Justus–Liebig–Universität Gießen

Multimodulare Arithmetik
Diplomarbeit in Mathematik

Vorgelegt von:
Oleg Lobachev
Hangelsteinring 15
35396 Gießen

Vorgelegt bei:
Prof. Dr. Tomas Sauer
Justus–Liebig–Universität Gießen

14. März 2007

Inhaltsverzeichnis

Einleitung

Die wesentliche Idee der multimodularen Arithmetik besteht darin, eine Berechnung mit großen Werten durch mehrere Berechnungen mit kleineren Wertebereichen zu ersetzen. Sind die „kleineren Werte“ klein genug, so bringt das auch komplexitätstheoretische Vorteile, aber man muss bedenken, wie man den gewünschten großen Wert wiedergewinnt. Das konstruktive Verfahren dazu basiert auf dem Chinesischen Restsatz. Die erste bekannte Anwendung des Restsatzes ist [8, Overview of Chinese mathematics], [6] von Sun Zi (400(?)–460(?)), die weitere Untersuchung des Chinesischen Restsatzes wurde von Qin Jiushao um 1247 durchgeführt. Einer der ersten europäischen Mathematiker, die sich mit dem Chinesischen Restsatz beschäftigt haben, war Leonardo Pisano Fibonacci.

Seit mindestens hundert Jahren ist man in fast jeder mathematischen Arbeit, die sich mit Restklassen beschäftigt, auf die Restklassenringe und somit die Ringtheorie angewiesen. Die Geburtsstunde der Ringtheorie [8, The development of Ring Theory] liegt im Jahre 1801, als Carl Friedrich Gauß die Kongruenzrelationen erforschte. Man kann mehrere Ringe zu einem größeren Ring zusammenfügen; das liefert die allgemeine Version des Chinesischen Restsatzes. Von 1844 bis 1847 lieferten G. Lamé und vor allem E. Kummer die ersten Ergebnisse der Idealtheorie. R. Dedekind führte die Wörter „Modul“ (1871) und „Körper“ ein. Der Name „Ring“ stammt von D. Hilbert. 1893 hat Hilbert seinen berühmten Basissatz bewiesen. Die Theorie der Zahlenringe und die Theorie der Polynomringe wurden um 1920 von E. Noether und W. Krull in eine axiomatische Theorie der abstrakten kommutativen Ringe zusammengeführt. Dieses Ergebnis wurde aber erst durch das klassische Werk „Moderne Algebra“ [15] von van der Waerden im Jahre 1930 weitgehend bekannt.

Die modernen Untersuchungen der multimodularen Arithmetik als einer weiteren Möglichkeit der computerbasierten Berechnungen sind bei Donald E. Knuth in seinem Meisterwerk *The Art of Computer Programming* [6] erwähnt. Die meisten Überlegungen der vorliegenden Arbeit sind auf *Methods and Applications of Error–Free Computation* [3] von R. T. Gregory und E. V. Krishnamurthy aufgebaut. Diese zitieren ihrerseits G. H. Hardy und E. M. Wright (1960), N. S. Szabó und R. I. Tanaka (1967), sowie D. M. Young und R. T. Gregory (1972–1973) und mehrere Arbeiten von J. A. Howell. Die Darstellung der rationalen Zahlen modulo m und die Möglichkeit der eindeutigen Rekonstruktion basiert auf den Arbeiten von P. S. Wang, sowie P. Kornerup und R. T. Gregory aus der Jahre 1983. Die multimodulare rationale Arithmetik, die im Folgenden als $\mathbb{M}_\beta$ bezeichnet wird, wurde nach [3] von D. Matula und C. Gregory erarbeitet. Mein Interesse für dieses Gebiet wurde durch eine Vorlesung „Computeralgebra“ [11] von T. Sauer geweckt.

Ziele und Aufteilung der Arbeit

Diese Arbeit verfolgt die folgende Ziele:

- *Eine Definition einer multimodularen rationalen Arithmetik*
- *der Beweis der Korrektheit* und

- *Bestimmen des „Gültigkeitsbereichs“* der multimodularen Arithmetik.

Sowohl [3], als auch ein modifiziertes Verfahren, führen zum ersten Ziel. Das zweite wurde durch den Beweise, dass die Vorwärtsabbildung in die modifizierte Arithmetik ein Homomorphismus ist, abgedeckt. Was die dritte Aussage angeht, so wurden die Grenzen, in denen die Arithmetik korrekt rechnet, nur implizit angegeben. Die genaueren Aussagen darüber bedürfen weiterer Forschungen, welche außerhalb des Rahmens dieser Arbeit liegt. Um die angestrebten Ziele zu erreichen, wurden folgende Ergebnisse erzielt:

- *Die Darstellung einer Teilmenge der rationalen Zahlen in einem Restklassenring,* sowie
- *eindeutige Wiederherstellung des Bruches aus dieser Darstellung.* Beide Ergebnisse erreicht man durch geschickte Anwendung des erweiterten euklidischen Algorithmus.
- Die Definition *einer ganzzahligen multimodularen Arithmetik* nach [3], sowie
- *eine rationale multimodulare Arithmetik.* Es wurde beides: das Ergebnis aus [3] und ein modifizierter Ansatz präsentiert.
- Es wurden *die Korrektheit der Addition und der Multiplikation* bewiesen. Anhand dieser kann man einen *Homomorphismus* aus einer Teilmenge der rationalen Zahlen in die zu der modifizierten rationalen multimodularen Arithmetik zugehörige Bildmenge samt Addition und Multiplikation darin angeben.
- Als Zusammenfassung diverser Ergebnisse aus dem Theorieteil folgt, dass in dieser Arbeit eine *endliche genaue rationale Arithmetik* $\mathbb{W}_\beta$ angegeben wurde, die *gegen Überläufe resistent ist:* Die Zwischenergebnisse mögen auch nicht in dem zulässigen Bereich liegen und somit nicht korrekt rückwärtsabbildbar sein. Solange aber das Endergebnis in dem zulässigen Bereich liegt, wird es korrekt wiederhergestellt.
- Es wurde ein *Beispiel für die Praxis–Anwendung* der multimodularen rationalen Arithmetik angegeben.

Die Kapitel 3 und 4 beschäftigen sich mit der multimodularen Arithmetik, jeweils ganzzahlig und rational. Letzteres präsentiert zwei verschiedene Darstellungsweisen, welche zu zwei verschiedenen, aber sehr ähnlichen Arithmetiken $\mathbb{M}_\beta$ und $\mathbb{W}_\beta$ führen. Die Transformation einer Teilmenge der rationalen Zahlen in eine durch die multimodulare Arithmetik verarbeitbare Form und die Wiederherstellung des Endergebnisses übernimmt der erweiterte euklidische Algorithmus. Diesen kann man in Kapitel 2 finden. Die im Satz 2.27 dargestellte Beobachtung des erweiterten euklidischen Algorithmus gibt der in der Kapitel 4 präsentierten Arithmetik eine sehr schöne Eigenschaft, die die Motivation zur Untersuchung dieser Klasse der Arithmetiken erheblich steigert. Der wesentliche Vorteil der multimodularen rationalen Arithmetik, die die Vorrausetzungen des Satzes 2.27 erfüllt, ist, dass diese nichts vor Überläufen zu befürchten hat. Die Zwischenergebnisse können beliebig groß – und somit unkorrekt – werden, solange das Endergebnis im zulässigen Bereich ist, wird dieses korrekt bestimmt. Der Aufwand hält sich dabei unter einer *a priori* bestimmten oberen Schranke. Ein weiterer Bonus ist, dass das Ergebnis *genau* berechnet wird. In einem praxisorientierten Teil der Arbeit gibt das Kapitel 5 ein Beispiel aus der numerischen linearen Algebra, in welchem die rationale multimodulare Arithmetik verwendet wird.

