-R+1} = x^{P+(-R)+1} = x^{P+1} \cdot x^{-R}$$
$$= (x_3 \circ v) \cdot (u \circ y_2) = x_3 \circ y_2$$

(zweite Induktionsvoraussetzung),

$$x_1 \circ y_3 = (x_1 \circ v) \cdot (u \circ y_3) = x^{P-1} \cdot x^{-R+1} = x^{P+(-R)} = x^P \cdot x^{-R}$$
$$= (x_2 \circ v) \cdot (u \circ y_2) = x_2 \circ y_2$$

(erste Induktionsvoraussetzung).

Damit sind die Voraussetzungen von (H') erfüllt. Nach Satz 11.1.5 gilt mit (H) auch (H'), so daß wir auf $x_2 \circ y_2 = x_3 \circ y_1$ schließen können. Daraus folgt schließlich die Induktionsbehauptung

$$x^{P+1} \cdot x^{-R-1} = (x_3 \circ v) \cdot (u \circ y_1) = x_3 \circ y_1 = x_2 \circ y_2$$
$$= x^P \cdot x^{-R} = x^{P+(-R)}.$$

Man zeigt leicht, daß durch diesen Induktionsschluß das Potenzgesetz für alle natürlichen Zahlen p und alle negativen ganzen Zahlen q bewiesen ist (vgl. Aufgabe 11.2.7.1).

Es bleibt noch der

Fall 4: Es gilt $x^p \cdot x^q = x^{p+q}$ für alle $p \in \mathbf{Z} \setminus \mathbf{N}$ und alle $q \in \mathbf{N}$.

Jetzt benutzen wir die Tatsache, daß in einer Loop, in der Linksinverse und Rechtsinverse eines Elementes jeweils zusammenfallen, $x = (x^{-1})^{-1}$ gilt. Mit dieser Beziehung und der Definition 11.2.2.c) kann dieser Fall auf den dritten Fall zurückgeführt werden:

$$x^p \cdot x^q = \left((x^{-1})^{-1}\right)^p \cdot (x^{-1})^{-q} = (x^{-1})^{-p} \cdot (x^{-1})^{-q} = (x^{-1})^{-(p+q)}.$$

Dabei können folgende Fälle eintreten:

a) $p + q > 0$; dann ist $(x^{-1})^{-(p+q)} = x^{p+q}$ nach Definition,

b) $p + q = 0$; dann ist $(x^{-1})^{-(p+q)} = (x^{-1})^0 = n = x^0 = x^{p+q}$,

c) $p + q < 0$; dann ist $(x^{-1})^{-(p+q)} = \left((x^{-1})^{-1}\right)^{p+q} = x^{p+q}$.

Damit ist Teil 11.2.4 des Beweises von Satz 11.2.3 durchgeführt.

Zu (11.2.5): Es bleibt also noch zu beweisen, daß die Loop $(I, \cdot)$ isomorph ist zur additiven Gruppe $(\mathbf{R}, +)$ der reellen Zahlen. Wir haben in Kapitel 9 bewiesen, daß

jede kommutative stetige Gruppe $(I, \cdot)$ mit $I \subsetneq \mathbf{R}$ zu $(\mathbf{R}, +)$ isomorph ist. Die Analyse dieses Beweises zeigt, daß Kommutativität und Assoziativität für beliebige Gruppenelemente gar nicht benutzt werden, sondern jeweils nur für Potenzen ein und desselben Gruppenelementes. Das wird durch das Potenzgesetz $\bigwedge_{x \in I} \bigwedge_{p,q \in \mathbf{Z}} x^p \cdot x^q = x^{p+q}$, das in Kapitel 9 aus Gruppeneigenschaften bewiesen worden ist, zum Ausdruck gebracht. Ist nämlich x eine beliebige Zahl mit $x \in I$ und sind p, q und r beliebige ganze Zahlen, so gilt

$$x^p \cdot x^q = x^{p+q} = x^{q+p} = x^q \cdot x^p$$

und

$$x^p \cdot (x^q \cdot x^r) = x^p \cdot x^{q+r} = x^{p+(q+r)} = x^{(p+q)+r} = x^{p+q} \cdot x^r = (x^p \cdot x^q) \cdot x^r.$$

Da jetzt für den hier zu führenden Beweis neben der Umkehrbarkeit und Kürzbarkeit der Operation $\cdot$ und der Existenz eines Inversen x^{-1} zu jedem x mit $x^{-1} \cdot x = x \cdot x^{-1} = n$ dieses Potenzgesetz zur Verfügung steht, kann der Teil des Beweises von Satz 9.1.2, in dem die Existenz des behaupteten Isomorphismus nachgewiesen wird, übertragen werden. Es ist der Reihe nach zu beweisen (der Leser gehe diesen Beweis zur Übung noch einmal durch und prüfe die dabei jeweils verwendeten Voraussetzungen):

A) Die Operation $\cdot$ ist in beiden Argumenten streng monoton wachsend. Hier muß vermerkt werden, daß man jetzt $k = l$ aus der Gleichung $t \cdot k = t \cdot l$ natürlich nicht durch Multiplikation mit t^{-1} von links folgern kann, da man dazu die Assoziativität braucht. Das folgt aber hier (wie auch dort) wegen der Kürzbarkeit von $\cdot$ direkt aus $t \cdot k = t \cdot l$.

B) Die Funktion $p_m : I \to I$ mit $p_m(v) := v^m$ ($m \in \mathbf{N}$, $m \geqq 1$) ist streng monoton wachsend und stetig.

C) Für jedes $v \in I$ ist die Folge (v^m) mit $m \in \mathbf{Z}$ monoton. Ist $v > n$, so ist (v^m) streng monoton wachsend. Ist $v < n$, so ist (v^m) streng monoton fallend. Ist $v = n$, dann ist (v^m) eine konstante Folge.

D) Es sei $v \in I =]\!]a, b[\![$. Ist $v > n$, so ist $\lim\limits_{m \to \infty} v^m = b$. Ist $v < n$, so ist $\lim\limits_{m \to \infty} v^m = a$.

E) Die Gleichung $v^m = u$ hat für jedes $u \in I$ und jedes $m \in \mathbf{N}^*$ eine eindeutig bestimmte Lösung $v \in I$. Wir setzen $u^{1/m} := v$ und $u^{k/m} := (u^{1/m})^k$ für $m \in \mathbf{N}^*$ und $k \in \mathbf{Z}$. Es gilt $(u^{1/m})^k = \left(u^{\frac{1}{m \cdot l}}\right)^{k \cdot l}$ für beliebige $l, m \in \mathbf{N}^*$ und $k \in \mathbf{Z}$.

Nach diesen Vorbereitungen können wir nun einen stetigen und streng monotonen Isomorphismus $\bar{h}$ von $(\mathbf{R}, +)$ auf $(I, \cdot)$ konstruieren:

a) Es sei $c \in I$ und $c \neq n$, etwa $c > n$. Dann wird definiert: $\bar{h}(0) := n$ und $\bar{h}(1) := c$.

b) $\bar{h}(m) := c^m$ für $m \in \mathbf{N}^*$.

c) $\bar{h}\left(\frac{1}{q}\right) := c^{1/q}$ und $\bar{h}\left(\frac{p}{q}\right) := c^{p/q}$ für $q \in \mathbf{N}^*$, $p \in \mathbf{Z}^*$.

Damit ist $\bar{h}$ für rationale Argumente eindeutig definiert, denn für beliebige rationale Zahlen $r = \frac{p}{q}$ ($p \in \mathbf{Z}$, $q \in \mathbf{N}^*$) gilt $\bar{h}\left(\frac{p}{q}\right) = \bar{h}\left(\frac{p \cdot s}{q \cdot s}\right)$ ($s \in \mathbf{N}^*$). Das folgt aus E) und der Definition von $\bar{h}$ für negative Argumente.

d) Die Funktion $\bar{h}\colon \mathbf{Q} \to I$ mit $\bar{h}(r) = c^r$ erfüllt

d_1) für alle rationalen Zahlen r_1, r_2 die Gleichung $\bar{h}(r_1 + r_2) = \bar{h}(r_1) \cdot \bar{h}(r_2)$ und ist

d_2) streng monoton wachsend für $c > n$ und streng monoton fallend für $c < n$.

e) Definition einer Funktion $\bar{h}\colon \mathbf{R} \to I$.

f) Für alle $w, z \in \mathbf{R}$ gilt $\bar{h}(w + z) = \bar{h}(w) \cdot \bar{h}(z)$.

g) Die Funktion $\bar{h}$ aus e) ist stetig.

Für die stetige und streng monotone Umkehrfunktion $h\colon I \to \mathbf{R}$ von $\bar{h}$ gilt nach f) $h(x \cdot y) = h(x) + h(y)$. (Man setze $x := \bar{h}(w)$ und $y := \bar{h}(z)$.)

Nach unseren Bemerkungen zu Beginn des Beweises von Satz 11.2.3 gibt es dann auch einstellige stetige und streng monotone Funktionen $h\colon I \to \mathbf{R}$, $f\colon I \to \mathbf{R}$ und $g\colon I \to \mathbf{R}$ mit $h(x \circ y) = h\big(F(x, y)\big) = f(x) + g(y)$ für alle $x, y \in I$. Damit ist der Satz 11.2.3 vollständig bewiesen.

Wir wollen uns noch einen Überblick über alle Skalenfunktionen h, f und g verschaffen (vgl. auch Bemerkung 9.1.5). Dazu seien F eine zweistellige Funktion und (h_1, f_1, g_1), (h_2, f_2, g_2) beliebige Tripel einstelliger Funktionen, die jeweils die in Satz 11.2.3 vorausgesetzten bzw. behaupteten Eigenschaften haben. Dann gilt also

$$\bigwedge_{x,y \in I} h_1\big(F(x, y)\big) = f_1(x) + g_1(y)$$

und

$$\bigwedge_{x,y \in I} h_2\big(F(x, y)\big) = f_2(x) + g_2(y).$$

Da die Funktionen h_1, f_1, g_1, h_2, f_2 und g_2 streng monoton sind und jeweils das Interfall I auf $\mathbf{R}$ abbilden, existieren ihre Umkehrfunktionen, die $\mathbf{R}$ auf I abbilden, so daß wir mit $f_2(x) = u$ und $g_2(y) = v$

$$h_1^{-1}\big(f_1(x) + g_1(y)\big) = F(x, y) \quad \text{und} \quad h_2^{-1}(u + v) = F\big(f_2^{-1}(u), g_2^{-1}(v)\big)$$

erhalten. Daraus folgt

$$h_2^{-1}(u + v) = F\big(f_2^{-1}(u), g_2^{-1}(v)\big) = h_1^{-1}\big(f_1 f_2^{-1}(u) + g_1 g_2^{-1}(v)\big)$$

oder[1])

$$h_1 h_2^{-1}(u + v) = f_1 f_2^{-1}(u) + g_1 g_2^{-1}(v).$$

Für $h_3 := h_1 h_2^{-1}$, $f_3 := f_1 f_2^{-1}$ und $g_3 := g_1 g_2^{-1}$ ergibt sich daraus die *Pexidersche Funktionalgleichung*

$$\bigwedge_{u,v \in \mathbf{R}} h_3(u + v) = f_3(u) + g_3(v),$$

die sich folgendermaßen auf die Cauchysche Funktionalgleichung zurückführen läßt (vgl. J. Aczél [3]):

Wir setzen $u = 0$, $v = t$ und dann $u = t$, $v = 0$ und erhalten

$$h_3(t) = f_3(0) + g_3(t) \quad \text{und} \quad h_3(t) = f_3(t) + g_3(0)$$

oder auch

$$g_3(t) = h_3(t) - f_3(0) \quad \text{und} \quad f_3(t) = h_3(t) - g_3(0).$$

[1]) Für die Nacheinanderausführung von Funktionen schreiben wir hier statt $f_1 \bullet f_2^{-1}$ kürzer $f_1 f_2^{-1}$.

Damit gilt

$$h_3(u+v) = h_3(u) - f_3(0) + h_3(v) - g_3(0),$$

woraus schließlich für die Funktion $h\colon \mathbf{R} \to \mathbf{R}$ mit $h(t) := h_3(t) - f_3(0) - g_3(0)$

$$\bigwedge_{u,v\in\mathbf{R}} h(u+v) = h(u) + h(v)$$

folgt. Wegen der Stetigkeit von h ist also $h(u) = c \cdot u$ für alle $u \in \mathbf{R}$. Durch einfache Umformungen erhält man daraus $f_1(x) = c \cdot f_2(x) + f_3(0)$, $g_1(x) = c \cdot g_2(x) + g_3(0)$ und $h_1(z) = c \cdot h_2(z) + f_3(0) + g_3(0)$.

Man prüft leicht nach, daß wie für h_2, f_2 und g_2 auch für die Funktionen h_1, f_1 und g_1

$$\bigwedge_{x,y\in I} h_1\big(F(x,y)\big) = f_1(x) + g_1(y)$$

gilt. Sind also h_2, f_2 und g_2 Skalenfunktionen von F, so erhält man durch alle Tripel (a, b, c) reeller Zahlen mit $c \neq 0$ auch alle Tripel von Skalenfunktionen für F.

In 9.1. haben wir die Funktionalgleichung

$$\bigwedge_{x,y\in\mathbf{R}} f(x+y) = F\big(f(x), f(y)\big)$$

gelöst (vgl. Satz 9.1.2). Für die Existenz nichtkonstanter stetiger Lösungen f dieser Funktionalgleichung war es notwendig und hinreichend, daß F über einem (nicht notwendig beschränkten) Intervall $]\!]a, b[\![$ eine kommutative stetige Gruppenoperation ist. Es ergab sich, daß $(]\!]a, b[\![, F)$ zur additiven Gruppe der reellen Zahlen isomorph ist.

Aus Satz 11.2.3 folgt nun direkt, daß jede stetige Gruppe $(]\!]a, b[\![, F)$ zur additiven Gruppe der reellen Zahlen isomorph ist, wobei $]\!]a, b[\![$ ein nicht notwendig beschränktes offenes Intervall reeller Zahlen ist. Stetige Gruppenoperationen F über solchen Intervallen erfüllen nämlich zunächst die Schließungsbedingung (H) (vgl. für das Folgende F. Radó [144]).

Es seien $x_1, x_2, x_3, y_1, y_2, y_3 \in I =]\!]a, b[\![$, und es gelte $F(x_1, y_2) = F(x_2, y_1)$ und $F(x_1, y_3) = F(x_2, y_2) = F(x_3, y_1)$. Es ist zu zeigen, daß dann auch $F(x_2, y_3) = F(x_3, y_2)$ gilt. Wählt man noch Zahlen $t_1, t_2 \in I$ mit $F(y_1, t_2) = F(y_2, t_1)$ (solche Zahlen existieren stets, da F eine Gruppenoperation ist), so erhält man

$$\begin{aligned} F\big(x_1, F(y_2, t_2)\big) &= F\big(F(x_1, y_2), t_2\big) = F\big(F(x_2, y_1), t_2\big) \\ &= F\big(x_2, F(y_1, t_2)\big) = F\big(x_2, F(y_2, t_1)\big) \\ &= F\big(F(x_2, y_2), t_1\big) = F\big(F(x_1, y_3), t_1\big) = F\big(x_1, F(y_3, t_1)\big). \end{aligned}$$

Also ist $F(y_2, t_2) = F(y_3, t_1)$. Mit diesem Ergebnis folgt nun

$$\begin{aligned} F\big(F(x_2, y_3), t_1\big) &= F\big(x_2, F(y_3, t_1)\big) = F\big(x_2, F(y_2, t_2)\big) \\ &= F\big(F(x_2, y_2), t_2\big) = F\big(F(x_3, y_1), t_2\big) \\ &= F\big(x_3, F(y_1, t_2)\big) = F\big(x_3, F(y_2, t_1)\big) = F\big(F(x_3, y_2), t_1\big) \end{aligned}$$

und damit schließlich $F(x_2, y_3) = F(x_3, y_2)$.

Also gibt es nach Satz 11.2.3 stetige und streng monotone Funktionen $h: I \to \mathbf{R}$, $f: I \to \mathbf{R}$ und $g: I \to \mathbf{R}$ mit

$$\bigwedge_{x,y \in I} F(x, y) = h^{-1}\big(f(x) + g(y)\big).$$

Das ergibt mit der Assoziativität von F

$$\bigwedge_{x,y,z \in I} fh^{-1}\big(f(x) + g(y)\big) - f(x) = gh^{-1}\big(f(y) + g(z)\big) - g(z).$$

Mit der rechten Seite dieser Gleichung hängt auch die linke Seite nicht von x ab, so daß sie sich als Funktion von $g(y)$ auffassen läßt:

$$\bigwedge_{x,y \in I} fh^{-1}\big(f(x) + g(y)\big) - f(x) = l\big(g(y)\big).$$

Daraus wird mit $k := fh^{-1}$, $r := f(x)$ und $s := g(y)$

$$\bigwedge_{r,s \in \mathbf{R}} k(r + s) = r + l(s)$$

und für festes $s = c$

$$\bigwedge_{r \in \mathbf{R}} k(r + c) = r + c_1 \quad \big(c_1 := l(c)\big).$$

Setzt man schließlich $r' = r + c$ und $c_2 = c_1 - c$, so folgt $\bigwedge_{r' \in \mathbf{R}} k(r') = r' + c_2$, d. h. $fh^{-1}(r') = r' + c_2$, woraus für $x := h^{-1}(r')$

$$\bigwedge_{x \in I} f(x) = h(x) + c_2$$

wird.

Analog erhält man

$$\bigwedge_{x \in I} g(x) = h(x) + c_3,$$

so daß aus (11.1.1)

$$\bigwedge_{x,y \in I} F(x, y) = h^{-1}\big(h(x) + h(y) + c_2 + c_3\big)$$

wird. Mit h durchläuft auch $\bar{h}(t) = h(t) + c_2 + c_3$ ganz $\mathbf{R}$; also gilt

$$\bigwedge_{x,y \in I} F(x, y) = \bar{h}^{-1}\big(\bar{h}(x) + \bar{h}(y)\big),$$

so daß (I, F) isomorph zu $(\mathbf{R}, +)$ ist.

Diese Betrachtungen lassen sich auf stetige und streng monotone Halbgruppen (I, F) ausdehnen. So wird in J. Aczél [3] bewiesen, und zwar wieder durch schrittweise Erweiterung des Potenzbegriffs, daß jede derartige Halbgruppe isomorph ist zu einer additiven Halbgruppe der reellen Zahlen. In der Sprache der Funktionalgleichungen sagt man dann auch, man habe unter den betreffenden Voraussetzungen die *Assoziativitätsgleichung*

$$\bigwedge_{x,y,z \in I} F\big(F(x, y), z\big) = F\big(x, F(y, z)\big)$$

gelöst. Dieses Ergebnis erhält man aber auch aus einem zum Satz 11.2.3 analogen Satz von F. RADÓ.

11.2.6. Satz. *Für jede zweistellige reellwertige Funktion F, die*

a) *über einem Gebiet $D(F) \subseteq \mathbf{R}^2$ definiert, dort stetig und in beiden Argumenten streng monoton ist und*

b) *in $D(F)$ die Schließungsbedingung* (H) *erfüllt,*

gibt es einstellige stetige und streng monotone Funktionen h, f und g mit

$$\bigwedge_{(x,y)\in D(F)} h\big(F(x,y)\big) = f(x) + g(y).$$

In diesem Satz (vgl. F. RADÓ [143] und [144]) sind die Funktionen F aus Satz 11.2.3 mit erfaßt. Die Umkehrbarkeit wird hier nicht mehr gefordert, und es werden viel allgemeinere Definitionsbereiche als dort zugelassen. Auf den Beweis des Satzes müssen wir hier verzichten. Es sei nur vermerkt, daß dieser Beweis mit ganz anderen Methoden geführt wird als der Beweis zu Satz 11.2.3. Man geht von einer Niveaulinie aus, konstruiert mit Hilfe von (H) weitere Niveaulinien und definiert für diese Linien die Funktionen h, f und g. Das so entstehende Netz von Niveaulinien wird dann, wieder mit (H), schrittweise verfeinert. Gleichzeitig wird der Definitionsbereich der Funktionen h, f und g erweitert.

Ist dann eine über einem offenen Intervall I definierte stetige, streng monotone und assoziative Operation F gegeben, so erfüllt sie wieder (H), so daß es aufgrund von Satz 11.2.6 streng monotone stetige Funktionen h, f und g mit

$$\bigwedge_{(x,y)\in D(F)} F(x,y) = h^{-1}\big(f(x) + g(y)\big)$$

gibt. Wie oben findet man dann wieder $h = f = g$. Natürlich ist der Wertebereich von h hier nicht mehr die Menge aller reellen Zahlen.

Auf diese Weise lassen sich noch weitere Funktionalgleichungen lösen (vgl. F. RADÓ [144]). So zeigt man z. B., daß (über offenen Intervallen streng monotone und stetige) bisymmetrische Operationen (vgl. Satz 4.2.10) die Schließungsbedingung (T) von THOMSEN erfüllen. Da (H) ein Spezialfall von (T) ist, läßt sich wieder Satz 11.2.6 anwenden, und man erhält

$$\bigwedge_{(x,y)\in D(F)} F(x,y) = h^{-1}\big(c_1 \cdot h(x) + c_2 \cdot h(y) + c_3\big)$$

(h stetig und streng monoton). Unter den angegebenen Voraussetzungen ist damit die Bisymmetriegleichung gelöst. Hierin sind die bekannten Mittelwerte als Spezialfälle enthalten (vgl. die Beispiele 5.1.10.1).

11.2.7. Aufgaben

1. Im dritten Fall des Beweisteiles (11.2.4.2) von Satz 11.2.3 auf S. 213 wurde behauptet, daß durch den dort angegebenen Induktionsschluß das Potenzgesetz $x^p \cdot x^q = x^{p+q}$ für alle natürlichen Zahlen p und für alle negativen ganzen Zahlen q bewiesen ist. Man beweise diese Behauptung. (Hinweis: Vgl. den entsprechenden Beweis im ersten Fall.)

2. Gegeben sei die Menge $M = \{a, b, c, d, e\}$. In M wird durch folgende Strukturtafel eine Operation $\cdot$ definiert:

$\cdot$	a	b	c	d	e
a	a	b	c	d	e
b	b	e	a	c	d
c	c	d	e	b	a
d	d	a	b	e	c
e	e	c	d	a	b

Man zeige, a) daß $(M, \cdot)$ eine Loop ist, b) daß es in $(M, \cdot)$ Elemente gibt, deren Links- und Rechtsinverse nicht zusammenfallen.

3. Man zeige, daß es für beliebige stetige und streng monotone Lösungen $f_1: \mathbf{R}_+^* \to \mathbf{R}$ und $f_2: \mathbf{R}_+^* \to \mathbf{R}$ der Funktionalgleichung $\bigwedge_{x,y \in \mathbf{R}_+^*} f(x \cdot y) = f(x) + f(y)$ eine reelle Zahl c gibt mit $f_1(x) = c \cdot f_2(x)$. (Hinweis: Aus 8.3. wissen wir, daß die Logarithmusfunktionen alle stetige Lösungen dieser Funktionalgleichung und damit alle Skalenfunktionen der Multiplikation reeller Zahlen sind. Bekanntlich (vgl. etwa R. Bittner, D. Ilse, S. Kubicek und W. Tietz [23]) gilt für zwei beliebige Funktionen $\log_a$, $\log_b$ $(a > 0, a \neq 1, b > 0, b \neq 1)$ die Beziehung $\log_a x = \frac{1}{\log_b a} \cdot \log_b x = c \cdot \log_b x$.)

12. Funktionale Charakterisierungen von Winkelfunktionen

12.1. Die Additions- und Subtraktionstheoreme der Sinus- und Kosinusfunktionen

Zur Motivation der folgenden Untersuchungen knüpfen wir an die Ausführungen in Kapitel 6 an. Dort wurde festgestellt, daß zwar aus den Subtraktionstheoremen (6.1.2) und (6.1.3) der „trigonometrische Pythagoras" (6.1.1), umgekehrt aus (6.1.1) aber weder (6.1.2) noch (6.1.3) folgt. Zum Beweis dafür wurde das Funktionenpaar (s, c) mit $s(x) := \sqrt{\frac{2}{\pi}}\, x$ und $c(x) := \sqrt{1 - \frac{2}{\pi}}\, x$ benutzt, das zunächst für $x \in \left[\!\left[0, \frac{\pi}{2} \right]\!\right]$ definiert war und dann auf die ganze Achse fortgesetzt wurde. Da für diese Funktionen die Subtraktionstheoreme nicht gelten, können sie mit den Funktionen sin und cos nicht identisch sein. Das erkennt man aber z. B. auch schon daran, daß die trigonometrischen Funktionen sin und cos in ihren Extremstellen differenzierbar sind, die Funktionen s und c aber nicht (die rechtsseitigen und linksseitigen Grenzwerte der Differenzenquotienten existieren jeweils, sind aber nicht gleich).

Nimmt man zu den Eigenschaften a) bis g) und (6.0.1) noch die Forderung hinzu, daß s und c überall differenzierbar sein sollen, so gibt es wieder Funktionen, die diese Eigenschaften zwar besitzen, die aber von den Sinus- und Kosinusfunktionen verschieden sind. Das ist auch dann noch der Fall, wenn man darüber hinaus Wendepunkte in den Nullstellen, Konkavität in den Intervallen $[\![2k\pi, (2k+1)\pi]\!]$ $(k \in \mathbf{Z})$ und Konvexität in den Intervallen $[\![(2k+1)\pi, 2(k+1)\pi]\!]$ für s fordert.[1]) Alle diese Funktionen, die mit den trigonometrischen Funktionen sin und cos also zahlreiche Eigenschaften gemeinsam haben und trotzdem mit diesen Funktionen nicht identisch sind, erfüllen nicht die Additions- und Subtraktionstheoreme von sin und cos. Das legt nun die Vermutung nahe, daß gerade diese Beziehungen für die trigonometrischen Funktionen „typisch" sind. Inwieweit diese Vermutung zutrifft, wird in 12.1. und 12.2. untersucht. Wir betrachten folgende Funktionalgleichungssysteme:

12.1.1. $\bigwedge_{x,y \in \mathbf{R}} s(x - y) = s(x)\, c(y) - s(y)\, c(x)$,

12.1.2. $\bigwedge_{x,y \in \mathbf{R}} c(x - y) = c(x)\, c(y) + s(x)\, s(y)$

und

12.1.3. $\bigwedge_{x,y \in \mathbf{R}} s(x + y) = s(x)\, c(y) + s(y)\, c(x)$,

[1]) Ausführliche Vergleiche dieser Art findet man in P. Wiechmann [179]. Darüber hinaus sei auf L. Vietoris [176] verwiesen.

12.1.4. $\bigwedge_{x,y\in\mathbf{R}} c(x+y) = c(x)\,c(y) - s(x)\,s(y)$

für Funktionen $s\colon \mathbf{R} \to \mathbf{R}$ und $c\colon \mathbf{R} \to \mathbf{R}$.

Wir zeigen zunächst, daß die beiden Systeme nicht äquivalent sind.

a) *Aus* (12.1.1) *und* (12.1.2) *folgen* (12.1.3) *und* (12.1.4). Vertauscht man in den Subtraktionstheoremen (12.1.1) und (12.1.2) x mit y, so erhält man

12.1.5. Für alle Lösungen s und c von (12.1.1) und (12.1.2) gilt: s ist ungerade und c ist gerade.

Es ist nämlich

$$s(y-x) = s(y)\,c(x) - s(x)\,c(y) = -\big(s(x)\,c(y) - s(y)\,c(x)\big) = -s(x-y),$$

also $s(-x) = -s(x)$ für alle $x \in \mathbf{R}$ und $c(y-x) = c(y)\,c(x) + s(y)\,s(x) = c(x-y)$, also $c(-x) = c(x)$ für alle $x \in \mathbf{R}$.

Ersetzt man jetzt in (12.1.1) und (12.1.2) y durch $-y$, so folgen (12.1.3) und (12.1.4).

b) *Umgekehrt folgen* (12.1.1) *und* (12.1.2) *nicht aus* (12.1.3) *und* (12.1.4). Zum Beweis dieser Behauptung geben wir das Funktionenpaar (s, c) mit $s(x) = a^x \sin bx$, $c(x) = a^x \cos bx$ ($a > 0$, $a \neq 1$, $b \neq 0$) an (vgl. O. Perron [131]), von dem man leicht zeigt, daß es (12.1.3) und (12.1.4) erfüllt, aber nicht (12.1.1) und (12.1.2).

Die Subtraktionstheoreme der Funktionen sin und cos sind also „stärker" als die Additionstheoreme. Deshalb benutzen wir zunächst die Gleichungen (12.1.1) und (12.1.2) zur funktionalen Charakterisierung dieser Funktionen, geben dann aber auch noch Charakterisierungen mit Hilfe der Additionstheoreme (12.1.3) und (12.1.4) und mit Hilfe von Kombinationen aus (12.1.1) bis (12.1.4) an (vgl. auch die Aufgaben 12.1.32.1 bis 12.1.32.5). In 12.2. werden wir nachweisen, daß man dabei ohne eine zusätzliche Forderung zu diesen Gleichungen nicht auskommt. Im vorliegenden Abschnitt wird dies die Differenzierbarkeit der Funktionen s und c sein. Es genügen aber auch schwächere Forderungen (vgl. 12.2.).

In diesem Zusammenhang sei auf eine Aufgabe im Mathematiklehrbuch der Klasse 12 [114], S. 96, hingewiesen, in der die Schüler aufgefordert werden, aus den Additionstheoremen der Sinus- und Kosinusfunktionen deren Subtraktionstheoreme herzuleiten. Die eben durchgeführten Betrachtungen zeigen, daß diese Aufgabe nur unter Benutzung zusätzlicher Eigenschaften der betreffenden Funktionen lösbar ist. (Die Schüler können die Geradheit bzw. die Ungeradheit der Funktionen benutzen.)

Wie wir aus Aufgabe 6.2.1 wissen, sind das Additionstheorem

$$\bigwedge_{x,y\in\mathbf{R}} f(x+y) = f(x)\cdot f(y)$$

und das Subtraktionstheorem

$$\bigwedge_{x,y\in\mathbf{R}} f(x-y) = \frac{f(x)}{f(y)}$$

der Exponentialfunktionen zueinander äquivalent. Dasselbe gilt für die Eigenschaften

$$\bigwedge_{x,y\in\mathbf{R}} f(x+y) = f(x) + f(y) \quad \text{und} \quad \bigwedge_{x,y\in\mathbf{R}} f(x-y) = f(x) - f(y)$$

der linearen Funktionen f mit $f(x) = c \cdot x$ $(c \in \mathbf{R})$ (vgl. Aufgabe 6.2.2). Das gilt übrigens allgemeiner für rationale Additions- und Subtraktionstheoreme (vgl. A. Kuwagaki [98], D. Ilse [75]).

12.1.6. Satz. *Es gibt genau zwei nichtkonstante Funktionen* $s\colon \mathbf{R} \to \mathbf{R}$ *und* $c\colon \mathbf{R} \to \mathbf{R}$, *die den Bedingungen* (12.1.1) *und* (12.1.2) *genügen*:

$$\bigwedge_{x,y\in\mathbf{R}} s(x-y) = s(x)\,c(y) - s(y)\,c(x),$$

$$\bigwedge_{x,y\in\mathbf{R}} c(x-y) = c(x)\,c(y) + s(x)\,s(y),$$

12.1.7. *s und c sind für alle $x \in \mathbf{R}$ differenzierbar,*

12.1.8. $s'(0) = a$ $(a \in \mathbf{R})$.

Beweis. Wir werden Funktionen s und c konstruieren und dann noch zeigen, daß sie mit den Sinus- und Kosinusfunktionen (den sogenannten „Kreisfunktionen" identisch sind. Zunächst wird eine Anzahl von Eigenschaften bewiesen, die allein aus den Subtraktionstheoremen (12.1.1) und (12.1.2) folgen. Mit (12.1.3), (12.1.4) und (12.1.5) liegen solche Eigenschaften bereits vor. Weiterhin gilt

12.1.9. $c(0) = 1$ und $s(0) = 0$.

Dazu setzt man $x = y = 0$ in (12.1.1) und (12.1.2): $s(0) = s(0)\,c(0) - s(0)\,c(0) = 0$; $c(0) = c(0)^2 + s(0)^2$, also $c(0)\big(1 - c(0)\big) = 0$. Damit gilt $s(0) = 0$ und entweder $c(0) = 0$ oder $c(0) = 1$.

Aus $c(0) = 0$ erhält man mit (12.1.2) für $y = 0$

$$c(x) = c(x)\,c(0) + s(x)\,s(0),$$

also $c(x) = 0$ für alle $x \in \mathbf{R}$. Damit ergibt sich aus (12.1.1) für $y = 0$

$$s(x) = s(x)\,c(0) - s(0)\,c(x) = 0 \qquad \text{für alle} \quad x \in \mathbf{R}.$$

Die beiden Funktionen s und c mit $s(x) = 0$ und $c(x) = 0$ jeweils für alle $x \in \mathbf{R}$ sind zwar Lösungen von (12.1.1) und (12.1.2), aber konstant, so daß sie nach Voraussetzung ausscheiden. Damit gilt (12.1.9).

Die Funktionen s und c mit $s(x) = 0$ und $c(x) = 1$ jeweils für alle $x \in \mathbf{R}$ sind ein weiteres konstantes Lösungspaar. Wegen (12.1.9) kann es keine weiteren konstanten Lösungspaare geben.

Mit (12.1.9) erhält man aus (12.1.2) für $y = x$, den sogenannten *trigonometrischen Pythagoras*:

12.1.10. $\bigwedge_{x\in\mathbf{R}} s(x)^2 + c(x)^2 = 1$

und daraus

12.1.11. $|s(x)| \leqq 1$ und $|c(x)| \leqq 1$ für alle $x \in \mathbf{R}$.

Für $y = x$ folgen aus (12.1.3) und (12.1.4) die sogenannten *Doppelwinkelformeln*

12.1.12. $\bigwedge_{x \in \mathbf{R}} s(2x) = 2s(x)\, c(x)$,

12.1.13. $\bigwedge_{x \in \mathbf{R}} c(2x) = c(x)^2 - s(x)^2 = 2c(x)^2 - 1$,

wobei die letzte Beziehung in (12.1.13) durch Einsetzen von (12.1.10) folgt.

Aus (12.1.4) und (12.1.2) folgt durch Subtraktion

$$\bigwedge_{x,y \in \mathbf{R}} c(x + y) - c(x - y) = -2s(x)\, s(y).$$

Für $x + y = \bar{x}$ und $x - y = \bar{y}$, also $\dfrac{\bar{x} + \bar{y}}{2} = x$ und $\dfrac{\bar{x} - \bar{y}}{2} = y$, erhält man daraus, wenn man anschließend noch $\bar{x}$ bzw. $\bar{y}$ in x bzw. y umbenennt:

12.1.14. $\bigwedge_{x,y \in \mathbf{R}} c(x) - c(y) = -2s\left(\dfrac{x + y}{2}\right) s\left(\dfrac{x - y}{2}\right)$.

Entsprechend ergibt sich aus (12.1.3) und (12.1.1)

12.1.15. $\bigwedge_{x,y \in \mathbf{R}} s(x) - s(y) = 2c\left(\dfrac{x + y}{2}\right) s\left(\dfrac{x - y}{2}\right)$.

Analog leitet man ab:

12.1.14.' $\bigwedge_{x,y \in \mathbf{R}} c(x) + c(y) = 2c\left(\dfrac{x + y}{2}\right) c\left(\dfrac{x - y}{2}\right)$,

12.1.15.' $\bigwedge_{x,y \in \mathbf{R}} s(x) + s(y) = 2s\left(\dfrac{x + y}{2}\right) c\left(\dfrac{x - y}{2}\right)$.

Jetzt benutzen wir die Forderungen (12.1.7) und (12.1.8) aus Satz 12.1.6 und bilden unter Verwendung von (12.1.15) für eine beliebige, aber fest gewählte Zahl $x \in \mathbf{R}$

$$s'(x) = \lim_{h \to 0} \frac{s(x + h) - s(x)}{h} = \lim_{h \to 0} \frac{2s\left(\dfrac{h}{2}\right) c\left(x + \dfrac{h}{2}\right)}{h}$$

$$= \lim_{h \to 0} \frac{s\left(\dfrac{h}{2}\right)}{\dfrac{h}{2}} \cdot \lim_{h \to 0} c\left(x + \frac{h}{2}\right).$$

Wegen $\lim\limits_{h \to 0} \dfrac{s\left(\dfrac{h}{2}\right)}{\dfrac{h}{2}} = s'(0) = a$ und $\lim\limits_{h \to 0} c\left(x + \dfrac{h}{2}\right) = c(x)$ (denn c ist differenzierbar, also stetig) gilt

12.1.16. $\bigwedge_{x \in \mathbf{R}} s'(x) = a \cdot c(x)$.