Die Motivation

> We know that the tail must wag the dog,
> for the horse is drawn by the cart;
> But the Devil whoops, as he whooped of old:
> "It's clever, but is it Art?"
>
> Rudyard Kipling, *The Conundrum of the Workshops*

Es gibt mehrere Arten, Zahlen in einem Computer darzustellen und diese zu handhaben. Einerseits kann man die ganzen Zahlen in einer Binärdarstellung als eine Bitfolge darstellen. Bis auf die verschiedenen Möglichkeiten der Darstellung der negativen Zahlen und *endianness*, liegt der Unterschied zwischen verschiedenen Ausprägungen dieser Darstellung in der zulässigen *Länge* der Zahlen. Wird diese beschränkt, so dass man effizient rechnen kann, so findet bei genügend großen Werten früher oder später ein *Überlauf* statt. Das Ergebnis nach dem Überlauf ist völlig und unwiderruflich falsch. Ein typisches Beispiel für das Schema sind die jeweilige Maschinentypen für ganze Zahlen, die man meist mit `int` (für *integer*) ansprechen kann. Wird die Länge der Zahlen nicht effektiv beschränkt, so erhält man eine *Langzahlarithmetik*. Diese hat den Überlauf (zumindest nicht für die Werte, die noch in den Computerspeicher passen) nicht zu befürchten. Allerdings sind wegen schnell wachsender Komplexität der Berechnungen die Implementierungen solcher Arithmetiken langsamer als andere Ansätze. Der Hauptvorteil der Darstellung, die Genauigkeit, kann aber schnell zu einem Nachteil werden. Es wird immer genau gerechnet, und die Komplexität der Rechnung steigt sehr schnell mit der Länge der Zahlen. Der Wunsch, *rational* zu rechnen, bringt in dieser Hinsicht keine Erleichterung, sondern eher zusätzliche Schwierigkeiten. Die Brüche werden natürlich als Paar von Zähler und Nenner dargestellt. Abgesehen von der natürlichen Frage „Wann kürzt man?“ kann man auch für betraglich geringe Werte eine sehr große nicht gekürzte Darstellung angeben:

$$\frac{99999999999999}{100000000000000} \approx 1.$$

Der andere Ansatz besteht darin, die *floating point*–Darstellung zu verwenden. Man stellt die Zahlen als eine n–stellige *Mantisse* samt m–stelligem *Exponent* dar, dem Vorzeichen wird meist ein extra Bit geschenkt. Der wesentliche Unterschied zwischen verschiedenen *floating point* Implementierungen besteht in der Länge n der Mantisse. Je länger die Mantisse ist, desto genauer ist die Arithmetik, der Definitionsbereich wird durch die Größe des Exponenten reguliert. Es ist aber zu bemerken, dass die *floating point*–Arithmetik *nie* genau ist. Dieser Ungenauigkeit verdankt man die Entwicklung der modernen Verfahren der Fehleranalyse und sogar der ganzen numerischen Mathematik: Ein altbewährtes Beispiel, das *Wilkinson Monster*

$$\prod_{k=1}^{20}(x-k),$$

liefert komplexwertige Nullstellen, falls es mit schlecht geeigneten Verfahren in der *floating point*–Arithmetik angegriffen wird, oder wenn die Arithmetik unpassend dimensioniert ist. Andererseits hat *floating point* auch seinen Reiz: die Größe der Zahlen ist beschränkt, man kann einen Berechnungsaufwand unter der bestimmten oberen Schranke und eine nachvollziehbare Fehlergröße erwarten. Der Aufwand für eine ***fl****oating point*–**Op**eration wird als *flop* bezeichnet und gern verwendet. Wie der Name sagt, es wird nur die „Rechenzeit“ gemessen, aber nicht die „Zugriffszeit“.

Man sieht, dass der Überlauf einer der wichtigsten Anstoßpunkte der endlich genaueren Arithmetik ist. Nimmt man den Überlauf in Kauf, so sind die Ergebnisse *ganz* und unwiederruflich falsch,

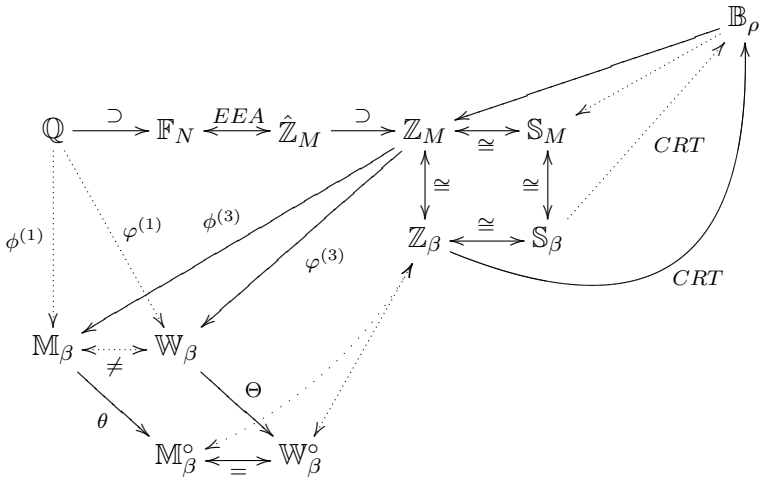
sobald einmal der Überlauf stattgefunden hat. Versucht man, anderweitig den Überlauf zu beseitigen und stets genau zu rechnen, so treibt man die Komplexität der Berechnung unnötig und sehr schnell hoch. Zwischen den zwei Alternativen liegt der alte Ansatz der Restklassenrechnung. Die ganzzahligen Restklassen der Primzahlen sind ein Körper. Man kann stets ein Ergebnis modulo der gewählten Primzahl erwarten. Dieses wird (modulo der Primzahl) exakt. Nun hat diese Vorgehensweise eine Schwäche: die Restklassenrechnung modulo *einer* Primzahl unterscheidet sich kaum von der gewöhnlichen Ganzzahlarithmetik. Diese Arbeit untersucht zwei Varianten der *mehrmodularen* Arithmetik, welche den obigen Nachteil nicht besitzen. Unter welchen Umständen man mit dem gewählten Ansatz der Skylla der Rechenfehler und der Charybdis der mangelhaften Effizienz entkommen kann, wird man dem folgenden Text entnehmen können.

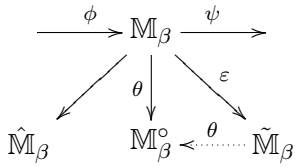
Danksagungen

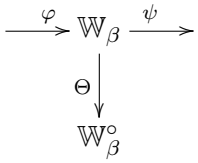
An dieser Stelle möchte ich mich herzlich bei Prof. Dr. Tomas Sauer für die Fragestellung, die zu dieser Arbeit geführt hat, sowie für die gute Betreuung dieser Diplomarbeit bedanken. Er hat vor einigen Jahren mein Interesse an der numerischen Mathematik erweckt, und diese Interesse bleibt stark. Ich bedanke mich bei Vitalij Klassen, Gisela Philippi, Jennifer Seith und Liane Velten, ohne deren Hilfe würde das Skript anders aussehen. Außerdem danke ich meinen Eltern für ihre Unterstützung.

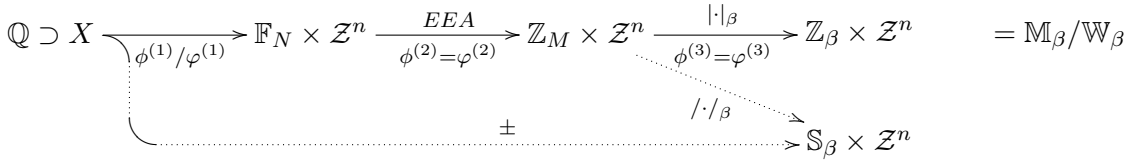
Bezeichnungen

Symbol	Bedeutung
$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$	die natürlichen, ganzen, rationalen, reellen Zahlen
$\mathbb{N}_0$	die natürlichen Zahlen inklusive Null
$\mathbb{Z}_m$	die Restklassen der ganzen Zahlen modulo m
$\hat{\mathbb{Z}}_m$	die Restklassen der rationalen Zahlen teilerfremd zu m
β	der Modulvektor
$\mathbb{Z}_\beta$	die Restklassen der ganzen Zahlen modulo Modulvektor β
$\lvert x\rvert_\mu$	die Restklasse von x in $\mathbb{Z}_\mu$, $\mu \in \{m, \beta\}$
$\mathbb{S}_m$	die symmetrischen Restklassen von $\mathbb{Z}$ modulo m
$\mathbb{S}_\beta$	die symmetrischen Restklassen von $\mathbb{Z}$ modulo Modulvektor β
$/x/_\mu$	ein Element von $\mathbb{S}_\mu$, $\mu \in \{m, \beta\}$
$\mathbb{B}_\rho$	die mixed-radix Darstellungen mit Basenvektor ρ
$\langle x\rangle_\rho$	die mixed-radix Darstellung von x in $\mathbb{B}_\rho$
$\hat{\mathbb{Q}} \subset \mathbb{Q}$	die Teilmenge von $\mathbb{Q}$ mit Brüchen teilerfremd zu m
$\mathbb{Q}_x$	eine Restklasse aus $\hat{\mathbb{Z}}_m$, enthält alle Brüche aus $\hat{\mathbb{Q}}$ kongruent zu $x \in \hat{\mathbb{Z}}_m$
$\mathbb{F}_N$	die Menge der Farey-Brüche der Ordnung N
$\mathbb{M}_\beta$	die Teilmenge von $\mathbb{Q}$, die modulo β *kanonisch* darstellbar ist
$\mathbb{W}_\beta$	die Teilmenge von $\mathbb{Q}$, die modulo β *nicht–kanonisch* darstellbar ist
$\left\lvert\frac{a}{b}\right\rvert_\beta$	ein Element von $\mathbb{M}_\beta$
$\left\lVert\frac{a}{b}\right\rVert_\beta$	ein Element von $\mathbb{W}_\beta$
ggT	der größter gemeinsamer Teiler
kgV	das kleinstes gemeinsames Vielfache
$\mathbb{X}^*$	die Menge der Einheiten in $\mathbb{X}$
EEA	erweiterter euklidischer Algorithmus
$x \leftarrow y$	der Variable x wird der Wert y zugewiesen
$f : x \to y$	die Abbildung f, die x auf y abbildet
$g_{\mathbb{X}}$	diejenige Abbildung g, die $\mathbb{X}$ auf $g(\mathbb{X})$ abbildet
$\mathcal{P}(X)$	die Potenzmenge der Menge X
$f \circ g$	die Abbildung f vor der Abbildung g, $g(f(\cdot))$
$\mathbf{x}$	ein Vektor
$\#\mathbf{x}$	die Länge des Vektors $\mathbf{x}$
$\mathbf{A}$	eine Matrix
$\mathcal{O}(\cdot)$	das Landau–Symbol


$\mathbb{B}_\rho$
$\mathbb{Q}$
$\supset$
$\mathbb{F}_N$
EEA
$\hat{\mathbb{Z}}_M$
$\supset$
$\mathbb{Z}_M$
$\cong$
$\mathbb{S}_M$
CRT
$\phi^{(1)}$
$\varphi^{(1)}$
$\phi^{(3)}$
$\varphi^{(3)}$
$\mathbb{Z}_\beta$
$\cong$
$\mathbb{S}_\beta$
CRT
$\mathbb{M}_\beta$
$\neq$
$\mathbb{W}_\beta$
Θ
θ
$\mathbb{M}^\circ_\beta$
$=$
$\mathbb{W}^\circ_\beta$


ϕ
$\mathbb{M}_\beta$
ψ
θ
ε
$\hat{\mathbb{M}}_\beta$
$\mathbb{M}^\circ_\beta$
θ
$\tilde{\mathbb{M}}_\beta$


φ
$\mathbb{W}_\beta$
ψ
Θ
$\mathbb{W}^\circ_\beta$


$\mathbb{Q} \supset X$
$\phi^{(1)}/\varphi^{(1)}$
$\mathbb{F}_N \times \mathcal{Z}^n$
EEA
$\phi^{(2)}=\varphi^{(2)}$
$\mathbb{Z}_M \times \mathcal{Z}^n$
$|\cdot|_\beta$
$\phi^{(3)}=\varphi^{(3)}$
$\mathbb{Z}_\beta \times \mathcal{Z}^n$
$= \mathbb{M}_\beta/\mathbb{W}_\beta$
$/\cdot/_\beta$
$\pm$
$\mathbb{S}_\beta \times \mathcal{Z}^n$

Abbildungsverzeichnis

Tabellenverzeichnis

Algorithmenverzeichnis

> "Begin at the beginning," the King said gravely, "and go on till you come to the end: then stop."
>
> Lewis Carroll, *Alice's Adventures in Wonderland*

Kapitel 1

Ringe und Restklassen

1.1 Ringe und Integritätsbereiche

Definition 1.1.

1. Ein kommutativer *Ring* R bzw. $(R, +, \cdot)$ ist eine Menge R, die eine kommutative Gruppe $(R, +)$ und eine kommutative Halbgruppe $(R, \cdot)$ mit Distributivgesetz enthält.

2. Ein (kommutativer) Ring heißt ein *Integritätsbereich*, falls dieser nullteilerfrei ist, also falls $\forall a \in R,\, a \neq 0\ \nexists b \in R,\, b \neq 0$ mit $ab = 0$.

3. Ein Element $e \in R$ mit $e^{-1} \in R$ wird als eine *Einheit* bezeichnet. Die Menge der Einheiten des Ringes R wird als R^* geschrieben.

Die Begriffe der Gruppe und Halbgruppe werden als bekannt vorausgesetzt. Das neutrale Element bezüglich der Addition wird mit 0 (Null) bezeichnet, das neutrale Element bezüglich Multiplikation mit 1 (Eins). Es werden ausschließlich *kommutative* Ringe betrachtet, im Folgenden ist ein Ring stets ein kommutativer Ring mit Eins.

Definition 1.2. Sei R ein Ring und $a, b \in R$. Das Element a *teilt* b, falls es ein $c \in R$ gibt, so dass $ac = b$. Man schreibt dafür $a \mid b$. Man bezeichnet gerne a als einen *Faktor* von b.

Definition 1.3 (Prim und koprim)**.** Sei R ein euklidischer Ring mit Eins.

1. Ein $p \in R$ für das kein weiteres $x \in R \setminus \{1, p\}$ existiert, das p teilt, ist ein *Primelement*. Für $R = \mathbb{Z}$ heißen die Primelemente *Primzahlen*.

2. Zwei Elemente a, $b \in R$ sind *koprim*, falls sie keine gemeinsame Faktoren außer 1 haben.

Im Folgenden werden einige Hilfsaussagen gezeigt, die trotz ihrer Einfachheit oft verwendet werden.

Lemma 1.4. *Seien a, b, c, m in Integritätsbereich R. Es gilt*

1. *$m \mid a$ und $m \mid b \Rightarrow m \mid (a + b)$.*

2. *Für alle $b \in R$: $m \mid a \Rightarrow m \mid ab$*

3. *$m \mid a \Leftrightarrow mb \mid ab$ für alle b.*

4. *Sind b und c koprim, so $b \mid a$ und $c \mid a \Rightarrow bc \mid a$.*

5. $a \mid b$ und $b \mid c \Rightarrow a \mid c$.

Beweis.

1. Sei $a = k_a m$ und $b = k_b m$. Nun $a + b = (k_a + k_b)m$.
2. a=$k_a m \Rightarrow ab = k_a bm =: k_{ab} m$.
3. $a = km \Leftrightarrow ab = kbm$.
4. $b \mid a \Rightarrow \exists k_1 \in R : a = k_1 b$, analog $c \mid a \Rightarrow \exists k_2 \in R : a = k_2 b$. Nun, wegen $b \nmid c$, aber $c \mid a$, gilt $c \mid k_1$, d. h. $\exists k_3 \in R : k_1 = k_3 c$. Somit ist aber $a = k_3 bc \Rightarrow bc \mid a$.
5. $b = k_b a$, $c = k_c b \Rightarrow c = k_c k_b a$.