Ganz entsprechend schließt man auf

12.1.17. $\bigwedge_{x \in \mathbf{R}} c'(x) = -a \cdot s(x)$.[1])

Für $a = 0$ erhält man $s'(x) = 0$ und $c'(x) = 0$ für alle $x \in \mathbf{R}$. Also ist $s(x) = c_1$ und $c(x) = c_2$ für alle $x \in \mathbf{R}$. Damit sind wir auf konstante Funktionen gestoßen, die aber ausgeschlossen wurden. Deshalb können wir im folgenden stets $a \neq 0$ voraussetzen.

Die Ergebnisse (12.1.16) und (12.1.17) beinhalten, daß s und c beliebig oft differenzierbar sind und daß für alle natürlichen Zahlen $n \geqq 0$

$$s^{(2n)}(x) = (-1)^n a^{2n} s(x), \qquad s^{(2n+1)}(x) = (-1)^n a^{2n+1} c(x),$$
$$c^{(2n)}(x) = (-1)^n a^{2n} c(x), \qquad c^{(2n+1)}(x) = (-1)^{n+1} a^{2n+1} s(x)$$

gilt.

Die Funktionen s und c erfüllen für jede natürliche Zahl $n \geqq 0$ die Voraussetzungen des Taylorschen Satzes (vgl. MfL Bd. 5, 3.3.3.). Die Folge der Restglieder (R_n) konvergiert in beiden Fällen gegen Null, was man mit (12.1.11) sofort erkennt. Also werden die Funktionen s und c durch ihre Taylorschen Reihen dargestellt. Wegen $s^{(2n)}(0) = 0$, $s^{(2n-1)}(0) = (-1)^n a^{2n+1}$ $(n \geqq 0)$ und $c^{(2n)}(0) = (-1)^n a^{2n}$, $c^{(2n+1)}(0) = 0$ $(n \geqq 0)$ gilt

12.1.18. $$s(x) = \sum_{n=0}^{\infty} (-1)^n \frac{(ax)^{2n+1}}{(2n+1)!} \text{ und } c(x) = \sum_{n=0}^{\infty} (-1)^n \frac{(ax)^{2n}}{(2n)!}$$

für alle reellen Zahlen x.

Damit haben wir gezeigt: Es gibt höchstens ein nichtkonstantes Funktionenpaar, das die Eigenschaften (12.1.1), (12.1.2), (12.1.7) und (12.1.8) aus Satz 12.1.6 besitzt. Alle bisher bewiesenen Aussagen gelten nur, wenn es solche Funktionen überhaupt gibt.

Nun stellt man mit Hilfe des Quotientenkriteriums leicht fest, daß die Potenzreihen (12.1.18) für alle $x \in \mathbf{R}$ absolut konvergieren. Also sind die durch diese Reihen gegebenen Funktionen s und c für alle reellen Zahlen definiert und differenzierbar, und es gilt $s'(0) = a$. Es bleibt noch nachzuweisen, daß diese Funktionen auch die Subtraktionstheoreme (12.1.1) und (12.1.2) erfüllen. Das kann man mit bekannten Regeln über das Rechnen mit Potenzreihen beweisen (vgl. z. B. K. Knopp [89]). Wir geben hier einen anderen Beweis an (vgl. J. D. Mancill [111]).

Durch Differentiation der Potenzreihen (12.1.18) findet man zunächst $s'(x) = a \cdot c(x)$ und $c'(x) = -a \cdot s(x)$. Bildet man die Hilfsfunktionen $g_1 : \mathbf{R} \to \mathbf{R}$ und $g_2 : \mathbf{R} \to \mathbf{R}$ mit

$$g_1(x) := s(x - y) - \big(s(x)\, c(y) - c(x)\, s(y)\big)$$

und

$$g_2(x) := c(x - y) - \big(c(x)\, c(y) + s(x)\, s(y)\big),$$

[1]) Die Subtraktionstheoreme bewirken also, daß aus der Differenzierbarkeit der Lösung s von (12.1.1) und (12.1.2) an der Stelle 0 die Differenzierbarkeit beider Lösungen für jedes $x \in \mathbf{R}$ folgt.

wobei y beliebig, aber fest gewählt sei, und differenziert die Funktion g: $\mathbf{R} \to \mathbf{R}$ mit $g(x) := g_1(x)^2 + g_2(x)^2$, so erhält man

$$g'(x) = 2g_1(x) \cdot a \cdot g_2(x) - 2g_2(x) \cdot a \cdot g_1(x) = 0$$

für alle x. Daraus folgt $g(x) = c_0$ ($c_0 \in \mathbf{R}$). Wegen $g(0) = 0$ ist $g(x) = 0$ und damit auch $g_1(x) = 0 = g_2(x)$ für alle $x \in \mathbf{R}$. Das bedeutet aber, daß die Funktionen s und c die Gleichungen (12.1.1) und (12.1.2) erfüllen. Damit ist Satz 12.1.6 bewiesen.[1])

Wir haben die Funktionen s und c rein analytisch definiert und wollen jetzt noch zeigen, daß sie mit den im Mathematikunterricht aufgrund von geometrischen Überlegungen eingeführten Funktionen sin und cos identisch sind. Bei diesen geometrischen Betrachtungen wird im Unterricht eine Reihe von Begriffen (wie Winkelmaß und Bogenlänge) anschaulich benutzt, deren exakte Behandlung einen nicht unerheblichen Aufwand mit sich bringt.

Da s und c den Subtraktionstheoremen (12.1.1) und (12.1.2) genügen, besitzen diese Funktionen alle daraus folgenden Eigenschaften. Für die Identifizierung dieser Funktionen mit den „Kreisfunktionen" sin und cos beweisen wir darüber hinaus noch einige weitere Eigenschaften.

12.1.19. Die Funktion c besitzt eine kleinste positive Nullstelle.

Zunächst zeigen wir, daß c überhaupt eine positive Nullstelle hat und betrachten dazu die Zahl $c\left(\frac{2}{a}\right)$, wobei $a = s'(0) = \lim\limits_{h \to 0} \frac{s(h)}{h}$ ist. Wir können $a \neq 0$ voraussetzen. Es sei $a > 0$. Wir erhalten aus (12.1.18)

$$c\left(\frac{2}{a}\right) = 1 - \left(\frac{2^2}{2!} - \frac{2^4}{4!}\right) - \left(\frac{2^6}{6!} - \frac{2^8}{8!}\right) - \cdots.$$

Nun ist $\frac{2^k}{k!} - \frac{2^{k+2}}{(k+2)!} > 0$ für alle $k \in \mathbf{N}^*$. Für $k = 1$ ist das wegen $2 > \frac{8}{6}$ richtig. Aus $\frac{2^k}{k!} > \frac{2^{k+2}}{(k+2)!}$ erhält man durch Multiplikation mit 2

$$\frac{2^{k+1}}{k!} > \frac{2^{k+3}}{(k+2)!}.$$

Wegen $\frac{1}{k+1} > \frac{1}{k+3} > 0$ folgt weiterhin

$$\frac{2^{k+1}}{k!} \cdot \frac{1}{k+1} = \frac{2^{k+1}}{(k+1)!} > \frac{2^{k+3}}{(k+2)!} \cdot \frac{1}{k+3} = \frac{2^{k+3}}{(k+3)!}$$

für jede natürliche Zahl $k \geqq 1$.

[1]) Von der in Satz 12.1.6 geforderten Differenzierbarkeit wurde nur die Stetigkeit der Funktionen s und c benutzt und die Differenzierbarkeit der Funktion s an der Stelle 0. Auch das läßt sich weiter abschwächen (vgl. 12.2.).

Damit ist

$$c\left(\frac{2}{a}\right) < 1 - \left(\frac{2^2}{2!} - \frac{2^4}{4!}\right) = -\frac{1}{3}, \quad \text{also } c\left(\frac{2}{a}\right) < 0.$$

Da c differenzierbar, also insbesondere stetig ist, gibt es wegen $c(0) = 1 > 0$ und $c\left(\frac{2}{a}\right) < 0$ eine reelle Zahl x_0 mit $c(x_0) = 0$ und $0 < x_0 < \frac{2}{a}$. Für den Fall, daß $a < 0$ ist, folgt $c(-x_0) = 0$ aus der Geradheit von c (siehe (12.1.5)).

Hätte c keine kleinste positive Nullstelle, dann gäbe es eine Folge (w_n) mit $w_n > 0$ für alle $n \in \mathbf{N}$, $\lim_{n\to\infty} w_n = 0$ und $c(w_n) = 0$ für alle $n \in \mathbf{N}$. Daraus würde $\lim_{n\to\infty} c(w_n) = 0$ folgen. Dann müßte aber wegen der Stetigkeit von c auch $c(0) = 0$ gelten; das steht aber im Widerspruch zu (12.1.9).

Es gibt also eine kleinste positive Nullstelle der Funktion c, die im folgenden mit w bezeichnet wird.

12.1.20. Die Funktionen s und c sind periodisch und besitzen die primitive Periode $4w$.[1])

Wegen $c(w) = 0$ folgt aus (12.1.10), daß entweder $s(w) = 1$ oder $s(w) = -1$ gilt.

Fall 1. $s(w) = 1$.

Wir benutzen die Additionstheoreme (12.1.3) und (12.1.4). Dann gilt der Reihe nach

12.1.21. $\bigwedge_{x\in\mathbf{R}} s(x + w) = s(x)\,c(w) + s(w)\,c(x) = c(x)$,

12.1.22. $\bigwedge_{x\in\mathbf{R}} s(x + 2w) = s\big((x + w) + w\big) = c(x + w) = c(x)\,c(w) - s(x)\,s(w)$
$= -s(x)$,

12.1.23. $\bigwedge_{x\in\mathbf{R}} s(x + 3w) = -s(x + w) = -c(x)$,

12.1.24. $\bigwedge_{x\in\mathbf{R}} s(x + 4w) = s\big((x + 2w) + 2w\big) = -s(x + 2w) = s(x)$,

12.1.25. $\bigwedge_{x\in\mathbf{R}} c(x + w) = c(x)\,c(w) - s(x)\,s(w) = -s(x)$,

12.1.26. $\bigwedge_{x\in\mathbf{R}} c(x + 2w) = -s(x + w) = -c(x)$,

12.1.27. $\bigwedge_{x\in\mathbf{R}} c(x + 3w) = -c(x + w) = s(x)$,

12.1.28. $\bigwedge_{x\in\mathbf{R}} c(x + 4w) = s(x + w) = c(x)$.

Die Zahl $4w$ ist also eine Periode der Funktionen s und c; die Gleichungen (12.1.21)

[1]) Zur Orientierung sei daran erinnert, daß die Funktion $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = \cos ax$ die kleinste positive Nullstelle $w = \frac{\pi}{2a}$ besitzt ($a > 0$). Dann ist $\cos\frac{2}{a} < 0$ wegen $\frac{\pi}{2a} < \frac{4}{2a} = \frac{2}{a} < \frac{\pi}{a}$, und $\frac{4\pi}{2a} = \frac{2\pi}{a}$ ist primitive Periode von f.

bis (12.1.28) werden auch „Quadrantenbeziehungen" genannt (vgl. etwa R. Bittner, D. Ilse, S. Kubicek und W. Tietz [23]).

Beide Funktionen müssen eine kleinste positive Periode besitzen. Gäbe es nämlich zu jeder positiven Periode eine noch kleinere positive, so würde daraus, zusammen mit der Stetigkeit, folgen, daß beide Funktionen konstant sind. Nimmt man z. B. die Existenz einer reellen Zahl x_0 an mit $x_0 > 0$ und $s(x_0) = b \neq 0$, so könnte man ja zu jeder Umgebung $U(0)$ der Zahl 0 eine Periode w_n und eine natürliche Zahl k derart angeben, daß $x_0 - k \cdot w_n \in U(0)$ ist und $s(x_0 - k \cdot w_n) = b$. Wegen der Stetigkeit von s wäre dann auch $s(0) = b$ im Widerspruch zu (12.1.9). Also bleibt nur $s(x) = 0$ für alle $x \in \mathbf{R}$. Die konstanten Lösungen hatten wir aber ausgeschlossen.

Wegen (12.1.21) haben s und c die gleiche primitive Periode. Wir zeigen, daß dies die Zahl $4w$ ist. Dazu nehmen wir an, daß es eine Zahl $p < 4w$ gäbe, die primitive Periode von s und c ist. Wegen (12.1.9) wäre dann $s(p) = s(0) = 0$ und $c(p) = c(0) = 1$. Aus (12.1.13) und (12.1.10) folgten $c(p) = c\left(\frac{p}{2}\right)^2 - s\left(\frac{p}{2}\right)^2$ und $s\left(\frac{p}{2}\right)^2 + c\left(\frac{p}{2}\right)^2 = 1$. Hieraus ergäbe sich $c\left(\frac{p}{2}\right)^2 = 1$ und daraus wiederum mit (12.1.10) $s\left(\frac{p}{2}\right)^2 = 0$, also $s\left(\frac{p}{2}\right) = 0$. Wäre nun $c\left(\frac{p}{2}\right) = 1$, so müßte

$$\bigwedge_{x \in \mathbf{R}} c\left(x + \frac{p}{2}\right) = c(x)\, c\left(\frac{p}{2}\right) - s(x)\, s\left(\frac{p}{2}\right) = c(x)$$

gelten. Das steht im Widerspruch dazu, daß p primitive Periode sein sollte. Wäre $c\left(\frac{p}{2}\right) = -1$, so würde aus (12.1.13) $c\left(\frac{p}{2}\right) = c\left(\frac{p}{4}\right)^2 - s\left(\frac{p}{4}\right)^2 = -1$ und (12.1.10) $c\left(\frac{p}{4}\right)^2 + s\left(\frac{p}{4}\right)^2 = 1$ folgen, so daß $c\left(\frac{p}{4}\right)^2 = 0$ wäre und damit $c\left(\frac{p}{4}\right) = 0$. Da nun $0 < \frac{\pi}{4} < w$ ist, steht dieses Ergebnis im Widerspruch dazu, daß w kleinste positive Nullstelle von c ist. Damit ist also $4w$ tatsächlich primitive Periode von s und c.

Fall 2: $s(w) = -1$. (Der Beweis erfolgt analog zum ersten Fall.)

Jetzt kann die angestrebte Interpretation der Funktionen s und c vorgenommen werden (vgl. J. D. Mancill [111], K. Knopp [89]). Wegen (12.1.5) und der eben bewiesenen Periodizität dieser Funktionen genügt es, diese Interpretation im Intervall $[\![0, 4w[\![$ durchzuführen und sich dabei auf den Fall $0 < a = \lim\limits_{h \to 0} \frac{s(h)}{h}$ zu beschränken.

In der euklidischen Ebene sei ein rechtwinkliges kartesisches Koordinatensystem gegeben. Wir betrachten den Kreis, dessen Mittelpunkt im Koordinatenursprung liegt und dessen Radius $r = \frac{1}{a}$ ist. Er hat die Gleichung

12.1.29. $x^2 + y^2 = \frac{1}{a^2}$.

Wir behaupten, daß dieser Kreis die Parameterdarstellung

12.1.30. $x = \frac{1}{a} \cdot c(t),\ y = \frac{1}{a} \cdot s(t)$ mit $0 \leqq t < 4w$

besitzt.

Wegen $x^2 + y^2 = \frac{1}{a^2}\left(c(t)^2 + s(t)^2\right) = \frac{1}{a^2}$ liegen alle durch (12.1.30) gegebenen Punkte (x, y) auf dem Kreis (12.1.29). Umgekehrt gibt es aber auch für jeden Punkt (x, y) des Kreises genau ein t, so daß $x = \frac{1}{a} \cdot c(t)$ und $y = \frac{1}{a} \cdot s(t)$ gilt.

Wir zeigen zunächst, daß die Funktionen c und s in $[\![0, w]\!]$ streng monoton fallend bzw. streng monoton wachsend sind. Aus der Stetigkeit von c, aus $c(0) = 1$ und daraus, daß w kleinste positive Nullstelle von c ist, folgt $c(x) > 0$ für alle x mit $0 < x < w$. Wegen $s'(x) = a \cdot c(x)$ ist dann s' in diesem Intervall positiv, also s selbst streng monoton wachsend (für $a > 0$ ist damit $s(w) = 1$). Da $c'(x) = -a \cdot s(x)$ gilt, ist c' in dem betrachteten Intervall negativ und damit c selbst streng monoton fallend (vgl. MfL Bd. 5, 3.1.4.). Also gibt es zu jedem Kreispunkt (x, y) im ersten Quadranten $\left(x \in \left[\!\left[\frac{1}{a}, 0\right]\!\right[, y \in \left[\!\left[0, \frac{1}{a}\right]\!\right]\right)$ genau ein $t \in [\![0, w]\!]$, so daß $x = \frac{1}{a} \cdot c(t)$ und $y = \frac{1}{a} \cdot s(t)$ gilt. Aufgrund der Quadrantenbeziehungen (12.1.21) bis (12.1.28) gibt es dann auch für jeden Punkt des Kreises im zweiten Quadranten genau ein $t \in [\![w, 2w]\!]$, für jeden Punkt aus dem dritten bzw. vierten Quadranten genau ein $t \in [\![2w, 3w]\!]$ bzw. genau ein $t \in [\![3w, 4w]\!]$ mit $x = \frac{1}{a} \cdot c(t)$ und $y = \frac{1}{a} \cdot s(t)$. Anschaulich heißt das: Durchläuft t das Intervall $[\![0, 4w[\![$, so durchläuft der zugehörige Punkt (x, y), in $\left(\frac{1}{a}, 0\right)$ beginnend, im mathematisch positiven Sinn den Kreis (12.1.29) genau einmal.

Darüber hinaus besteht zwischen dem Parameter t und den Punkten (x, y) des Kreises folgender Zusammenhang:

12.1.31. Der Kreisbogen vom Punkt $\left(\frac{1}{a}, 0\right)$ bis zum Punkt $\left(\frac{1}{a} \cdot c(T), \frac{1}{a} \cdot s(T)\right)$ hat die Länge T.

Bezeichnet man nämlich die Länge dieses Bogens mit l, so gilt (vgl. MfL Bd. 5' 4.4.1.)

$$l = \int_0^T \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

Mit (12.1.16) und (12.1.17) ergibt sich

$$\frac{dx}{dt} = \frac{1}{a}\left(-a \cdot s(t)\right) = -s(t), \quad \frac{dy}{dt} = \frac{1}{a}\left(a \cdot c(t)\right) = c(t)$$

und daraus wegen (12.1.10)

$$l = \int_0^T \sqrt{\dot{s}(t)^2 + \dot{c}(t)^2}\, dt = \int_0^T 1 dt = T \quad \text{(siehe Abb. 60).}$$

Speziell für $T = 4w$ ergibt sich $l = 4w = \frac{2\pi}{a}$, weil $2\pi r = \frac{2\pi}{a}$ der Umfang des Kreises ist. Ist außerdem $a = 1$, so wird $4w = 2\pi$ und $w = \frac{\pi}{2}$ (das ist bekanntlich die kleinste positive Nullstelle der Funktion cos).

Es sei nun ein beliebiges Winkelmaß gegeben. Bei dieser Messung sei s das Maß des gestreckten Winkels und α_s das eines beliebigen Winkels. Dann gilt (siehe Abb. 61) $\frac{\alpha_s}{T} = \frac{s}{\frac{\pi}{a}}$, d. h. $\alpha_s = \frac{a \cdot s}{\pi} \cdot T$. Wählt man jetzt $a = \frac{\pi}{s}$, also $r = \frac{1}{a} = \frac{s}{\pi}$, so wird $\alpha_s = T$. Geht man also zu diesem Kreis über, so ist α_s das Maß des Bogens T (in Längeneinheiten mit dem $\frac{s}{\pi}$-ten Teil des Radius als Einheit[1])). Ist beispielsweise für $s = 180°$ das Maß eines Winkels $\alpha_{180°} = x°$, so gilt also $x° = \frac{a \cdot 180°}{\pi} \cdot \hat{x}$, und für $a = \frac{\pi}{180°}$ wird $x° = \hat{x}$.

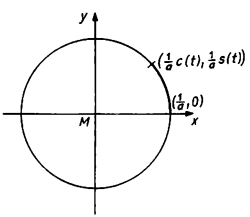


Abb. 60

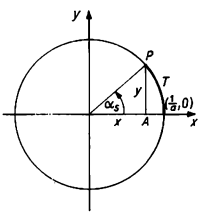


Abb. 61

Im Kreis mit $r = \frac{s}{\pi}$ gilt nach Definition der Sinus- und Kosinusfunktionen (siehe Abb. 61) $\sin \alpha_s = \frac{y}{r} = \frac{\pi}{s} \cdot y$ und $\cos \alpha_s = \frac{x}{r} = \frac{\pi}{s} \cdot x$, also $y = \frac{s}{\pi} \cdot \sin \alpha_s$ und $x = \frac{s}{\pi} \cdot \cos \alpha_s$. Andererseits war $y = \frac{1}{a} \cdot s(t)$ und $x = \frac{1}{a} \cdot c(t)$, also in diesem Kreis $y = \frac{s}{\pi} \cdot s(t)$ und $x = \frac{s}{\pi} \cdot c(t)$. Wir erhalten durch Vergleich die angestrebte Identifizierung $s(t) = \sin \alpha_s$ und $c(t) = \cos \alpha_s$.

Unsere Betrachtungen vor Satz 12.1.6 haben gezeigt, daß die Sinus- und Kosinusfunktionen nicht als differenzierbare Lösungen ihrer Additionstheoreme cha-

[1]) Diese Einheit ist gleich dem $2s$-ten Teil des Kreisumfangs, und das ist $\frac{2\pi r}{2s} = \frac{r}{\frac{s}{\pi}}$.

rakterisiert werden können. Nun kann man zeigen, daß die Funktionen $s\colon \mathbf{R} \to \mathbf{R}$ und $c\colon \mathbf{R} \to \mathbf{R}$ mit $s(x) = a^x \sin bx$ und $c(x) = a^x \cos bx$ $(a > 0)$ (neben den Funktionen s und c mit $s(x) = 0$ und $c(x) = 0$ für alle $x \in \mathbf{R}$) die einzigen differenzierbaren Lösungen der Additionstheoreme sind (vgl. etwa D. Ilse [76]). Das legt die Vermutung nahe, daß man eine Charakterisierung der Winkelfunktionen sin und cos erhält, wenn man noch eine Zusatzforderung stellt, durch die $|s(x)| \leqq 1$ und $|c(x)| \leqq 1$ für alle Argumente erreicht wird. Das ist in der Tat der Fall und wird z. B. durch (12.1.10) gesichert. Andere derartige Zusatzforderungen, die dasselbe leisten, sind z. B. $\bigwedge_{x \in \mathbf{R}} c(-x) = c(x)$ (denn c mit $c(x) = a^x \cos bx$ ist für $a \neq 1$ nicht gerade) oder $\lim_{x \to 0} \frac{c(x) - 1}{x} = 0$ (denn für c mit $c(x) = a^x \cos bx$ gilt $\lim_{x \to 0} \frac{c(x) - 1}{x} = \ln a \neq 0$, wenn $a \neq 1$).

Weitere Charakterisierungsmöglichkeiten erhält man z. B. durch geeignete Kombinationen eines Additions- oder Subtraktionstheorems mit anderen Eigenschaften (vgl. L. Vietoris [176], P. Wiechmann [179]).

Auf Charakterisierungen durch Differentialgleichungen gehen wir hier nicht ein. Es sei nur erwähnt, daß in M. Barner [19] eine Charakterisierung mit Hilfe der Beziehungen (12.1.16) und (12.1.17) (für $a = 1$) vorgenommen wird.

12.1.32. Aufgaben

1. Durch Angabe geeigneter Funktionen sind die folgenden Aussagen über die Additions- bzw. Subtraktionstheoreme (12.1.1), (12.1.2), (12.1.3) und (12.1.4) zu beweisen:

a) Jedes der Systeme ((12.1.1), (12.1.3)), ((12.1.3), (12.1.4)) und ((12.1.1), (12.1.4)) ist unabhängig.

b) Aus keinem der in a) angegebenen Systeme folgt (12.1.2).

c) (12.1.3) ist unabhängig von dem System ((12.1.1), (12.1.4)).

d) (12.1.1) ist unabhängig von dem System ((12.1.3), (12.1.4)).

e) (12.1.4) ist unabhängig von dem System ((12.1.1), (12.1.3)).

f) Aus der Gültigkeit von (12.1.3) und der Geradheit von c folgt nicht die Gültigkeit von (12.1.4).

g) Aus der Gültigkeit von (12.1.4) und der Ungeradheit von s folgt nicht die Gültigkeit von (12.1.3).

2. Man zeige:

a) Für jede Lösung (s, c) mit $s\colon \mathbf{R} \to \mathbf{R}$ und $c\colon \mathbf{R} \to \mathbf{R}$ von (12.1.1) gilt: s ist ungerade.

b) Jede Lösung (s, c) mit $s\colon \mathbf{R} \to \mathbf{R}$ und $c\colon \mathbf{R} \to \mathbf{R}$ von (12.1.3), bei der s ungerade ist, ist auch Lösung von (12.1.1).

c) Jede Lösung (s, c) mit $s\colon \mathbf{R} \to \mathbf{R}$ und $c\colon \mathbf{R} \to \mathbf{R}$ von (12.1.3), bei der c gerade ist, ist auch Lösung von (12.1.1).

d) Jede Lösung (s, c) mit $s\colon \mathbf{R} \to \mathbf{R}$ und $c\colon \mathbf{R} \to \mathbf{R}$ von (12.1.4), bei der s ungerade ist, ist auch Lösung von (12.1.1).

e) Jede Lösung (s, c) mit $s\colon \mathbf{R} \to \mathbf{R}$ und $c\colon \mathbf{R} \to \mathbf{R}$ von (12.1.4), bei der c gerade ist, ist auch Lösung von (12.1.2).

3. Man zeige:

a) Die Systeme ((12.1.2), (12.1.3)), ((12.1.1), (12.1.2)) und ((12.1.2), (12.1.4)) sind äquivalent.

b) (12.1.2) ist äquivalent dem System ((12.1.1), (12.1.3), (12.1.4)).

4. Man zeige: (s, c) mit $s: \mathbf{R} \to \mathbf{R}$ und $c: \mathbf{R} \to \mathbf{R}$ ist genau dann eine Lösung von (12.1.2), wenn es die folgenden Bedingungen erfüllt:

a) (12.1.1), c ist gerade, $\bigwedge_{x \in \mathbf{R}} c(0) = c(x)\, c(-x) - s(x)\, s(-x)$,

$c(x - y) = c(x)\, c(y)$ für alle x, y mit $s(x) = s(y) = 0$.

b) (12.1.1), (12.1.10), c ist gerade für alle x mit $s(x) = 0$,

$c(x - y) = c(x)\, c(y)$ für alle x, y mit $s(x) = s(y) = 0$.

c) (12.1.3), $\bigwedge_{x \in \mathbf{R}} s(x)^2 = \frac{1}{2}(1 - c(2x))$, $\bigwedge_{x \in \mathbf{R}} c(x)^2 = \frac{1}{2}(1 + c(2x))$.

d) (12.1.4), (12.1.10).

5. Man beweise: Die Funktionen $s: \mathbf{R} \to \mathbf{R}$ und $c: \mathbf{R} \to \mathbf{R}$ mit $s(x) = \sin ax$ und $c(x) = \cos ax$ $(a = 0)$ sind die einzigen nichtkonstanten Funktionen, für die folgende Bedingungen gelten: (12.1.3), (12.1.4), (12.1.10), s und c sind für alle $x \in \mathbf{R}$ differenzierbar, $s'(0) = a$. (Hinweis: Wenn der Reihe nach bewiesen wird: s und c haben keine gemeinsamen Nullstellen; $c(0) = 1$, $s(0) = 0$; $\bigwedge_{x,h \in \mathbf{R}} s(x + h) - s(x) = 2s\left(\frac{h}{2}\right) c\left(x + \frac{h}{2}\right)$ und $\bigwedge_{x,h \in \mathbf{R}} c(x + h) - c(x) = -2s\left(\frac{h}{2}\right) c\left(x + \frac{h}{2}\right)$; s ist ungerade; $s'(x) = a \cdot c(x)$, $c'(x) = -a \cdot s(x)$, ergibt sich ein Anschluß an Satz 12.1.6.)

12.2. Äquivalente Regularitätsforderungen bei der funktionalen Charakterisierung der Sinus- und Kosinusfunktionen

In 12.1. haben wir bewiesen, daß die Sinus- und Kosinusfunktionen die einzigen (nichtkonstanten) differenzierbaren Lösungen ihrer Subtraktionstheoreme sind. Wir haben aber auch gesehen, daß die beiden Subtraktionstheoreme gleichwertig ersetzt werden können. (In Aufgabe 12.1.32.5 z. B. durch die Additionstheoreme dieser Funktionen und die Eigenschaft $\bigwedge_{x \in \mathbf{R}} s(x)^2 + c(x)^2 = 1$.) Es gibt noch eine ganze Reihe weiterer zu den Subtraktionstheoremen äquivalenter Gleichungssysteme (siehe P. Wiechmann [179], woraus auch einige der Aufgaben 12.1.32 entnommen sind). Es sei noch erwähnt, daß von den in 12.1. betrachteten Gleichungen (12.1.1) bis (12.1.4), die von vielen Autoren untersucht worden sind (weitere Literaturangaben findet man in J. Aczél [3], [5]), das Subtraktionstheorem (12.1.2) in dem Sinne „am stärksten" ist, daß es im wesentlichen allein schon ausreicht, die Sinus- und Kosinusfunktionen zu charakterisieren (vgl. etwa H. E. Vaughan [175], S. I. Novoselov [126]).

Wie wir aus 12.1. wissen, leisten dies die Additionstheoreme (12.1.3) und (12.1.4) nicht. Auch das Subtraktionstheorem (12.1.1) hat noch andere differenzierbare Lösungen, wie das Paar (s, c) mit $s(x) = x$ und $c(x) = x + 1$ für alle reellen Zahlen x zeigt (vgl. P. Wiechmann [179]).

Nach den Untersuchungen über die Äquivalenz von Funktionalgleichungssystemen bei der Charakterisierung von Winkelfunktionen (auf die wir in 12.3. noch einmal zurückkommen werden), wollen wir jetzt ein solches System festhalten, z. B. die beiden Subtraktionstheoreme, und nach Eigenschaften fragen, die zur Differenzierbarkeit äquivalent sind. Diese Frage wird einmal durch entsprechende Untersuchungen bei der Charakterisierung anderer elementarer Funk-

tionen in den Kapiteln 7 und 8 nahegelegt, zum anderen aber auch durch die Tatsache, daß für den Beweis von Satz 12.1.6 von der für s und c geforderten Differenzierbarkeit nur die Stetigkeit beider Funktionen und die Differenzierbarkeit der Funktion s an der Stelle 0 gebraucht wurde. Vielleicht genügt noch weniger?

Zunächst überzeugen wir uns, daß man nicht ohne eine zusätzliche Forderung zu den Subtraktionstheoremen auskommt. Diese Gleichungen haben nämlich auch unstetige Lösungen, die in dem Sinne „total unstetig" sind, daß ihre Graphen den Streifen $B = \{(x, y): x \in \mathbf{R} \wedge -1 \leqq y \leqq 1\}$ der x,y-Ebene dicht bedecken. Zum Beweis dieser Behauptung betrachten wir das für alle reellen Zahlen definierte Funktionenpaar (s, c) mit $s(x) = \sin\big(f(x)\big)$, $c(x) = \cos\big(f(x)\big)$, wobei f eine beliebige unstetige Lösung der Cauchyschen Funktionalgleichung

$$\bigwedge_{x,y \in \mathbf{R}} f(x + y) = f(x) + f(y)$$

ist. Für jede Lösung dieser Funktionalgleichung gilt

$$\bigwedge_{x,y \in \mathbf{R}} f(x - y) = f(x) - f(y).$$

Das folgt aus $f\big(x + (-x)\big) = f(0) = f(x) + f(-x)$; damit ist $\bigwedge_{x \in \mathbf{R}} f(-x) = -f(x)$. Also gilt für s und c

$$\begin{aligned}\bigwedge_{x,y \in \mathbf{R}} \sin\big(f(x - y)\big) &= \sin\big(f(x) - f(y)\big) \\ &= \sin\big(f(x)\big) \cdot \cos\big(f(y)\big) - \sin\big(f(y)\big) \cdot \cos\big(f(x)\big),\end{aligned}$$

und das ist (12.1.1). Entsprechend gilt auch (12.1.2). Weiterhin sind die Funktionen s und c unstetig und bedecken außerdem den Streifen B dicht. Ist nämlich $(x_0, y_0) \in B$ beliebig gewählt, dann gibt es ein $x_1 \in \mathbf{R}$, so daß $y_0 = \sin x_1$ ist. Wegen der Stetigkeit der Funktion sin gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, so daß für jedes x gilt: Wenn $|x - x_1| < \delta$, so $|\sin x - y_0| < \varepsilon$. Nach 7.2. bedecken die Graphen der unstetigen Lösungen der Cauchyschen Grundgleichung die Ebene dicht. Insbesondere gibt es in jeder Umgebung von x_0 Zahlen $\bar{x}$ derart, daß $|f(\bar{x}) - x_1| < \delta$ gilt. Dann ist aber auch $\big|\sin\big(f(\bar{x})\big) - y_0\big| < \varepsilon$. Entsprechendes gilt für die Funktion c.

Man kann zeigen, daß alle Lösungen (c, s) der Gleichung (12.1.2) von der Form $c(x) = \cos\big(f(x)\big)$ und $s(x) = \sin\big(f(x)\big)$ sind, wobei f Lösung der Kongruenz

$$\bigwedge_{x,y \in \mathbf{R}} f(x + y) \equiv f(x) + f(y) \pmod{2\pi}$$

ist (vgl. J. G. VAN DER CORPUT [36]).

Jede Forderung an die Lösungen von ((12.1.1), (12.1.2)), die (anschaulich gesprochen) verhindert, daß die Graphen der Funktionen in B dicht liegen, sichert also deren Stetigkeit.

Im folgenden Satz werden einige derartige Forderungen, die man auch als Regularitätsforderungen bezeichnet, zusammengestellt und deren Äquivalenz bewiesen (vgl. auch M. KRAFFT [91], E. HOPF [69], E. GRUHLKE [63]).

12.2.1. Satz. *Für alle nichtkonstanten Lösungen* $s\colon \mathbf{R} \to \mathbf{R}$ *und* $c\colon \mathbf{R} \to \mathbf{R}$ *der Funktionalgleichungen* (12.1.1) *und* (12.1.2)

$$\bigwedge_{x,y\in\mathbf{R}} s(x-y) = s(x)\,c(y) - s(y)\,c(x),$$

$$\bigwedge_{x,y\in\mathbf{R}} c(x-y) = c(x)\,c(y) + s(x)\,s(y)$$

sind folgende Eigenschaften äquivalent:

(1) *Es gibt ein Intervall* $[\![0, a_1[\![$ $(a_1 \in \mathbf{R}_+{}^*)$, *in dem die Funktion c streng monoton fallend ist.*

(2) *Es gibt ein Intervall* $[\![0, a_2[\![$ $(a_2 \in \mathbf{R}_+{}^*)$, *in dem die Funktion c monoton fallend ist.*

(3) *Es gibt ein Intervall* $[\![0, a_3[\![$ $(a_3 \in \mathbf{R}_+{}^*)$, *in dem die Funktion s streng monoton ist.*

(4) *Es gibt ein Intervall* $[\![0, a_4[\![$ $(a_4 \in \mathbf{R}_-{}^*)$, *in dem die Funktion s monoton ist.*

(5) *Es gibt ein Intervall* $[\![0, a_5[\![$ $(a_5 \in \mathbf{R}_+{}^*)$, *in dem s nicht das Vorzeichen wechselt.*

(6) *Die Funktion s ist an der Stelle* 0 *stetig.*

(7) *Die Funktion s ist an der Stelle* 0 *rechtsseitig stetig.*

(8) *Es gibt eine Stelle* $x_0 \in \mathbf{R}$, *an der s rechtsseitig stetig ist.*

(9) *Die Funktion c ist an der Stelle* 0 *stetig.*

(10) *Die Funktion c ist für alle* $x \in \mathbf{R}$ *stetig.*

(11) *Die Funktion s ist an der Stelle* 0 *differenzierbar.*

(12) *Die Funktion s ist für jedes* $x \in \mathbf{R}$ *differenzierbar.*

(13) *Die Funktionen s und c sind für alle* $x \in \mathbf{R}$ *differenzierbar.*

Zum Beweis benutzen wir Ideen aus H. E. Vaughan [175], E. Gruhlke [63]. Wir zeigen, daß (2) aus (1) folgt, (3) aus (2) folgt usw. und schließlich (1) aus (13) folgt. Dabei können wir die in 12.1. allein aus (12.1.1) und (12.1.2) (also ohne Regularitätsforderung) abgeleiteten Eigenschaften (12.1.3), (12.1.4), (12.1.5), (12.1.9) bis (12.1.15′) von s und c benutzen. Wir zeigen zunächst:

12.2.2. Aus (12.1.1) und (12.1.2) folgt: Wenn $\lim\limits_{x\to 0+} c(x)$ existiert, so ist s an der Stelle 0 stetig.

Ersetzt man in der Gleichung (12.1.3) x durch $2x$ und y durch x und beachtet (12.1.12) $\bigwedge\limits_{x\in\mathbf{R}} s(2x) = 2s(x)\,c(x)$ und (12.1.13) $\bigwedge\limits_{x\in\mathbf{R}} c(2x) = 2c(x)^2 - 1$, so erhält man

12.2.3. $\bigwedge\limits_{x\in\mathbf{R}} s(3x) = s(x)\left(4c(x)^2 - 1\right)$.

Es sei nun $\lim\limits_{x\to 0+} c(x) = g$. Für diese Zahl g gilt wegen (12.1.13) $g = 2g^2 - 1$, also $g = -\frac{1}{2}$ oder $g = 1$. Für $g = -\frac{1}{2}$ ergibt sich aus (12.2.3) und (12.1.11) $\lim\limits_{x\to 0+} s(3x) = 0$. Andererseits folgt aber nun aus (12.1.10), $\lim\limits_{x\to 0+} s(x)^2 = \frac{3}{4}$, so daß der Fall $g = -\frac{1}{2}$ nicht eintreten kann. Für $g = 1$ erhält man aus (12.1.10) $\lim\limits_{x\to 0+} s(x)^2$

$= \lim_{x \to 0+} \left(1 - c(x)^2\right) = 0$, also auch $\lim_{x \to 0+} s(x) = 0$. Wegen (12.1.5) $\wedge$ $s(-x) = -s(x)$ $\forall x \in \mathbf{R}$ gilt dann $\lim_{x \to 0} s(x) = 0$, so daß wegen (12.1.9) $s(0) = 0$ die Funktion s an der Stelle 0 stetig ist.

Wir kommen nun zum Beweis der Implikationskette.

(1) $\Rightarrow$ (2). (2) ist eine unmittelbare Folge aus (1).

(2) $\Rightarrow$ (3). Es sei (s, c) eine Lösung von ((12.1.1), (12.1.2)) und c im Intervall $[\![0, a_2[\![$ monoton fallend. Es ist zu zeigen, daß es dann ein Intervall $[\![0, a_3[\![$ gibt, in dem s streng monoton ist.[1]) Zunächst zeigen wir

a) Es gibt eine rechtsseitige Umgebung $[\![0, a_2'[\![$ von 0, in der s nicht das Vorzeichen wechselt.

b) Es gibt eine rechtsseitige Umgebung $]\!]0, a_2''[\![$ von 0, in der $0 < c(x) < 1$ ist.

Zu a). Würde s in jeder rechtsseitigen Umgebung von 0 das Vorzeichen wechseln, so gäbe es Zahlen $e, f \in [\![0, a_2[\![$ mit $e - f$, $e + f \in [\![0, a_2[\![$ und $s(e)\, s(f) < 0$. Damit würde man aus (12.1.2) $c(e - f) = c(e)\, c(f) + s(e)\, s(f)$ und (12.1.4) $c(e + f) = c(e)\, c(f) - s(e)\, s(f)$ die Ungleichung $c(e - f) < c(e)\, c(f) < c(e + f)$ erhalten. Das steht aber im Widerspruch zu (2).

Zu b). Wenn c in $[\![0, a_2[\![$ monoton fallend ist, gibt es eine Zahl $a_2'' > 0$ mit $c(a_2'') > 0$. Gäbe es nämlich eine solche Zahl a_2'' nicht, so wäre $c(x) \leqq 0$ in einem Intervall $]\!]0, \bar{a}_2[\![$. Wir wählen wieder Zahlen $e, f \in]\!]0, \bar{a}_2[\![$ mit $e - f$, $e + f \in]\!]0, \bar{a}_2[\![$. Dann folgt, wieder aus (12.1.2) und (12.1.4), daß $s(e)\, s(f) \leqq 0$ und $s(e)\, s(f) \geqq 0$, also $s(e)\, s(f) = 0$ gelten würde. Es sei o. B. d. A. $s(e) = 0$. Das hätte $c(e - f) = 0 = c(e + f)$ zur Folge. Da c monoton fallend ist, wäre $c(e) = 0$ für alle $x \in [\![e - f, e + f]\!]$, also auch $c(e) = 0$ und damit $c(0) = c(e - e) = c(e)\, c(e) + s(e)\, s(e) = 0$ im Widerspruch zu (12.1.9) $c(0) = 1$. Wegen der Monotonie von c ist $c(x) > 0$ für alle $x \in [\![0, a_2''[\![$. Es gilt aber auch $c(x) < 1$ für alle $x \in]\!]0, a_2''[\![$. Anderenfalls müßte es wegen des monotonen Fallens von c ein x_0 geben mit $0 < x_0 < a_2''$ und $c(x) = 1$ für alle $x \in [\![0, x_0[\![$. Da die konstante Lösung c mit $c(x) = 1$ für alle $x \in \mathbf{R}$ ausgeschlossen ist, gibt es ein $x_1 > 0$ mit $c(x_1) \neq 1$ und $s(x_1) \neq 0$ wegen $c(x_1)^2 + s(x_1)^2 = 1$. Dann ist auch $c\left(\frac{x_1}{2}\right) \neq 1$. Aus $c\left(\frac{x_1}{2}\right) = 1$ würde nämlich $s\left(\frac{x_1}{2}\right) = 0$ und daraus wegen (12.1.13) $c(x_1) = c\left(\frac{x_1}{2}\right)^2 - s\left(\frac{x_1}{2}\right)^2 = 1$ folgen. Also gibt es ein $k \in \mathbf{N}$ mit $\frac{x_1}{2^k} < x_0$ und $c\left(\frac{x_1}{2^k}\right) \neq 1$ im Widerspruch zur Annahme, daß $c(x) = 1$ für alle $x \in [\![0, x_0[\![$ gilt. Damit ist b) bewiesen.

Es sei nun $a_3 := \min(a_2', a_2'')$. Dann gilt also $0 < c(x) < 1$ für alle $x \in]\!]0, a_3[\![$ und wegen (12.1.10) $s(x) \neq 0$ für alle $x \in]\!]0, a_3[\![$. Da s in $]\!]0, a_3[\![$ nicht das Vorzeichen wechselt, gilt also entweder $s(x) > 0$ oder $s(x) < 0$ in diesem Intervall. Jetzt können wir zeigen, daß s in $[\![0, a_3[\![$ streng monoton ist.

[1]) Die Funktion s kann in einer geeigneten rechtsseitigen Umgebung von 0 streng monoton wachsend, aber auch streng monoton fallend sein. Das hängt von der Zahl $a = s'(0)$ ab, die positiv oder negativ sein kann. Darauf haben die angegebenen Voraussetzungen aber keinen Einfluß. Aus der Ungeradheit der Sinusfunktion folgt $\sin((-a)\, x) = -\sin ax$ für $a > 0$. Für $a > 0$ ist s monoton wachsend, für $a < 0$ monoton fallend.

Es seien nämlich $e, f \in [\![0, a_3[\![$ beliebig mit $e < f$ und $f - e$, $f + e \in [\![0, a_3[\![$ gegeben. Dann ist nach (12.1.15) $s(f) - s(e) = 2c\left(\frac{f+e}{2}\right) s\left(\frac{f-e}{2}\right)$, also $s(f) - s(e) > 0$, wenn $s(x) > 0$ ist für alle $x \in]\!]0, a_3[\![$ und $s(f) - s(e) < 0$, wenn $s(x) < 0$ für alle $x \in]\!]0, a_3[\![$ ist. In jedem Fall ist s in $[\![0, a_3[\![$ streng monoton.

(3) $\Rightarrow$ (4). (4) ist eine unmittelbare Folge von (3).

(4) $\Rightarrow$ (5). Nach Voraussetzung ist s in einem Intervall $[\![0, a_4[\![$ monoton. Wenn s monoton wachsend ist, so können wegen (12.1.9) $s(0) = 0$ keine negativen Funktionswerte von s in $[\![0, a_4[\![$ auftreten. Wenn s monoton fallend ist, so können aus dem gleichen Grund keine positiven Funktionswerte von s in $[\![0, a_4[\![$ auftreten. In beiden Fällen gibt es also keinen Vorzeichenwechsel in $[\![0, a_4[\![$.

(5) $\Rightarrow$ (6). Die Funktion s möge also in einem Intervall $[\![0, a_5[\![$ nicht das Vorzeichen wechseln. Wir wählen beliebige Zahlen $e, f \in [\![0, a_5[\![$ mit $0 < e < f < a_5$. Für diese Zahlen gilt dann $0 < \frac{f-e}{2} < \frac{f+e}{2} < a_5$. Damit folgt aus (12.1.14) $c(f) - c(e) = -2s\left(\frac{f+e}{2}\right) s\left(\frac{f-e}{2}\right)$, daß $c(f) - c(e) \leqq 0$ ist. Also ist c in $[\![0, a_5[\![$ monoton fallend. Wenn nun (x_n) eine monoton fallende Nullfolge ist, dann ist $\big(c(x_n)\big)$ eine monoton wachsende Folge, die wegen (12.1.11) $\bigwedge_{x \in \mathbb{R}} |c(x)| \leqq 1$ auch beschränkt ist, also konvergiert, d. h. $\lim_{x \to 0+} c(x)$ existiert. Daraus folgt nach (12.2.2), daß s an der Stelle 0 stetig ist.

(6) $\Rightarrow$ (7). (7) ist eine unmittelbare Folge von (6).

(7) $\Rightarrow$ (8). (8) ist eine unmittelbare Folge von (7).

(8) $\Rightarrow$ (9). Es sei s in x_0 rechtsseitig stetig. Falls $s(x_0) \neq 0$ ist, so folgt aus (12.1.15') (für $x = x_0$)

$$\lim_{y \to x_0 -} \big(s(x_0) + s(y)\big) = 2 \lim_{y \to x_0 +} \left(s\left(\frac{x_0 + y}{2}\right) \cdot c\left(\frac{x_0 - y}{2}\right)\right),$$

daß

$$2s(x_0) = 2s(x_0) \cdot \lim_{y \to x_0 +} c\left(\frac{x_0 - y}{2}\right)$$

gilt und damit auch $\lim_{y \to x_0 +} c\left(\frac{x_0 - y}{2}\right) = 1$, d. h. $\lim_{x \to 0-} c(x) = 1$. Da c gerade ist (siehe (12.1.5)), erhält man $\lim_{x \to 0} c(x) = 1$, so daß also wegen $c(0) = 1$ die Funktion c an der Stelle 0 stetig ist. Falls $s(x_0) = 0$ ist, so folgt aus (12.1.10), daß $c(x_0) \neq 0$ ist. Damit erhält man aus (12.1.3)

$$\lim_{x \to 0+} s(x_0 + x) = c(x_0) \lim_{x \to 0+} s(x) + s(x_0) \lim_{x \to 0+} c(x),$$

daß $\lim_{x \to 0+} s(x) = 0$ ist. Wegen (12.1.10) existiert dann auch $\lim_{x \to 0+} c(x)^2$ und wegen (12.1.13) $\lim_{x \to 0+} c(2x)$, was die Existenz von $\lim_{x \to 0+} c(x)$ impliziert. Daraus folgt schließ-

lich (vgl. Beweis von (12.2.2)), daß $\lim_{x\to 0} c(x) = 1$, daß also c wegen $c(0) = 1$ an der Stelle 0 stetig ist.

(9) $\Rightarrow$ (10). Es sei c an der Stelle 0 stetig. Dann folgt aus (12.2.2) sofort, daß auch s an der Stelle 0 stetig ist. Für eine beliebige, aber feste Zahl x_0 ist nach (12.1.14)

$$|c(x_0) - c(x)| = 2\left|s\left(\frac{x_0 + x}{2}\right)\right| \cdot \left|s\left(\frac{x_0 - x}{2}\right)\right|.$$

Wegen (12.1.11) $\wedge_{x\in\mathbf{R}} |s(x)| \leqq 1$ gilt dann auch $|c(x_0) - c(x)| \leqq 2\left|s\left(\frac{x_0 - x}{2}\right)\right|$, also ist $\lim_{x\to x_0} |c(x_0) - c(x)| = 0$, d. h. $\lim_{x\to x_0} c(x) = c(x_0)$. Die Funktion c ist damit an der Stelle x_0 stetig.

(10) $\Rightarrow$ (11). Es sei c für alle $x \in \mathbf{R}$ stetig. Wir zeigen zunächst, daß dann auch s für alle $x \in \mathbf{R}$ stetig ist. Da c insbesondere an der Stelle 0 stetig ist, folgt aus (12.2.2), daß s ebenfalls an der Stelle 0 stetig ist. Damit erhält man für eine beliebige Zahl x_0 aus (12.1.15) unter Beachtung von (12.1.11) $\wedge_{x\in\mathbf{R}} |c(x)| \leqq 1$, daß

$$\begin{aligned}\lim_{x\to x_0} |s(x_0) - s(x)| &= 2 \lim_{x\to x_0} \left(\left|c\left(\frac{x_0 + x}{2}\right)\right| \cdot \left|s\left(\frac{x_0 - x}{2}\right)\right|\right) \\ &\leqq 2 \lim_{x\to x_0} \left|s\left(\frac{x_0 - x}{2}\right)\right| = 0,\end{aligned}$$

also $\lim_{x\to x_0} s(x) = s(x_0)$. Damit ist s für alle $x_0 \in \mathbf{R}$ stetig. Aus der Stetigkeit von s folgt nun, daß es ein Intervall $]\!]0, a_6[\![$ $(a_6 \in \mathbf{R}_+^*)$ gibt mit $s(x) \neq 0$ für alle $x \in]\!]0, a_6[\![$. Zum Beweis dieser Behauptung stellen wir zunächst fest, daß die Nullstellen von s bezüglich der Addition eine Untergruppe von $(\mathbf{R}, +)$ bilden. Sind nämlich e und f reelle Zahlen mit $s(e) = 0 = s(f)$, so erhält man aus (12.1.1) $s(e - f) = s(e)\, c(f) - s(f)\, c(e) = 0$. Die Annahme, daß es keine Zahl a_6 mit den angegebenen Eigenschaften gibt, hat die Existenz beliebig kleiner Nullstellen zur Folge. Wenn also e und f mit $e < f$ Nullstellen von s sind, dann ist auch $f - e > 0$ eine Nullstelle, und es gibt eine Nullstelle x_0 mit $0 < x_0 < f - e$. Durch Addition von e ergibt sich aus dieser Ungleichung $e < x_0 + e < f$. Auch $x_0 + e$ ist eine Nullstelle, denn die Nullstellen von s bilden ja bezüglich der Addition eine Gruppe, also liegen die Nullstellen von s sogar dicht. Wegen der Stetigkeit von s ist dann aber jede relle Zahl Nullstelle. Wir hatten aber die konstanten Lösungen ausgeschlossen. Also gibt es das behauptete Intervall.

Für die Zahlen $x, h \in \mathbf{R}$ und $k \in \mathbf{N}^*$ mit $x = k \cdot h$ und $0 < k \cdot h < a_6$ gilt (mit (12.1.15))

$$\begin{aligned}s\left(\frac{h}{2}\right) \cdot \sum_{v=1}^{k} c(v \cdot h) &= \sum_{v=1}^{k} c\left(\frac{(2v-1)\,h}{2} + \frac{h}{2}\right) s\left(\frac{h}{2}\right) \\ &= \frac{1}{2} \sum_{v=1}^{k} \left(s\left(\frac{(2v+1)\,h}{2}\right) - s\left(\frac{(2v-1)\,h}{2}\right)\right) \\ &= \frac{1}{2}\left(s\left(\frac{(2k+1)\,h}{2}\right) - s\left(\frac{h}{2}\right)\right) = \frac{1}{2}\left(s\left(x + \frac{h}{2}\right) - s\left(\frac{h}{2}\right)\right).\end{aligned}$$

Daraus folgt $\left(s\left(\frac{h}{2}\right) \neq 0(!)\right)$

$$\sum_{\nu=1}^{k} c(\nu \cdot h)\, h = \left(s\left(x + \frac{h}{2}\right) - s\left(\frac{h}{2}\right)\right) \frac{\frac{h}{2}}{s\left(\frac{h}{2}\right)}.$$

Für $h \to 0+$ erhält man auf der linken Seite eine Riemannsche Summenfolge, die wegen der Stetigkeit von c konvergiert. Dieser Grenzwert ist von Null verschieden, da $c(x) > 0$ für hinreichend kleine x gilt (das folgt aus der Stetigkeit von c und $c(0) = 1$). Da weiterhin auch s stetig ist und $s(0) = 0$ gilt, erhält man

$$\int_0^x c(t)\, dt = s(x) \cdot \lim_{h \to 0+} \frac{\frac{h}{2}}{s\left(\frac{h}{2}\right)}.$$

Da auch $s(x) \neq 0$ in $]0, a_0[$ ist, existiert der Grenzwert $\lim_{h \to 0+} \frac{\frac{h}{2}}{s\left(\frac{h}{2}\right)}$, und dieser Grenzwert ist von Null verschieden. Die Funktion s ist also an der Stelle 0 rechtsseitig differenzierbar. Wegen der Ungeradheit von s ist diese Funktion in 0 sogar differenzierbar.

Es gibt in der Literatur verschiedene Beweise dafür, daß aus der Stetigkeit der Lösungen von ((12.1.1), (12.1.2)) deren Differenzierbarkeit folgt. Wir verweisen in diesem Zusammenhang noch auf L. Vietoris [177] (siehe auch J. Aczél [3], [5], H. Lenz [105], D. Ilse [76]), wo das Problem auf die Cauchysche Funktionalgleichung $\bigwedge_{x,y \in \mathbb{R}} f(x + y) = f(x) + f(y)$ zurückgeführt wird. Man kann z. B. auch durch Addition von (12.1.2) und (12.1.4) zur *d'Alembertschen Funktionalgleichung* $\bigwedge_{x,y \in \mathbb{R}} c(x + y) + c(x - y) = 2c(x)\, c(y)$ übergehen und dann aus der Tatsache, daß c als stetige Funktion eine Stammfunktion besitzt, die Differenzierbarkeit von c beweisen (vgl. J. Aczél [3]). Daraus folgt unter Benutzung von (12.1.14) die Differenzierbarkeit von s an der Stelle 0.

(11) $\Rightarrow$ (12). Es sei s an der Stelle 0 differenzierbar und x_0 eine beliebige reelle Zahl. Dann erhält man aus (12.1.15)

$$\lim_{h \to 0} \frac{s(x_0 + h) - s(x_0)}{h} = \lim_{h \to 0} \frac{2s\left(\frac{h}{2}\right) c\left(x_0 + \frac{h}{2}\right)}{h} = \lim_{h \to 0} \frac{s\left(\frac{h}{2}\right)}{\frac{h}{2}} \cdot \lim_{h \to 0} c\left(x_0 + \frac{h}{2}\right).$$

Da die Funktion s an der Stelle 0 differenzierbar ist, ist sie dort auch stetig. Nun haben wir im Beweisschritt (8) $\Rightarrow$ (9) gezeigt, daß (sogar) aus der rechtsseitigen

Stetigkeit von s an einer beliebigen Stelle die Stetigkeit von c an der Stelle 0 folgt. Wie wir ebenfalls schon bewiesen haben ((9) $\Rightarrow$ (10)), ist dann c an jeder Stelle stetig. Da außerdem $\lim\limits_{h\to 0} \dfrac{s\left(\frac{h}{2}\right)}{\frac{h}{2}}$ nach Voraussetzung existiert, folgt aus der letzten Gleichung die Differenzierbarkeit von s an der Stelle x_0 und damit an jeder Stelle.

(12) $\Rightarrow$ (13). Die Funktion s sei für alle $x \in \mathbf{R}$ differenzierbar und x_0 eine beliebige reelle Zahl. Aus (12.1.14) folgt sofort

$$\lim_{h\to 0} \frac{c(x_0+h)-c(x_0)}{h} = -\lim_{h\to 0} \frac{s\left(\frac{h}{2}\right)}{\frac{h}{2}} \cdot \lim_{h\to 0} s\left(x_0+\frac{h}{2}\right).$$

Da s überall differenzierbar ist, existieren beide Grenzwerte auf der rechten Seite. Das bedeutet, daß c in x_0 und damit für jedes $x \in \mathbf{R}$ differenzierbar ist.

(13) $\Rightarrow$ (1). Zum Beweis dieses letzten Schrittes seien nun s und c Funktionen, die für alle $x \in \mathbf{R}$ differenzierbar sind. Dann ist zu zeigen, daß es ein Intervall $[\![0, a_1[\![$ ($a_1 \in \mathbf{R}_+^*$) gibt, in dem c streng monoton fallend ist. Im Beweisschritt (10) $\Rightarrow$ (11) haben wir aus der Stetigkeit von s und c abgeleitet, daß s nicht beliebig kleine Nullstellen haben kann. Also gibt es ein Intervall $]\!]0, a_1[\![$ ($a_1 \in \mathbf{R}_+^*$), in dem s keine Nullstellen hat. Aus der Stetigkeit von s folgt daraus weiter, daß entweder $s(x) > 0$ oder $s(x) < 0$ jeweils für alle $x \in]\!]0, a_1[\![$ gilt. Es seien nun mit $e, f \in]\!]0, a_1[\![$ und $0 < e < f < a_1$ zwei Zahlen gegeben. Dann sind auch die Zahlen $\dfrac{f-e}{2}$ und $\dfrac{f+e}{2}$ aus dem Intervall $]\!]0, a_1[\![$, und man erhält aus (12.1.14) $c(f) - c(e) = -2s\left(\dfrac{f+e}{2}\right) s\left(\dfrac{f-e}{2}\right) < 0$, so daß also c in $]\!]0, a_1[\![$ und damit auch in $[\![0, a_1[\![$ streng monoton fallend ist. Damit ist Satz 12.2.1. vollständig bewiesen.

12.2.4. Aufgabe. Man zeige, daß für alle Lösungen $s: \mathbf{R} \to \mathbf{R}$ und $c: \mathbf{R} \to \mathbf{R}$ von ((12.1.1), (12.1.2)) gilt: Es gibt ein Intervall $[\![0, b[\![$ ($b \in \mathbf{R}_+^*$) mit $c(x) > 0$ für alle $x \in [\![0, b[\![$ genau dann, wenn $\operatorname{sgn} s(x) = \operatorname{sgn} s(2x)$ für $x \in [\![0, b[\![$ gilt.

12.3. Die Funktionalgleichungen $f(x+y) + f(x-y) = 2f(x)f(y)$ und $f(x+y)f(x-y) = f(x)^2 - f(y)^2$

In den Additions- und Subtraktionstheoremen der Funktionen sin und cos treten die beiden Funktionen jeweils gemeinsam auf. Das führte in 12.1. zu funktionalen Charakterisierungen dieser Funktionen durch Systeme von zwei Funktionalgleichungen, wobei in jeder dieser Gleichungen zwei unbekannte Funktionen vorkamen. Wie wir in den Kapiteln 7 und 8 gesehen haben, können die linearen Funktionen, die Exponential- und Logarithmusfunktionen jeweils mit Hilfe *einer* Funktionalglei-

chung mit *einer* unbekannten Funktion charakterisiert werden. Wir wollen in diesem Kapitel zeigen, daß eine solche Charakterisierung sowohl für die Sinusfunktionen als auch für die Kosinusfunktionen ebenfalls möglich ist.

Derartige Gleichungen findet man folgendermaßen: Addiert man die Gleichungen (12.1.2) und (12.1.4), so erhält man

12.3.1. $\bigwedge_{x,y\in\mathbf{R}} c(x+y)+c(x-y)=2c(x)\,c(y)$,[1])

also eine Gleichung, in der nur eine Funktion auftritt. Multipliziert man die Gleichungen (12.1.1) und (12.1.3), so erhält man bei Benutzung von (12.1.10)

12.3.2. $\bigwedge_{x,y\in\mathbf{R}} s(x+y)\,s(x-y)=s(x)^2-s(y)^2$,

also wieder eine Gleichung, in der nur eine Funktion auftritt.

Im folgenden werden wir alle stetigen Lösungen der Gleichungen (12.3.1) bzw. (12.3.2) bestimmen. Daraus werden sich dann unmittelbar die angestrebten Charakterisierungen ergeben. Auf mögliche Abschwächungen der Stetigkeitsforderung wollen wir hier nicht mehr eingehen. Wie in 12.2. überzeugt man sich leicht davon, daß die Gleichungen (12.3.1) bzw. (12.3.2) auch unstetige Lösungen besitzen. Wenn $f\colon \mathbf{R}\to\mathbf{R}$ eine unstetige Lösung der Funktionalgleichung $\bigwedge_{x,y\in\mathbf{R}} f(x+y)=f(x)+f(y)$ ist, dann sind die Funktionen $c\colon \mathbf{R}\to\mathbf{R}$ und $s\colon \mathbf{R}\to\mathbf{R}$ mit $c(x)=\cos\big(f(x)\big)$ bzw. $s(x)=\sin\big(f(x)\big)$ unstetige Lösungen von (12.3.1) bzw. (12.3.2).

Wir wenden uns zunächst der d'Alembertschen Funktionalgleichung zu. Einfache Rechnungen zeigen, daß neben den Funktionen $c\colon \mathbf{R}\to\mathbf{R}$ mit $c(x)=\cos ax$ $(a\in\mathbf{R})$ auch die Funktionen $c\colon \mathbf{R}\to\mathbf{R}$ mit $c(x)=\cosh ax:=\frac{1}{2}(e^{ax}+e^{-ax})$ $(a\in\mathbf{R})$ Lösungen von (12.3.1) sind. Daß dies zusammen mit der Lösung c, für die $c(x)=0$ für alle $x\in\mathbf{R}$, schon alle stetigen Lösungen von (12.3.1) sind, ist von zahlreichen Autoren mit unterschiedlichen Methoden (z. B. durch Rückführung auf die Subtraktionstheoreme, Beweis der Differenzierbarkeit aus der Stetigkeit und Rückführung auf die Schwingungsgleichung $y''=ay$ oder Entwicklung in Potenzreihen) bewiesen worden.[2])

Wir betrachten also stetige Funktionen $f\colon \mathbf{R}\to\mathbf{R}$ mit (12.3.1) $f(x+y)+f(x-y)=2f(x)\,f(y)$ für alle $x,y\in\mathbf{R}$.

12.3.3. Alle konstanten Lösungen von (12.3.1) sind die Funktionen $f\colon \mathbf{R}\to\mathbf{R}$ mit $f(x)=0$ und $f(x)=1$.

Setzt man nämlich $f(x)=k$ $(k\in\mathbf{R})$, so folgt aus (12.3.1) unmittelbar $2k=2k^2$,

[1]) Diese Gleichung wurde von J. D'ALEMBERT (1717–1783) bereits im Jahre 1750 gelöst und wird *d'Alembertsche Funktionalgleichung* genannt.

[2]) Man vergleiche etwa J. D'ALEMBERT [13], JU. M. GAJDUK [59], A. L. CAUCHY [34], P. HOPF [69], E. PICARD [133], J. ACZÉL [3], [5].

also $k = 0$ oder $k = 1$. Die triviale Lösung mit $f(x) = 0$ für alle $x \in \mathbf{R}$ wird im folgenden außer acht gelassen. Dann gilt

12.3.4. $f(0) = 1$.

Für $x = y = 0$ erhält man aus (12.3.1) $f(0) = f(0)^2$, also $f(0) = 0$ oder $f(0) = 1$. Aus $f(0) = 0$ folgt aber $f(x) = 0$ für alle $x \in \mathbf{R}$. Das erkennt man sofort, wenn man $y = 0$ in (12.3.1) setzt.

12.3.5. Jede Lösung von (12.3.1) ist gerade.

Das folgt für $x = 0$ aus (12.3.1) unter Benutzung von (12.3.4).

Wir bestimmen jetzt im folgenden Satz 12.3.6 alle stetigen Lösungen von (12.3.1), die Nullstellen besitzen, und dann in Satz 12.3.13 alle stetigen Lösungen von (12.3.1), die keine Nullstellen besitzen.

12.3.6. Satz. *Alle stetigen nichtkonstanten Lösungen* $f\colon \mathbf{R} \to \mathbf{R}$ *der d'Alembertschen Funktionalgleichung, die (mindestens) eine Nullstelle besitzen, sind die Funktionen* f *mit* $f(x) = \cos ax$ $(a \in \mathbf{R}^*)$.

Beweis. Einen relativ kurzen Beweis, dessen Grundgedanke von A. L. Cauchy in [34] entwickelt wurde, findet man in J. Aczél [3]. Wir folgen hier einem Beweisgedanken von Ju. M. Gajduk in [59], bei dem auch solche bekannten Eigenschaften der Kosinusfunktionen wie die Quadrantenbeziehungen und die Periodizität aus (12.3.1) abgeleitet werden. Das ermöglicht einen Vergleich mit den entsprechenden Ableitungen aus den Subtraktionstheoremen in 12.1.[1])

Wegen (12.3.4) und (12.3.5) können wir voraussetzen, daß f eine positive Nullstelle besitzt. Zusammen mit der Stetigkeit folgt daraus, daß auch eine kleinste positive Nullstelle existiert (anderenfalls müßte ja wegen der Stetigkeit $f(0) = 0$ gelten). Sie werde mit $\frac{w}{2}$ bezeichnet. Dann gilt

12.3.7. $\bigwedge_{x \in \mathbf{R}} f(x + w) = -f(x)$.

Dazu ersetzen wir in (12.3.1) x durch $\frac{w}{2} + x$ und y durch $\frac{w}{2}$ und erhalten

$$\bigwedge_{x \in \mathbf{R}} f(w + x) + f(x) = 2f\left(\frac{w}{2} + x\right) f\left(\frac{w}{2}\right) = 0,$$

also $\bigwedge_{x \in \mathbf{R}} f(w + x) = -f(x)$. Insbesondere ist $f(w) = -f(0) = -1$.

12.3.8. Jede Lösung von (12.3.1) mit der kleinsten positiven Nullstelle $\frac{w}{2}$ ist periodisch und besitzt die primitive Periode $2w$.

[1]) Wir ziehen diese ausführlicheren Betrachtungen möglichen knapperen Beweisführungen vor, um die gegenseitige Ableitbarkeit von Eigenschaften der elementaren Funktionen immer wieder und in verschiedenen Zusammenhängen studieren zu können.

Zunächst folgt aus (12.3.7), wenn man $x + w$ für x setzt:

$$\bigwedge_{x \in \mathbf{R}} f(x + 2w) = -f(x + w) = f(x).$$

Insbesondere ist $f(2w) = 1$. Die Zahl $2w$ ist auch die kleinste positive Periode von f. Gäbe es nämlich eine kleinere Periode $2c$ $(0 < 2c < 2w)$, so müßte $f(2c) = f(2c + 2w) = f(2w) = 1$ gelten. Das würde für $x = y = c$ aus (12.3.1) $f(2c) + f(0) = 2f(c)^2$ liefern, so daß also $1 = f(c)^2$ wäre.

Fall 1: $f(c) = 1$. In diesem Fall folgt zunächst aus (12.3.5) und (12.3.7)

12.3.9. $$f\left(\frac{c+w}{2}\right) = f\left(-\frac{c+w}{2}\right) = -f\left(w - \frac{c+w}{2}\right) = -f\left(\frac{w-c}{2}\right).$$

Aus (12.3.1) erhält man für $x = \dfrac{w+c}{2}$ und $y = \dfrac{w-c}{2}$

$$f(w) + f(c) = 2f\left(\frac{w+c}{2}\right) f\left(\frac{w-c}{2}\right).$$

Wegen $f(w) = -1$ und (12.3.9) ist damit $f\left(\dfrac{w-c}{2}\right) = 0$ im Widerspruch dazu, daß $\dfrac{w}{2}$ kleinste positive Nullstelle ist.

Fall 2: $f(c) = -1$. Das liefert für $x = y = \dfrac{c}{2}$ aus (12.3.1) $f(c) + f(0) = 2f\left(\dfrac{c}{2}\right)^2$, also $f\left(\dfrac{c}{2}\right) = 0$. Es ist aber $0 < \dfrac{c}{2} < \dfrac{w}{2}$, also ebenfalls ein Widerspruch.

12.3.10. Es gilt $\bigwedge_{x \in \mathbf{R}} |f(x)| \leqq 1$.

Setzt man in (12.3.1) zunächst $x = y$ und dann $x = y = \dfrac{w}{2} - x$, so erhält man für alle $x \in \mathbf{R}$ die Beziehungen $f(2x) = 2f(x)^2 - 1$ und unter Ausnutzung von (12.3.7) und (12.3.5) $f(2x) = 1 - 2f\left(\dfrac{w}{2} - x\right)^2$, also $f\left(\dfrac{w}{2} - x\right)^2 = 1 - f(x)^2$, also $|f(x)| \leqq 1$.

12.3.11. Die Funktionalgleichung (12.3.1) besitzt höchstens eine stetige Lösung mit der kleinsten positiven Nullstelle $\dfrac{w}{2}$.

Es seien $f_1 : \mathbf{R} \to \mathbf{R}$ und $f_2 : \mathbf{R} \to \mathbf{R}$ zwei derartige Lösungen von (12.3.1). Dann gibt es ein $b \in \mathbf{R}^*$ mit $f_1(b) = f_2(b)$ (etwa $b = \dfrac{w}{2}$ oder $b = w$). Daraus erhält man durch vollständige Induktion $f_1(kb) = f_2(kb)$ für alle $k \in \mathbf{N}$. Gilt nämlich $f_1(mb) = f_2(mb)$ für alle $m \in \mathbf{N}$ mit $0 < m \leqq l$ für eine beliebige, aber fest gewählte Zahl $l \in \mathbf{N}$, dann gilt auch $f_1\big((l+1)\,b\big) = f_2\big((l+1)\,b\big)$ (ordnungstheoretische Induktion, vgl. MfL

Bd. 1, 3.4.). Zum Beweis der Induktionsbehauptung setzen wir in (12.3.1) $x = lb$ und $y = b$:

$$f_1\big((l+1)\,b\big) + f_1\big((l-1)\,b\big) = 2f_1(lb)\,f_1(b) = 2f_2(lb)\,f_2(b)$$
$$= f_2\big((l+1)\,b\big) + f_2\big((l-1)\,b\big).$$

Durch nochmalige Benutzung der Induktionsvoraussetzung folgt daraus

$$f_1\big((l+1)\,b\big) = f_2\big((l+1)\,b\big).$$

Die Funktionen f_1 und f_2 haben auch für alle Argumente $\frac{b}{2^m}$ $(m \in \mathbf{N})$ gleiche Funktionswerte. Auch das beweisen wir durch vollständige Induktion. Dazu wird zunächst $f_1\left(\frac{b}{2}\right) = f_2\left(\frac{b}{2}\right)$ als Induktionsanfang benötigt. Für $x = y = \frac{b}{2}$ folgt aus (12.3.1) $f_1(b) + f_1(0) = 2f_1\left(\frac{b}{2}\right)^2$ und $f_2(b) + f_2(0) = 2f_2\left(\frac{b}{2}\right)^2$, also

12.3.12. $f_1\left(\frac{b}{2}\right)^2 = f_2\left(\frac{b}{2}\right)^2$.