□

1.2 Die Restklassen $\mathbb{Z}_m$

Betrachtet man die arithmetischen Operationen auf dem Ring $\mathbb{Z}$, so stellt man schnell fest: sehr wenige Elemente haben ein multiplikatives Inverses. Jedoch kann man eine vernünftige Division einführen – die Division mit Rest. Diese wird folgendermaßen ausgeführt: Man führt den Divisionsschritt aus, solange man noch kann, und sobald der Divisionsschritt mit dem Ergebnis in $\mathbb{Z}$ unmöglich wird, liefert man als Ergebnis beides – den bis zu diesem Moment errechneten *Quotienten* und den Divisionsrest. Letzteren bezeichnet man auch als einen *Rest*. Ebenso schnell stellt man fest, dass die Reste Äquivalenzklassen, die sogenannten *Restklassenringe* $\mathbb{Z}_m$ – die eigentliche Restklassen $\{0, \ldots, m-1\}$ samt Addition und Multiplikation modulo m, bilden. Diese haben einige bemerkenswerte Eigenschaften:

- $\mathbb{Z}_m \cong \mathbb{Z}/m\mathbb{Z}$,
- $\mathbb{Z}_m$ ist ein Ring,
- $\mathbb{Z}_m$ ist sogar ein Körper, falls m prim ist.

Im folgenden Abschnitt werden einige dieser Tatsachen gezeigt. Im Falle „$\mathbb{Z}_m$ ist ein Körper“ haben *alle* Elemente von $\mathbb{Z}_m$ eine multiplikative Inverse. Man schreibt $\mathbb{Z}_p$ dafür.

Außerdem ist $\mathbb{Z}_m$ nicht die einzige Möglichkeit, Restklassen zu bilden. Falls man die Forderungen an den Quotienten und den Rest verändert, bekommt man die eng verwandte, aber doch verschiedene Restklassenfamilie $\mathbb{S}_m$, die *symmetrischen* Restklassen. Diese werden im Abschnitt 1.3 behandelt.

1.2.1 Kongruenz modulo m

Definition 1.5 (Kongruenz)**.** Seien a, b und $m \in \mathbb{Z}$ mit $m > 1$. Teilt m die Differenz $b - a$, so ist b *kongruent* zu a modulo m, in Zeichen:

$$b \equiv a (\text{mod } m).$$

Ist dies nicht der Fall, so ist b *nicht kongruent* zu a modulo m, in Zeichen

$$b \not\equiv a (\text{mod } m).$$

Die Relation von kongruenten a und b (mod m) heißt *Kongruenz*.

Satz 1.6 (Eigenschaften der Kongruenzrelation)**.** *Seien a, b, $m \in \mathbb{Z}$, $m > 1$. Die Kongruenzrelation $\equiv$ ist*

1. symmetrisch*. Es sind äquivalent:*
   (a) $a \equiv b (\mathrm{mod}\ m)$
   (b) $b \equiv a (\mathrm{mod}\ m)$
   (c) $a - b \equiv 0 (\mathrm{mod}\ m)$
2. transitiv*:* $a \equiv b (\mathrm{mod}\ m)$ und $b \equiv c (\mathrm{mod}\ m) \Rightarrow a \equiv c (\mathrm{mod}\ m)$.
3. additiv*: Für x, $y \in \mathbb{Z}$ gilt* $a \equiv b (\mathrm{mod}\ m)$ und $c \equiv d (\mathrm{mod}\ m) \Rightarrow ax + cy \equiv bx + dy (\mathrm{mod}\ m)$.
4. multiplikativ*:* $a \equiv b (\mathrm{mod}\ m)$ und $c \equiv d (\mathrm{mod}\ m) \Rightarrow ac \equiv bd (\mathrm{mod}\ m)$.

Beweis. Aus der Definition 1.5 folgt:

1. (a) $\Leftrightarrow m \mid (a-b) \Leftrightarrow m \mid (b-a) \Leftrightarrow$ (b) sowie (c) $\Leftrightarrow a - b \equiv 0 (\mathrm{mod}\ m) \Leftrightarrow m \mid (0 - (a-b)) \Leftrightarrow m \mid (b-a) \Leftrightarrow$ (b).
2. $m \mid (a-b)$ und $m \mid (b-c) \Rightarrow m \mid (a-b) + (b-c) \Leftrightarrow m \mid (a-c)$.
3. Unter mehrfacher Verwendung des Lemmas 1.4 folgt $a \equiv b (\mathrm{mod}\ m) \Rightarrow m \mid (a-b) \Rightarrow$ auch $m \mid (a-b)x$ für $x \in \mathbb{Z}$. Analog mit $y \in \mathbb{Z}$: $c \equiv d (\mathrm{mod}\ m) \Rightarrow m \mid (c-d)y$. Es folgt $m \mid ((a-b)x + (c-d)y)$, somit $m \mid ((ax+cy) - (bx+dy)) \Leftrightarrow ax + cy \equiv bx + dy (\mathrm{mod}\ m)$.
4. Seien $a = b + k_1 m$ und $c = d + k_2 m$. Dann ist $ac = (b + k_1 m)(d + k_2 m) = bd + m(k_1 d + k_2 c + k_1 k_2 m)$.

□

Die in dem Beweis von Satz 1.6.(4) verwendete Idee wird formal als Proposition 1.8.(1) dargestellt.

1.2.2 Eine Abbildung in $\mathbb{Z}_m$

Fordert man, dass für $a = km + r$ stets $0 \leq r < m$ gilt, so kann man diese Restklassen samt Addition und Multiplikation modulo m mit $\{0, \ldots, m-1\} =: \mathbb{Z}_m$ identifizieren.

Definition 1.7 (unimodulare Vorwärtsabbildung)**.** Man schreibt

$$\begin{aligned} |\cdot|_m &: \mathbb{Z} \to \mathbb{Z}_m \\ |\cdot|_m &: b \to r, \end{aligned} \tag{1.1}$$

wobei $r \in \mathbb{Z}_m$ so gewählt ist, dass

$$b \equiv r (\mathrm{mod}\ m) \tag{1.2}$$

gilt.

Proposition 1.8. *Seien a, b, $m \in \mathbb{Z}$, $m > 1$. Es gilt*

1. *Für alle $a \in \mathbb{Z}$ gilt $|a|_m = 0 \Leftrightarrow \exists k \in \mathbb{Z}$ mit $a = km$,*
2. $\big||a|_m\big|_m = |a|_m$,
3. *$|a|_m$ ist eindeutig,*
4. $|a|_m = |b|_m \Leftrightarrow a \equiv b (\mathrm{mod}\ m)$.

Bemerkung 1.9. In $a = km+r$ schreibt man oft den Quotienten k als $k = [a/m]$ bzw. als $k = \lfloor a/m \rfloor$, falls man Wert darauf legt, dass $km \leq a$ ist. Dabei ist $r = (a/m - \lfloor a/m \rfloor)\, k$.

Beweis der Proposition 1.8. Man verwendet in diesem Beweis die Schreibweise $x = k_x m + r_x$ mit $0 \leq r_x < m$ für die ganzzahlige Division mit Rest. Offensichtlich ist $|x|_m = r_x$.

1. Man formuliert die Behauptung um: „$\forall a \in \mathbb{Z} : |a|_m = 0 \Leftrightarrow m \mid a$“. Nun ist einerseits im Falle $m \mid a$: $a = k_a m + 0$, also $r_a = 0$ bzw. $|a|_m = 0$. Andererseits, falls $|a|_m =: r_a = 0$ ist, so reduziert sich die Darstellung $a = k_a m + r_a$ auf $a = k_a m$, also $m \mid a$.

2. Da $0 \leq |a|_m < m$, ist $|a|_m = 0m + |a|_m$.

3. Seien $x \neq y$ zwei verschiedene Darstellungen von $|a|_m$. Dann gilt $a = k_x m + x$ und $a = k_y m + y$. Ist $k_x = k_y$, so ist die Behauptung bewiesen. Angenommen $k_x \neq k_y$, o. B. d. A. ist $k_y = k_x + k$ mit $k \in \mathbb{Z}$, $k > 0$ (sonst k_x und k_y vertauschen). Es folgt $a = (k_x + k)m + y = k_x m + x$, also $x = km + y \geq m$. Es gilt aber $0 \leq x, y < m$ und $k, m > 0$, Widerspruch.

4. Seien $r_a := |a|_m$ und $r_b := |b|_m$ mit $a = k_a m + r_a$ und $b = k_b m + r_b$. Nach 1. ist der Ausdruck $a - b = (k_a - k_b)m + (r_a - r_b)$ kongruent Null modulo m genau dann, wenn $r_a = r_b$.