Nun können wir wegen (12.3.8) $\bigwedge_{x\in\mathbf{R}} f(x+2w) = f(x)$ und (12.3.7) $\bigwedge_{x\in\mathbf{R}} f(x+w) = -f(x)$ voraussetzen, daß $0 < b \leqq w$ ist. Dann gilt aber für $\frac{b}{2}$ sogar $0 < \frac{b}{2} \leqq \frac{w}{2}$. Da $\frac{w}{2}$ kleinste positive Nullstelle, $f(0) = 1$ und f stetig ist, muß jede Lösung f von (12.3.1) zwischen 0 und $\frac{w}{2}$ nicht negativ sein. Damit folgt aus (12.3.12) $f_1\left(\frac{b}{2}\right) = f_2\left(\frac{b}{2}\right)$.

Jetzt sei $m \in \mathbf{N}$ beliebig, und es gelte $f_1\left(\frac{b}{2^m}\right) = f_2\left(\frac{b}{2^m}\right)$. Dann gilt auch $f_1\left(\frac{b}{2^{m+1}}\right) = f_2\left(\frac{b}{2^{m+1}}\right)$. Für $x = y = \frac{b}{2^{m+1}}$ folgt nämlich aus (12.3.1)

$$2f_1\left(\frac{b}{2^{m+1}}\right)^2 = f_1\left(\frac{b}{2^m}\right) + 1 \quad \text{und} \quad 2f_2\left(\frac{b}{2^{m+1}}\right)^2 = f_2\left(\frac{b}{2^m}\right) + 1$$

und daraus

$$f_1\left(\frac{b}{2^{m+1}}\right) = f_2\left(\frac{b}{2^{m+1}}\right)$$

wegen $0 < \frac{b}{2^{m+1}} \leqq \frac{b}{2}$ und wegen der Positivität von f_1 und f_2 zwischen 0 und $\frac{b}{2}$.

Insgesamt ist damit $f_1\left(\frac{k}{2^m}\,b\right) = f_2\left(\frac{k}{2^m}\,b\right)$ für alle $k, m \in \mathbf{N}$.

Ist x eine beliebige positive reelle Zahl, dann gibt es eine Folge von Dualbrüchen $\left(\frac{n(m)}{2^m}\right)$ mit $\lim\limits_{m\to\infty} \frac{n(m)}{2^m} = \frac{x}{b}$. Damit folgt aus der Stetigkeit von f_1 und f_2

$$\lim_{m\to\infty} f_1\left(\frac{n(m)}{2^m}\,b\right) = f_1\left(\frac{x}{b}\,b\right) = f_1(x) \quad \text{und} \quad \lim_{m\to\infty} f_2\left(\frac{n(m)}{2^m}\,b\right) = f_2\left(\frac{x}{b}\,b\right) = f_2(x),$$

also $f_1(x) = f_2(x)$.

Wegen der Geradheit aller Lösungen von (12.3.1) stimmen dann f_1 und f_2 für jedes $x \in \mathbf{R}$ überein.

Da also die d'Alembertsche Funktionalgleichung höchstens eine stetige Lösung mit vorgegebener kleinster positiver Nullstelle $\frac{w}{2}$ besitzt, andererseits aber $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = \cos\left(\frac{\pi}{w}x\right)$ eine solche Funktion ist, ist dies die einzige derartige Funktion.[1]) Sie ist durch die genannten Eigenschaften funktional charakterisiert, und die Funktionen $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = \cos ax$ ($a \in \mathbf{R}^*$) sind, wie behauptet, alle nichtkonstanten stetigen Lösungen von (12.3.1), die Nullstellen besitzen.

12.3.13. Satz. *Alle stetigen Lösungen der d'Alembertschen Funktionalgleichung, die keine Nullstellen besitzen, sind die Funktionen $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = \cosh ax$ ($a \in \mathbf{R}$)*[2]).

Beweis. Die Eigenschaften (12.3.4) und (12.3.5) gelten für alle Lösungen von (12.3.1), können also auch für den Beweis benutzt werden. Wir zeigen zunächst: Wenn f eine nichtkonstante stetige Lösung von (12.3.1) ist, die keine Nullstellen besitzt, dann ist $f(x) > 1$ für alle $x \in \mathbf{R}^*$.

Wir stellen fest, daß es keine Zahl x_0 geben kann mit $f(x_0) < 0$. Sonst hätte nämlich f als stetige Funktion doch Nullstellen, da es ja Zahlen x mit $f(x) > 0$ (z. B. $f(0) = 1$) gibt. Es gibt darüber hinaus aber auch keine reelle Zahl $b \neq 0$ mit $0 < f(b) \leqq 1$. Anderenfalls ließe sich nämlich eine reelle Zahl w so finden, daß $f(b) = \cos\left(\frac{\pi}{w}b\right)$ $\left(0 < b < \frac{w}{2}\right)$ ist, und f würde nach dem Beweis von Satz 12.3.6 mit dieser Kosinusfunktion f für jedes $x \in \mathbf{R}$ übereinstimmen, hätte also doch wieder Nullstellen.

Lassen wir die Lösung $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = 1$ beiseite, so können wir also insbesondere $f(1) > 1$ voraussetzen. Man zeigt nun wie im Beweis zu Satz 12.3.6, daß Lösungen f_1 und f_2, die an einer Stelle übereinstimmen, an jeder Stelle gleiche Werte haben. Da f hier immer positiv ist, entfallen die dort notwendigen Fallunterscheidungen.

Ist $f(1) = c > 1$ gegeben, so gibt es genau eine Zahl $a \in \mathbf{R}$ mit $\frac{1}{2}(e^a + e^{-a}) = c$. Also ist f identisch mit der Funktion $g\colon \mathbf{R} \to \mathbf{R}$, für die $g(x) = \cosh ax$ gilt. Damit ist der Beweis von Satz 12.3.13 beendet.

Die d'Alembertsche Funktionalgleichung hat damit genau die folgenden stetigen Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ mit

(1) $f(x) = 0$,

(2) $f(x) = \cos ax$ ($a \in \mathbf{R}$),

(3) $f(x) = \cosh ax$ ($a \in \mathbf{R}$).

Dabei ist in (2) und (3) für $a = 0$ die Lösung f mit $f(x) = 1$ enthalten.

[1]) Für $w = \pi$ erhält man also $f(x) = \cos x$.

[2]) Für $a = 0$ ist hierin die konstante Lösung $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = 1$ enthalten.

Bevor wir im folgenden alle stetigen Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ der Funktionalgleichung (12.3.2)

$$\bigwedge_{x,y\in\mathbf{R}} f(x+y)\,f(x-y) = f(x)^2 - f(y)^2$$

bestimmen, bemerken wir zunächst, daß die Sinusfunktionen[1]), die hyperbolischen Sinusfunktionen $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = \sinh ax := \frac{1}{2}(e^{ax} - e^{-ax})$ $(a \in \mathbf{R})$ und die linearen Funktionen $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = ax$ $(a \in \mathbf{R})$ Lösungen dieser Gleichung sind, was man durch Einsetzen bestätigt.

Im folgenden Satz 12.3.14 werden wir beweisen, daß das im wesentlichen auch alle stetigen Lösungen sind. In dem Beweis werden die Lösungen von (12.3.2) auf die Lösungen der Differentialgleichung $y'' = c \cdot y$ $(c \in \mathbf{R})$ (vgl. auch MfL Bd. 5, 5.3.) zurückgeführt.[2])

12.3.14. Satz. *Alle nichtkonstanten stetigen Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ der Funktionalgleichung* (12.3.2) *sind die Funktionen f mit $f(x) = ax$, $f(x) = b \sin ax$ und $f(x) = b \sinh ax$ $(a, b \in \mathbf{R}^*)$.*

Beweis. Die Gleichung (12.3.2) hat als konstante Lösung nur $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = 0$. Für $f(x) = k$ folgt nämlich aus (12.3.2) $k^2 = k^2 - k^2 = 0$, also $k = 0$. Diese Lösung ist in den anderen Lösungen von (12.3.2) für $a = 0$ enthalten.

Für alle Lösungen f von (12.3.2) gilt

12.3.15. $f(0) = 0$ $(x = y = 0$ in (12.3.2)),

12.3.16. $\bigwedge_{x\in\mathbf{R}} f(-x) = -f(x)$.

Aus (12.3.2) ergibt sich für $x = 0$ wegen (12.3.15) $\bigwedge_{y\in\mathbf{R}} f(y)\,f(-y) = -f(y)^2$, woraus sich die Behauptung für $f(y) \neq 0$ ergibt. Es sei nun $f(y) = 0$. Wir setzen $x = -y$ in (12.3.2) und erhalten $\bigwedge_{y\in\mathbf{R}} f(0)\,f(-2y) = f(-y)^2 - f(y)^2$, woraus sich $f(-y) = 0$ ergibt, so daß auch in diesem Fall die Behauptung gilt.

Wegen (12.3.15) und (12.3.16) genügt es zunächst, die Lösungen von (12.3.2) für $x \in \mathbf{R}_-{}^*$ zu kennen. Hat nun f keine Nullstellen $w \neq 0$, so ist $f(x) \neq 0$ für alle $x \in \mathbf{R}_+{}^*$. Hat f Nullstellen $w \neq 0$, so ist entweder $f(x) = 0$ für alle x, oder f hat eine kleinste positive Nullstelle. Da nämlich mit x und y auch $x - y$ Nullstelle ist (das folgt aus (12.3.2), wenn man x durch $x - y$ ersetzt), bilden die Nullstellen bezüglich der Addition eine Untergruppe von $(\mathbf{R}, +)$. Damit lägen die Nullstellen von f dicht, wenn f keine kleinste Nullstelle hätte, woraus wieder mit der Stetigkeit von f sogar $f(x) = 0$ für alle x folgen würde (vgl. 12.2., Beweisschritt (10) $\Rightarrow$ (11)).

---

[1]) Obwohl weder $\sin(x+y) = \sin x + \sin y$ noch $\sin(x-y) = \sin x - \sin y$ für alle $x, y, \in \mathbf{R}$ gilt, ergibt sich durch Multiplikation dieser beiden Gleichungen die Gleichung $\sin(x+y)\sin(x-y) = \sin^2 x - \sin^2 y$, die für alle $x, y \in \mathbf{R}$ gilt.

[2]) Man vergleiche auch R. A. Rosenbaum und S. L. Segal [148], W. H. Wilson [181] und J. Aczél [3]. In [3] z. B. wird die Lösung von (12.3.2) auf die Lösung der d'Alembertschen Funktionalgleichung zurückgeführt. Eine Charakterisierung der Sinusfunktion durch (12.3.2) ohne Benutzung von Differenzierbarkeitseigenschaften wird in S. Thiel [169] durchgeführt.

Nun zeigt sich, daß eine nichtkonstante stetige Lösung f mit der kleinsten positiven Nullstelle w_0 für alle reellen Zahlen festgelegt ist, sobald sie auf dem Intervall $[\![0, w_0]\!]$ bekannt ist.

12.3.17. Für alle Lösungen f von (12.3.2) und $w \in \mathbf{R}^*$ gilt $\bigwedge_{x\in\mathbf{R}} f(x+2w) = f(x)$ genau dann, wenn $f(w) = 0$ ist (R. A. Rosenbaum und S. L. Segal [148]).

a) Es sei $x \in \mathbf{R}$ beliebig und $w \in \mathbf{R}^*$ mit $f(w) = 0$. Wir setzen in (12.3.2) $y = w$, und ersetzen dann wir x durch $x + w$ und y durch x. Das ergibt $f(x+w)f(x-w) = f(x)^2 - f(w)^2 = f(x)^2$ und $f(2x+w)f(w) = 0 = f(x+w)^2 - f(x)^2$, also

12.3.18. $\bigwedge_{x\in\mathbf{R}} f(x+w)^2 = f(x)^2$.

Damit gilt auch $\bigwedge_{x\in\mathbf{R}} f(x+w)f(x-w) = f(x+w)^2$. Für $f(x+w) \neq 0$ ergibt sich daraus $f(x+w) = f(x-w)$. Das bedeutet $f(x+2w) = f(x)$. Für $f(x+w) = 0$ erhält man aus (12.3.18) $f(x) = 0$, also $f(x+w) = f(x)$, d. h. $f(x+2w) = f(x+w) = f(x)$.

b) Es sei $w \in \mathbf{R}^*$ eine Zahl, so daß $\bigwedge_{x\in\mathbf{R}} f(x+2w) = f(x)$ gilt. Jetzt ersetzen wir in (12.3.2) x durch $x + w$ und y durch w und dann x durch $x + w$ und y durch x und erhalten

12.3.19. $\bigwedge_{x\in\mathbf{R}} f(x+2w)f(x) = f(x+w)^2 - f(w)^2$,

12.3.20. $\bigwedge_{x\in\mathbf{R}} f(2x+w)f(w) = f(x+w)^2 - f(x)^2$.

Unter Beachtung der Voraussetzung folgt aus (12.3.19) $\bigwedge_{x\in\mathbf{R}} f(w)^2 = f(x+w)^2 - f(x)^2$ und damit aus (12.3.20) $\bigwedge_{x\in\mathbf{R}} f(w)^2 = f(2x+w)f(w)$. Wäre nun $f(w) \neq 0$, so würde $\bigwedge_{x\in\mathbf{R}} f(w) = f(2x+w)$ gelten, und f wäre damit konstant. Also gilt $f(w) = 0$.

12.3.21. Es sei $w \in \mathbf{R}^*$ und $f(w) = 0$. Dann gilt: Wenn $f\left(\frac{w}{2}\right) = 0$, so $\bigwedge_{x\in\mathbf{R}} f(x+w) = f(x)$, und wenn $f\left(\frac{w}{2}\right) \neq 0$, so $\bigwedge_{x\in\mathbf{R}} f(x+w) = -f(x)$ (vgl. R. A. Rosenbaum und S. L. Segal [148]).

Es sei w eine Zahl mit den vorausgesetzten Eigenschaften. Dann ist nach (12.3.18) $\bigwedge_{x\in\mathbf{R}} f(x+w)^2 = f(x)^2$, also $\bigwedge_{x\in\mathbf{R}} \big(f(x+w) = f(x) \vee f(x+w) = -f(x)\big)$. Ist nun noch $f\left(\frac{w}{2}\right) = 0$, so folgt aus (12.3.17) die Behauptung. Ist $f\left(\frac{w}{2}\right) \neq 0$, so folgt ebenfalls aus (12.3.17), daß $f(x+w) \neq f(x)$ ist, also notwendigerweise $f(x+w) = -f(x)$.

Da wir konstante Lösungen ausgeschlossen haben, können wir also bei der Bestimmung der nichtkonstanten stetigen Lösungen von (12.3.2) voraussetzen, daß f entweder für $x \in \mathbf{R}_+$ oder für $x \in]\!]0, w_0[\!]$ das Vorzeichen nicht wechselt. In jedem Fall gibt es eine Zahl $b \in \mathbf{R}_+^*$, so daß $\int_0^b f(x)\,dx$ verschieden von Null ist.

12.3.22. Jede stetige Lösung $f\colon \mathbf{R} \to \mathbf{R}$ von (12.3.2) ist differenzierbar (vgl. S. KUREPA [94]).

Wir ersetzen in (12.3.2) x durch $\frac{x+y}{2}$ und y durch $\frac{x-y}{2}$ und erhalten

$$\bigwedge_{x,y\in\mathbf{R}} f(x)\,f(y) = f\left(\frac{x+y}{2}\right)^2 - f\left(\frac{x-y}{2}\right)^2.$$

Wir integrieren über y:

$$\textbf{12.3.23.}\quad \bigwedge_{x\in\mathbf{R}} f(x)\int_0^b f(y)\,dy = \int_0^b f\left(\frac{x+y}{2}\right)^2 dy - \int_0^b f\left(\frac{x-y}{2}\right)^2 dy,$$

wobei die Zahl b so gewählt wurde, daß $\int_0^b f(y)\,dy \neq 0$ ist.

Auf der rechten Seite von (12.3.23) setzen wir jetzt $v = \frac{x+y}{2}$ im ersten Integral, $v = \frac{x-y}{2}$ im zweiten Integral und erhalten

$$\textbf{12.3.24.}\quad \bigwedge_{x\in\mathbf{R}} f(x)\int_0^b f(y)\,dy = \int_{x/2}^{(x+b)/2} 2f(v)^2\,dv + \int_{x/2}^{(x-b)/2} 2f(v)^2\,dv.$$

Aufgrund des Hauptsatzes der Differential- und Integralrechnung für stetige Funktionen sind die Funktionen (in x) auf der rechten Seite von (12.3.24) (nach x) differenzierbar. Damit ist auch die Funktion auf der linken Seite von (12.3.24) und deshalb auch f selbst (nach x) differenzierbar.[1])

12.3.25. Jede nichtkonstante stetige Lösung $f\colon \mathbf{R} \to \mathbf{R}$ von (12.3.2) ist für alle $x \in \mathbf{R}$ mit $f(x) \neq 0$ zweimal differenzierbar und genügt der Differentialgleichung $f''(x) = c \cdot f(x)$ $(c \in \mathbf{R})$.

Zum Beweis benutzen wir Überlegungen von R. A. ROSENBAUM und S. L. SEGAL [148] und F. TISSERAND [172]. Wir differenzieren (12.3.2) nach x und setzen anschließend $x = y$. Das ergibt $f'(x) = \frac{f(2x)}{f(x)} \cdot \frac{f'(0)}{2}$, woraus gleich folgt, daß f zweimal differenzierbar ist. Es sei jetzt $f\colon \mathbf{R} \to \mathbf{R}$ eine zweimal differenzierbare Funktion, die der Gleichung (12.3.2) genügt. Wir differenzieren nach x:

$$f'(x+y)\,f(x-y) + f(x+y)\,f'(x-y) = 2f(x)\,f'(x)$$

und anschließend nach y:

$$f''(x+y)\,f(x-y) - f'(x+y)\,f'(x-y) + f'(x+y)\,f'(x-y) - f(x+y)\,f''(x-y) = 0.$$

[1]) Mit entsprechenden Überlegungen haben wir bereits in den Kapiteln 7 und 8 von der Stetigkeit auf die Differenzierbarkeit der Lösungen bestimmter Funktionalgleichungen geschlossen. Andererseits zeigte sich in Kapitel 9 und wird sich auch in Kapitel 13 zeigen, daß dies nicht für jede Funktionalgleichung möglich ist.

Für f gilt also

$$f''(x+y)\,f(x-y) = f(x+y)\,f''(x-y)$$

und damit

$$\frac{f''(x+y)}{f(x+y)} = \frac{f''(x-y)}{f(x-y)}.$$

Daraus folgt $\dfrac{f''(x)}{f(x)} = \text{const}$, also gilt $f''(x) = c \cdot f(x)$, und zwar für alle $x > 0$, falls f keine Nullstellen $w \neq 0$ hat, bzw. für alle $x \in]0, w_0[$, wobei w_0 die kleinste positive Nullstelle von f ist.

Zur Lösung dieser Differentialgleichung (vgl. MfL Bd. 5, 5.3.) unterscheiden wir folgende Fälle:

Fall 1: $c = 0$. Dann ist $f''(x) = 0$, also $f(x) = a \cdot x + d$. Daraus folgt wegen $f(0) = 0$ dann $f(x) = ax$.

Fall 2: $c \neq 0$. Der Ansatz $f(x) = e^{\lambda x}$ liefert $\lambda^2 = c$, also $\lambda_{1,2} = \pm\sqrt{c}$. Hier haben wir noch einmal zu unterscheiden $\alpha)$ $c > 0$ und $\beta)$ $c < 0$. Im Fall $\alpha)$ erhalten wir $f(x) = c_1 e^{\sqrt{c}x} + c_2 e^{-\sqrt{c}x}$. Wegen $f(0) = 0$ ist $c_1 + c_2 = 0$, also $c_2 = -c_1$, so daß $f(x) = c_1\left(e^{\sqrt{c}x} - e^{-\sqrt{c}x}\right) = 2c_1 \cdot \sinh \sqrt{c}\,x$ gilt. Durch Vertauschung von λ_1 und λ_2 ergibt sich $f(x) = 2c_1 \sinh\left(-\sqrt{c}\right)x$. Insgesamt ist also $f(x) = b \cdot \sinh ax$ $(a, b \in \mathbf{R}^*)$. Im Fall $\beta)$ ist $\pm\sqrt{c} = \pm i\sqrt{-c}$, und wir erhalten

$$f(x) = c_1 \operatorname{Re}\left(e^{i\sqrt{-c}x}\right) + c_2 \operatorname{Im}\left(e^{-i\sqrt{-c}x}\right) = c_1 \cos\sqrt{-c}\,x + c_2 \sin\sqrt{-c}\,x.$$

Diesmal folgt aus $f(0) = 0$, daß $c_1 = 0$ ist, so daß also $f(x) = c_2 \sin\sqrt{-c}\,x$ gilt. Es ergibt sich nach Vertauschung von λ_1 und λ_2

$$f(x) = b \cdot \sin ax \;(a, b \in \mathbf{R}^*).$$

Das bisher für x mit $f(x) \neq 0$ erzielte Ergebnis gilt auch für die Nullstellen von f, da f stetig sein sollte. Damit ist Satz 12.3.14 vollständig bewiesen.

12.4. Anwendungen der d'Alembertschen Funktionalgleichung in der Geometrie und in der Physik

Zur mathematischen Beschreibung der Statik des Massenpunktes benutzt man den Vektorraum der an einem Punkt angreifenden Kräfte (vgl. z. B. Physik, Lehrbuch für Klasse 9 [132] oder E. Grimsehl [62]). Dabei wird einer geeigneten Längeneinheit eine geeignete Krafteinheit umkehrbar eindeutig zugeordnet, so daß jeder in einem festen Punkt A angreifenden Kraft eindeutig ein von A ausgehender Pfeil (eine gerichtete Strecke) entspricht und umgekehrt jedem derartigen Pfeil eine solche Kraft. Die Maßzahl der Länge des Pfeils ist dann gleich der Maßzahl der zugeord-

neten Kraft und umgekehrt. Diese Zahl wird wie üblich Betrag der Kraft genannt.

Werden diese Betrachtungen in der euklidischen Geometrie durchgeführt, so wird die Resultierende zweier Kräfte, deren zugeordnete Pfeile nicht in einer Geraden liegen, mit Hilfe der „Parallelogrammregel" bestimmt: Man zeichnet die den gegebenen Kräften $\vec{K}_1$ und $\vec{K}_2$[1]) entsprechenden Pfeile, die wir ebenfalls mit $\vec{K}_1$ und $\vec{K}_2$ bezeichnen wollen, und konstruiert das durch diese Pfeile eindeutig bestimmte Parallelogramm („Parallelogramm der Kräfte"). Die Diagonale, die vom gemeinsamen Anfangspunkt A der Pfeile $\vec{K}_1$ und $\vec{K}_2$ zur gegenüberliegenden Ecke des Parallelogramms führt, ist der der Resultierenden $\vec{R}$ von $\vec{K}_1$ und $\vec{K}_2$ zugeordnete Pfeil. Definiert man noch in bekannter Weise die Addition von Kräften gleicher und entgegengesetzter Richtung (deren Pfeile in einer Geraden liegen) und die Vervielfachung von Kräften mit reellen Zahlen, so erhält man den Vektorraum der an einem Punkt angreifenden Kräfte bzw. den Vektorraum der von A ausgehenden Pfeile. Beide sind zueinander isomorph, so daß wir die Kräfte mit den ihnen zugeordneten Pfeilen identifizieren können. Hat man die zu $\vec{K}_1$ und $\vec{K}_2$ gehörende Resultierende $\vec{R}$ bestimmt, so befinden sich die Kräfte $-\vec{R}$, $\vec{K}_1$ und $\vec{K}_2$ im Gleichgewicht.[2])

Die folgenden Betrachtungen führen wir im Vektorraum der in einem Punkt angreifenden Kräfte, die alle in ein und derselben Ebene liegen, durch. Dabei legen wir die euklidische Geometrie der Ebene zugrunde. Alle benutzten Eigenschaften dieses Vektorraumes setzen wir hier als bekannt voraus, beweisen sie also nicht. Bei der Konstruktion der Resultierenden zweier Kräfte mit Hilfe der Parallelogrammregel spielt das euklidische Parallelenaxiom eine wesentliche Rolle. Wir wollen jetzt einige bekannte Eigenschaften der Addition von Kräften ((12.4.2) bis (12.4.6)) zusammenstellen und dann zeigen, daß die Zusammensetzung von Kräften mit Hilfe des Parallelogramms der Kräfte die einzige zweistellige Operation im zugrunde gelegten Vektorraum ist, die diese Eigenschaften besitzt. Wir erhalten damit also eine funktionale Charakterisierung der Kräfteaddition. Wir werden dabei erkennen, daß diese Charakterisierung ohne Benutzung eines Parallelenaxioms durchgeführt werden kann. Dadurch können alle dabei angestellten Betrachtungen in die absolute Geometrie und in die Lobačevskijsche Geometrie übernommen werden.[3]) Wir werden auch in diesen Geometrien durch (12.4.2) bis (12.4.6) jeweils den Vektorraum der an einem Punkt angreifenden Kräfte, die alle in ein und derselben Ebene liegen, definieren

---

[1]) Wir bezeichnen die Kräfte, also die Elemente des betrachteten Vektorraumes mit großen lateinischen Buchstaben und einem darüber gesetzten Pfeil (z. B. $\vec{K}$). Sind A bzw. B Anfangs- bzw. Endpunkt einer Kraft, so werden wir auch $\vec{AB}$ schreiben. Die Beträge dieser Kräfte bezeichnen wir mit den betreffenden Buchstaben ohne Pfeil, gelegentlich auch durch senkrechte Striche (z. B. K oder $|\vec{K}|$). Die Winkelmessung sei stets so gewählt, daß der rechte Winkel das Maß $\frac{\pi}{2}$ hat.

[2]) $-\vec{R}$ liegt mit $\vec{R}$ in derselben Geraden, hat die zu $\vec{R}$ entgegengesetzte Richtung und denselben Betrag wie $\vec{R}$.

[3]) Alle Folgerungen aus den Inzidenz-, Anordnungs- und Bewegungsaxiomen sowie dem Stetigkeitsaxiom ergeben die *absolute Geometrie*. Nimmt man zu dieser Geometrie das euklidische bzw. das Lobačevskijsche Parallelenaxiom hinzu, so ergibt sich die *euklidische* bzw. die *Lobačevskijsche Geometrie* (siehe MfL Bd. 17, Kap. 1. A. P. Norden [125]).

und damit die Statik des Massenpunktes beschreiben können. Es wird sich weiter zeigen, daß man die Statik des starren Körpers nicht mehr ohne Benutzung eines Parallelenaxioms beschreiben kann. Auf diese Weise erhalten wir schließlich einen Zugang zur Lobačevskijschen Geometrie.[1])

12.4.1. Satz. *Im Vektorraum der in einer euklidischen Ebene liegenden und in einem Punkt A angreifenden Kräfte gibt es genau eine zweistellige Operation +, die zwei Kräften $\vec{K}_1$ und $\vec{K}_2$ (den Komponenten) genau eine Kraft $\vec{R}$ (die Resultierende) zuordnet, mit folgenden Eigenschaften:*

12.4.2. *Der Betrag der Resultierenden von Kräften, die in einer Geraden liegen, ist gleich der Summe bzw. der Differenz der Beträge dieser Kräfte, wenn diese Kräfte gleiche bzw. entgegengesetzte Richtung haben. Die Resultierende liegt ebenfalls in dieser Geraden und hat im ersten Fall dieselbe Richtung wie die Komponenten. Haben die Komponenten im zweiten Fall verschiedene Beträge, so hat die Resultierende die Richtung der Komponente mit dem größeren Betrag. Haben sie gleiche Beträge, so ist die Resultierende $\vec{0}$ ($\vec{0}$ hat den Betrag 0 und keine bestimmte Richtung). Es ist $\vec{K} + \vec{0} = \vec{0} \dotplus \vec{K} = \vec{K}$ für alle Kräfte $\vec{K}$.*

12.4.3. *Liegen die Komponenten nicht in einer Geraden, so liegt die Resultierende im Inneren des von den Kräften gebildeten Winkels*[2]).

12.4.4. *Der Betrag R der Resultierenden zweier Kräfte von gleichem Betrag K ist eine stetige Funktion der Beträge dieser Kräfte und der Größe des von ihnen eingeschlossenen Winkels. Hat dieser Winkel die Größe 2α, so sei $R := F(K, \alpha)$. Der Definitionsbereich der Funktion F ist die Menge aller Paare (K, α) mit $K \in \mathbf{R}_+$, $\alpha \in \left[\!\left[0, \frac{\pi}{2} \right]\!\right]$. Es gelte $F \neq 0$ für $K \neq 0$ und $0 < \alpha < \frac{\pi}{2}$.*

12.4.5. *Die Operation + ist assoziativ und kommutativ.*

12.4.6. *Die Resultierende bildet mit den Komponenten ein gegen Spiegelung an einer Geraden durch A invariantes System.*[3])

Beweis. Wir zeigen: Wenn die Komponenten $\vec{K}_1$ und $\vec{K}_2$ von $\vec{0}$ verschieden sind und außerdem die Größe des von ihnen eingeschlossenen Winkels zwischen 0 und π liegt, so ist die durch die Parallelogrammregel definierte Operation + die einzige Operation mit den Eigenschaften (12.4.3) bis (12.4.6). Die Sonderfälle (12.4.2) lassen sich als Grenzfälle der Parallelogrammregel auffassen.

---

[1]) Wir folgen in diesem Abschnitt den entsprechenden Ausführungen in E. Picard [133], die sich auf S. D. Poisson [137] und J. Andrade [15] stützen. Bemerkungen zur historischen Entwicklung dieser Untersuchungen werden z. B. im Vorwort von S. Straszewicz [166] gemacht. Dort und in J. Aczél [3], [5] befinden sich auch weitere Literaturhinweise.

[2]) Die Größe dieses Winkels ist kleiner als π.

[3]) Gilt also $\vec{R} = \vec{K}_1 + \vec{K}_2$ und wird die aus diesen Kräften bestehende Figur an einer Geraden gespiegelt, die den Punkt A enthält, und sind $\vec{K}_1'$, $\vec{K}_2'$ und $\vec{R}'$ die Bilder von $\vec{K}_1$, $\vec{K}_2$ bzw. $\vec{R}$, so gilt $\vec{R}' = \vec{K}_1' + \vec{K}_2'$.

Es genügt, im Beweis nur Kräfte $\vec{K}_1$ und $\vec{K}_2$ mit von Null verschiedenem Betrag zu betrachten. Wir behandeln der Reihe nach die Fälle:

(a) Die gegebenen Kräfte schließen einen Winkel der Größe 2α mit $0 < \alpha < \frac{\pi}{2}$ ein und haben gleiche Beträge.

(b) Die gegebenen Kräfte haben beliebige Beträge und schließen einen rechten Winkel ein.

(c) Die gegebenen Kräfte sind beliebig.

Zu (a): Für die in A angreifenden Komponenten $\vec{K}_1$ und $\vec{K}_2$ gelte also $K_1 = K_2 =: K$. Der von ihnen eingeschlossene Winkel habe die Größe 2α. In diesem Fall liegt die Resultierende $\vec{R}$ auf der Winkelhalbierenden des eingeschlossenen Winkels. Spiegelt man nämlich an der Winkelhalbierenden, so erhält man $\vec{K}_1' = \vec{K}_2$ und $\vec{K}_2' = \vec{K}_1$. Dann folgt aus (12.4.6), (12.4.3) und (12.4.5) $\vec{R}' = \vec{K}_1' + \vec{K}_2' = \vec{K}_2 + \vec{K}_1 = \vec{K}_1 + \vec{K}_2 = \vec{R}$. Das kann nicht gelten, wenn $\vec{R}$ nicht auf der Winkelhalbierenden liegt.[1]) Damit ist nur noch der Betrag von $\vec{R}$ zu bestimmen. Nach (12.4.4) ist

12.4.7. $R = F(K, \alpha)$,

wobei F eine zweistellige stetige Funktion mit $K \in \mathbf{R}_+^*$ und $\alpha \in \left[\!\left[0, \frac{\pi}{2} \right]\!\right]$ ist, für die wegen (12.4.2) $F(K, 0) = 2K$ und $F\left(K, \frac{\pi}{2}\right) = 0$ gilt.

Wir zeigen zunächst, daß R linear von K abhängt. Dazu betrachten wir die in Abb. 62 dargestellte Konstellation von Kräften $\vec{K}_1'$, $\vec{K}_2'$, $\vec{K}_1''$ und $\vec{K}_2''$, die alle von A ausgehen, mit $K_1 := K_1' = K_1''$ und $K_2 := K_2' = K_2''$. Es gelte $\vec{R}_1 := \vec{K}_1' + \vec{K}_1'$

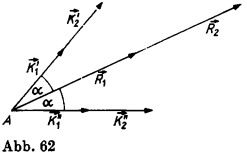


Abb. 62

und $\vec{R}_2 := \vec{K}_2' + \vec{K}_2''$. Die Resultierenden $\vec{R}_1$ und $\vec{R}_2$ liegen auf der Winkelhalbierenden und haben dieselbe Richtung. Also ist nach (12.4.2) $R_1 + R_2$ der Betrag der Kraft $\vec{R}_1 + \vec{R}_2$. Für die Gesamtresultierende $\vec{R}$ der in A angreifenden Kräfte gilt dann mit (12.4.5)

12.4.8. $\vec{R} = \vec{R}_1 + \vec{R}_2 = (\vec{K}_1' + \vec{K}_1'') + (\vec{K}_2' + \vec{K}_2'') = (\vec{K}_1' + \vec{K}_2') + (\vec{K}_1'' + \vec{K}_2'')$.

Nun ist nach (12.4.7) einerseits $R_1 = F(K_1, \alpha)$, $R_2 = F(K_2, \alpha)$ und andererseits, wenn man (12.4.8) beachtet, $R = F(K_1 + K_2, \alpha) = F(K_1, \alpha) + F(K_2, \alpha)$.

[1]) Es würde hier genügen zu fordern, daß die Resultierende auf der Winkelhalbierenden liegt. In Hinblick auf (b) haben wir diese „Symmetrieforderung" in der stärkeren Form (12.4.6) gewählt.

Für festes $\alpha \in \left]\!\left]0, \frac{\pi}{2}\right[\!\right[$ genügt F also bezüglich K der Cauchyschen Grundgleichung, wobei F für $K \in \mathbf{R}_+^*$ definiert und für diese K stetig ist. Also gilt $F(K, \alpha) = K \cdot F(1, \alpha)$ (vgl. Bemerkung 7.4.3). Zur Vereinfachung der Schreibweise setzen wir $f(\alpha) := \frac{1}{2} F(1, \alpha)$, so daß wir für die Resultierende von Kräften mit gleichem Betrag

$$12.4.9. \quad \bigwedge_{K \in \mathbf{R}_+^*} \bigwedge_{\alpha \in \left]\!\left]0, \pi/2\right[\!\right[} R = 2K \cdot f(\alpha)$$

erhalten, wobei $f: \left]\!\left]0, \frac{\pi}{2}\right[\!\right[\to \mathbf{R}$ eine stetige Funktion ist. Zur Bestimmung dieser Funktion f gehen wir von einem System von vier Kräften $\vec{K}_1'$, $\vec{K}_1''$, $\vec{K}_2'$ und $\vec{K}_2''$ aus, die alle denselben Betrag K haben und von einem Punkt A ausgehen. Weiterhin

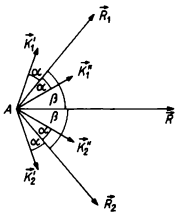


Abb. 63

schließen jeweils $\vec{K}_1'$ und $\vec{K}_1''$ bzw. $\vec{K}_2'$ und $\vec{K}_2''$ einen Winkel der Größe 2α $\left(\alpha \in \left]\!\left]0, \frac{\pi}{2}\right[\!\right[\right)$ ein. Es sei $\vec{R}_1 := \vec{K}_1' + \vec{K}_1''$, $\vec{R}_2 := \vec{K}_2' + \vec{K}_2''$ und $\vec{R} := \vec{R}_1 + \vec{R}_2$. Dann ist $R_1 = R_2 =: R$. Der Winkel zwischen $\vec{R}_1$ und $\vec{R}_2$ habe die Größe 2β (siehe Abb. 63), wobei $\beta \in \left[\!\left[0, \frac{\pi}{2}\right]\!\right]$ ist. Wir können $\vec{K}_1'$, $\vec{K}_1''$, $\vec{K}_2'$ und $\vec{K}_2''$ so anordnen, daß sowohl $\beta \in \left]\!\left]0, \frac{\pi}{2}\right[\!\right[$ als auch $\alpha + \beta \in \left]\!\left]0, \frac{\pi}{2}\right[\!\right[$ und $\beta - \alpha \in \left]\!\left]0, \frac{\pi}{2}\right[\!\right[$ ist. Mit (12.4.5) gilt

$$\begin{aligned}\vec{R} &= \vec{R}_1 + \vec{R}_2 = (\vec{K}_1' + \vec{K}_1'') + (\vec{K}_2' + \vec{K}_2'') \\ &= (\vec{K}_1' + \vec{K}_2') + (\vec{K}_1'' + \vec{K}_2'').\end{aligned}$$

Nach (12.4.4) und (12.4.9) erhält man $2\vec{R} \cdot f(\beta) = 2K \cdot f(\beta + \alpha) + 2K \cdot f(\beta - \alpha)$, woraus wegen $\vec{R} = 2K \cdot f(\alpha)$ folgt, daß für f über dem Intervall $\left]\!\left]0, \frac{\pi}{2}\right[\!\right[$ die d'Alembertsche Funktionalgleichung gilt: $2f(\alpha) \cdot f(\beta) = f(\beta + \alpha) + f(\beta - \alpha)$.