□

1.2.3 Die Arithmetik

> ... the different branches of Arithmetic – Ambition, Distraction, Uglification, and Derision.
>
> Lewis Carroll, *Alice's Adventures in Wonderland*

Man bildet den *Restklassenring* $\mathbb{Z}_m$ aus den Restklassen $\{0, \ldots, m-1\}$ von $\mathbb{Z}$ modulo m. Die Operationen in $\mathbb{Z}_m$ bildet man, indem man die Addition bzw. die Multiplikation in $\mathbb{Z}$ ausführt und anschließend das Ergebnis modulo m reduziert. Im Folgenden bezeichnet man mit $\oplus$ bzw. $\odot$ die Addition bzw. die Multiplikation in $\mathbb{Z}_m$ entsprechend.

Satz 1.10 (Arithmetik in $\mathbb{Z}_m$, Teil 1)**.** *Seien a, b, $m \in \mathbb{Z}$, $m > 1$. Es gilt*

$$a \oplus b := |a + b|_m = \big||a|_m + |b|_m\big|_m = \big||a|_m + b\big|_m = \big|a + |b|_m\big|_m \tag{1.3}$$

$$a \odot b := |ab|_m = \big||a|_m |b|_m\big|_m = \big||a|_m b\big|_m = \big|a|b|_m\big|_m \tag{1.4}$$

Beweis. Sei $* \in \{+, \cdot\}$. Es folgt aus Proposition 1.8 (1): Setzt man $|a|_m =: r_a$ mit $0 \leq r_a < m$ und $a = k_a m + r_a$, so ist $|a|_m = r_a$ mit Proposition 1.8 (2). Mit derselben und 1.8 (1) folgt, $\big|a * |b|_m\big|_m = |(k_a m + r_a) * r_b|_m$.

- Ist „$* = +$“, so $|k_a m + r_a + r_b| = \big||k_a m|_m + |r_a|_m + |r_b|_m\big|_m = \big||r_a|_m + |r_b|_m\big|_m = |r_a + r_b|$.
- Im Fall „$* = \cdot$“ folgt $|r_b k_a m + r_a r_b|_m = \big||r_b k_a m|_m + |r_a r_b|_m\big|_m = |r_a r_b|_m$.

Andere Teile der Behauptung ergeben sich, indem die Rollen von a und b vertauscht bzw. die bereits bewiesenen Teile zweimal angewendet werden. □

Beispiel 1.11. *Der Satz 1.10 zeigt, dass der Zeitpunkt der Berechnung des Restes irrelevant ist, so dass folgende Berechnungen gleichwertig sind:*

$$\begin{aligned}
|24+38|_3 &= |62|_3 = 2\\
|24+38|_3 &= \big||24|_3+|38|_3\big|_3 = |0+2|_3 = 2\\
|24+38|_3 &= \big||24|_3+38\big|_3 = |0+38|_3 = 2\\
|24+38|_3 &= \big|24+|38|_3\big|_3 = |24+2|_3 = |26|_3 = 2.
\end{aligned}$$

Satz 1.12. *Das System* $(\mathbb{Z}_m, \oplus, \odot)$ *ist ein* endlicher kommutativer Ring mit Eins*, wobei*

- $\oplus$ *die Addition modulo* m*,*
- $\odot$ *die Multiplikation modulo* m *bezeichnet, und*
- $m > 1$ *ist.*

Den Beweis in der allgemeinen Form kann man in jedem Algebrabuch finden, z. B. in [4].

Bemerkung 1.13. Das *additive Inverse* von a modulo m ist $-a := |-a|_m = m-a$, denn: $|a+(-a)|_m = |a+(m-a)|_m = |m|_m = 0$.

Korollar 1.14 (Arithmetik in $\mathbb{Z}_m$, Teil 2)**.** *Die* Subtraktion *in* $\mathbb{Z}_m$ *wird für* a*,* b *in* $\mathbb{Z}_m$ *als*

$$|a-b|_m = |a+(-b)|_m\,,$$

definiert.

Bemerkung 1.15 (Schnelle Multiplikation in Restklassenringen)*.* Zu der schnellen Multiplikation modulo m: [1, S. 5] verweist auf [9], in dieser Arbeit wurde die Multiplikation in $\mathbb{Z}_p$ ohne die anschließende Division präsentiert. Dies *könnte* man bei der Implementierung verwenden, in dieser Arbeit an sich wird es nicht verwendet.

Allerdings ist $\mathbb{Z}_m$ nicht bloß ein Ring, sondern viel mehr.

Satz 1.16. *Der endliche kommutative Ring* $(\mathbb{Z}_m, \oplus, \odot)$ *ist ein* endlicher Körper *genau dann, wenn* m *eine Primzahl ist.*

Beweisidee. $a \in \mathbb{Z}_m$ ist eine Einheit $\Leftrightarrow \mathrm{ggT}(a, m) = 1$. †

1.3 Symmetrische Reste modulo m

> Numero deus impari gaudet.
> Gott freut sich der ungeraden Zahlen.
>
> Vergil, *Eklogen, VIII, 75*

Es wird eine transparente Darstellung für die *negativen* ganzen Zahlen benötigt. Eine der Möglichkeiten, dies zu bewerkstelligen, sind die symmetrischen Reste.

Definition 1.17 (Symmetrische Reste modulo m)**.**

1. Sei $m > 1$ ungerade[1], definiere die *Menge der symmetrischen Reste modulo m* als

$$\mathbb{S}_m := \left\{ -\frac{m-1}{2}, \ldots, -2, -1, 0, 1, 2, \ldots, \frac{m-1}{2} \right\}. \tag{1.5}$$

2. Die Abbildung

$$\begin{aligned} /\cdot/_m &: \mathbb{Z} \to \mathbb{S}_m \\ /\cdot/_m &: b \mapsto s \end{aligned} \tag{1.6}$$

ist definiert durch

$$\begin{aligned} b &\equiv s (\mathrm{mod}\ m) \\ -\frac{m}{2} &< s < \frac{m}{2}. \end{aligned}$$

3. Der Wert $/b/_m$ heißt der *symmetrische Rest* von b modulo m.

Satz 1.18.

1. $(\mathbb{S}_m, \oplus, \odot)$ ist ein endlicher kommutativer Ring mit Eins.

2. Ist m prim, so ist $(\mathbb{S}_m, \oplus, \odot)$ ein endlicher Körper.

3. Es gilt

$$(\mathbb{S}_m, \oplus, \odot) \cong (\mathbb{Z}_m, \oplus, \odot). \tag{1.7}$$

Beweis. Es genügt (1.7) zu zeigen. Definiere $\sigma : \mathbb{Z}_m \to \mathbb{S}_m$ und $\varsigma : \mathbb{S}_m \to \mathbb{Z}_m$ mit

$$\sigma(|a|_m) = /a/_m = \begin{cases} |a|_m & \text{falls } 0 \le |a|_m < \frac{m}{2} \\ |a|_m - m & \text{sonst} \end{cases}$$

$$\varsigma(/a/_m) = |a|_m = \begin{cases} /a/_m & \text{falls } 0 \le /a/_m < \frac{m}{2} \\ /a/_m + m & \text{sonst.} \end{cases}$$

Die Elemente zwischen 0 und $\lfloor m/2 \rfloor$ werden identisch aufeinander abgebildet. Die Standarddarstellung hat noch $\lfloor m/2 \rfloor$ Elemente von $\lceil m/2 \rceil$ bis $m-1$, die symmetrische Darstellung hat noch $\lfloor m/2 \rfloor$ Elemente von -1 bis $-\lfloor m/2 \rfloor$, zwischen denen σ und ς Bijektionen sind. □

Beispiel 1.19. *Die Idee des Beweises für $\mathbb{Z}_5$ und $\mathbb{S}_5$:*

$\mathbb{Z}_5$	0	1	2	3	4
$\mathbb{S}_5$	0	1	2	-2	-1

[1] Denn die symmetrischen Reste modulo gerade Zahlen sind nicht wirklich symmetrisch. Außerdem will man eigentlich die Restklassen*körper* haben, und die meisten Primzahlen sind ungerade. Der Ausnahmefall $m = 2$ ist kaum passend, meist braucht man *große* Restklassen, was auch heißt – große Primzahlen. In [11] sollen diese bei der Implementierung „fast so groß“, wie ein Maschinenwort sein, und $2^{32}, 2^{64} \gg 2$.