Da die Funktion f stetig ist und $\frac{\pi}{2}$ ihre kleinste positive Nullstelle ist (vgl. (12.4.2) und (12.4.4)), gilt $f(x) = \cos x$ (vgl. Satz 12.3.6).[1])

Insgesamt erhalten wir für die Resultierende $\vec{R}$ zweier Kräfte $\vec{K}_1$ und $\vec{K}_2$, die denselben Betrag K haben und einen Winkel mit der Größe 2α $\left(\alpha \in \left[\!\left[0, \frac{\pi}{2} \right]\!\right]\right)$ einschließen:

12.4.10. $R = 2K \cdot \cos \alpha$.

Es sei ausdrücklich hervorgehoben, daß wir für die Ableitung von (12.4.10) aus (12.4.2) bis (12.4.6) das euklidische Parallelenaxiom nicht benutzt haben. Interpretiert man (12.4.10) jedoch in der euklidischen Geometrie der Ebene, so ist die Parallelogrammregel für den Fall (a) bewiesen (vgl. Abb. 64).

Zu (b): Es seien jetzt $\vec{K}_1$ und $\vec{K}_2$ Kräfte mit beliebigen Beträgen, der von ihnen eingeschlossene Winkel habe das Maß $\frac{\pi}{2}$, und es sei $\vec{R} = \vec{K}_1 + \vec{K}_2$. Wir spiegeln das aus den Kräften $\vec{K}_1$, $\vec{K}_2$ und $\vec{R}$ bestehende System an der durch $\vec{K}_1$ gegebenen Geraden (vgl. Abb. 65). Wegen (12.4.6) gilt dann $\vec{R}' = \vec{K}_1 + (-\vec{K}_2)$ und somit $\vec{R} + \vec{R}' = 2\vec{K}_1$ wegen (12.4.5). Da $\vec{R}'$ das Spiegelbild von $\vec{R}$ ist, haben beide Kräfte den glei-

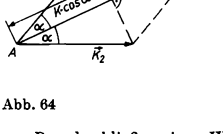


Abb. 64

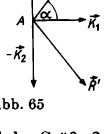


Abb. 65

chen Betrag R und schließen einen Winkel der Größe 2α ein. Damit liegt die Resultierende $2\vec{K}_1$ auf der Geraden, auf der $\vec{K}_1$ liegt, und nach (12.4.10) ergibt sich nun $K_1 = R \cdot \cos \alpha$. Eine Spiegelung des aus den Kräften $\vec{K}_1$, $\vec{K}_2$ und $\vec{R}$ bestehenden Systems an der durch $\vec{K}_2$ gegebenen Geraden liefert schließlich $K_2 = R \cdot \cos\left(\frac{\pi}{2} - \alpha\right)$, also $K_2 = R \cdot \sin \alpha$. Insgesamt erhalten wir

12.4.11. $K_1 = R \cdot \cos \alpha$ und $K_2 = R \cdot \sin \alpha$.

Durch (12.4.11) sind bei gegebenen zueinander rechtwinkligen Kräften $\vec{K}_1$ und $\vec{K}_2$ sowohl die Größe des Winkels α (also die Richtung der Resultierenden) als auch der

[1]) Einfache Abänderungen des Beweises von Satz 12.3.6 zeigen, daß die d'Alembertsche Funktionalgleichung, die dort über $\mathbf{R}$ gelöst wurde, auch über $\left[\!\left[0, \frac{\pi}{2} \right]\!\right]$ nur die auf dieses Intervall eingeschränkten stetigen Lösungen f mit $f(x) = \cos ax$ $(a \neq 0)$, $f(x) = \cosh ax$ $(a \neq 0)$, $f(x) = 1$ und $f(x) = 0$ jeweils für alle $x \in \left[\!\left[0, \frac{\pi}{2} \right]\!\right]$ besitzt (vgl. auch Bemerkung 7.4.3).

Betrag der Resultierenden eindeutig bestimmt. In der euklidischen Geometrie beschreiben diese Formeln die Parallelogrammregel.

Zu (c): Diesen Fall führen wir auf den Fall (b) zurück. Es seien also $\vec{K}_1$ und $\vec{K}_2$ zwei in einem Punkt A angreifende Kräfte. Wir zerlegen eine der gegebenen Kräfte (z. B. $\vec{K}_2$ in Abb. 66) so in rechtwinklige Komponenten $\vec{K}_2'$ und $\vec{K}_2''$, daß $\vec{K}_2'$ in der Geraden von $\vec{K}_1$ liegt. Damit gilt nach (12.4.5) $\vec{R} = \vec{K}_1 + \vec{K}_2 = \vec{K}_1 + (\vec{K}_2' + \vec{K}_2'') = (\vec{K}_1 + \vec{K}_2') + \vec{K}_2''$, womit $\vec{R}$ die Resultierende zweier zueinander rechtwinkliger

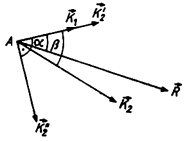


Abb. 66

Kräfte ist. Nach (12.4.11) gelten für K_2' bzw. K_2'' die Beziehungen $K_2' = K_2 \cdot \cos\beta$ und $K_2'' = K_2 \cdot \sin\beta$ (wobei β die Größe des Winkels zwischen $\vec{K}_1$ und $\vec{K}_2$ ist). Insgesamt erhält man daraus mit (12.4.11) und (12.4.2) für K_1, K_2 und R:

12.4.12. $K_1 + K_2 \cdot \cos\beta = R \cdot \cos\alpha$ und $K_2 \cdot \sin\beta = R \cdot \sin\alpha$,

wobei α die Größe des Winkels zwischen $\vec{K}_1$ und $\vec{R}$ ist. Wir bemerken, daß durch (14.12) auch die Komponenten festgelegt sind, wenn R, α und β gegeben sind. In der euklidischen Geometrie beschreiben diese Formeln die Parallelogrammregel.

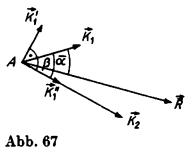


Abb. 67

Wir überzeugen uns noch davon, daß R und α davon unabhängig sind, welche der gegebenen Kräfte $\vec{K}_1$ oder $\vec{K}_2$ in rechtwinklige Komponenten zerlegt wird. Es sei also $\vec{K}_1$ so in rechtwinklige Komponenten $\vec{K}_1'$ und $\vec{K}_1''$ zerlegt, daß $\vec{K}_1''$ in der Geraden von $\vec{K}_2$ liegt. Die Resultierende $\vec{\bar{R}}$ von $\vec{K}_1$ und $\vec{K}_2$ habe den Betrag $\bar{R}$ und bilde mit $\vec{K}_1$ den Winkel $\bar{\alpha}$ (vgl. Abb. 67). Dann gilt

12.4.13. $K_1 \cdot \cos\beta + K_2 = \bar{R} \cdot \cos(\beta - \bar{\alpha})$ und $K_1 \cdot \sin\beta = \bar{R} \cdot \sin(\beta - \bar{\alpha})$.

Nun ergibt sich aus (12.4.12)

$$\textbf{12.4.14.}\quad \tan\alpha = \frac{K_2 \cdot \sin\beta}{K_1 + K_2 \cdot \cos\beta}$$

und aus (12.4.13)

$$\textbf{12.4.15.}\quad \tan(\beta - \bar{\alpha}) = \frac{K_1 \cdot \sin\beta}{K_1 \cdot \cos\beta + K_2}.$$

Die letzte Beziehung formen wir äquivalent um:

$$\tan\beta - \tan\bar{\alpha} = (1 + \tan\bar{\alpha}\cdot\tan\beta)\cdot\frac{K_1\cdot\sin\beta}{K_1\cdot\cos\beta + K_2}$$

oder

$$\tan\beta - \frac{K_1\cdot\sin\beta}{K_1\cdot\cos\beta + K_2} = \tan\bar{\alpha}\left(1 + \frac{K_1\cdot\sin\beta}{K_1\cdot\cos\beta + K_2}\cdot\tan\beta\right),$$

also

$$\frac{K_1\cdot\sin\beta\cos\beta + K_2\cdot\sin\beta - K_1\cdot\sin\beta\cos\beta}{K_1\cdot\cos^2\beta + K_2\cdot\cos\beta}$$
$$= \tan\bar{\alpha}\left(\frac{K_1\cdot\cos^2\beta + K_2\cdot\cos\beta + K_1\cdot\sin^2\beta}{K_1\cdot\cos^2\beta + K_2\cdot\cos\beta}\right)$$

und schließlich

12.4.16. $$\tan\bar{\alpha} = \frac{K_2\cdot\sin\beta}{K_1 + K_2\cdot\cos\beta}.$$

Wegen $0 < \beta < \pi$ können wir voraussetzen, daß $0 < \alpha < \frac{\pi}{2}$ gilt. Anderenfalls benutzen wir den Winkel zwischen $\vec{R}$ und $\vec{R}_2$ als Richtungswinkel von $\vec{R}$. Damit liefert der Vergleich von (12.4.14) und (12.4.16) $\alpha = \bar{\alpha}$. Es gilt aber auch $R = \bar{R}$. Aus der zweiten Gleichung von (12.4.13) folgt nämlich

12.4.17. $$\bar{R} = \frac{K_1\cdot\sin\beta}{\sin(\beta - \alpha)},$$

wobei wir $\alpha = \bar{\alpha}$ benutzen (Man überlegt sich, daß mit $\vec{K}_1 \neq \vec{0}$, $\vec{K}_2 \neq \vec{0}$ und $\beta \neq 0$ auch $\beta \neq \alpha$ ist.). Aus (12.4.12) erhält man durch Einsetzen von K_2 aus der zweiten Gleichung in die erste Gleichung

$$K_1 + R\cdot\frac{\sin\alpha}{\sin\beta}\cdot\cos\beta = R\cdot\cos\alpha,$$

also

$$K_1\cdot\sin\beta = R(\cos\alpha\sin\beta - \sin\alpha\cos\beta)$$

oder

$$R = \frac{K_1\cdot\sin\beta}{\sin(\beta - \alpha)}$$

und damit wegen (12.4.17) $R = \bar{R}$.

Damit haben wir insgesamt gezeigt, daß es höchstens eine zweistellige Operation $+$ gibt, die die Eigenschaften (12.4.2) bis (12.4.6) besitzt. Da nun die durch die Parallelogrammregel bestimmte Addition von Kräften die Eigenschaften (12.4.2) bis (12.4.6) hat, wenn man als Grenzfälle einbezieht, daß die Komponenten in einer Geraden liegen bzw. daß mindestens eine Komponente $\vec{0}$ ist, ist Satz 12.4.1 bewiesen.

Wir setzen nun eine Ebene voraus, in der die Axiome der absoluten Geometrie gelten. Dort betrachten wir die Menge der von einem festen Punkt A ausgehenden Pfeile (gerichtete Strecken). Die eineindeutige Zuordnung zwischen diesen Pfeilen und den in A angreifenden Kräften wird wie im euklidischen Fall definiert, ebenso die reellen Vielfachen und der Betrag einer Kraft. In den Eigenschaften (12.4.2) bis (12.4.6) wird nirgends Bezug genommen auf Parallelität, so daß wir den folgenden Satz formulieren:

12.4.18. Satz. *In der Menge der in einem Punkt A einer Ebene der absoluten Geometrie angreifenden und in dieser Ebene liegenden Kräfte ist durch die Eigenschaften* (12.4.2) *bis* (12.4.6) *genau eine zweistellige Operation $+$ definiert.*[1])

Beweis. Der Beweis, daß es höchstens eine Operation gibt, die den Eigenschaften (12.4.2) bis (12.4.6) genügt, braucht nicht mehr geführt zu werden, denn im entsprechenden Teil des Beweises von Satz 12.4.1 wurden nur Tatsachen benutzt, die auch in der absoluten Geometrie gelten. Es bleibt noch zu zeigen, daß die durch (12.4.12) definierte Operation[2]), wenn man sie durch Hinzunahme der Sonderfälle (12.4.2) auf alle Komponenten stetig erweitert, die genannten Eigenschaften (12.4.2) bis (12.4.6) besitzt. Diese Sonderfälle sind wieder Grenzfälle der durch (12.4.12) definierten Operation. Das folgt aus (12.4.12), besonders leicht aber aus (12.4.19) — siehe weiter unten.

Wir wollen hier nur die Assoziativität der durch (12.4.12) definierten Operation beweisen. (Die anderen Beweise sind offensichtlicher.)

Zunächst folgt aus (12.4.12) durch Quadrieren und anschließendes Addieren

$$\textbf{12.4.19.}\quad R^2 = K_1^2 + K_2^2 + 2K_1K_2 \cdot \cos\beta.$$

Die zweite Gleichung von (12.4.12) ist gleichbedeutend mit

$$\textbf{12.4.20.}\quad \frac{K_2}{R} = \frac{\sin\alpha}{\sin\beta}.$$

Das sind in der euklidischen Geometrie der *Kosinussatz* und der *Sinussatz* für Dreiecke. Zeichnet man in der euklidischen Ebene das zu $\vec{K}_1$, $\vec{K}_2$ und $\vec{R}$ gehörende Parallelogramm, so erhält man die entsprechenden Dreiecke mit den Seitenlängen K_1, K_2 und R. Die Formeln (12.4.19) und (12.4.20) sind in der absoluten Geometrie natürlich keine Sätze über Beziehungen in Dreiecken. Zwar kann man auch hier die durch $\vec{K}_1$, $\vec{K}_2$ und $\vec{R}$ gebildete Figur zu Dreiecken ergänzen, erhält dabei aber im allgemeinen eine andere Situation als in der euklidischen Geometrie. Die für die Anwendung des

[1]) Die Menge dieser Kräfte bildet mit dieser Operation und der Multiplikation von reellen Zahlen mit Kräften sogar einen reellen Vektorraum.

[2]) Durch (12.4.12) ist in der Tat eine Operation für $\vec{K}_1 \neq \vec{0}$ und $\vec{K}_2 \neq \vec{0}$ definiert, wobei β die Größe des Winkels zwischen den durch $\vec{K}_1$ und $\vec{K}_2$ gegebenen Geraden ist, denn bei gegebenen $K_1 \neq 0$, $K_2 \neq 0$ und $\beta \in]0, \pi[$ ergeben sich aus (12.4.12) eindeutig bestimmte Zahlen $\alpha \in \left]0, \frac{\pi}{2}\right[$ und R, durch die die Resultierende $\vec{R}$ eindeutig festgelegt ist.

Kosinus- und des Sinussatzes benötigten Aussagen über die Gleichheit von Stufen- und Wechselwinkeln an geschnittetenen Parallelen (vgl. Abb. 68) sind zum euklidischen Parallelenaxiom äquivalent (vgl. MfL Bd. 17, 2.1.2.).

Nun zum Beweis der Assoziativität. Es seien drei beliebige von A ausgehende Kräfte $\vec{K}_1$, $\vec{K}_2$ und $\vec{K}_3$ gegeben, die in der gleichen Ebene liegen. Wir führen folgende Bezeichnungen ein (vgl. Abb. 69): $\vec{R}_1 := \vec{K}_1 + \vec{K}_2$; $\vec{R}_2 := \vec{K}_2 + \vec{K}_3$; $\vec{R} := \vec{R}_1 + \vec{K}_3 = (\vec{K}_1 + \vec{K}_2) + \vec{K}_3$; $\vec{R}' := \vec{K}_1 + \vec{R}_2 = \vec{K}_1 + (\vec{K}_2 + \vec{K}_3)$; $\beta_1, \beta_2, \beta_3, \beta_4$ sind in dieser

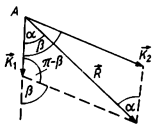


Abb. 68

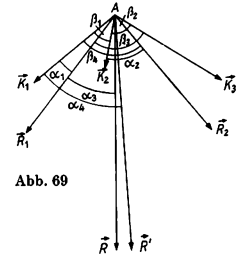


Abb. 69

Reihenfolge die Größen der Winkel zwischen $\vec{K}_1$ und $\vec{K}_2$, $\vec{K}_2$ und $\vec{K}_3$, $\vec{R}_1$ und $\vec{K}_3$, $\vec{K}_1$ und $\vec{R}_2$; $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ sind in dieser Reihenfolge die Größen der Winkel zwischen $\vec{K}_1$ und $\vec{R}_1$, $\vec{K}_2$ und $\vec{R}_2$, $\vec{R}_1$ und $\vec{R}$, $\vec{K}_1$ und $\vec{R}'$. Für die Winkel β_v $(v = 1, 2, 3, 4)$ und α_v $(v = 1, 2)$ gilt

12.4.21. $\beta_3 = \beta_1 + \beta_2 - \alpha_1$ und $\beta_4 = \beta_1 + \alpha_2$,

so daß aus (12.4.19)

$$R^2 = {R_1}^2 + {K_3}^2 + 2R_1K_3 \cdot \cos\beta_3,$$
$$R'^2 = {K_1}^2 + {R_2}^2 + 2K_1R_2 \cdot \cos\beta_4$$

folgt. Daraus erhält man durch nochmalige Benutzung von (12.4.19) für ${R_1}^2$ und ${R_2}^2$ und durch Einsetzen für β_3 und β_4

$$R^2 = {K_1}^2 + {K_2}^2 + {K_3}^2 + 2K_1K_2 \cdot \cos\beta_1 + 2R_1K_3 \cdot \cos(\beta_1 + \beta_2 - \alpha_1),$$
$$R'^2 = {K_1}^2 + {K_2}^2 + {K_3}^2 + 2K_2K_3 \cdot \cos\beta_2 + 2K_1R_2 \cdot \cos(\beta_1 + \alpha_2).$$

Nun gilt $R^2 = R'^2$ und damit wegen $R > 0$ auch $R = R'$ genau dann, wenn

12.4.22. $K_1K_2 \cdot \cos\beta_1 + R_1K_3 \cdot \cos(\beta_1 + \beta_2 - \alpha_1)$

und

12.4.23. $K_2K_3 \cdot \cos\beta_2 + K_1R_2 \cdot \cos(\beta_1 + \alpha_2)$

gleich sind. Man erhält aus (12.4.22) durch Anwendung der Subtraktions- und Additionstheoreme für cos und sin sowie durch (12.4.12) $K_1 + K_2 \cdot \cos\beta_1 = R_1 \cdot \cos\alpha_1$

und $K_2 \cdot \sin \beta_1 = R_1 \cdot \sin \alpha_1$

$$\begin{aligned} K_1K_2 \cdot \cos \beta_1 + R_1K_3 \Big(& \cos (\beta_1 + \beta_2) \cdot \frac{K_1 + K_2 \cdot \cos \beta_1}{R_1} \\ & + \sin (\beta_1 + \beta_2) \cdot \frac{K_2 \cdot \sin \beta_1}{R_1} \Big) \\ = K_1K_2 \cdot \cos \beta_1 & + K_1K_3 \cdot \cos (\beta_1 + \beta_2) \\ & + K_2K_3 \big(\cos \beta_1 \cos (\beta_1 + \beta_2) + \sin \beta_1 \sin (\beta_1 + \beta_2)\big) \\ = K_1K_2 \cdot \cos \beta_1 & + K_1K_3 \cdot \cos (\beta_1 + \beta_2) + K_2K_3 \cdot \cos \beta_2. \end{aligned}$$

Dieses Ergebnis erhält man durch entsprechende Umformungen auch aus (12.4.23), wenn man diesmal (12.4.12) in der Form $K_2 + K_3 \cdot \cos \beta_2 = R_2 \cdot \cos \alpha_2$ und $K_3 \cdot \sin \beta_2 = R_2 \cdot \sin \alpha_2$ benutzt.

Es bleibt noch zu zeigen, daß $\vec{R}$ und $\vec{R}'$ auch diesselbe Richtung haben, daß also $\alpha_4 = \alpha_1 + \alpha_3$ gilt (vgl. Abb. 69). Wir gehen wieder von (12.4.12) aus, wobei wir diese Beziehungen einmal für R_1 und K_3 (in (12.4.24)) und einmal für K_1 und R_2 (in (12.4.25)) formulieren und dabei gleich $R' = R$ setzen:

12.4.24. $R_1 + K_3 \cdot \cos (\beta_1 + \beta_2 - \alpha_1) = R \cdot \cos \alpha_3$;

$K_3 \cdot \sin (\beta_1 + \beta_2 - \alpha_1) = R \cdot \sin \alpha_3$,

12.4.25. $K_1 + R_2 \cdot \cos (\beta_1 + \alpha_2) = R \cdot \cos \alpha_4$;

$R_2 \cdot \sin (\beta_1 + \alpha_2) = R \cdot \sin \alpha_4$.

Aus (12.4.25) folgt durch Anwendung des Additionstheorems von sin und durch Einsetzen von

$$\cos \alpha_2 = \frac{K_2 + K_3 \cdot \cos \beta_2}{R_2} \quad \text{und} \quad \sin \alpha_2 = \frac{K_3 \cdot \sin \beta_2}{R_2},$$

wobei diese beiden Gleichungen wieder aus (12.4.12) (für K_2, K_3 und R_2) gewonnen wurden:

12.4.26. $R \cdot \sin \alpha_4 = K_2 \cdot \sin \beta_1 + K_3 \cdot \sin (\beta_1 + \beta_2)$.

Jetzt beweisen wir mit den Formeln (12.4.12) und (12.4.19), daß

12.4.27. $R \cdot \sin (\alpha_1 + \alpha_3) = K_2 \cdot \sin \beta_1 + K_3 \cdot \sin (\beta_1 + \beta_2)$

gilt, woraus dann $\alpha_4 = \alpha_1 + \alpha_3$ folgt, da wir wieder $0 < \alpha_4 < \frac{\pi}{2}$ voraussetzen können. Zunächst ist mit (12.4.12) für $\sin \alpha_1$, $\cos \alpha_1$, $\sin \alpha_3$ und $\cos \alpha_3$

$$\begin{aligned} \sin (\alpha_1 + \alpha_3) &= \sin \alpha_1 \cos \alpha_3 + \cos \alpha_1 \sin \alpha_3 \\ &= \frac{K_2 \cdot \sin \beta_1}{R_1} \cdot \frac{R_1 + K_3 \cdot \cos \beta_3}{R} \\ &\quad + \frac{K_1 + K_2 \cdot \cos \beta_1}{R_1} \cdot \frac{K_3 \cdot \sin \beta_3}{R}. \end{aligned}$$

Setzt man hierin für β_3 aus (12.4.21) ein und wendet die entsprechenden Additions- bzw. Subtraktionstheoreme der Funktionen sin und cos an und setzt dann aus (12.4.12) für $\cos \alpha_1$ und $\sin \alpha_1$ ein, so erhält man

$$\begin{aligned} & R_1 R \cdot \sin(\alpha_1 + \alpha_3) \\ &= R_1 K_2 \cdot \sin \beta_1 + \frac{K_2^2 K_3}{R_1} \cdot \sin(\beta_1 + \beta_2) \cdot (\sin^2 \beta_1 + \cos^2 \beta_1) \\ &\quad + \frac{K_1^2 K_3}{R_1} \cdot \sin(\beta_1 + \beta_2) + 2 \frac{K_1 K_2 K_3}{R_1} \cdot \cos \beta_1 \sin(\beta_1 + \beta_2) \\ &= R_1 K_2 \cdot \sin \beta_1 + \frac{K_3}{R_1} (K_1^2 + K_2^2) \cdot \sin(\beta_1 + \beta_2) \\ &\quad + 2 \frac{K_1 K_2 K_3}{R_1} \cdot \cos \beta_1 \sin(\beta_1 + \beta_2). \end{aligned}$$

Wenn man jetzt schließlich aus (12.4.19) $K_1^2 + K_2^2 = R_1^2 - 2K_1 K_2 \cdot \cos \beta_1$ einsetzt, erhält man (12.4.27). Damit beenden wir die Ausführungen zu Satz 12.4.18.

Unsere bisherigen Überlegungen haben gezeigt, daß sich die Statik des Massenpunktes in der absoluten, in der euklidischen und in der Lobačevskijschen Geometrie rechnerisch nicht unterscheiden. Wir haben ja für die Ableitung der Formeln (12.4.19) und (12.4.20) aus (12.4.3) bis (12.4.6) kein Parallelenaxiom benötigt. In der euklidischen Ebene ist durch die Parallelogrammregel eine bequeme zeichnerische Bestimmung der Resultierenden gegeben.

Versucht man aber diese Betrachtungen auf die Statik des starren Körpers zu übertragen, so ergeben sich wesentliche Unterschiede. Darauf werden wir nun eingehen. Dabei wird sich eine nochmalige Anwendung der d'Alembertschen Funktionalgleichung ergeben.

Wir betrachten im folgenden zwei Kräfte $\vec{K}_1$ und $\vec{K}_2$, die an einem starren Körper in den Punkten A_1 bzw. A_2 angreifen und in der gleichen Ebene liegen. Dabei sind die Fälle $A_1 = A_2$ und $A_1 \neq A_2$ zu unterscheiden. Falls $A_1 \neq A_2$ ist, sei stets vorausgesetzt, daß $\vec{K}_1$ und $\vec{K}_2$ auf ein und derselben Seite der Geraden $g(A_1 A_2)$ liegen.

In der euklidischen Geometrie wird bekanntlich die Resultierende folgendermaßen bestimmt:

a) Wenn $A_1 = A_2$ ist, dann wird die Resultierende direkt mit Hilfe der Parallelogrammregel bestimmt. Wenn $A_1 \neq A_2$ ist, beide Kräfte auf ein und derselben Seite der Geraden $g(A_1 A_2)$ liegen und wenn sich die durch $\vec{K}_1$ und $\vec{K}_2$ gegebenen Geraden (diese Geraden werden auch als *Wirkungslinien* der Kräfte $\vec{K}_1$ bzw. $\vec{K}_2$ bezeichnet) in einem Punkt S schneiden, so „verschiebt"[1]) man die Kräfte (Pfeile) $\vec{K}_1$ und $\vec{K}_2$, so daß ihre Angriffspunkte in S liegen, bestimmt dort mit Hilfe der Parallelogrammregel die Resultierende (vgl. Abb. 70).

b) Wenn sich die Wirkungslinien von $\vec{K}_1$ und $\vec{K}_2$ nicht schneiden, dann sind sie in der euklidischen Geometrie der Ebene zueinander parallel. Dann gibt es eine Gerade, die senkrecht zu beiden Wirkungslinien ist und diese in den Punkten B_1 und B_2 schneidet. Diese Gerade sei Wirkungslinie für zwei Kräfte $\vec{L}$ und $-\vec{L}$, die ebenfalls

[1]) Diese „Verschiebung" oder „Linienflüchtigkeit" wird weiter unten (S. 261) präzisiert.

in B_1 bzw. B_2 angreifen. Wir verschieben die Kräfte $\vec{K}_1$ und $\vec{K}_2$ so, daß ihre Angriffspunkte B_1 bzw. B_2 sind. Nun bestimmen wir die Resultierenden $\vec{K}_1' = \vec{K}_1 + \vec{L}$ und $\vec{K}_2' = \vec{K}_2 + (-\vec{L})$. Die Wirkungslinien dieser Resultierenden schneiden sich in einem Punkt S, es sei denn, die Komponenten $\vec{K}_1$ und $\vec{K}_2$ liegen auf verschiedenen Seiten von $g(A_1A_2)$, und es gilt $K_1 = K_2$. Existiert der Punkt S, so kann man dann weiter bei der Bestimmung der Resultierenden $\vec{R}$ wie unter a) verfahren (vgl. Abb. 71). Mit Sätzen der euklidischen Geometrie (Ähnlichkeitssätze) gilt

$$\frac{|\overline{BB_1}|}{|\overline{SB}|} = \frac{L}{K_1} \quad \text{und} \quad \frac{|\overline{BB_2}|}{|\overline{SB}|} = \frac{L}{K_2}, \quad \text{also} \quad \frac{|\overline{BB_1}|}{|\overline{BB_2}|} = \frac{K_2}{K_1}.$$

Außerdem gilt $R = K_1 + K_2$. Die Resultierende $\vec{R}$ ist von den „Hilfskräften" $\vec{L}$ und $-\vec{L}$ unabhängig. Diese Kräfte heben sich in ihrer Wirkung auf. Bei diesen Überlegungen spielt das euklidische Parallelenaxiom eine entscheidende Rolle.[1])

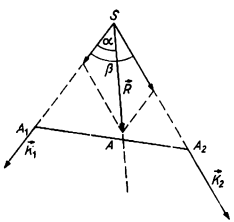


Abb. 70

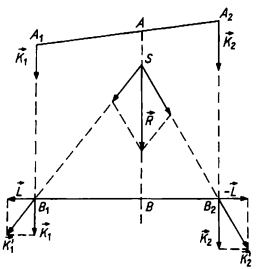


Abb. 71

Wir wollen jetzt einige Folgerungen untersuchen, die sich ergeben, wenn man die eben skizzierten Methoden der Behandlung der Statik des starren Körpers auch in die absolute und in die Lobačevskijsche Geometrie überträgt. Dabei streben wir keine Vollständigkeit an. Wir wollen für einige Spezialfälle die Bedingungen untersuchen, denen drei in einer Ebene liegende und an einem starren Körper angreifende Kräfte $\vec{K}_1$, $\vec{K}_2$ und $\vec{R}$ genügen müssen, damit sich der Körper im Gleichgewicht befindet. (Das ergibt einen interessanten Zugang zur Lobačevskijschen Funktion und eine weitere Anwendung der d'Alembertschen Funktionalgleichung.)

Vorher sei an folgende Ergebnisse der ebenen Lobačevskijschen Geometrie erinnert:[2]) Sind eine Gerade g und ein Punkt $P \notin g$ gegeben, so gibt es durch P zu g

[1]) Bekanntlich (vgl. MfL Bd. 17) ist das euklidische Parallelenaxiom äquivalent zu der Tatsache, daß es zwei nichtkongruente Dreiecke gibt, die in allen Winkeln übereinstimmen. Die Ähnlichkeitslehre ist also eine Folgerung des euklidischen Parallelenaxioms. In der Lobačevskijschen Geometrie sind Dreiecke, die in allen Winkeln übereinstimmen, stets kongruent. Die Kongruenzsätze (wsw), (sww) und (ssw) gelten bereits in der absoluten Geometrie. Davon werden wir noch Gebrauch machen.

[2]) Für eine gründliche Information sei auf MfL Bd. 17, A. P. Norden [125] und die Enzyklopädie eine Elementarmathematik [49] verwiesen.

genau zwei *Randparallelen* (wir wollen sie hier kurz *Parallelen* zu g nennen). Diese beiden Geraden „trennen" die Geraden durch P, die g schneiden, von denen, die g nicht schneiden (vgl. Abb. 72). Sie liegen symmetrisch zum Lot von P auf g und schneiden g ebenfalls nicht. Sind p_1 und p_2 die beiden Parallelen zu g durch P, so bilden sie einen Winkel der Größe $2x$, in dessen Innerem g verläuft. Der Winkel

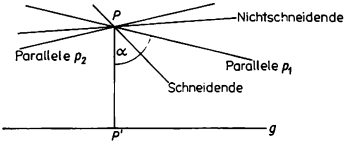


Abb. 72

zwischen dem Lot PP' und einer Parallelen heißt *Parallelwinkel*. Dieser Winkel hängt von dem Abstand des Punktes P von g ab. Man schreibt für seine Größe $\alpha = \pi(\overline{PP'})$ und nennt π die *Lobačevskijsche Funktion*, die für alle positiven Zahlen definiert und deren Wertebereich das Intervall $\left]\!\left]0, \frac{\pi}{2}\right[\!\right[$ ist. In der euklidischen Geometrie ist $\pi(x) = \frac{\pi}{2}$ für alle $x \in \mathbf{R}_+^*$. Ist die Gerade h durch einen Punkt P parallel zu einer Geraden g, so ist sie in jedem ihrer Punkte parallel zu g.[1]) Umgekehrt ist auch g zu h parallel.

Wir führen unsere Untersuchungen zunächst in der ebenen absoluten Geometrie durch, benutzen somit kein Parallelenaxiom. Im Unterschied zur Statik des Massenpunktes genügt es jetzt nicht, jeder Kraft nur einen Pfeil zuzuordnen. Wir fassen alle Pfeile einer Geraden mit gleicher Richtung und gleicher Länge zu einer Klasse zusammen und ordnen jeder derartigen Klasse umkehrbar eindeutig eine Kraft zu. Zwei Pfeile auf verschiedenen Geraden liegen also nie in der gleichen Klasse. Die Maßzahl der Länge eines jeden Pfeils der Klasse ist gleich der Maßzahl der Kraft, und jeder dieser Pfeile hat dieselbe Richtung wie die zugeordnete Kraft. Die Kraft kann durch jeden Pfeil der ihr zugeordneten Pfeilklasse repräsentiert werden. Einzelne Repräsentanten werden wie üblich wie die Kraft selbst bezeichnet. Ganz im Sinne der Statik des starren Körpers wird also durch Änderung des Betrages oder durch Änderung der Richtung einer Kraft i. allg. ihre Wirkung geändert, nicht aber durch Änderung ihres Angriffspunktes innerhalb ihrer Wirkungslinie (in der Physik sprich man in diesem Zusammenhang auch von „Linienflüchtigkeit").

Wir wollen jetzt versuchen, auch ohne Benutzung des euklidischen Parallelenaxioms die Zusammensetzung von Kräften am starren Körper auf die in der Statik des Massenpunktes zurückzuführen. Dazu werden wir auch hier zu Repräsentanten übergehen, die einen gemeinsamen Anfangspunkt S haben und dann den in S beginnenden Repräsentanten der Resultierenden bestimmen. Dafür wird die Gültigkeit

[1]) Sind also zwei Geraden zueinander parallel, so braucht man nicht zu sagen, in welchem Punkt, wohl aber, in welcher Richtung sie zueinander parallel sind.

der Forderungen (12.4.2) bis (12.4.6)[1]) vorausgesetzt, so daß wir insbesondere die Formeln (12.4.10) und (12.4.12) (für Repräsentanten mit gemeinsamem Anfangspunkt) benutzen können.

Nun zu den angekündigten Spezialfällen.

Die angreifenden Kräfte $\vec{K}_1$ und $\vec{K}_2$ haben beliebige Beträge, ihre Wirkungslinien schneiden sich in einem Punkt S. Wir finden die Resultierende $\vec{R} = \vec{K}_1 + \vec{K}_2$[2]) über die in S angreifenden Repräsentanten wie bei der Statik des Massenpunktes mit (12.4.12) $K_1 + K_2 \cdot \cos\beta = R \cdot \cos\alpha$ und $K_2 \cdot \sin\beta = R \cdot \sin\alpha$.

Es seien jetzt $\vec{K}_1$ und $\vec{K}_2$ Kräfte mit gleichem Betrag K und mit Repräsentanten, die in A_1 bzw. A_2 angreifen und auf ein und derselben Seite der Geraden $g(A_1A_2)$ liegen. Die Wirkungslinien von $\vec{K}_1$ und $\vec{K}_2$ mögen senkrecht auf $g(A_1A_2)$ stehen, haben also keinen gemeinsamen Schnittpunkt. Wir lassen daher in A_1 bzw. A_2 wieder Hilfskräfte $\vec{L}$ bzw. $-\vec{L}$ angreifen (vgl. Abb. 73), deren Betrag so gewählt sei, daß sich die Wirkungslinien der Resultierenden $\vec{K}_1' := \vec{K}_1 + \vec{L}$ und $\vec{K}_2' := \vec{K}_2 + (-\vec{L})$ schneiden.

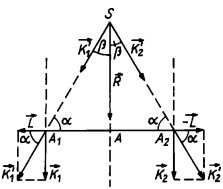


Abb. 73

Eine solche Wahl von $\vec{L}$ ist auch in der absoluten Geometrie möglich. Es gibt ja sowohl durch A_1 als auch durch A_2 Geraden, die etwa die Mittelsenkrechte der Strecke $\overline{A_1A_2}$ schneiden, wobei wir außerdem voraussetzen können, daß ihre Schnittwinkel mit $g(A_1A_2)$ in A_1 bzw. A_2 kongruent sind, also dieselbe Größe α besitzen. Da die entstehenden Dreiecke aufgrund des Kongruenzsatzes (wsw) kongruent sind, schneiden sich die so gewählten Wirkungslinien von $\vec{K}_1'$ bzw. $\vec{K}_2'$ ebenfalls, und ihr Schnittpunkt S liegt auf der Mittelsenkrechten der Strecke $\overline{A_1A_2}$. Nach (12.4.11) kann nun $\vec{L}$ so gewählt werden, daß die Schnittwinkel in den Punkten A_1 bzw. A_2 jede beliebige Größe α mit $0 < \alpha < \frac{\pi}{2}$ annehmen.

Durch ähnliche Überlegungen folgt: Wählt man $\vec{L}$ und damit die Winkelgröße α so, daß sich die Wirkungslinien von $\vec{K}_1'$ und $\vec{K}_2'$ schneiden, so liegt ihr Schnittpunkt S

[1]) Es ist lediglich zu beachten, daß sich die dort getroffenen Aussagen über Kräfte und Resultierende jetzt entsprechend auf Repräsentanten mit gemeinsamem Anfangspunkt beziehen.

[2]) Wir benutzen hier und im folgenden für die Zusammensetzung von Kräften am starren Körper dasselbe Zeichen „+" wie in der Statik des Massenpunktes, obwohl es sich um eine neu zu definierende (partielle) Operation in der Menge der Pfeilklassen handelt.

stets auf der Mittelsenkrechten der Strecke $\overline{A_1A_2}$. Nach (12.4.11) gilt

$$K_1' = \frac{K_1}{\sin\alpha} = \frac{K}{\sin\alpha}, \quad K_2' = \frac{K_1}{\sin\alpha} = \frac{K}{\sin\alpha}, \quad \text{also} \quad K' := K_1' = K_2'.$$

Die Resultierende $\vec{R}$ der Kräfte $\vec{K}_1'$ und $\vec{K}_2'$ liegt also auf der Winkelhalbierenden des Winkels, den die Repräsentanten der Kräfte $\vec{K}_1'$ und $\vec{K}_2'$ mit dem gemeinsamen Anfangspunkt S bilden. Damit steht die Wirkungslinie von $\vec{R}$ senkrecht auf der Geraden $g(A_1A_2)$ und geht durch den Mittelpunkt der Strecke $\overline{A_1A_2}$. Für R gilt dann nach (12.4.10) $R = 2K' \cdot \cos\beta$, also

12.4.28. $$R = 2K \cdot \frac{\cos\beta}{\sin\alpha}.$$

Wir zeigen jetzt, wie man von (12.4.28) ausgehend, durch zwei für die Statik des starren Körpers sinnvolle Forderungen (vgl. (12.4.29) und (12.4.30)) Zugänge zur euklidischen und zur Lobačevskijschen Geometrie erhalten kann. Die Forderungen

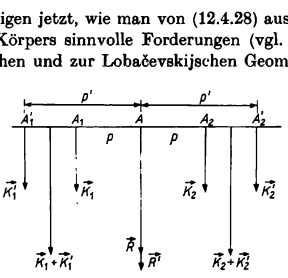


Abb. 74

(12.4.29) und (12.4.30) sind äquivalent zu bestimmten Grundformeln dieser Geometrien, die ihrerseits wiederum äquivalent zum jeweiligen Parallelenaxiom sind.

12.4.29. Der Betrag $R = 2K \cdot \frac{\cos\beta}{\sin\alpha}$ der Resultierenden ist von den Hilfskräften $\vec{L}$ und $-\vec{L}$ unabhängig.

Zur Formulierung der Forderung (12.4.30) betrachten wir folgende Konstellation von Kräften, die an einem starren Körper angreifen (vgl. Abb. 74): Es seien $\vec{K}_1$, $\vec{K}_2$, $\vec{K}_1'$ und $\vec{K}_2'$ Kräfte gleichen Betrages, die in den Punkten A_1, A_2, A_1' bzw. A_2' angreifen. Wir setzen $K := K_1 = K_2 = K_1' = K_2'$ und können voraussetzen, daß alle Angriffspunkte auf der Geraden $g(A_1A_2)$ liegen. Die Wirkungslinien der Kräfte sollen $g(A_1A_2)$ rechtwinklig schneiden. Außerdem seien die Kräfte so gewählt, daß die Repräsentanten mit den Anfangspunkten A_1, A_2, A_1' bzw. A_2' sämtlich auf ein und derselben Seite der Geraden $g(A_1A_2)$ liegen. Die Resultierenden $\vec{R} := \vec{K}_1 + \vec{K}_2$ und $\vec{R}' := \vec{K}_1' + \vec{K}_2'$ liegen dann auf einer Geraden, die $g(A_1A_2)$ rechtwinklig im Mittelpunkt der Strecke $\overline{A_1A_2}$ schneidet. Die Resultierenden $\vec{K}_1 + \vec{K}_1'$ und $\vec{K}_2 + \vec{K}_2'$ liegen auf Geraden, die die Strecken $\overline{A_1'A_1}$ bzw. $\overline{A_2A_2'}$ jeweils im Mittelpunkt rechtwinklig schneiden.

Es gilt $\vec{R} + \vec{R}' = (\vec{K}_1 + \vec{K}_2) + (\vec{K}_1' + \vec{K}_2')$. Die Resultierende $\vec{R}'' := (\vec{K}_1 + \vec{K}_1') + (\vec{K}_2 + \vec{K}_2')$ liegt ebenfalls in der Wirkungslinie von $\vec{R}$ und $\vec{R}'$. Wir fordern nun

12.4.30. Die Gesamtwirkungen $\vec{R} + \vec{R}'$ und $\vec{R}''$ der angreifenden Kräfte sind gleich. Es gilt also $\vec{R} + \vec{R}' = \vec{R}''$.

Wegen $R = 2K$ sind (12.4.29) und (12.4.30) in der euklidischen Geometrie erfüllt und folgen also aus dem Satz, daß die Summe der Innenwinkelgrößen eines Dreiecks gleich π ist. Umgekehrt folgt dieser Satz schon aus (12.4.29), denn mit $\frac{\cos\beta}{\sin\alpha} = 1$ ist $\cos\beta = \sin\alpha$, also $\sin\left(\frac{\pi}{2} - \beta\right) = \sin\alpha$, d. h. $\frac{\pi}{2} = \alpha + \beta$ für jedes rechtwinklige Dreieck, woraus der erwähnte Satz sofort für beliebige Dreiecke folgt. Dieser Satz ist aber äquivalent zum euklidischen Parallelenaxiom (vgl. MfL Bd. 17, 2.1.2.).

Welche geometrischen Folgerungen haben nun (12.4.29) und (12.4.30)? Zur Beantwortung dieser Frage setzen wir noch $p := \frac{1}{2}|(\overline{A_1A_2})|$, $p' := \frac{1}{2}|(\overline{A_1'A_2'})|$ und $\frac{\cos\beta}{\sin\alpha} =: c(p)$, wobei c eine stetige Funktion sei. Damit wird zum Ausdruck gebracht, daß der Betrag der Resultierenden in (12.4.28) zwar nicht von der Hilfskraft $\vec{L}$ abhängen soll, daß er jedoch von p abhängen könnte. Falls eine solche Abhängigkeit vorliegt, soll sie stetig sein (auch das ist eine physikalisch sinnvolle Forderung).

Mit (12.4.28) und den Bezeichnungen aus Abb. 74 ist $R = 2K \cdot c(p)$, $R' = 2K \cdot c(p')$ und

$$R'' = 2\,|\vec{K}_1 + \vec{K}_1'| \cdot c\left(\frac{p' + p}{2}\right) = 2 \cdot 2K \cdot c\left(\frac{p' - p}{2}\right) c\left(\frac{p' + p}{2}\right)$$

und damit nach (12.4.30)

$$2K \cdot c(p) + 2K \cdot c(p') = 4K \cdot c\left(\frac{p' - p}{2}\right) c\left(\frac{p' + p}{2}\right)$$

für $p', p \in \mathbf{R}_+^*$ und $p' - p > 0$. Setzen wir jetzt $x := \frac{p' + p}{2}$ und $y := \frac{p' - p}{2}$, so folgt

$$\bigwedge_{x,y \in \mathbf{R}_+^*} c(x + y) + c(x - y) = 2c(x)\,c(y).$$

Die gesuchte Funktion c ist also stetige Lösung der d'Alembertschen Funktionalgleichung, so daß die Funktionen $c: \mathbf{R}_+^* \to \mathbf{R}$ mit $c(x) = \cosh ax$ $(a \neq 0)$, $c(x) = \cos ax$ $(a \neq 0)$, $c(x) = 1$ und $c(x) = 0$ in Frage kommen (vgl. die Fußnote auf S. 253). Damit erhalten wir im ersten Fall $\frac{\cos\beta}{\sin\alpha} = \cosh\frac{p}{k}$, im zweiten Fall $\frac{\cos\beta}{\sin\alpha} = \cos\frac{p}{k}$ und im dritten Fall $\frac{\cos\beta}{\sin\alpha} = 1$ ($c(x) = 0$ für alle $x \in \mathbf{R}_+^*$ scheidet schon aus physikalischen Gründen aus).

Im ersten Fall handelt es sich um eine Grundformel der Lobačevskijschen (oder hyperbolischen) Geometrie und im zweiten Fall (das sei hier nur angemerkt) um eine Grundformel der Riemannschen (oder elliptischen) Geometrie. Im dritten Fall er-

halten wir noch einmal den euklidischen Fall. Die Zahl k gibt jeweils eine Länge an und ist ein charakteristischer Parameter der betreffenden Geometrie.

Wir wollen auf den ersten Fall noch näher eingehen und aus

12.4.31. $$\frac{\cos\beta}{\sin\alpha} = \cosh\frac{p}{k}$$

einige Sätze der Lobačevskijschen Geometrie ableiten. Zunächst folgt aus $\frac{\cos\beta}{\sin\alpha} > 1$, daß $\sin\left(\frac{\pi}{2} - \beta\right) > \sin\alpha$ und damit $\frac{\pi}{2} > \alpha + \beta$ ist. Da man jedes Dreieck in zwei rechtwinklige Dreiecke zerlegen kann, folgt daraus $\pi > \alpha + \beta + \gamma$, und das ist äquivalent zum Lobačevskijschen Parallelenaxiom (vgl. MfL Bd. 17, 2.1.2.). Weiterhin folgen aus $R = 2K \cdot \cosh\frac{p}{k}$ die Forderungen (12.4.29) und (12.4.30).

Als nächstes bestimmen wir die Lobačevskijsche Funktion π. Die Hilfskraft $\vec{L}$ (vgl. Abb. 73) sei so gewählt, daß $\alpha = \pi(p)$ ist. Wir bilden eine Folge von Repräsentanten $(\vec{L}_v)$ mit $\lim\limits_{v\to\infty} \vec{L}_v = \vec{L}$ und betrachten die zugehörigen Folgen von Winkelgrößen (α_v) und (β_v). Dann gilt $\lim\limits_{v\to\infty} \alpha_v = \pi(p)$, $\lim\limits_{v\uparrow\infty} \beta_v = 0$ und wegen der Beziehung (12.4.31) $\lim\limits_{v\to\infty} \frac{\cos\beta_v}{\sin\alpha_v} = c(p)$. Man erhält

$$c(p) = \lim_{v\to\infty} \frac{\cos\beta_v}{\sin\alpha_v} = \frac{1}{\sin\pi(p)}.$$

Insgesamt ist also

12.4.32. $$\cosh\frac{p}{k} = \frac{1}{\sin\pi(p)} \left(= \frac{\cos\beta}{\sin\alpha}\right).$$

Aus (12.4.32) folgt

$$\cos^2\pi(p) = 1 - \sin^2\pi(p) = 1 - \frac{1}{\cosh^2\frac{p}{k}} = \tanh^2\frac{p}{k}.$$

Wegen $0 < \pi(p) < \frac{\pi}{2}$ und $\frac{p}{k} > 0$ ist $\cos\pi(p) = \tanh\frac{p}{k}$ und damit

$$\tan\frac{\pi(p)}{2} = \frac{1 - \cos\pi(p)}{\sin\pi(p)} = \left(1 - \tanh\frac{p}{k}\right)\cosh\frac{p}{k}$$

$$= \cosh\frac{p}{k} - \sinh\frac{p}{k} = e^{-p/k}.$$

Daraus folgt für die Lobačevskijsche Funktion π

$$\pi(p) = 2 \cdot \arctan e^{-p/k}.$$

Wir kehren noch einmal zur Statik des starren Körpers zurück und bestimmen die Resultierende zweier (im Lobačevskijschen Sinne) in gleicher Richtung paralleler

Kräfte gleichen Betrages. Es seien g_1 und g_2 zueinander senkrechte Geraden mit dem Schnittpunkt A und A_1 bzw. A_2 Punkte auf g_1 auf verschiedenen Seiten von A mit $l(\overline{A_1A}) = l(\overline{A_2A}) = p$. Weiterhin seien g_3 und g_4 Geraden durch A_1 bzw. A_2, die jeweils zu g_2 in derselben Richtung parallel sind. Dann sind g_3 und g_4 ebenfalls zueinander parallel (in derselben Richtung) (vgl. die Fußnote auf S. 261) (vgl. Abb. 75). Die Kräfte $\vec{K}_1$ bzw. $\vec{K}_2$ mit dem Betrag $K := K_1 = K_2$ liegen in g_3 bzw. g_4. Zur Bestimmung der Resultierenden $\vec{R}$ von $\vec{K}_1$ und $\vec{K}_2$ arbeiten wir wieder mit Hilfskräften

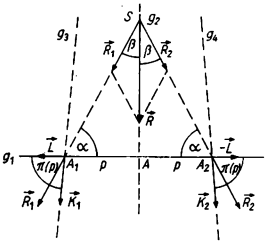


Abb. 75

$\vec{L}$ und $-\vec{L}$, deren Wirkungslinie die Gerade g_1 sei (vgl. Abb. 75). Die Hilfskräfte seien so gewählt, daß sich die Wirkungslinien von $\vec{R}_1 := \vec{K}_1 + \vec{L}$ und $\vec{R}_2 := \vec{K}_2 + (-\vec{L})$ in einem Punkt S schneiden. S liegt auf der Geraden g_2, die gleichzeitig Wirkungslinie der Resultierenden $\vec{R} := \vec{R}_1 - \vec{R}_2$ ist. Das folgt aus der Kongruenz (sww) der Dreiecke $\triangle SA_1A$ und $\triangle SAA_2$. Da $\vec{R}_1$ und $\vec{R}_2$ denselben Betrag $\vec{R}$ haben, gilt nach (12.4.10) $R = 2\vec{R} \cdot \cos\beta$. Außerdem ist nach (12.4.12) $K \cdot \sin\pi(p) = \vec{R} \cdot \sin\alpha$, also $\vec{R} = \dfrac{K \cdot \sin\pi(p)}{\sin\alpha}$. Damit ergibt sich $R = 2K \cdot \sin\pi(p)\dfrac{\cos\beta}{\sin\alpha}$, woraus schließlich mit (12.4.32) $R = 2K$ folgt. Man erhält also auch in der Lobačevskijschen Geometrie den Betrag der Resultierenden zweier paralleler Kräfte mit gleichen Beträgen durch Addition der Beträge.

Nun zu einigen Sätzen der Lobacevskijschen Trigonometrie.

12.4.33. Satz. *In jedem rechtwinkligen Dreieck $\triangle ABC$ mit den Kathetenlängen a, b und der Hypotenusenlänge c gilt* $\cosh\dfrac{c}{k} = \cosh\dfrac{a}{k} \cdot \cosh\dfrac{b}{k}$ $(k \in \mathbb{R}_+^*)$.

Dieser Satz entspricht dem Satz von PYTHAGORAS in der euklidischen Geometrie. Zum Beweis ergänzen wir das bei C rechtwinklige Dreieck $\triangle ABC$ durch ein bei A rechtwinkliges Dreieck $\triangle ABD$ (vgl. Abb. 76). Dann gilt nach (12.4.31)

$$\cosh\frac{c}{k} = \frac{\cos\gamma}{\sin\beta}, \quad \cosh\frac{b}{k} = \frac{\cos\gamma}{\sin\left(\dfrac{\pi}{2} - \alpha\right)} = \frac{\cos\gamma}{\cos\alpha}, \quad \cosh\frac{a}{k} = \frac{\cos\alpha}{\sin\beta}.$$

Aus diesen Beziehungen folgt unmittelbar die Behauptung.

Die folgenden Sätze beweisen wir für spitzwinklige Dreiecke. Man überlegt sich leicht, daß sie auch für stumpfwinklige Dreiecke gelten.

12.4.34. Satz. *Für beliebige Dreiecke $\triangle ABC$ gilt der (sogenannte) Sinussatz*

$$\frac{\sinh\frac{a}{k}}{\sin\alpha} = \frac{\sinh\frac{b}{k}}{\sin\beta} = \frac{\sinh\frac{c}{k}}{\sin\gamma} \quad (k \in \mathbf{R}_+^*).$$

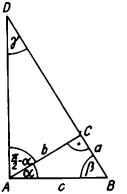


Abb. 76

Zum Beweis betrachten wir zunächst ein rechtwinkliges Dreieck $\triangle ABC$ (vgl. Abb. 76). Aus Grundbeziehungen der Winkelfunktionen und der hyperbolischen Funktionen sowie aus (12.4.31) und Satz 12.4.33 folgt

12.4.35. $$\sinh^2\frac{b}{k} = \cosh^2\frac{b}{k} - 1 = \frac{\cos^2\beta}{\sin^2\alpha} - 1 = \frac{1 - (\sin^2\beta + \sin^2\alpha)}{\sin^2\alpha},$$

12.4.36. $$\sinh^2\frac{a}{k} = \frac{1 - (\sin^2\beta + \sin^2\alpha)}{\sin^2\beta},$$

12.4.37. $$\sinh^2\frac{c}{k} = \cosh^2\frac{c}{k} - 1 = \cosh^2\frac{a}{k}\cosh^2\frac{b}{k} - 1$$
$$= \frac{\cos^2\alpha\cdot\cos^2\beta}{\sin^2\beta\cdot\sin^2\alpha} - 1 = \frac{1 - (\sin^2\beta + \sin^2\alpha)}{\sin^2\beta\cdot\sin^2\alpha}.$$

Da alle vorkommenden Funktionswerte positiv sind, gilt nach (12.4.35) bis (12.4.37)

$$\sinh\frac{c}{k}\cdot\sin\alpha\cdot\sin\beta = \sinh\frac{a}{k}\cdot\sin\beta = \sinh\frac{b}{k}\cdot\sin\alpha$$

oder

$$\sinh\frac{c}{k} = \frac{\sin\frac{a}{k}}{\sin\alpha} = \frac{\sinh\frac{b}{k}}{\sin\beta}.$$

Es sei jetzt $\triangle ABC$ ein beliebiges Dreieck mit den Bezeichnungen aus Abb. 77, h sei die Länge der Höhe von C auf $\overline{AB}$. Dann folgt aus den eben erhaltenen Beziehungen

$$\sinh\frac{b}{k} = \frac{\sinh\frac{h}{k}}{\sin\alpha}, \quad \sinh\frac{a}{k} = \frac{\sinh\frac{h}{k}}{\sin\beta}, \quad \text{also} \quad \frac{\sinh\frac{b}{k}}{\sin\beta} = \frac{\sinh\frac{a}{k}}{\sin\alpha}.$$

Führt man dieselben Überlegungen mit einer anderen Höhe durch, so erhält man die Behauptung.

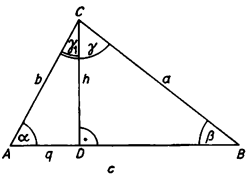


Abb. 77

12.4.38. Satz. *Für beliebige Dreiecke* $\triangle$ *ABC gilt der Seitenkosinussatz*

$$\cosh\frac{a}{k} = \cosh\frac{b}{k}\cosh\frac{c}{k} - \sinh\frac{b}{k}\sinh\frac{c}{k}\cos\alpha,$$

$$\cosh\frac{b}{k} = \cosh\frac{a}{k}\cosh\frac{c}{k} - \sinh\frac{a}{k}\sinh\frac{c}{k}\cos\beta,$$

$$\cosh\frac{c}{k} = \cosh\frac{a}{k}\cosh\frac{b}{k} - \sinh\frac{a}{k}\sinh\frac{b}{k}\cos\gamma \qquad (k \in \mathbf{R}_+^*).$$

Wir beweisen hier nur die erste Gleichung. Im rechtwinkligen Dreieck $\triangle$ *DBC* (vgl. Abb. 77) ist gemäß Satz 12.4.33

$$\cosh\frac{a}{k} = \cosh\frac{h}{k}\cosh\frac{c-q}{k},$$

d. h.

$$\cosh\frac{a}{k} = \cosh\frac{h}{k}\left(\cosh\frac{c}{k}\cosh\frac{q}{k} - \sinh\frac{c}{k}\sinh\frac{q}{k}\right).$$

Aus dieser Gleichung folgt die Behauptung, wenn man $\cosh\frac{b}{k} = \cosh\frac{h}{k}\cosh\frac{q}{k}$ einsetzt und dann $\cosh\frac{h}{k} = \frac{\cos\alpha}{\sin\gamma_1}$ und $\sinh\frac{b}{k} = \frac{\sinh\frac{q}{k}}{\sin\gamma_1}$ beachtet.

Ohne Beweis sei noch mitgeteilt, daß aus den Sätzen 12.4.34 und 12.4.38 der *Winkelkosinussatz* folgt:

12.4.39. Satz. *Für beliebige Dreiecke* $\triangle$ *ABC gilt*

$$\cos\alpha = -\cos\beta\cdot\cos\gamma + \sin\beta\cdot\sin\gamma\cdot\cosh\frac{a}{k},$$

$$\cos\beta = -\cos\alpha\cdot\cos\gamma + \sin\alpha\cdot\sin\gamma\cdot\cosh\frac{b}{k},$$

$$\cos\gamma = -\cos\alpha\cdot\cos\beta + \sin\alpha\cdot\sin\beta\cdot\cosh\frac{c}{k} \qquad (k \in \mathbf{R}_+^*).$$

Dieser Satz besagt, daß man in der Lobačevskijschen Geometrie die Seiten eines Dreiecks aus den drei Innenwinkeln berechnen kann. Dreiecke, die in ihren Winkeln übereinstimmen, sind ja auch kongruent.

Es sei noch darauf hingewiesen, daß zwischen den trigonometrischen Beziehungen in der ebenen Lobačevskijschen Geometrie, den entsprechenden Beziehungen in der sphärischen Trigonometrie und den auch im Mathematikunterricht behandelten trigonometrischen Beziehungen der ebenen euklidischen Geometrie enge Zusammenhänge bestehen (vgl. A. P. Norden [125] und Enzyklopädie der Elementarmathematik [49]).

13. Lösungen der Funktionalgleichung $f(x + yf(x)) = f(x) \cdot f(y)$

Bei der Untersuchung der Untergruppen der zentralaffinen Gruppe einer Ebene wurde S. Gołąb zu der Funktionalgleichung $f(x + yf(x)) = f(x) \cdot f(y)$ geführt. Diese Funktionalgleichung ist von 1959 bis 1969 umfassend von S. Gołąb und A. Schinzel [61], C. Popa [138], Z. Daróczy [40], P. Javor [81, 82] und S. Wołodzko [183] untersucht worden.[1]) Dabei ergab sich, daß es sowohl Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ dieser Funktionalgleichung gibt, die stetig, aber nicht differenzierbar sind, als auch solche, die meßbar, aber nicht stetig sind. Solche Fälle traten bei den in den Kapiteln 7, 8 und 12 diskutierten Funktionalgleichungen nicht auf. Die folgenden Ausführungen verdeutlichen auch eine Tendenz der heutigen Entwicklung bei der Untersuchung von Funktionalgleichungen, die Betrachtung von Lösungen in Strukturen wie Gruppen, Vektorräumen und topologischen Vektorräumen, in denen die Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ als Spezialfälle enthalten sind. Bei dieser Betrachtungsweise bedingen sich algebraische und analytische Arbeitsmethoden.

Wir werden zunächst einige Eigenschaften der Lösungen dieser Funktionalgleichung erarbeiten, die für das folgende immer wieder benötigt werden.

13.1. Einige Eigenschaften der Lösungen von $f(x + yf(x)) = f(x) \cdot f(y)$

In diesem Abschnitt werden Eigenschaften von Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ der Funktionalgleichung

$$\text{13.1.1.} \quad \bigwedge_{x,y \in \mathbf{R}} f(x + yf(x)) = f(x) \cdot f(y)$$

bewiesen. Die Beweise werden so geführt, daß sie sich sofort auf Funktionen $f\colon V \to K$ übertragen lassen, wobei V die Menge der Elemente eines Vektorraumes $\mathfrak{V}$ über einem (kommutativen) Körper $\mathfrak{K}$ (mit der Trägermenge K) ist.

13.1.2. Lemma. *Die einzigen konstanten Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ von* (13.1.1) *sind die Funktionen f_1 und f_2 mit $f_1(x) = 0$ und $f_2(x) = 1$.*

Beweis. Durch Einsetzen überzeugt man sich, daß f_1 und f_2 Lösungen von (13.1.1) sind. Es sei nun f mit $f(x) = c$ eine konstante Lösung von (13.1.1). Dann folgt aus (13.1.1) $c = c^2$, also $c_1 = 0$ und $c_2 = 1$.

[1]) In J. Aczél [5] findet man ebenfalls einige Ausführungen.

13.1.3. Lemma. *Für jede Lösung $f\colon \mathbf{R} \to \mathbf{R}$, die verschieden ist von f_1 mit $f_1(x) = 0$, ist $f(0) = 1$.*

Beweis. Es sei x_0 eine reelle Zahl mit $f(x_0) \neq 0$. Wir setzen in (13.1.1) $x = x_0$ und $y = 0$ und erhalten $f(x_0) = f(x_0) \cdot f(0)$, also $f(0) = 1$.

13.1.4. Lemma. *Für jede Lösung $f\colon \mathbf{R} \to \mathbf{R}$ von (13.1.1) gilt: Wenn zwei reelle Zahlen x_1 und x_2 mit $x_1 \neq x_2$ und $f(x_1) = f(x_2) \neq 0$ existieren, dann ist $x_2 - x_1$ eine Periode von f.*

Beweis. Es ist zu zeigen: Für x_1 und x_2 mit den genannten Eigenschaften gilt $f(x + x_2 - x_1) = f(x)$ für alle $x \in \mathbf{R}$. Aus (13.1.1) mit $y = \dfrac{x - x_1}{f(x_1)}$ unter Berücksichtigung, daß $f(x_1) = f(x_2)$ ist, ergibt sich

$$\begin{aligned} f(x + x_2 - x_1) &= f(x_2 + x - x_1) = f\left(x_2 + \frac{x - x_1}{f(x_1)} f(x_1)\right) \\ &= f\big(x_2 + yf(x_1)\big) = f\big(x_2 + yf(x_2)\big) \\ &= f(x_2) \cdot f(y) = f(x_1) \cdot f(y) \\ &= f\big(x_1 + yf(x_1)\big) = f\left(x_1 + \frac{x - x_1}{f(x_1)} f(x_1)\right) \\ &= f(x_1 + x - x_1) = f(x). \end{aligned}$$

13.1.5. Lemma. *Für jede Lösung $f\colon \mathbf{R} \to \mathbf{R}$ mit $f \neq f_1$ von (13.1.1) gilt: $p \in \mathbf{R}$ ist Periode von f genau dann, wenn $f(p) = 1$ ist.*

Beweis. Es sei p eine Periode von f. Dann gilt $f(x + p) = f(x)$ für alle $x \in \mathbf{R}$. Für $x = 0$ ergibt sich $f(p) = f(0)$, also wegen Lemma 13.1.3 $f(p) = 1$. Es sei nun a eine reelle Zahl mit $f(a) = 1$. Dann gilt wegen (13.1.3) für alle $x \in \mathbf{R}$

$$f(x + a) = f(a + x) = f\big(a + xf(a)\big) = f(a) \cdot f(x) = f(x),$$

also ist a eine Periode von f.

13.2. Reellwertige Lösungen von $f(x + yf(x)) = f(x) \cdot f(y)$

13.2.1. Satz. *Die Funktionen $f\colon \mathbf{R} \to \mathbf{R}$ mit $f(x) = 1 + mx$ ($m \in \mathbf{R}^*$) sind die einzigen nichtkonstanten differenzierbaren Lösungen von (13.1.1).*

Beweis (siehe S. Gołąb und A. Schinzel [61]). Die Funktionen f mit $f(x) = 1 + mx$ sind differenzierbar und erfüllen (13.1.1), denn es ist

$$\begin{aligned} f\big(x + yf(x)\big) &= 1 + m\big(x + y(1 + mx)\big) = 1 + mx + my + m^2xy \\ &= (1 + mx) \cdot (1 + my) = f(x) \cdot f(y). \end{aligned}$$

Es sei nun $f\colon \mathbf{R} \to \mathbf{R}$ eine nichtkonstante differenzierbare Lösung von (13.1.1). Wir differenzieren f in (13.1.1) nach x und erhalten

$$f'\big(x + yf(x)\big) \cdot \big(1 + yf'(x)\big) = f'(x) \cdot f(y).$$

Für $x = 0$ und $f'(0) = m$ ergibt sich $f'\big(yf(0)\big)\,(1 + my) = mf(y)$. Wegen Lemma 13.1.3 ergibt sich $f'(y)\,(1 + my) = mf(y)$, und für $y \neq -\frac{1}{m}$ und $f(y) \neq 0$ erhält man die Differentialgleichung $\frac{f'(y)}{f(y)} = \frac{m}{1 + my}$, deren allgemeine Lösung f mit $f(y) = c(1 + my)$ ist. Für $y = 0$ ergibt sich $f(0) = c$, also wegen Lemma 13.1.3 $c = 1$ und somit $f(x) = 1 + mx$ für $x \neq -\frac{1}{m}$. Wegen der Differenzierbarkeitsforderung an f gilt $f\left(-\frac{1}{m}\right) = 0$.

Bevor wir die stetigen Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ von (13.1.1) angeben, beweisen wir einige Eigenschaften stetiger Lösungen von (13.1.1).

13.2.2. Lemma. *Stetige Lösungen $f\colon \mathbf{R} \to \mathbf{R}$ von (13.1.1), die periodisch sind, sind konstant.*

Beweis. Es sei $f\colon \mathbf{R} \to \mathbf{R}$ eine stetige Funktion, die periodisch ist. Dann gibt es Zahlen t_1 und t_2 mit $t_1 \neq t_2$ und $f(t_1) = f(t_2)$. In dem Intervall $[\![t_1, t_2]\!]$ findet man nun Zahlen x_1' und x_2' mit $x_1' \neq x_2'$ und $f(x_1') = f(x_2')$ derart, daß $x_2' - x_1'$ beliebig klein ausfällt. Ist f außerdem noch eine von f_1 verschiedene Lösung von (13.1.1), dann besitzt f wegen Lemma 13.1.4 beliebig kleine Perioden. Daher gibt es für jedes $x_0 \in \mathbf{R}$ eine Zahlenfolge (x_k), deren Glieder aus der Menge der Perioden von f sind, mit $\lim\limits_{k\to\infty} x_k = x_0$. Wegen $f(x_k) = 1$ (siehe Lemma 13.1.5) gilt wegen der Stetigkeit von f auch $f(x_0) = 1$, womit sich f als konstant erweist.

13.2.3. Lemma. *Für jede stetige nichtkonstante Lösung $f\colon \mathbf{R} \to \mathbf{R}$ von (13.1.1) gilt: Ist $x \neq 0$ und $f(x) \neq 0$, dann ist $\frac{x}{f(x) - 1}$ konstant.*

Beweis. Angenommen, es gibt Zahlen x, y mit $x \neq y$ und $xyf(x)\,f(y) \neq 0$, für die $\frac{x}{f(x) - 1} \neq \frac{y}{f(y) - 1}$ gilt. Dann ist $x + yf(x) \neq y + xf(y)$ und

$$f\big(x + yf(x)\big) = f(x) \cdot f(y) = f(y) \cdot f(x) = f\big(y + xf(y)\big) \neq 0.$$

Also ist f wegen Lemma 13.1.4 periodisch und wegen Lemma 13.2.2 konstant, was im Widerspruch zur Voraussetzung steht.

Nun können wir einen Überblick über alle stetigen Lösungen von (13.1.1) geben.

13.2.4. Satz. *Die Funktionen $f\colon \mathbf{R} \to \mathbf{R}$ mit*

a) $f(x) = 0$,

b) $f(x) = 1 + mx$ $(m \in \mathbf{R})$,

c) $f(x) = \begin{cases} 1 - \dfrac{x}{a} & \text{für} \quad x \in]\!]-\infty, a]\!], \\ 0 & \text{für} \quad x \in [\![a, \infty[\![, \end{cases}$ $(a \in \mathbf{R}_+{}^*)$,

$$\text{d) } f(x) = \begin{cases} 0 & \text{für} \quad x \in]\!]-\infty, b]\!], \\ 1 - \dfrac{x}{b} & \text{für} \quad x \in [\![b, \infty[\![, \end{cases} \qquad (b \in \mathbf{R} \setminus \mathbf{R}_+)$$

sind die einzigen stetigen Lösungen von (13.1.1).

Beweis (siehe S. GOŁĄB und A. SCHINZEL [61]). Man überzeugt sich durch Einsetzen, daß die genannten stetigen Funktionen Lösungen von (13.1.1) sind. Es sei f nun eine stetige nichtkonstante Lösung von (13.1.1) und $A = \{x: f(x) \neq 0\}$. Für alle $x \in A$ mit $x \neq 0$ gilt wegen Lemma 13.2.3 $\dfrac{x}{f(x)-1} = c$, also $f(x) = 1 + \dfrac{x}{c}$. Unter Berücksichtigung von Lemma 13.1.3 und der Definition der Menge A ergibt sich

$$f(x) = \begin{cases} 1 + \dfrac{x}{c} & \text{für} \quad x \in A, \\ 0 & \text{für} \quad x \notin A, \end{cases}$$

was wegen der Stetigkeitsforderung an f nur auf die unter b), c) und d) genannten Funktionen führt.

13.2.5. Bemerkung. Während die unter a) und b) in Satz 13.2.4 genannten Funktionen auch differenzierbar sind, sind die unter c) und d) genannten Funktionen an genau einer Stelle nicht differenzierbar.

S. GOŁĄB und A. SCHINZEL geben in [61] auch meßbare Lösungen von (13.1.1) an, die nicht stetig sind. Sie können aber keinen Überblick über alle meßbaren Lösungen von (13.1.1) geben. Dies gelingt C. POPA [138], indem er einen algebraischen Aspekt in die Untersuchung einbringt. Er definiert eine Operation $\circ: \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ durch $x \circ y := x + yf(x)$, wobei $f: \mathbf{R} \to \mathbf{R}$ eine Lösung von (13.1.1) ist. Nun ergeben sich folgende Aussagen, deren Beweis wir dem Leser überlassen:

13.2.6. Lemma. $(\mathbf{R}, \circ)$ *ist eine Halbgruppe, in der die Zahl* 0 *ein rechts-neutrales Element ist.*

13.2.7. Lemma. *Ist* $f \neq f_1$ *eine Lösung von* (13.1.1), *dann ist* $(A, \circ)$ *eine Gruppe.*

13.2.8. Lemma. *Es sei* $f \mid A$ *die Einschränkung*[1]) *von* f *auf die Menge* A. *Dann ist* $f \mid A$ *ein Homomorphismus der Gruppe* $(A, \circ)$ *auf die Gruppe* $\mathfrak{G} = (f(A), \cdot)$. *Insbesondere ist die Menge* P *der Perioden von* $f \mid A$ *der Kern des Homomorphismus* $f \mid A$.

Wegen Lemma 13.2.8 ist es für die Bestimmung der meßbaren Lösungen von (13.1.1) hinreichend, die meßbaren Mengen reeller Zahlen zu bestimmen, die bezüglich der Multiplikation eine Gruppe bilden. C. POPA bewies in [138]

13.2.9. Satz. *Ist* $(G, \cdot)$ *eine multiplikative Gruppe reeller Zahlen und* G *meßbar, dann ist* $G = \mathbf{R}^*$ *oder* $G = \mathbf{R}_+^*$, *oder das Maß von* G *ist Null.*

Auf einen Beweis dieses Satzes verzichten wir hier.