Kapitel 2

Der euklidische Algorithmus

> In a dead fantasy–mathematical world where points are just material nothings, you can divide anything *ad infinitum*...
>
> Liona Fan–Chiang

Der euklidische Algorithmus ist einer der ältesten bekannten Algorithmen, aber im Gegensatz zu den anderen zeitgenössischen Ablaufvorschriften enthält der euklidische Algorithmus eine `while`–Schleife, was ihn zum ersten „echten“ Algorithmus krönt. (Andere Beispiele für die „Ur–Algorithmen“ sind in [6, Abschnitt 4.5.2] zu finden.) Etwas mehr zur Geschichte des Algorithmus und seinen Verbesserungen ist in Abschnitt 2.1.2 zu finden. Der „eigentliche“ euklidische Algorithmus hat seine Bedeutung nicht verloren, man wird beispielsweise den Bezug zwischen dem euklidischen Algorithmus und einem anderen alten, schönen, mathematischen Konzept – den Kettenbrüchen – darstellen. Jedoch gibt es einige Erweiterungen. Man kann den Algorithmus allgemein auf Integritätsbereichen mit etwas mehr Struktur einführen. Man kann auch den Algorithmus erweitern, indem man zusätzliche Spalten hinzufügt. Diese Vorgehensweise führt zu dem erweiterten euklidischen Algorithmus, welcher im Abschnitt 2.2.1 präsentiert wird. Dieser Algorithmus ist eine effiziente Methode der Berechnung der Bézout–Identität, was das effiziente Auffinden der multiplikativen Inversen in Restklassenringen ermöglicht. Bleibt man bei $\mathbb{Z}_m$, so kann man mit dem erweiterten euklidischen Algorithmus eine bestimmte Teilmenge der rationalen Zahlen in $\mathbb{Z}_m$ abbilden (und, was nicht unwichtig ist, diese Darstellung invertieren und zurück in $\mathbb{Q}$ abbilden). Das ist das entscheidende Element bei der Konstruktion der endlichen rationalen Arithmetik, die unter gewissen Vorraussetzungen exakt rechnet. Anfangs wird der euklidische Algorithmus betrachtet.

2.1 Der euklidische Algorithmus

2.1.1 ggT und kgV

Definition 2.1 (ggT)**.** Der *größte gemeinsame Teiler* $\mathrm{ggT}(a, b)$ zweier ganzer Zahlen a und b ist der eindeutige nichtnegative Generator (erzeugendes Element) der von a und b erzeugten additiven Untergruppe von $\mathbb{Z}$.

Intuitiv ist der ggT ein Teiler von a und b, der unter der zu präzisierenden Bedingung maximal ist. Man kann auch nach [15,14] den ggT als ein Ideal[1], das eine *Summe* zweier oder mehrerer Ideale

[1]Nach [15]: Sei R ein kommutativer Ring. Man bezeichnet $I \subset R$ als ein *Ideal* von R, falls aus a, $b \in I$ folgt $a + b \in I$ und falls für $a \in I$, $x \in R$ beliebig folgt $ax \in I$.

ist, also als das von der Vereinigungsmenge der Ideale erzeugte Ideal definieren, man muss sich dabei gar nicht auf $\mathbb{Z}$ einschränken. Dabei ist jedoch die Eindeutigkeit zu beachten [11].

Eine beliebige ganze Zahl $a \geq 0$ lässt sich stets als ein Produkt von Potenzen der Primzahlen darstellen, dies besagt der „Fundamentalsatz der Arithmetik“ [6, Abschnitt 4.5.2]:

$$a = \prod_{p \text{ prim}} p^{a_p} = 2^{a_2} 3^{a_3} 5^{a_5} \dots$$

So ist auch für $a, b \geq 0$

$$\begin{aligned} \mathrm{ggT}(a,b) &= \prod_{p \text{ prim}} p^{\min(a_p, b_p)}, \\ \mathrm{kgV}(a,b) &= \prod_{p \text{ prim}} p^{\max(a_p, b_p)}. \end{aligned}$$

D. Knuth [6, Abschnitt 4.5.2] gibt auch die Konvention $\mathrm{ggT}(0,0) = 0$ an. Es gilt $\mathrm{ggT}(a,b) = \mathrm{ggT}(b,a)$, $\mathrm{ggT}(a,b) = \mathrm{ggT}(-a,b)$, $\mathrm{ggT}(a,0) = |a|$ und nach [1] wegen $\mathrm{ggT}(a,0) = \mathrm{ggT}(0,a) = |a|$ und $\mathrm{ggT}(a,b) = \mathrm{ggT}(|a|,|b|)$ kann man stets annehmen, dass a, b nicht negativ sind.

Bemerkung 2.2 (ggT und koprim). Die Zahlen $a, b \in \mathbb{Z}$ sind *koprim* genau dann, wenn $\mathrm{ggT}(a,b) = 1$. Das bezeichnet man auch als „a und b sind *teilerfremd*“.

2.1.2 Der Algorithmus, „basic version“

Algorithmus 1 (euklidischer Algorithmus)**.**
Eingabe: *a, b*

1. *Solange $b \neq 0$, wiederhole 2.*
2. *Setze $(a,b) \leftarrow (a \bmod b, a)$.*

Ergebnis: $b = \mathrm{ggT}(a,b)$

Man operiert dabei auf den *Paaren* der Zahlen, die „Programmierervariante“ von diesem Schritt ist $t \leftarrow a \bmod b$, $b \leftarrow a$, $a \leftarrow t$.

Die Geschichte und Evolution des euklidischen Algorithmus von der Originalversion aus den „Elementen“ bis zu dem binären euklidischen Algorithmus findet sich in [6, Abschnitt 4.5.2]. Beim binären euklidischen Algorithmus [1, Algorithmus 1.3.5] versucht man so viele Zweierpotenzen, wie möglich abzuspalten und anschließend wiederherzustellen: $2 \mid a$, $2 \mid b \Rightarrow \mathrm{ggT}(a,b) = 2\,\mathrm{ggT}(a/2, b/2)$. Es gibt aber auch weitere Verbesserungen des euklidischen Algorithmus. Wählt man nicht die Standard–, sondern die symmetrische Restklassendarstellung (siehe 1.3), so erhöht man drastisch [11] die „Konvergenzgeschwindigkeit“ des Algorithmus in *worst case*. Dieses Ergebnis wird im Abschnitt 2.3.2 diskutiert.

Man müsste auch beweisen, dass der Algorithmus 1 tatsächlich den ggT liefert. Der formale Beweis ist in der Proposition 2.4 angegeben. Allerdings, um sich mit dem Algorithmus vertraut zu machen, betrachte man die folgende Beobachtung.

Lemma 2.3. *Für zwei ganze Zahlen a, b und ein beliebiges ganzes c gilt*

$$\mathrm{ggT}(a,b) = \mathrm{ggT}(a - cb, b). \tag{2.1}$$

Beweis. Sei $x := \mathrm{ggT}(a, b)$. So $x \mid a$ und $x \mid b$. Also auch $x \mid cb$ und daher $x \mid (a - cb)$, damit ist x ein gemeinsamer Teiler von $a - cb$ und b. Die Maximalität von x folgt unmittelbar aus der von $\mathrm{ggT}(a, b)$. □

Nun geschieht beim Algorithmus 1 nichts anderes als die Berechnung des ominösen $a - cb$, was, wie man soeben gesehen hat, den ggT nicht verändert. Der euklidische Algorithmus ist auch tatsächlich ein Algorithmus, denn er terminiert in endlich vielen Schritten. Dazu bemerkt man, dass die „Größe“ von a und b in jedem Schritt des Algorithmus verringert wird. Um dies genauer zu beschreiben, muss man den Begriff der euklidischen Funktion einführen. Die genauere Handhabung dieser Aussage findet sich in der Proposition 2.7.