[1]) Zur Definition siehe z. B. MfL Bd. 4, 1.3.

Ist $f(A) = \mathbf{R}^*$, dann ergeben sich die nichtkonstanten differenzierbaren Lösungen von (13.1.1). Ist $f(A) = \mathbf{R}_+^*$, dann ergeben sich die stetigen Lösungen von (13.1.1), die nicht überall differenzierbar sind. Ist das Maß von $f(A)$ Null, dann erhält man meßbare Lösungen von (13.1.1), die unstetig oder konstant sind. Daher ergibt sich

13.2.10. Satz. *Ist $f: \mathbf{R} \to \mathbf{R}$ eine meßbare Lösung von* (13.1.1), *dann ist f entweder stetig oder f ist fast überall Null.*

13.2.11. Beispiel. Die Dirichletfunktion f mit $f(x) = \begin{cases} 1 & \text{für } x \in \mathbf{Q}, \\ 0 & \text{für } x \in \mathbf{R} \setminus \mathbf{Q} \end{cases}$ ist eine Lösung von (13.1.1), denn

x	y	$f(x)$	$f(y)$	$f(x) \cdot f(y)$	$f(x + yf(x))$
rat.	rat.	1	1	1	1
rat.	irr.	1	0	0	0
irr.	rat.	0	1	0	0
irr.	irr.	0	0	0	0

Da $A = \mathbf{Q}$, das Maß von $\mathbf{Q}$ gleich Null und $f \mid A$ konstant ist, ist die Dirichletfunktion eine meßbare Funktion, die offensichtlich nicht stetig ist.

Wir geben nun ein Beispiel für eine unbeschränkte und nicht meßbare Lösung $f: \mathbf{R} \to \mathbf{R}$ von (13.1.1) an (siehe auch J. Aczél [5]). Dabei benutzen wir einige Eigenschaften von unstetigen Lösungen $g: \mathbf{R} \to \mathbf{R}$ der Cauchyschen Grundgleichung $\bigwedge_{x,y \in \mathbf{R}} g(x + y) = g(x) + g(y)$, die in 7.3. behandelt wurden. Es sei $\mathbf{R}_B$ eine Basis der reellen Zahlen. Für jedes $x \in \mathbf{R}_B$ wählen wir $g(x)$ rational, wobei nicht alle Werte Null sein sollen. Dann sind alle Funktionswerte von g rational (siehe 7.3.). Außerdem ist $g(rx) = rg(x)$ für alle $r \in \mathbf{Q}$ und alle $x \in \mathbf{R}$ (siehe Bemerkung 7.1.5). Daher gilt auch $g\big(g(x)\, y\big) = g(x) \cdot g(y)$. Die Funktion $f: \mathbf{R} \to \mathbf{R}$ mit $f(x) = 1 + g(x)$ ist wegen der entsprechenden Eigenschaften der Funktion g über keinem Intervall beschränkt und über keinem Intervall meßbar. Wegen

$$\begin{aligned} f\big(x + yf(x)\big) &= 1 + g\big(x + y\big(1 + g(x)\big)\big) = 1 + g\big(x + y + g(x)\, y\big) \\ &= 1 + g(x) + g(y) + g\big(g(x)\, y\big) = 1 + g(x) + g(y) + g(x) \cdot g(y) \\ &= \big(1 + g(x)\big) \cdot \big(1 + g(y)\big) = f(x) \cdot f(y) \end{aligned}$$

ist f eine Lösung von (13.1.1).

Wir gehen nun dazu über, Lösungen von $f\big(x + yf(x)\big) = f(x) \cdot f(y)$ in allgemeineren Strukturen zu untersuchen.

13.3. Die allgemeine Lösung von $f(x + yf(x)) = f(x) \cdot f(y)$

Das von C. Popa [138] benutzte Konzept griff P. Javor [81] auf. Es gelang ihm, ein notwendiges und hinreichendes Kriterium für die Existenz von Lösungen $f: V \to K$ von

13.3.1. $\bigwedge_{x,y \in V} f\big(x + yf(x)\big) = f(x) \cdot f(y)$

anzugeben, wobei $\mathfrak{V} = (V, K)$ ein Vektorraum über einem (kommutativen) Körper $\mathfrak{K}$ mit der Trägermenge K ist. Es gilt

13.3.2. Satz. *Wenn eine Funktion $f\colon V \to K$ eine nichtkonstante Lösung von $f\big(x + yf(x)\big) = f(x) \cdot f(y)$ für $x, y \in V$ ist, dann existieren eine Untergruppe $(N, \circ) = (N, +)$ von $(V, +)$, eine Untergruppe $(M, \cdot)$ von $(K\backslash\{0\}, \cdot)$ und eine Funktion $w\colon M \to A$ $(A = \{x\colon f(x) \neq 0\})$ mit den Eigenschaften*

a) $MN = N$[1]).

b) $w(m_1 m_2) - w(m_1) - m_1 w(m_2) \in N$ *für alle* $m_1, m_2 \in M$.

c) *Für alle* $m \in M$ *gilt: Wenn* $w(m) \in N$, *so* $m = 1$.

Beweis. Es sei $f\colon V \to K$ eine nichtkonstante Lösung von (13.3.1). Als Menge N wählen wir die Menge $P = \{x\colon f(x) = 1\}$. $(P, \circ)$ erweist sich als Gruppe. Wegen $x \circ y = x + yf(x) = x + y \cdot 1 = x + y$ gilt $(P, \circ) = (P, +)$. Als Menge M wählen wir $f(A)$. Nun ist $\big(f(A), \cdot\big)$ eine Gruppe (siehe Lemma 13.2.8). Als $w\colon f(A) \to A$ wählen wir eine Funktion, für die $f\big(w(m)\big) = m$ für alle $m \in f(A)$ gilt. Wir zeigen jetzt, daß $f(A)\,P = P$, also die Eigenschaft a) gilt. Wegen Lemma 13.2.8 ist $(P, +)$ ein Normalteiler von $(A, \circ)$. Daher gilt $P \circ x = x \circ P$ für alle $x \in A$. Für alle $p \in P$ und für alle $x \in A$ gilt $p \circ x = p + xf(p) = p + x \cdot 1 = p + x$. Also ist $P \circ x = P + x$. Weiterhin ist $x \circ P = x + Pf(x)$, denn für alle $p \in P$ und für alle $x \in A$ ist $x \circ p = x + pf(x)$. Damit gilt $P + x = x + Pf(x)$, also $P + x = x + Pf(A)$ für alle $x \in A$. Für $x = 0$ ergibt sich die Behauptung.

Wir zeigen jetzt Eigenschaft b). Es sei w eine Funktion mit $f\big(w(m)\big) = m$ für alle $m \in f(A)$. Zu zeigen ist, daß es für alle $m_1, m_2 \in f(A)$ ein $p \in P$ gibt mit $p = w(m_1 m_2) - w(m_1) - m_1 w(m_2)$. Es sei $w(m_1) = x_1$ und $w(m_2) = x_2$ mit $x_1, x_2 \in A \setminus P$. Da $(P, \circ)$ ein Normalteiler von $(A, \circ)$ ist, existiert ein $p \in P$ mit $w(m_1 m_2) = p \circ (x_1 \circ x_2)$. Wegen $(P, \circ) = (P, +)$ gilt somit auch $w(m_1 m_2) = p + (x_1 \circ x_2)$. Nach Definition von $\circ$ ergibt sich $w(m_1 m_2) = p + x_1 + x_2 f(x_1)$, also

$$w(m_1 m_2) = p + w(m_1) + w(m_2)\, f\big(w(m_1)\big) = p + w(m_1) + w(m_2)\, m_1,$$

was der Behauptung entspricht.

Es sei nun $w(m) \in P$. Dann gilt einerseits $f\big(w(m)\big) = m$ und andererseits $f\big(w(m)\big) = 1$, also $m = 1$. Das ist Eigenschaft c).

Damit ist Satz 13.3.2 bewiesen. Es sei nun mit $\mathfrak{V} = (V, K)$ ein Vektorraum über einem (kommutativen) Körper $\mathfrak{K}$ mit der Trägermenge K gegeben. Dann gilt

13.3.3. Satz. *Sind eine Untergruppe $(N, +)$ von $(V, +)$, eine Untergruppe $(M, \cdot)$ von $(K \setminus \{0\}, \cdot)$ und eine Funktion $w\colon M \to V$ gegeben, für die die folgenden Bedingungen erfüllt sind:*

a) $MN = N$,

b) $w(m_1 m_2) - w(m_1) - m_1 w(m_2) \in N$ *für alle* $m_1, m_2 \in M$,

c) *Für alle* $m \in M$ *gilt: Wenn* $w(m) \in N$, *so* $m = 1$,

[1]) $MN := \{m \cdot n\colon m \in M \wedge n \in N\}$. Dabei ist $\cdot$ die in jedem Vektorraum definierte Verknüpfung zwischen den Vektoren und den Körperelementen.

dann ist die Funktion $f\colon V \to K$ *mit*

$$f(x) := \begin{cases} m & \text{für } x \in N + w(m), \\ 0 & \text{sonst} \end{cases}$$

eine Lösung von $f(x + yf(x)) = f(x) \cdot f(y)$ *für alle* $x, y \in V$.

Beweis. Zunächst zeigen wir, daß f korrekt definiert ist, d. h., daß für jedes $x \in V$ der Funktionswert $f(x)$ eindeutig bestimmt ist. Dafür genügt es zu zeigen, daß $x \in V$ nicht in zwei verschiedenen Rechtsnebenklassen bezüglich $(N, +)$ liegen kann. Wir zeigen: Wenn $n_1 + w(m_1) = n_2 + w(m_2)$, dann $m_1 = m_2$.

Es sei $n_1 + w(m_1) = n_2 + w(m_2)$. Wir wählen $m_3 \in M$ so, daß $m_2 = m_1 m_3$ ist. Wegen b) ist dann $w(m_1 m_3) - w(m_1) - m_1 w(m_3) \in N$, also $w(m_2) - w(m_1) - m_1 w(m_3) \in N$. Daher ist auch $n_1 - n_2 - m_1 w(m_3) \in N$, was $m_1 w(m_3) \in N$ nach sich zieht. Wegen a) ist $w(m_3) \in N$, was wegen c) $m_3 = 1$ bedeutet. Daher ist $m_2 = m_1$.

Nun zeigen wir, daß die in Satz 13.3.3 definierte Funktion f eine Lösung von (13.3.1) ist. Es seien $x_1, x_2 \in V$. Dann sind folgende zwei Fälle möglich:

(1) $\quad f(x_1) = 0.$

Dann ist $x_1 + x_2 f(x_1) = x_1$ und $f(x_1 + x_2 f(x_1)) = f(x_1) = 0$ und $f(x_1) \cdot f(x_2) = 0$. Also ist (13.3.1) erfüllt, unabhängig von $f(x_2)$.

(2) $\quad f(x_1) \neq 0.$

Dann ist $f(x_1) = m_1$ mit $m_1 \in M$, d. h. $x_1 \in N + w(m_1)$, also gibt es ein $n_1 \in N$ mit $x_1 = n_1 + w(m_1)$.

Es ergeben sich die beiden Fälle $\alpha)$ $f(x_2) \neq 0$ und $\beta)$ $f(x_2) = 0$.

$\alpha)$ Ist $f(x_2) \neq 0$, dann gibt es entsprechend $n_2 \in N$ und $m_2 \in M$ mit $x_2 = n_2 + w(m_2)$. Nun gilt

$$\begin{aligned} x_1 + x_2 f(x_1) &= n_1 + w(m_1) + (n_2 + w(m_2))\, m_1 \\ &= n_1 + w(m_1) + n_2 m_1 + m_1 w(m_2). \end{aligned}$$

Wegen a) existiert ein $n_3 \in N$ mit $n_2 m_1 = n_3$. Ist nun n_4 durch $n_4 := n_1 + n_3$ definiert, so ergibt sich $x_1 + x_2 f(x_1) = n_4 + w(m_1) + m_1 w(m_2)$. Wegen b) existiert ein $n_5 \in N$ mit $n_5 = w(m_1 m_2) - w(m_1) - m_1 w(m_2)$. Also ergibt sich nun $x_1 + x_2 f(x_1) \in N + w(m_1 m_2)$. Wegen der Definition von f ist daher

$$f(x_1 + x_2 f(x_1)) = m_1 m_2 = f(x_1) \cdot f(x_2),$$

und (13.3.1) ist erfüllt.

$\beta)$ Nun ist zu zeigen, daß $f(x_1 + x_2 f(x_1)) = 0$ ist. Wir nehmen an, daß $f(x_1 + x_2 f(x_1)) \neq 0$ ist. Dann existieren ein $m \in M$ und ein $n \in N$ mit $x_1 + x_2 f(x_1) = n + w(m)$. Wir wählen $m_2 \in M$ derart, daß $m = m_1 m_2$ ist mit $m_1 = f(x_1)$. Dann ist $x_1 + x_2 f(x_1) = n + w(m_1 m_2)$, also $x_1 + x_2 m_1 = n + w(m_1 m_2)$, also $x_2 m_1 = n - n_1 + w(m_1 m_2) - w(m_1)$. Aus b) ergibt sich, daß ein $n_6 \in N$ existiert mit $x_2 m_1 = n_6 + m_1 w(m_2)$ und somit $x_2 = \dfrac{n_6}{m_1} + w(m_2)$. Wegen a) ist $x_2 = n_7 + w(m_2)$ und $n_7 \in N$, und wegen

der Definition von f ist $f(x_2) = m_2$. Wegen $m_2 \in M$ ist $m_2 \neq 0$, und es ergibt sich ein Widerspruch.

Damit ist Satz 13.3.3 vollständig bewiesen.

Durch die Sätze 13.3.2 und 13.3.3 haben wir einen vollständigen Überblick über die Lösungen $f\colon V \to K$ von (13.3.1).

13.3.4. Bemerkung. Etwa gleichzeitig und unabhängig von P. Javor erhielt S. Wołodzko [183] auf etwas andere Weise ein den Sätzen 13.3.2 und 13.3.3 entsprechendes Ergebnis.

13.4. Die stetigen Lösungen von $f(x + yf(x)) = f(x) \cdot f(y)$ für einen Hilbertraum

Z. Daróczy [40] untersuchte die stetigen Lösungen $f\colon H \to \mathbf{R}$ von

13.4.1. $\bigwedge_{x,y \in H} f(x + yf(x)) = f(x) \cdot f(y)$,

wobei H ein Hilbertraum, also ein vollständiger Vektorraum mit einem Skalarprodukt über dem Körper der reellen Zahlen ist. Ist $x, y \in H$, dann bezeichnet $\|x\|$ die Norm von x und (x, y) das Skalarprodukt von x und y.[1]) Zunächst beweisen wir das

13.4.2. Lemma. *Ist $f\colon H \to \mathbf{R}$ eine nichtkonstante stetige Lösung von $f(x + yf(x)) = f(x) \cdot f(y)$ für alle $x, y \in H$, dann ist $P = \{x\colon f(x) = 1\}$ ein vollständiger Unterraum von H.*

Beweis. Zu zeigen ist a) die topologische Abgeschlossenheit von P und b) die algebraische Abgeschlossenheit von P.

Zu a). Wegen der Stetigkeit von f gilt: Ist $x_k \in P$ und $\lim_{k\to\infty} x_k = x_0$, so ist $\lim_{k\to\infty} f(x_k) = f(x_0)$. Wegen $f(x_k) = 1$ für $x_k \in P$ ist auch $\lim_{k\to\infty} f(x_k) = 1$, und somit ist $f(x_0) = 1$. Das heißt aber $x_0 \in P$.

Zu b). Da $(P, +)$ eine Gruppe ist (siehe Beweis von Satz 13.3.2) und so mit $x, y \in P$ auch $x - y \in P$ ist, bleibt nur zu zeigen, daß $sp \in P$ für alle $s \in \mathbf{R}$ und für alle $p \in P$ gilt.

Für $p = 0$ ist die Behauptung trivial. Es sei nun $p \neq 0$ ein Element aus P. Wir betrachten die Funktion $c_p\colon \mathbf{R} \to \mathbf{R}$ mit $c_p(s) := f(sp)$. Da f stetig ist, ist auch c_p stetig. Für $s, t \in \mathbf{R}$ gilt nun

$$c_p(s + tc_p(s)) = f(sp + tpf(sp)) = f(sp) \cdot f(tp) = c_p(s) \cdot c_p(t),$$

[1]) Eine ausführliche Darstellung der Eigenschaften von Hilberträumen findet man z. B. in H. v. Mangoldt und K. Knopp [112], aus dem die hier benutzte Terminologie stammt, wobei insbesondere für die Menge der Elemente eines Vektorraumes und den Vektorraum selbst keine unterschiedlichen Symbole benutzt werden. Einen Spezialfall eines Hilbertraumes findet man auch in MfL Bd. 3, Kap. 7, angedeutet.

da f eine Lösung von (13.1.4) ist. Das heißt, daß c_p eine stetige Lösung von (13.1.1) ist. Weiterhin gilt $c_p(0) = f(0p) = f(0) = f(0 + p) = f(p) = c_p(1) = 1$. Wegen Lemma 13.1.4 ist 1 eine Periode von c_p. Wegen Lemma 13.2.2 ist c_p daher konstant. Also ist $c_p(s) = 1$ für alle $s \in \mathbf{R}$ und damit $f(sp) = 1$, was zu zeigen war.

13.4.3. Satz. *Ist $f\colon H \to \mathbf{R}$ eine nichtkonstante stetige Lösung von*

$$\bigwedge_{x,y \in H} f\big(x + yf(x)\big) = f(x) \cdot f(y),$$

dann gilt für f entweder

$$f(x) = 1 + (x, a) \quad \text{oder} \quad f(x) = \begin{cases} 0 & \text{für} \quad 1 + (x, a) \leqq 0, \\ 1 + (x, a) & \text{für} \quad 1 + (x, a) > 0, \end{cases}$$

wobei a ein beliebiges, aber festes Element aus H ist.

Beweis (siehe Z. Daróczy [40]). Wir zeigen zuerst, daß f mit $f(x) = 1 + (x, a)$ eine Lösung der Funktionalgleichung (13.4.1) ist. Durch Fallunterscheidung läßt sich entsprechend die andere Möglichkeit zeigen. Es ist

$$\begin{aligned} f\big(x + yf(x)\big) &= 1 + \big(x + yf(x), a\big) = 1 + (x, a) + f(x) \cdot (y, a) \\ &= 1 + (x, a) + \big(1 + (x, a)\big) \cdot (y, a) \\ &= 1 + (x, a) + (y, a) + (x, a) \cdot (y, a) \\ &= \big(1 + (x, a)\big)\big(1 + (y, a)\big) = f(x) \cdot f(y). \end{aligned}$$

Es sei nun $f\colon H \to \mathbf{R}$ eine nichtkonstante stetige Lösung der Funktionalgleichung. Es sei $P^\perp$ das orthogonale Komplement zu P, d. h. die Menge aller zu P orthogonalen Vektoren aus H. Jedes $x \in H$ läßt sich dann auf eindeutige Weise als $x = p + x'$ mit $p \in P$ und $x' \in P^\perp$ darstellen (siehe H. v. Mangoldt und K. Knopp [112], Nr. 119, Satz 2). Dann gilt

$$f(x) = f(p + x') = f\big(p + x'f(p)\big) = f(p) \cdot f(x') = f(x').$$

Da f eine nichtkonstante stetige Lösung ist, ist f auch eine nichtkonstante stetige Lösung für $x' \in P^\perp$. Weiterhin ist f nicht periodisch über $P^\perp$, denn ist $p' \in P^\perp$ Periode, dann gilt wegen Lemma 13.1.5 auch $p' \in P$. Wegen $P \cap P^\perp = \{0\}$ ist $p' = 0$ einzige Periode von f über $P^\perp$.

Wir bestimmen nun die nichtkonstanten stetigen Lösungen von

$$\bigwedge_{x,y \in P^\perp} f\big(x + yf(x)\big) = f(x) \cdot f(y).$$

Dazu zerlegen wir $P^\perp$ in zwei Mengen

$$A' = \{x' : x' \in P^\perp \wedge f(x') \neq 0\}$$

und

$$B' = \{x' : x' \in P^\perp \wedge f(x') = 0\}.$$

B' ist eine abgeschlossene Menge, da B' das Urbild der abgeschlossenen Menge $\{0\}$

bei der stetigen Funktion f ist. A' ist als Komplementärmenge von B' eine offene Menge. Für alle $x' \in A'$ mit $x' \neq 0$ ist $\dfrac{x'}{f(x') - 1}$ konstant (siehe Lemme 13.2.3), also ist

13.4.4. $$\frac{x'}{f(x') - 1} = \frac{a}{\|a\|^2},$$

wobei a ein festes Element aus $P^\perp$ ist. Wir zeigen jetzt: Durch Hinzunahme eines Punktes wird A' zu einer abgeschlossenen Menge. Da A' eine offene Menge ist, existiert ein Häufungspunkt x_0' mit $x_0' \in B'$. Dann existiert eine Folge (x_k') mit $x_k' \in A'$ und $\lim\limits_{k\to\infty} x_k' = x_0'$. Wegen der Stetigkeit von f gilt dann $\lim f(x_k') = f(x_0') = 0$, was wegen (13.4.4) $x_0' = -\dfrac{a}{\|a\|^2}$ nach sich zieht. Also ist $A' \overset{k\to\infty}{\cup} \left\{-\dfrac{a}{\|a\|^2}\right\}$ eine abgeschlossene Menge.

Für die Elemente aus A', die vom Nullvektor verschieden sind, ergibt sich aus (13.4.4) $x' = \dfrac{f(x') - 1}{\|a\|^2}\, a$. Damit ist jedes Element x' aus A' mit $x' \neq 0$ ein reelles Vielfaches des Vektors a.

Aufgrund der bisherigen Ergebnisse bestehen für A' nur zwei Möglichkeiten:

$$A_1' = \left\{ma: -\infty < m < -\frac{1}{\|a\|^2}\right\} \cup \left\{ma: -\frac{1}{\|a\|^2} < m < +\infty\right\}$$

und

$$A_2' = \left\{ma: -\frac{1}{\|a\|^2} < m < +\infty\right\}.$$

In jedem der beiden Fälle ergibt sich für $x' \in A'$ $(x' \neq 0)$ aus (13.4.4) $x'\,\|a\|^2 = \big(f(x') - 1\big)\,a$ und wegen $x' = ma$ und $\|a\|^2 = (a, a)$ auch $(a, a)\, ma = \big(f(x') - 1\big)\,a$, also $(ma, a)\, a = \big(f(x') - 1\big)\,a$. Somit ist $(x', a)\, a = \big(f(x') - 1\big)\,a$, woraus sich $(x', a) = f(x') - 1$ ergibt, also $f(x') = 1 + (x', a)$. Wegen $f(0) = 1 = 1 + (0, a)$ ordnet sich $x' = 0$ ein.

Ist $A' = A_1'$, dann ist $A' = P^\perp \setminus \left\{-\dfrac{a}{\|a\|^2}\right\}$ und $f\left(-\dfrac{a}{\|a\|^2}\right) = 0$. Wegen

$$1 + \left(-\frac{a}{\|a\|^2}, a\right) = 1 + \left(-\frac{1}{\|a\|^2}\right) \cdot (a, a)$$
$$= 1 + \left(-\frac{1}{\|a\|^2}\right) \|a\|^2 = 1 - 1 = 0$$

gilt $f(x') = 1 + (x', a)$ für alle $x' \in P^\perp$.

Ist $A' = A_2'$, dann ist $f(x') = 1 + (x', a)$ für $x' \in A'$ und $f(x') = 0$ für $x' \in B'$. Wegen $1 + (x', a) > 0$ genau dann, wenn $1 + (ma, a) > 0$, genau dann, wenn $1 + m(a, a) > 0$, genau dann, wenn $1 + m\,\|a\|^2 > 0$, genau dann, wenn $m > -\dfrac{1}{\|a\|^2}$, ist f auf $P^\perp$ durch

$$f(x') = \begin{cases} 0 & \text{für} \quad 1 + (x', a) \leqq 0, \\ 1 + (x', a) & \text{für} \quad 1 + (x', a) > 0 \end{cases}$$

definiert. Damit sind alle nichtkonstanten stetigen Lösungen auf $P^\perp$ gefunden.

Da sich jedes $x \in H$ eindeutig durch $x = p + x'$ mit $p \in P$ und $x' \in P^{\perp}$ darstellen läßt, gilt für alle $x \in H$ mit $f(x) \neq 0$

$$f(x) = f(p + x') = f(x') = 1 + (x', a)$$
$$= 1 + (x - p, a) = 1 + (x, a) - (p, a).$$

Wegen $p \in P$ und $a \in P^{\perp}$ ist $(p, a) = 0$, woraus sich die Behauptung von Satz 13.4.3 ergibt, denn für alle $x \in H$ mit $f(x) = 0$ ist auch $f(x') = 0$.

13.5. Aufgaben

1. Es sei $(V, +)$ durch

$+$	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

gegeben, und $\mathfrak{K}$ sei der Körper mit genau zwei Elementen. Durch

$K \cdot V \rightarrow$ $\downarrow$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5

ist eine Abbildung von $K \times V$ in V gegeben. $\mathfrak{V} = (V, K)$ ist dann ein Vektorraum über dem Körper $\mathfrak{K}$. Man zeige, daß die Funktionen $f_i\colon V \to K$ $(i = 1, 2, 3, 4, 5)$ mit

$$f_1(x) = \begin{cases} 1 & \text{für } x = 0 \\ 0 & \text{für } x \in \{1, 2, 3, 4, 5\} \end{cases}, \quad f_2(x) = \begin{cases} 1 & \text{für } x \in \{0, 3\} \\ 0 & \text{für } x \in \{1, 2, 4, 5\} \end{cases},$$

$$f_3(x) = \begin{cases} 1 & \text{für } x \in \{0, 2, 4\} \\ 0 & \text{für } x \in \{1, 3, 5\} \end{cases}, \quad f_4(x) = 1, \quad f_5(x) = 0$$

die einzigen Lösungen von $\bigwedge_{x,y \in V} f(x + yf(x)) = f(x) \cdot f(y)$ sind.

2. Man bestätige die Aussagen von Satz 13.2.4 mit Hilfe der Sätze 13.3.2 und 13.3.3.

Literatur

[1] Aczél, J.: The notion of mean values, Norske Vid. Selsk. Forh. **19** (1946), 83–86.
[2] Aczél, J.: A solution of some problems of K. Borsuk and L. Janossy, Acta phys. Acad. Sci. Hung. **4** (1955), 351–362.
[3] Aczél, J.: Vorlesungen über Funktionalgleichungen und ihre Anwendungen, Birkhäuser Verlag, Basel–Stuttgart/VEB Deutscher Verlag der Wissenschaften, Berlin 1961.
[4] Aczél, J.: Quasigroups, nets, and nomograms, Advances in Math. **1** (1965), 383–450.
[5] Aczél, J.: Lectures on functional equations and their applications, Academic Press, New York–London 1966.
[6] Aczél, J.: Conditions for a loop to be a group and for a groupoid to be a semigroup, Amer. math. Monthly **76** (1969), 547–549.
[7] Aczél, J., V. D. Belousov and M. Hosszú: Generalized associativity and bisymmetry on quasigroups, Acta math. Acad. Sci. Hung. **11** (1960), 127–136.
[8] Aczél, J., and P. Erdös: The non-existence of a Hamel-basis and the general solution of Cauchy's functional equation for nonnegative numbers, Publ. math. Debrecen **12** (1965), 259–263.
[9] Aczél, J., und I. Fenyö: Über die Theorie der Mittelwerte, Acta Sci. math. Szeged **11** (1948), 239–245.
[10] Aczél, J., und H. Kiesewetter: Über die Reduktion der Stufe einer Klasse von Funktionalgleichungen, Publ. math. Debrecen **5** (1957/58), 348–363.
[11] Aczél, J., G. Pickert und F. Radó: Nomogramme, Gewebe und Quasigruppen, Mathematica Cluj **2** (25) (1960), 5–24.
[12] Albert, A. A.: Quasigroups I, Trans. Amer. math. Soc. **54** (1943), 507–519.
[13] d'Alembert, J.: Addition au Mémoire sur la courbe que forme une corde tendue mise en vibration, Hist. Acad. Berlin 1750, pp. 355–360.
[14] Alexiewicz, A., et W. Orlicz: Remarques sur l'équation fonctionelle $f(x + y) = f(x) + f(y)$, Fundamenta math. **33** (1945), 314–315.
[15] Andrade, J.: Leçons de mécanique physique, Paris 1897.
[16] Arbib, M. A.: Algebraische Theorie abstrakter Automaten, formaler Sprachen und Halbgruppen, Akademie-Verlag, Berlin 1973 (Übersetzung aus dem Englischen).
[17] Artzy, R.: Cayley-Diagrams of binary systems, Duke math. J. **28** (1961), 491–495.
[18] Artzy, R.: Isotopy and parastrophy of quasigroups, Proc. Amer. math. Soc. **14** (1963), 429–431.
[19] Barner, M.: Differential- und Integralrechnung I, de Gruyter, Berlin 1961.
[20] Белоусов, В. Д.: Системы квазигрупп с обобщенными тождествами, Успехи матем. наук **20** (1965), № 1 (121), 75–146.
[21] Белоусов, В. Д.: Основй теории квазигрупп и луп, Наука, Москва 1967.
[22] Белоусов, В. Д.: Алгебраические сети и квазигруппы, Штиинца, Кишинев 1971.
[23] Bittner, R., D. Ilse, S. Kubicek und W. Tietz: Kompendium der Mathematik, Volk und Wissen Volkseigener Verlag, Berlin 1969.
[24] Blaschke, W., und G. Bol: Geometrie der Gewebe, Springer-Verlag, Berlin 1938.
[25] Boggs, H., and G. Y. Rainich: Note on group postulates, Bull. Amer. math. Soc. **43** (1937), 81–84.

[26] Bol, G.: Gewebe und Gruppen, Math. Ann. **114** (1937), 414–431.
[27] Borůvka, O.: Grundlagen der Gruppoid- und Gruppentheorie, VEB Deutscher Verlag der Wissenschaften, Berlin 1960 (engl. Übersetzung Birkhäuser Verlag, Basel–Stuttgart 1975).
[28] Brandt, H.: Verallgemeinerung des Gruppenbegriffs, Math. Ann. **96** (1927), 360–366.
[29] Bruck, R. H.: Some results in the theory of quasigroups, Trans. Amer. math. Soc. **55** (1944), 19–52.
[30] Bruck, R. H.: Contributions to the theory of loops, Trans. Amer. math. Soc. **60** (1946), 245–354.
[31] Bruck, R. H.: Finite nets. I. Numerical invariants, Canad. J. Math. **3** (1951), 94–107.
[32] Bruck, R. H.: A survey of binary systems, Springer-Verlag, Berlin–Göttingen–Heidelberg 1958.
[33] Bruck, R. H.: What ist a loop? Studies in modern algebra, vol. 2, Prentice-Hall, Englewood Cliffs 1963, pp. 59–99.
[34] Cauchy, A. L.: Cours d'analyse de l'Ecole Polytechnique, 1. Analyse algébrique, V. Paris 1821.
[35] Clifford, A. H., and G. B. Preston: The algebraic theory of semigroups, vol. 1, Amer. math. Soc. Surveys No. 7, 1964.
[36] van der Corput, J. G.: Goniometrische Functies gekarakteriseerd door een Funktionaalbetrekking, Euclides **17** (1940), 194–201.
[37] Coxeter, H. S. M., and W. O. J. Moser: Generators and relations for discrete groups, Springer-Verlag, Berlin 1965.
[38] Čulik, K.: Poznámky k teorii operaci, Časopis pro pěstování matematiky **88** (1958), 473–474.
[39] Darboux, M. G.: Sur la composition des forces en statique, Bull. Sci. math. (1) **9** (1875), 281–288.
[40] Daróczy, Z.: Az $f(x + yf(x)) = f(x)\,f(y)$ függvenyegyenlet folytonos megaldasairol Hilbert-terekben, Mat. Lapok **17** (1966), 339–343.
[41] Daróczy, Z., und L. Losonczi: Über die Erweiterung der auf einer Punktmenge additiven Funktionen, Publ. math. Debrecen **14** (1967), 239–245.
[42] Делоне, Б. Н., и Д. А. Райков: Аналитическая геометрия, том I, Москва–Ленинград 1948.
[43] Dembowski, P.: Finite geometries, Springer-Verlag, Berlin–Heidelberg–New York 1968.
[44] Dénes, J., and A. D. Keedwell: Latin squares and their applications, Akadémiai Kiadó, Budapest 1974.
[45] Dieudonné, J.: Grundzüge der modernen Analysis, 2. Aufl., Friedr. Vieweg + Sohn, Braunschweig/VEB Deutscher Verlag der Wissenschaften, Berlin 1972 (Übersetzung aus dem Englischen).
[46] Doyen, J., and A. Rosa: A bibliography and survey of Steiner systems, Bolletino U.M.I. (4) **7** (1973), 392–419.
[47] Drobot, V.: Discontinuous solutions of the functional equation $f(x + y) = H(f(x), f(y))$, Funkcialaj Ekvacioj **16** (1973), 165–168.
[48] Engel, A.: Anwendungen der Analysis zur Konstruktion mathematischer Modelle, Der Mathematikunterricht **17** (1971), No. 3, 5–56.
[49] Enzyklopädie der Elementarmathematik, Band V, VEB Deutscher Verlag der Wissenschaften, Berlin 1971 (Übersetzung aus dem Russischen).
[50] Etherington, I. M. H.: Transposed algebras, Proc. Edinb. math. Soc. (2) **7** (1945), 104–121.
[51] Etherington, I. M. H.: Quasigroups and cubic curves, Proc. Edinb. math. Soc. (2) **14** (1965), 273–291.
[52] Evans, T.: Homomorphisms of non-associative systems, J. London math. Soc. **24** (1949), 254–260.
[53] Falconer, E.: Isotopy invariants in quasigroups, Trans. Amer. math. Soc. **151** (1970), 511–526.

[54] Faragó, T.: Contribution to the definition of a group, Publ. math. Debrecen **3** (1953), 133–137.
[55] Fisher, R. A., and F. Yates: The 6 × 6 latin squares, Proc. Camb. phil. Soc. **30** (1934), 492–507.
[56] Frink, O.: Symmetric and self-distributive systems, Amer. math. Monthly **62** (1955), 697–707.
[57] Frolov, M.: Recherches sur les permutations carrées, J. math. Spéc. (3) **4** (1890), 8–11.
[58] Fuchs, L.: Abelian groups, Publishing House of the Hungarian Academy of Sciences, Budapest 1958.
[59] Гайдук, Ю. М.: К вопросу об аналитическом и геометрическом определениях тригонометрических функций, Математика в школе 1953, № 4, 1–7.
[60] Gołąb, S.: Theorie der Quasigruppen, Potsdamer Forschungen, Reihe B, Heft 3, 1974, 137–151.
[61] Gołąb, S., et A. Schinzel: Sur l'équation fonctionelle $f(x + yf(x)) = f(x)\,f(y)$, Publ. math. Debrecen **6** (1959), 113–125.
[62] Grimsehl, E.: Lehrbuch der Physik. 1. Band, B. G. Teubner Verlagsgesellschaft, Leipzig 1957.
[63] Gruhlke, E.: Einführung der Sinus- und Kosinusfunktionen unter Verwendung von Funktionalgleichungen, Diplomarbeit, Humboldt-Universität zu Berlin 1977.
[64] Hamel, G.: Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung $f(x + y) = f(x) + f(y)$, Math. Ann. **60** (1905), 459–462.
[65] Hammel, A.: Verifying the associative property for finite groups, Math. Teacher **61** (1968), 136–139.
[66] Hammer, F. D.: Verifying associativity, Amer. math. Monthly **87** (1980), 409–410.
[67] Hausmann, B. A., and O. Ore: Theory of quasigroups, Amer. J. Math. **59** (1937), 983 to 1004.
[68] Hofmann, K. H.: Topologische Loops, Math. Z. **70** (1958), 13–37, 125–155, 213–230.
[69] Hopf, E.: Über die Funktionalgleichungen der trigonometrischen und hyperbolischen Funktionen, Sitz.-Ber. math.-nat. Abt. bayer. Akad. Wiss. 1945–1946, S. 167–173.
[70] Hosszú, M.: Some functional equations, related with the associative law, Publ. math. Debrecen **3** (1953/54), 205–214.
[71] Hosszú, M.: On the functional equation of transitivity, Acta. Sci. math. Szeged **15** (1953/54), 203–208.
[72] Huntington, E. V.: Simplified definition of a group, Bull. Amer. math. Soc. **8** (1901/02), 296–300.
[73] Ilse, D.: Über funktionale Charakterisierungen der direkten Proportionalität $f(x) = cx$, Math. Schule **9** (1971), 16–37.
[74] Ilse, D.: Bemerkungen zu den Additions- und Multiplikationstheoremen einiger elementarer Funktionen, Math. Schule **10** (1972), 193–205.
[75] Ilse, D.: Über rationale Additionstheoreme und verwandte Funktionalgleichungen, Math. Schule **11** (1973), 392–403.
[76] Ilse, D.: Zur funktionalen Charakterisierung der Winkelfunktionen, Math. Schule **14** (1976), 87–96.
[77] Ilse, D.: Zur Proportionalität von Größen, Math. Schule **15** (1977), 358–359.
[78] Ilse, D., und I. Lehmann: Schulmathematische Potenzen der Theorie der Funktionalgleichungen und der Theorie der Quasigruppen, Wiss. Z. Berlin, Math.-Nat. R. **23** (1974), 463–469.
[79] Ilse, D., und I. Lehmann: Über strukturelle Betrachtungen in der Mathematik und deren Nutzen für den Mathematikunterricht, Wiss. Z. Berlin, Math.-Nat. R. **27** (1978), 639–643.
[80] Ilse, D., und I. Lehmann: Zur Berücksichtigung struktureller Arbeitsweisen im Mathematikunterricht, Wiss. Z. Berlin, Math.-Nat. R. **27** (1978), 645–655.
[81] Javor, P.: On the general solution of the functional equation $f(x + yf(x)) = f(x)\,f(y)$, Aequ. math. **1** (1968), 235–238.