Proposition 2.4 ([11, Proposition 2.20])**.** *Der euklidische Algorithmus berechnet genau den ggT.*

2.1.3 Die euklidische Funktion, euklidische Ringe

Definition 2.5 (Funktionen und Ringe, [11, Definition 2.16])**.** Ein Integritätsbereich R ist ein *euklidischer Ring*, falls eine *euklidische Funktion* $d : R \to \mathbb{N} \cup \{-\infty\}$ existiert, dass es für alle $a, b \in R$, $b \neq 0$, stets einen *Quotienten* c und einen *Rest* $r \in R$ gibt mit

$$a = bc + r, \quad d(r) < d(b). \tag{2.2}$$

Bemerkung 2.6 (zu Definition 2.5).

1. $(\mathbb{Z}, |\cdot|)$ ist ein euklidischer Ring.
2. Für $\mathbb{Z}$ ist der Betrag $|\cdot|$ ein Beispiel der euklidischen Funktion, für univariate Polynome – die Gradfunktion.
3. Es gibt viele euklidische Funktionen für einen Ring, man kann aber immer [11, Bemerkung 2.17.3] eine punktweise *minimale* euklidische Funktion bilden. Zur Vereinfachung der Darstellung wird im Folgenden stets mit minimalen euklidischen Funktionen gearbeitet.
4. Der Wert $-\infty$ kann (muss aber nicht) nur bei $d(0)$ angenommen werden.

Proposition 2.7 (Euklidischer Algorithmus ist ein Algorithmus)**.** *Der euklidische Algorithmus 1 terminiert nach einer endlichen Anzahl von Schritten.*

Beweis. Der euklidische Algorithmus 1 läuft in $\mathbb{Z}$ ab, es ist ein euklidischer Ring mit der zugehörigen euklidischen Funktion d. In jedem Schnitt des Algorithmus wird die Division mit Rest ausgeübt, somit ist $d(b) < d(a)$ in jedem Schritt. Die Menge $\mathbb{N} \cup \{-\infty\}$ hat aber ein Minimum. □

Beweis der Proposition 2.4. Man schreibt den „euklidischen Schritt“ des Algorithmus 1 als

$$r_{i+1} = |r_{i-1}|_{r_i}, \quad i \in \mathbb{N}, \tag{2.3}$$

mit $r_0 = a$, $r_1 = b$ und bezeichnet den im i–ten Schritt berechneten Wert $\lfloor r_{i-1}/r_i \rfloor$ als $q_i \in \mathbb{Z}$. Dann ist $r_{i+1} = r_{i-1} - q_i r_i$. Daraus folgt $\mathrm{ggT}(a, b) \mid r_i$ für $i \in \mathbb{N}_0$. Da aber der euklidische Algorithmus nach endlich vielen Schritten terminiert, gibt es ein $n \in \mathbb{N}_0$ mit $r_{n+1} = 0$, aber $r_n \neq 0$. Nun kann man die Formel (2.3) „rückwärts aufrollen“

$$r_{i-1} = q_i r_i + r_{i+1}, \quad i = n - 1, \ldots, 1.$$

Also $r_n \mid r_{n-1}$ und auch $r_n \mid r_{n-2}$, $r_n \mid r_{n-3}$, $\ldots$, $r_n \mid r_1$ und $r_n \mid r_0$, d. h. $r_n \mid b$, $r_n \mid a$. Folglich ist $r_n = \mathrm{ggT}(a, b)$. □

In dem Beweis von Proposition 2.7 kann man $(\mathbb{Z}, |\cdot|)$ durch einen beliebigen euklidischen Ring ersetzen. Das nach $(\mathbb{Z}, |\cdot|)$ bekannteste Beispiel für einen euklidischen Ring sind die univariaten Polynome mit der Gradfunktion, also $(\Pi, \deg)$. In folgendem Beispiel wird der euklidische Algorithmus auf diesem Ring präsentiert, alle Einträge wurden *normalisiert*, sodass der Leitkoeffizient stets Eins ist.

Beispiel 2.8 (EA in Π)**.** *Berechne den* ggT *von* $5x^4-6x^3+4x^2+8x-11$ *und* $3x^3-8x^2+13x-7$.

i	$\deg r_i$	q_i	r_i
1	4		$x^4-\frac{6}{5}x^3+\frac{4}{5}x^2+\frac{8}{5}x-\frac{11}{5}$
2	3		$x^3-\frac{8}{3}x^2+\frac{13}{3}x+\frac{7}{3}$
3	2	$x+\frac{22}{15}$	$x^2-\frac{109}{17}x+\frac{55}{17}$
4	1	$x+\frac{191}{51}$	$x-\frac{1392}{2419}$
5	0	$x-\frac{240007}{41123}$	$\boxed{1}$

2.2 Erweiterter euklidischer Algorithmus

Man kann den euklidischen Algorithmus erweitern. Fasst man die Eingabe des euklidischen Algorithmus als eine 1×2 Matrix auf, so wird diese Matrix bei dem erweiterten euklidischen Algorithmus erweitert, indem zu ihr zusätzliche Spalten hinzugefügt werden. Es werden hier zwei– und dreispaltige Varianten des Algorithmus betrachtet.

2.2.1 Der Algorithmus

Im Folgenden wird gezeigt, wie der euklidische Algorithmus erweitert werden kann, wobei die Bezeichnung „erweiterter euklidischer Algorithmus“ mit EEA abgekürzt wird.

Algorithmus 2 (Erweiterter euklidischer Algorithmus)**.**
Eingabe: *Die Matrix*

$$\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}.$$

1. *Setze* $i \leftarrow 1$.
2. *Solange* $a_{i+1,1} \neq 0$*, wiederhole die Schritte 2–7, ansonsten terminiere.*
3. *Setze*
$$q \leftarrow \left\lfloor \frac{a_{i,1}}{a_{i+1,1}} \right\rfloor$$
4. *Setze*
$$a_{i+2,1} \leftarrow a_{i,1} - qa_{i+1,1} = a_{i,1} \bmod a_{i+1,1}.$$
5. *Setze*
$$a_{i+2,2} \leftarrow a_{i,2} - qa_{i+1,2}.$$
6. *Inkrementiere den Zähler:* $i \leftarrow i+1$.
7. *Gehe zu dem Schritt 2.*

Ergebnis: *die Zeile* $(a_{i-1,1}, a_{i-1,2})$.

Bemerkung 2.9. Es existiert auch eine dreispaltige Variante des Algorithmus, wobei die Vorschrift für die dritte Spalte analog zu der für die zweite ist

$$a_{i+2,3} \leftarrow a_{i,3} - qa_{i+1,3}$$

mit $q = [a_{i,1}/a_{i+1,1}]$. Für die genauere Beschreibung dieser Variante der EEA siehe [1, Abschnitt 1.3.2], insbesondere [1, Algorithmus 1.3.6].

Korollar 2.10. *Der EEA terminiert.*

Beweis. Aus der Proposition 2.7 folgt: die „erste Spalte“ von EEA terminiert. □

2.2.2 Lemma von Bézout

Satz 2.11 (Lemma von Bézout)**.** *Sei R ein euklidischer Ring. Dann existieren zu allen $0 \neq a, b \in R$ ein x und y in R, dass*

$$ax + by = \mathrm{ggT}(a, b) \tag{2.4}$$

gilt.

Den Beweis kann man in den meisten Büchern über Algebra bzw. Zahlentheorie, wie z. B. in [4, Proposition 5.3, S. 62], finden.

Korollar 2.12 ([3, Theorem 6.15])**.** *Sei $\mathrm{ggT}(b, m) = 1$. Ist $mx + by = 1$, so ist $b^{-1} (\mathrm{mod}\ m) = |y|_m$.*

Korollar 2.13 (Berechnung)**.** *Die dreispaltige Variante des erweiterten euklidischen Algorithmus (Bemerkung 2.9) mit der Startmatrix*

$$\begin{pmatrix} a & 1 & 0 \\ b & 0 & 1 \end{pmatrix}$$

liefert gerade $g := \mathrm{ggT}(a, b)$ und die Bézout–Koeffizienten x, y mit $ax + by = g$. Das Ergebnis steht in der letzten Zeile der EEA in der Form (g, x, y).

Beispiel 2.14. *Wähle $a = 166$ und $b = 154$. Dann ist die Startmatrix des EEA*

$$\begin{pmatrix} 166 & 1 & 0 \\ 154 & 0 & 1 \end{pmatrix}.$$

Nun ergibt sich

166	1	0
154	0	1
12	1	−1
10	−12	13
2	13	−14
0		

Es ist: $165 \cdot 13 + 154 \cdot (-14) = 2$.