[82] JAVOR, P.: Continuous solutions of the functional equation $f(x + yf(x)) = f(x)\,f(y)$, Proc. Intern. Sympos. on Topology and its Application (Herceg.-Novi 1968), Savez Drustava Mat. Fiz. Astron. Belgrad 1969, pp. 206–209.

[83] KAMKE, E.: Mengenlehre, Walter de Gruyter & Co., Berlin 1947.

[84] KESTELMAN, H.: On the functional equation $f(x + y) = f(x) + f(y)$, Fundamenta math. **34** (1947), 144–147.

[85] KIESEWETTER, H.: Ein Zusammenhang zwischen assoziativen und zyklisch-bisymmetrischen Verknüpfungen, Aequ. math. **4** (1970), 83–88.

[86] KIRKMAN, T. P.: On a problem in combinations, Camb. & Dublin math. J. **2** (1847), 191–204.

[87] KIRSCH, A.: Über die Veranschaulichung einfacher Gruppenhomomorphismen, Der Mathematikunterricht **11** (1965), Heft 1, 54–67.

[88] KIRSCH, A.: Die Funktionalgleichung $f(x + y) = f(x) + f(y)$ als Thema für den Oberstufenunterricht. Math.-phys. Semesterberichte **24** (1977), 172–187.

[89] KNOPP, K.: Theorie und Anwendung der unendlichen Reihen, Springer-Verlag, Berlin–Heidelberg 1947.

[90] КОЛЯГИН, Ю. М.: О функциональных уравнениях, Математика в школе 1959, № 5, 4–8.

[91] KRAFFT, M.: Herleitung der trigonometrischen Funktionen aus ihren Funktionalgleichungen, Deutsche Math. **4** (1939), 194–201.

[92] KRONECKER, L.: Auseinandersetzung einiger Eigenschaften der Klassenzahl idealer complexer Zahlen, Monatsber. königl. preußischen Akad. Wiss. Berlin 1870, S. 881–889.

[93] KUCZMA, M.: Functional Equations in a single variable, PWN, Warschau 1968.

[94] KUREPA, S.: Functional equation $F(x + y)\,F(x - y) = F^2(x) - F^2(y)$ in n-dimensional vectorspace, Monatshefte für Mathematik **64** (1960), 321–329.

[95] KUROŠ, A. G.: Vorlesungen über allgemeine Algebra, B. G. Teubner Verlagsgesellschaft, Leipzig 1964 (Übersetzung aus dem Russischen).

[96] KURSAWE, K.: Stetige monotone Funktionen ohne Differenzierbarkeitsintervalle, Prax. Math. **16** (1974), 281–284.

[97] KURZWEIL, H.: Endliche Gruppen, Springer-Verlag, Berlin–New York 1977.

[98] KUWAGAKI, A.: Sur l'équation fonctionelle $f(x + y) = R(f(x), f(y))$, Mém. Coll. Sci. Univ. Kyoto (A, Math.) **26** (1951), 139–144.

[99] LEHMANN, I.: Axiomatische Charakterisierung abelscher Gruppen mit Hilfe der Umkehroperationen, Wiss. Z. Berlin, Math.-Nat. R. **23** (1974), 471–478.

[100] LEHMANN, I.: Über die Vielfalt der Operationen, die im derzeitigen Mathematikunterricht explizite oder implizite vorkommen, Math. Schule **14** (1976), 463–472.

[101] LEHMANN, I.: Umkehrbarkeit und Umkehroperationen von Operationen, Math. Schule **16** (1978), 71–81.

[102] LEHMANN, I.: Quasigruppen-Identitäten und spezielle abelsche Gruppen, Wiss. Z. Berlin, Math.-Nat. R. **27** (1978), 657–668.

[103] LEHMANN, I.: Die vier Grundrechenoperationen aus ungewohnter Sicht, Math. Schule **17** (1979), 11–15.

[104] LEHMANN, I., und W. SCHULZ: Kapriolen der Null, Math. Schule **19** (1981), 266–268, 430–452; **20** (1982), 29–32.

[105] LENZ, H.: Grundlagen der Elementarmathematik, 3. Aufl., VEB Deutscher Verlag der Wissenschaften, Berlin/Carl Hanser Verlag, München–Wien 1975.

[106] LINDNER, CH. C., and D. STEEDLEY: On the number of conjugates of a quasigroup, Algebra Universalis **5** (1975), 191–196.

[107] LORENZEN, P.: Ein vereinfachtes Axiomensystem für Gruppen, J. reine angew. Math. **182** (1940), 50.

[108] LUGOWSKI, H.: Grundzüge der Universellen Algebra, BSB B. G. Teubner Verlagsgesellschaft, Leipzig 1976.

[109] LUGOWSKI, H., und H. WEINERT: Grundzüge der Algebra, Teil 1: Allgemeine Gruppentheorie, 4. Aufl., B. G. Teubner Verlagsgesellschaft, Leipzig 1968.

[110] Мальцев, А. И.: К общей теорий алгебраических систем, Мат. сборник **35** (1954), 3—20.
[111] Mancill, J. D.: The sine and cosine functions, Math. Mag. **36** (1963), 302—311.
[112] v. Mangoldt, H., und K. Knopp: Einführung in die höhere Mathematik, Band IV, (F. Lösch), 3. Aufl., S. Hirzel Verlag, Leipzig/S. Hirzel, Stuttgart 1979.
[113] Mathematik-Lehrbuch für Klasse 3. Volk und Wissen Volkseigener Verlag, Berlin 1966.
[114] Mathematik-Lehrbuch für die erweiterte Oberschule, Klasse 12, Volk und Wissen Volkseigener Verlag, Berlin 1970.
[115] Mathematik-Lehrbuch für Klasse 9, 8. Aufl., Volk und Wissen Volkseigener Verlag, Berlin 1977.
[116] Mathematik-Lehrbuch für Klasse 8, 7. Aufl., Volk und Wissen Volkseigener Verlag, Berlin 1977.
[117] Meyer zur Capellen, W.: Leitfaden der Nomographie, Springer-Verlag, Berlin—Göttingen—Heidelberg 1953.
[118] Minetti, S.: Sulla equazione funzionale $f(x + y) = f(x) f(y)$, R. C. Accad. Lincei (5) **31** (1921), 12—15, 202—203.
[119] Mitschke, A., and H. Werner: On groupoids representable by vector spaces over finite fields, Arch. Math. **24** (1973), 14—20.
[120] Murdoch, D. C.: Quasi-groups which satisfy certain generalized associative laws, Amer. J. Math. **61** (1939), 509—522.
[121] Murdoch, D. C.: Structure of abelian quasi-groups, Trans. Amer. math. Soc. **49** (1941), 392—409.
[122] Natanson, I. P.: Theorie der Funktionen einer reellen Veränderlichen, 5. Aufl., Akademie-Verlag, Berlin 1981 (Übersetzung aus dem Russischen).
[123] Neagu, M.: Solutia cea mai generala a ecuatiilor functionale de tipul teoremelor de aditiune, Bul. Sti. Tehn. Politehn. Inst. Timisoara, Ser. Mat.-Fiz.-Mec. Teor. Apl. **17** (31) (1972), fasc. 2, 123—129.
[124] Netto, E.: Lehrbuch der Combinatorik, B. G. Teubner Verlagsgesellschaft, Leipzig 1901.
[125] Norden, A. P.: Elementare Einführung in die Lobatschewskische Geometrie, VEB Deutscher Verlag der Wissenschaften, Berlin 1958 (Übersetzung aus dem Russischen).
[126] Новоселов, С. И.: Специальный курс тригонометрии, Сов. Наука. Москва 1957.
[127] Osgood, W. F.: Lehrbuch der Funktionentheorie, 4. Aufl., 1. Band, Teubner, Leipzig—Berlin 1923.
[128] Ostrowski, A.: Über die Funktionalgleichung der Exponentialfunktionen und verwandte Funktionalgleichungen, Jber. dtsch. Math. Ver. **38** (1929), 54—62.
[129] Di Paola, J. W.: When is a totally symmetric loop a group? Amer. math. Monthly **76** (1969), 249—252, 312.
[130] Pentkowski, M. W.: Nomographie, Akademie Verlag, Berlin 1953 (Übersetzung aus dem Russischen).
[131] Perron, O.: Über Additions- und Subtraktionstheoreme, Arch. Math. Phys. **28** (1919/20), 97—100.
[132] Physik-Lehrbuch für Klasse 9, 6. Aufl., Volk und Wissen Volkseigener Verlag, Berlin 1975.
[133] Picard, E.: Leçons sur quelques équations fonctionelles avec des applications à divers problèmes d'analyse et de physique mathématique, Gauthier-Villars, Paris 1928.
[134] Pickert, G.: Projektive Ebenen, Springer-Verlag, Berlin—Göttingen—Heidelberg 1955.
[135] Pickert, G.: Analytische Geometrie, Geest & Portig, Leipzig 1963.
[136] Pickert, G.: Einführung in die endliche Geometrie, Klett-Verlag, Stuttgart 1974.
[137] Poisson, S. D.: Traité de mécanique, Paris, Bachelier, 1811.
[138] Popa, C.: Sur l'équation fonctionelle $f(x + yf(x)) = f(x) f(y)$, Ann. Polon. Math. **17** (1965), 193—198.
[139] Продан, Н. И.: Некоторые вопросы теории группоидов с делением, Вопросы теории квазигрупп и луп, Кишинев 1971, стр. 104—110.

[140] Пухарев, Н. К.: Некоторые свойства регулярных конечных плоскостей Лобачевского, Перм. Гос. Унив. Учен. Зап. Мат. **103** (1963), 61—63.
[141] Пухарев, Н. К.: Об A_n^k-алгебрах и регулярных конечных плоскостях, Сибир. мат. Ж. **6** (1965), 892—899.
[142] Пухарев, Н. К.: Геометрические вопросы некоторых медиальных квазигрупп, Сибир. мат. Ж. **9** (1968), 891—897.
[143] Radó, F.: Ecuatii functionale in legatura cu nomografia, Studii cerc. mat. Cluj **9** (1958), 249—319.
[144] Radó, F.: Equations fonctionelles caractérisant les nomogrammes avec trois échelles rectilignes, Mathematica Cluj **1** (24) (1959), 143—166.
[145] Rådström, H.: Nagra elementära funktionalekvationer och Hilberts femte problem, Nordisk Mat. Tidskr. **3** (1955), 129—147.
[146] Rätz, J.: Begründung und Charakterisierung der reellen Logarithmusfunktionen, Elem. Math. **20** (1965), 122—128.
[147] Reiss, M.: Über eine Steinersche combinatorische Aufgabe, J. reine angew. Math. **56** (1859), 326—344.
[148] Rosenbaum, R. A., et S. L. Segal: A functional equation characterizing the sine, Math. Gaz. **44** (1960), 97—105.
[149] Rossier, P.: Equations fonctionelles et mathématiques supérieures élémentaires, Elem. Math. **1** (1946), 81—87.
[150] Saaty, L.: Modern nonlinear equations, McGraw-Hill Book Co., New York—St. Louis—San Francisco—Toronto—London—Sydney 1967.
[151] Sade, A.: Quasigroupes, Marseille 1950.
[152] Sade, A.: Quasigroupes obéissant à certaines lois, Rev. Fac. Sci. Univ. Istanbul, Sér. A, **22** (1957), 151—184.
[153] Sade, A.: Quasigroupes parastrophiques. Expressions et identités, Math. Nachr. **20** (1—2) (1959), 73—106.
[154] Sade, A.: Quasigroupes demi-symétriques 1. Ann. Soc. Sci. Bruxelles 79, II, Ser. **1**, 1965, 133—143.
[155] Saks, S.: Theory of the integral, Hafner Publishing Company, New York 1937.
[156] Schadach, D. J.: Biomathematik II, Akademie-Verlag Berlin/Pergamon Press, Oxford/Vieweg + Sohn, Braunschweig 1971.
[157] Schönhardt, E.: Über lateinische Quadrate und Unionen, J. reine angew. Math. **163** (1930), 183—230.
[158] Schröder, E.: Lehrbuch der Arithmetik und Algebra für Lehrer und Studierende, Band 1, Teubner, Leipzig 1873.
[159] Schulz, W.: Schulmathematische Untersuchungen zur Behandlung der Exponential- und Logarithmusfunktionen im Mathematikunterricht, Dissertation, Humboldt-Universität zu Berlin, 1978.
[160] Schulz, W.: Ein gruppentheoretisches Kriterium für die Lösbarkeit der Funktionalgleichung $f(x + y) = F(f(x), f(y))$, Beitr. Alg. Geom. **10** (1980), 141—143.
[161] Shaw, J. B.: On parastrophic algebras, Trans. Amer. math. Soc. **16** (1915), 361 to 370.
[162] Siemon, H.: Loops der 3-Gewebe von Verknüpfungstafeln, Beitr. Math. Unterr. 1971, Schroedel, Hannover 1972, S. 220—227.
[163] Sierpiński, W.: Sur one propriété des fonctions de M. Hamel, Fundamenta math. **5** (1924), 334—336.
[164] Stein, S. K.: On the foundations of quasigroups, Trans. Amer. math. Soc. **85** (1957), 228—256.
[165] Steiner, J.: Combinatorische Aufgabe, J. reine angew. Math. **45** (1853), 181—182.
[166] Straszewicz, S.: Sur la trigonométrie de Lobatschewsky, Ann. Polon. Math. **3** (1956—57), 225—239.
[167] Szamkolowicz, L.: On the problem of existence of finite regular planes, Colloqu. Math. **9** (1962), 245—250.

[168] TAYLOR, M. A.: R- and T-groupoids: a generalization of groups, Aequ. math. **11** (1974), 318–319; **12** (1975), 242–248.
[169] THIEL, S.: Eine funktionale Charakterisierung der Sinusfunktion, Wiss. Z. Berlin, Math.-Nat. R. **27** (1978), 669–672.
[170] THOMSEN, G.: Topologische Fragen der Differentialgeometrie, XII. Schnittpunktsätze in ebenen Geweben, Abh. Math. Sem. Univ. Hamburg **7** (1939), 99–106.
[171] THURSTON, H. A.: Certain congruences on quasigroups, Proc. Amer. math. Soc. **3** (1952), 10–12.
[172] TISSERAND, F.: Recueil complémentaire d'exercices sur le calcul infinitésimal, Gauthier-Villars, Paris 1877.
[173] TOYODA, K.: On axioms of mean transformations and automorphic transformations of abelian groups, Tôhoku Math. J. **46** (1940), 239–251.
[174] TRACEWELL, T. N.: A general theory of mediality in groupoids, Doct. diss., Berkeley 1964.
[175] VAUGHAN, H. E.: Characterization of the sine and cosine, Amer. math. Monthly **62** (1955), 707–713.
[176] VIETORIS, L.: Zur Kennzeichnung des Sinus und verwandter Funktionen durch Funktionalgleichungen, J. reine angew. Math. **186** (1944–49), 1–15.
[177] VIETORIS, L.: Vom Grenzwert $\lim_{x \to 0} \frac{\sin x}{x}$, Elem. Math. **12** (1957), 8–10.
[178] WEBER, H.: Die allgemeinen Grundlagen der Galois'schen Gleichungstheorie, Math. Ann. **43** (1893), 521–549.
[179] WIECHMANN, P.: Axiomatisches Arbeiten mit elementaren Funktionen und zweistelligen Operationen, Dissertation, Humboldt-Universität zu Berlin, 1982.
[180] WILSON, E. B.: Note on the function satisfying the functional relation $\varphi(u)\,\varphi(v) = \varphi(u + v)$, Ann. Math. (2) **1** (1899–1900), 47–48.
[181] WILSON, W. H.: On certain related functional equations, Bull. Amer. math. Soc. **26** (1920), 300–312.
[182] WISWESSER, G.: Zur Definition von Funktionen aus ihren Funktionaleigenschaften, Math.-Naturw. Unterr. **28** (1975), 95–98.
[183] WOLODZKO, S.: Solution générale de l'équation fonctionelle $f(x + yf(x)) = f(x)\,f(y)$, Aequ. math. **2** (1968/69), 12–29.
[184] WUSSING, H.: Die Genesis des abstrakten Gruppenbegriffes, VEB Deutscher Verlag der Wissenschaften, Berlin 1969 (engl. Übersetzung MIT Press, Cambridge (Mass.)–London 1984).
[185] YOUNG, G. S.: The linear functional equation, Amer. math. Monthly **65** (1958), 37–38.
[186] ZEITLER, H.: Über Klassen spezieller Blockpläne, Prax. Math. **13** (1971), 225–230, 259–265.
[187] ZEITLER, H.: Konstruktion spezieller Steiner-Tripel-Systeme, Math.-phys. Semesterber. **21** (1974), 206–233.
[188] ZEITLER, H.: Endliche Geometrien, Binärkodes und Versuchspläne, Math.-Naturwiss. Unterr. **29** (1976), 83–95.
[189] ZERMELO, E.: Beweis, daß jede Menge wohlgeordnet werden kann (Aus einem an Herrn HILBERT gerichteten Brief), Math. Ann **59** (1904), 514–516.
[190] ZERMELO, E.: Neuer Beweis für die Möglichkeit einer Wohlordnung, Math. Ann. **65** (1908), 107–128.

Namenverzeichnis

Abel, N. H. 33
Aczél, J. 5, 23, 37, 55, 116, 146, 149, 156, 205, 210, 218, 240, 241, 245, 250, 270
Albert, A. A. 30, 102, 103, 104, 109
d'Alembert, J. 240
Alexiewicz, A. 152
Andrade, J. 250
Artzy, R. 17
Asser, G. 23, 76, 77

Barner, M. 231
Belousov, V. D. 22, 23, 47, 55, 69, 97, 99, 116, 134, 205
Blaschke, W. 136
Boggs, H. 23
Bol, G. 30, 136
Brandt, H. 14, 38
Brianchon, Ch. J. 235
Bruck, R. H. 22, 23, 97, 108, 116, 117, 133, 136

Cantor, G. 188
Cauchy, A. 33, 166, 240, 241
Cayley, A. 16, 17, 33

Daróczy, Z. 156, 270, 277
Dehn, M. 17
Dénes, J. 22, 55, 69, 92, 116
Doyen, J. 132

Erdös, P. 156
Etherington, I. M. H. 23
Euler, L. 26, 33
Evans, T. 23

Falconer, E. 114
Fenyö, I. 116
Flachsmeyer, J. 76, 77
Frolov, M. 37
Fuchs, L. 190, 191

Gajduk, Ju. M. 240
Galois, E. 33
Gauss, C. F. 33
Golab, S. 205, 270, 273
Gruhlke, E. 234

Hamel, G. 149, 154
Hausmann, B. A. 20
Hopf, P. 240
Hosszú, M. 124
Huntington, E. V. 33

Javor, P. 270, 274, 277
Jordan, C. 33

Keedwell, A. D. 22, 55, 69, 92, 116
Kiesewetter, H. 129
Kirkman, T. P. 132
Kirsch, A. 94
Klein, F. 33
Knopp, K. 152, 277
Kronecker, L. 33

Lagrange, J. L. 33
Lebesgue, H. L. 151
Lie, S. 33
Lorenzen, P. 23
Losonczi, L. 156

Mal'cev, A. I. 23
v. Mangoldt, H. 152, 277
Meyer zur Capellen, W. 202
Moufang, R. 22
Murdoch, D. C. 117

Natanson, I. P. 189
Netto, E. 132
Neumann, B. H. 23
Norden, A. P. 260

Ore, O. 20
Orlicz, W. 152

di Paola, J. W. 135
Pentkowski, M. W. 202
Picard, E. 240, 250
Pickert, G. 23, 130

Poisson, S. D. 250
Popa, C. 270, 273, 274
Prodan, N. I. 99
Prohaska, L. 76, 77
Pucharev, N. K. 133

Radó, F. 219
Rätz, J. 177
Rainich, G. Y. 23
Reidemeister, K. 37
Reiss, M. 132
Riemann, B. 151
Rosa, A. 132
Rosenbaum, R. A. 245, 247
Ruffini, P. 33

Sade, A. 49, 55, 133
Saks, S. 189
Schinzel, A. 270, 273
Schönhardt, E. 23, 30, 57
Schröder, E. 23, 55
Segal, S. L. 245, 247
Shaw, J. B. 49
Stein, S. K. 63, 64, 65, 69, 117
Steiner, J. 131, 132
Straszewicz, S. 250

Thurston, H. A. 23
Tisserand, F. 247
Toyoda, K. 111, 113, 115, 117, 127, 128, 129

Vaugham, H. E. 234
Vietoris, L. 221, 238

Weber, H. 33
Wiechmann, P. 221
Wilson, E. B. 171
Wilson, W. H. 245
Wolodzko, S. 270, 277
Wussing, H. 33

Zeitlee, H. 131, 132, 135
Zermelo, E. 154

Sachverzeichnis

Abbildung 15, 76
--, bijektive 17, 77
--, eineindeutige 43
--, homomorphe 88
--, identische 43
--, injektive 42, 77
--, inverse 44
-, isomorphe 91
, isotope 99
, kanonische 78
; Kern 77, 78
, surjektive 17, 42, 77
Abel-Graßmann-Identität 67
abelsche Gruppe 33, 37, 41, 45, 111, 117
Abgeschlossenheit einer Operation 15
Ableitbarkeit 14, 121, 144
absolute Geometrie 249
abstraktes Gruppoid 91
Abstraktionsprinzip 76
Addition von Kräften 249
Additionstheorem 147, 167, 182, 191, 222
Additivität 147, 162
affine Ebene 131
d'Alembertsche Funktionalgleichung 238, 240, 244
algebraische Struktur 13, 14, 22, 32, 41, 42
Anamorphose 201
—, parallele 201
Äquivalenzklasse 76
Äquivalenzrelation 51, 76
—, (mit einer Operation) verträgliche 82
— en; Hauptsatz 76, 81
arithmetisches Gewichtsmittel 120
— Mittel 15, 18, 23, 93, 97, 102
assoziative Quasigruppe 34
Assoziativität 13, 33, 35, 64, 67
—, zyklische 67
Assoziativitätsgleichung 218
Automorphismus 91
Autotopismus 100

Basis von $\mathfrak{R}$ über $\mathfrak{Q}$ 154
bijektive Abbildung 17, 77
bijektiver Homomorphismus 91
Bild, homomorphes 89
binäre Operation 14
— Relation 75
Bisymmetrie 13, 68, 111, 116, 117
Bisymmetriegleichung 118, 219
Bogen 16
Brandtsches Gruppoid 14

Cauchysche Funktionalgleichung 93, 147, 157, 216
- Grundgleichung 93, 147
Cayley-Diagramm 17
Cayley-Tafel 16
Charakterisierung, funktionale 144, 146

Dehnsches Gruppenbild 17
de-Morganscher Isomorphismus 93
D-Gruppoid 19, 21, 41, 42, 45
Diagramm 79
—, kommutatives 79, 88
direkte Proportionalität 159
— Summe 191
Dirichletsche Funktion 274
dividierbare Gruppe 190, 191, 193
Durchschnitt 19

Ebene, affine 131
-. endliche reguläre 131
. komplexe 28
--. Lobačevskijsche 131
--. projektive 131
eineindeutige Abbildung 43
Eingewandtes Produkt 68
Elastizität 46, 67
Element, individuelles links-neutrales 29
—, —rechts-neutrales 29
—, inverses 29, 34
—, links-inverses 31
—, links-neutrales 29, 31, 32
—, neutrales 13, 29, 31
—, rechts-inverses 31
—, rechts-neutrales 29, 31, 32

endliche reguläre Ebene 131
-s Gruppoid 15, 21
Endomorphismus 88
euklidische Geometrie 249
Exponentialfunktion 146, 165, 168, 169, 172, 177
—, verallgemeinerte 179, 187

Faktorgruppoid 81
Faktormenge 76, 87
Farbe 17
Fluchtlinientafel 199
Funktion, Dirichletsche 274
—, Lobačevskijsche 260, 261, 265
—, quasilineare 118
funktionale Charakterisierung 144, 146
Funktionalgleichung 146
—, d'Alembertsche 238, 240, 244
— der Assoziativität 218
— der Bisymmetrie 118, 219
—, Cauchysche 93, 147, 157, 216
—, Jensensche 93, 102
—, Pexidersche 102, 216

Geometrie, absolute 249
—, euklidische 249
—, Lobačevskijsche 249
—, projektive 20, 202
geometrisches Mittel 18, 23, 93
gerichteter Graph 16
Gesetz, Hookesches 159, 161
Gewebe 136, 137, 200
—; Geschlecht 133
—; Ordnung 137
Gewichtsmittel, arithmetisches 120
—, quasilineares 119
Graph, gerichteter 16
—, kantengefärbter 17
—, streng zusammenhängender 27, 41
größter gemeinsamer Teiler 19
Grundgleichung, Cauchysche 93, 147
Gruppe 33, 34, 36, 38, 41, 45, 111
—, abelsche 33, 37, 41, 45, 111, 117
—, dividierbare 190, 191, 193
—, kommutative 33
—, symmetrische 54, 61, 92
—, torsionsfreie 191, 192
Gruppenbild, Dehnsches 17
Gruppengraph 17, 40
Gruppentafel 16
Gruppoid 14, 41
—, abstraktes 91
—, Brandtsches 14
— mit Division 19, 21, 41, 42, 45
—, endliches 15, 21
Gruppoid mit Kürzungsregel 21, 41, 43, 45
—, totalsymmetrisches 72
—, unendliches 15
—, unipotentes 31
-e; Homomorphiesatz 96

Halbgruppe 21, 33, 34, 36, 40, 41, 44, 45, 46
Halbsymmetrie 69, 72, 121, 122
—, linksseitige 46, 60, 61, 67, 69, 72
—, rechtsseitige 46, 60, 61, 67, 69, 72
halbsymmetrische Operation 60
Hamelbasis 154
harmonisches Mittel 18, 22
Hauptisotopismus 100
Hauptsatz über Kongruenzrelationen in Gruppen 85, 86, 87
— über Äquivalenzrelationen 76, 81
Hilbertraum 277
homomorphe Abbildung 88
-s Bild 89
Homomorphie 89
Homomorphiesatz für Gruppoide 96
Homomorphismus 88
—, bijektiver 91
—, injektiver 91
—, kanonischer 95
—; Kern 93, 94
—, surjektiver 91
— von Stone 97
Hookesches Gesetz 159, 161
HS-Gruppoid 122
HS-Quasigruppe 122

Idempotenz 67
identische Abbildung 43
Identität 24
—, parastrophe 64
Indextafel 106
individuelles links-neutrales Element 29
— rechts-neutrales Element 29
injektive Abbildung 42, 77
-r Homomorphismus 91
inverse Abbildung 44
-s Element 29, 34
Inzidenzrelation 130
Inzidenzstruktur 130
isomorphe Abbildung 91
Isomorphie 91
Isomorphismus 91, 177
—, de-Morganscher 93
isotope Abbildung 99
Isotopie 99, 102
Isotopismus 99, 199

Jensensche Funktionalgleichung 93, 102

Kanonische Abbildung 78
-r Homomorphismus 95
kantengefärbter Graph 17
Kegelschnitt 20
Kern einer Abbildung 77, 78
— eines Homomorphismus 93, 94
K-Gruppoid 21, 41, 43
Kleinsche Vierergruppe $\mathfrak{V}_4$ 42, 67, 75, 90, 92, 97, 136
kleinstes gemeinsames Vielfaches 19
Knoten 16
Kommutativität 13, 32, 45, 56, 66, 67
— eines Diagramms 79, 88
kompatibel 82
Komplex 85
komplexe Ebene 28
Komplexprodukt 85
Kongruenzklasse 82
Kongruenzrelation 82
—, normale 83
-en in Gruppen (Hauptsatz) 85, 86, 87
Kosinusfunktion 144, 241
Kosinussatz 256
Kräfteaddition 249
kürzbare Operation 21
Kürzbarkeit 20, 22
Kürzungsregel 20

Lateinisches Quadrat 26, 27
LH-Isotop 105
Light-Test 37
links-abelsche Selbstdistributivität 68
Linksdivision 23
Links-Halbsymmetrie 46, 60, 61, 67, 69, 72
links-inverses Element 31
Links-Loop 30, 41, 45
Linksnebenklasse 85
links-neutrales Element 29, 31, 32
Links-Permutabilität 38, 45, 67
Links-Schweitzer-Identität 68
Links-Selbstdistributivität 46
Links-Systemregel 46, 65, 66, 67
Links-Transitivität 64, 68
Links-Translation 42
Lobačevskij-Ebene 131
Lobačevskijsche Funktion 260, 261, 265
— Geometrie 249
— Trigonometrie 266
Logarithmengesetze 145
Logarithmusfunktion 146, 165, 172, 177, 198
—, verallgemeinerte 179, 184, 187
Loop 30, 31, 41, 44, 45
—, totalsymmetrische 107, 114
Loop-Hauptisotop 105
Loop-Isotop 110

Mathematikunterricht 5, 6, 14, 34, 55, 73, 121, 203, 226
Menge, wohlgeordnete 154
Meßbarkeit 152
Mittel, arithmetisches 15, 18, 23, 93, 97, 102
—, geometrisches 18, 23, 93
—, harmonisches 18, 22
—, quadratisches 18
Mittelpunktskonstruktion 13, 23, 24
Mittel-Transitivität 68
Mittelwert, quasiarithmetischer 119
—, quasilinearer 119
Monoid 33, 41, 44, 45
Moufang-Loop 22, 107, 114

Nacheinanderausführung von Abbildungen 43
— von Isotopismen 102
Netztafel 199
Neumann-Identität 68
neutrales Element 13, 29, 31
Nichtableitbarkeit 14, 121
Niveaulinie 200
Nomogramm 199
normale Kongruenzrelation 83
— Unterquasigruppe 87
Normalteiler 85

Operation 14
—, binäre 14
—, halbsymmetrische 60
—, kürzbare 21
—, parastrophe 49
—, totalsymmetrische 72
—, umkehrbare 19
—, unipotente 31
Operationstabelle 15
operationstreu 88, 147
Ordnung eines Gewebes 137
— eines Gruppoids 15

Parallele Anamorphose 201
Parallelogrammregel 249, 253, 254, 259
parastrophe Identität 64
— Operationen 49
— Quasigruppe 49
-s lateinisches Quadrat 57
Parastrophie 51, 56
Parastrophismus 50, 53
Periode, primitive 227
Permutabilität, linksseitige 38, 45, 67
—, rechtsseitige 45, 67
Permutation 26
Pexidersche Funktionalgleichung 102, 216
Potenz 118, 166, 180, 210

Potenzfunktion 144, 170
Potenzieren 19, 21, 29, 165
primitive Periode 227
Produkt, Eingewandtes 68
projektive Ebene 131
— Geometrie 20, 202
Proportionalität, direkte 159

Quadrat, lateinisches 26, 27
—, parastrophes lateinisches 57
quadratisches Mittel 18
quasiarithmetischer Mittelwert 119
quasilineare Funktion 118
-r Mittelwert 119
-s Gewichtsmittel 119
Quasigruppe 22, 24, 27, 41, 43, 45
—, assoziative 34
—, parastrophe 49
—, totalsymmetrische 121, 122
Quasigruppenoperation 25, 26

Rechenstab 198
Rechteck-Regel 38, 39
rechts-abelsche Selbstdistributivität 68
Rechtsdivision 23
Rechts-Halbsymmetrie 46, 60, 61, 67, 69, 72
rechts-inverses Element 31
Rechts-Loop 30, 41, 45
Rechtsnebenklasse 85
rechts-neutrales Element 29, 31, 32
Rechts-Permutabilität 45, 67
Rechts-Schweitzer-Identität 68
Rechts-Selbstdistributivität 46
Rechts-Systemregel 46, 65, 66, 67
Rechts-Transitivität 64, 68
Rechts-Translation 42
Regularitätsforderung 159, 191, 233
Reidemeister-Bedingung 35, 36, 110, 111, 204
Relation, binäre 75

Satz von Aczél 210
— von Albert 109
— von Toyoda 113
Schleife 16
Schließungsbedingung 204, 205, 217
Schließungssatz 110, 111, 204
Schröder-I-Identität 67
Schröder-II-Identität 67
Schweitzer-Identität, linksseitige 68
—, rechtsseitige 68
Sechseckbedingung (von Brianchon) 111, 203—205
Seitenkosinussatz 268
Selbstdistributivität 74
Selbstdistributivität, links-abelsche 68
—, linksseitige 46
—, rechts-abelsche 68
—, rechtsseitige 46
Sinusfunktion 141, 142, 144, 245
Sinussatz 256, 267
Statik des Massenpunktes 248
— des starren Körpers 250
Stein-I-Identität 67
Stein-II-Identität 67
Stein-III-Identität 67
Steiner-Tripel-System 131, 133
streng zusammenhängender Graph 27, 41
Struktur, algebraische 13, 14, 22, 32, 41, 42
Strukturtafel 16
Subtraktionstheorem 144, 221, 222
Summe, direkte 191
surjektive Abbildung 17, 42, 77
-r Homomorphismus 91
symmetrische Gruppe $\mathfrak{S}_3$ 54, 61, 92
Systemregel, linksseitige 46, 65, 66, 67
—, rechtsseitige 46, 65, 66, 67

Tarski-Identität 68
Teiler, größter gemeinsamer 19
Thomsen-Bedingung 37, 110, 111
Torsionselement 191, 192
torsionsfreie Gruppe 191, 192
Torsionsgruppe 191
Totalsymmetrie 72, 121, 122
totalsymmetrische Loop 107, 114
— Operation 72
— Quasigruppe 121, 122
-s Gruppoid 72
Trägermenge 14
Transitivität, linksseitige 65, 68
—, rechtsseitige 65, 68
Translation 32, 42
—, linksseitige 42
—, rechtsseitige 42
Trigonometrie, Lobačevskijsche 266
trigonometrischer Pythagoras 142, 223
TS-Gruppoid 122
TS-Loop 126
TS-Quasigruppe 122

Umkehrbare Operation 19
Umkehrbarkeit 20, 22
Umkehroperation(en) 23
unendliches Gruppoid 15
unipotente Operation 31
-s Gruppoid 31
Unipotenz 31, 32, 67
Untergruppe 85

Untergruppoid 89
Unterquasigruppe 87
—, normale 87

Veblen-Young-Konfiguration 134
verallgemeinerte Exponentialfunktion 179, 187
— Logarithmusfunktion 179, 184, 187
Vereinigung 19
verträgliche Äquivalenzrelation 82
Vielfaches, kleinstes gemeinsames 19

Wachstumsprozeß 165
Winkelkosinussatz 268
wohlgeordnete Menge 154
Wohlordnungssatz 154

Zahlenkongruenz 83
Zerfallsprozeß 165
zyklische Assoziativität 67