2.2.3 Multiplikative Inverse und Brüche modulo m

Man kann die multiplikativen Inversen in euklidischen Ringen, insbesondere in $\mathbb{Z}_m$, effizient mit dem erweiterten euklidischen Algorithmus berechnen. Sind ganze Zahlen b und m koprim, so existiert die Inverse.

Lemma 2.15 (Multiplikative Inverse)**.** *Ist* $\mathrm{ggT}(m, b) = 1$ *und* $0 \neq b \in \mathbb{Z}_m$*, so gibt es ein* $c \in \mathbb{Z}_m$ *mit*

$$|bc|_m = |cb|_m = 1,$$

eine multiplikative Inverse von b modulo m*, im Zeichen*

$$c = b^{-1} (\mathrm{mod}\ m).$$

Beweis. Für $b \in \mathbb{Z}_m$ liefert Korollar 2.12 die Bedingung $\mathrm{ggT}(b, m) = 1$ für die Existenz der Inversen. Die Äquivalenz der linken und rechten Inversen folgt sofort, denn $\mathbb{Z}_m$ ist ein kommutativer Ring. □

Bemerkung 2.16 (Arithmetik in $\mathbb{Z}_m$, Teil 3). Korollar 2.12 gibt die Bedingung zur Existenz der multiplikativen Inversen an: Zu $0 \neq b \in \mathbb{Z}_m$ existiert ein eindeutiges ($\mathbb{Z}_m$ ist ein Ring!) $b^{-1} \in \mathbb{Z}_m$ genau dann, wenn b und m *koprim* sind.

Definition 2.17 („Einkodieren“ der Brüche)**.** Man kann

$$a \oslash b = \left|\frac{a}{b}\right|_m = |ab^{-1}|_m \tag{2.5}$$

schreiben.

Bemerkung 2.18 (Wichtiges zu der Definition 2.17). $|ab^{-1}|_m$ ist eine ganze Zahl. Genauer gesagt ist dieser Ausdruck kongruent zu einer ganzen Zahl x in $\mathbb{Z}_m$. Die Menge der zu x modulo m kongruenten Brüchen wird mit $\mathbb{Q}_x$ bezeichnet, vgl. die Definition 2.26 (2). Man kann analog zu der obigen Definition auch eine *symmetrische* „Kodierung“ der Brüche einführen.

Beispiel 2.19. $|3/8|_5 = |3 \cdot 8^{-1}|_5 = |3 \cdot 2|_5 = 1$.

Es folgt ein (triviales) technisches Resultat, das später verwendet wird.

Lemma 2.20. *Sei* $\frac{a}{b}$ *ein Bruch und sei* m *teilerfremd mit* b*.*

1. *Die* additive Inverse *von* $\left|\frac{a}{b}\right|_m$ *ist* $\left|-ab^{-1}\right|_m$*.*
2. *Ist noch* a *teilerfremd mit* m*, so ist* $\left|\frac{b}{a}\right|_m$ *die* multiplikative Inverse *von* $\left|\frac{a}{b}\right|_m$*.*

Beweis. Setze $\left|-\frac{a}{b}\right|_m := \left|(-a)(b^{-1})\right|_m = \left|-ab^{-1}\right|_m$. Nun ist $\left|-\frac{a}{b}\right|_m \oplus \left|\frac{a}{b}\right|_m = \left|\frac{a}{b} - \frac{a}{b}\right|_m = |0|_m$ und $\left|\frac{a}{b}\right|_m \odot \left|\frac{b}{a}\right|_m = |ab^{-1}|_m \odot |ba^{-1}|_m = |ab^{-1}ba^{-1}|_m = |1|_m$. □

2.2.4 Farey–Brüche

> Farey is immortal because he failed to understand a theorem which Haros had proved perfectly fourteen years before.
>
> G. H. Hardy, *A Mathematician's Apology*

John Farey war ein Geologe, dessen Entdeckung der Farey–Brüche seine einzige mathematische Leistung war [8, Biography of J. Farey]. G. H. Hardy schrieb: „Farey als Geologen wurden zwanzig

Zeilen im *Dictionary of National Biography* gewidmet. Als Geologe ist er vergessen und die einzige [mathematische] Arbeit, die nicht in Vergessenheit geraten ist, wurde nicht aufgelistet.“ 1816 veröffentlichte Farey einen Artikel *On a curious property of vulgar fractions*, in dem er die nach ihm benannte Sequenz einführte und auf die folgende „sonderbare Eigenschaft“ hinwies.

Definition 2.21 (Farey–Brüche)**.** Die Brüche $\frac{a}{b} \in \mathbb{Q}$ mit $|a| \leq N$, $|b| \leq N$ für ein festes N bilden die Menge der *Farey–Brüche* der Ordnung N, in Zeichen: $\mathbb{F}_N$.

Beispiel 2.22.

$$\mathbb{F}_3 = \left\{ -3, -2, -\frac{3}{2}, -1, -\frac{2}{3}, -\frac{1}{2}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1, \frac{3}{2}, 2, 3 \right\}.$$

Dabei ist anzumerken, dass in einigen Quellen, z. B. [17, Farey Sequence], nur positive Farey–Brüche, oder auch nur die Einträge, die kleiner als Eins sind, betrachtet werden. Manchmal wird die Sequenz der Farey–Brüche als Farey–Sequenz bzw. Farey–Reihe bezeichnet.

Proposition (Fareys „sonderbare Eigenschaft“)**.** *Zähler und Nenner eines Farey–Bruches sind Summe von jeweils Zähler und Nenner der benachbarten Brüche in der Farey–Sequenz.*

Ein Beispiel dazu: $\frac{1}{2} \in \mathbb{F}_3$, so sind die benachbarten Brüche $\frac{1}{3}$ und $\frac{2}{3}$. Tatsächlich ist $\frac{1+2}{3+3} = \frac{3}{6} = \frac{1}{2}$.

Lemma 2.23. *Sind x und y in $\mathbb{F}_N$, so ist*

$$x + y \in \mathbb{F}_{2N^2} \text{ und } xy \in \mathbb{F}_{N^2} \tag{2.6}$$

Beweis. Seien $\frac{a}{b}$ und $\frac{c}{d}$ in $\mathbb{F}_N$, also $|a|, b, |c|, d \leq N$. Es ist $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd} \leq \frac{|ac+bd|}{bd}$, und $|ac + bd| \leq |ac| + |bd| \leq 2N^2$, $bd \leq N^2$, sowohl $\frac{a}{b}\frac{c}{d} \leq \frac{|ac|}{bd}$, und $|ac| \leq N^2, bd \leq N^2$. □

Farey selbst hat keinen Beweis angegeben, jedoch wurde die „sonderbare Eigenschaft“ in gleichem Jahr von A. L. Cauchy bewiesen. Die am Anfang dieses Abschnittes zitierte Aussage von G. H. Hardy ist laut MacTutor [8] etwas zu harsch und ungenau. Obwohl Haros tatsächlich in 1802 eine Arbeit über die Approximation der dezimalen Zahlen mit gewöhnlichen Brüchen veröffentlicht hatte, in der er tatsächtlich die Farey–Brüche und die „sonderbare Eigenschaft“ verwendete, hat Haros keine allgemeine Aussage getroffen und keinen Beweis dafür angegeben.

2.2.5 EEA und Brüche modulo m

Das Folgende beruht auf einer Aussage von Kornerup, die [3] entnommen wurde und die in Satz 2.25 dargestellt ist. Dies gibt die Möglichkeit, die Abbildung EEA von $\mathbb{F}_N$ auf $\hat{\mathbb{Z}}_m$ angeben und verschafft einem den Grundsatz für endliches genaues rationales Rechnen. Diese Abbildung kann man mit folgendem Algorithmus berechnen.

Algorithmus 3 (EEA mit dem „Einkodieren“ der Brüche)**.**
Eingabe: *$\frac{r}{s}$, m.*

1. *Starte EEA (Algorithmus 2) mit der Startmatrix*

$$\begin{pmatrix} m & 0 \\ s & r \end{pmatrix}.$$

2. *Bekomme das Paar (y, x) als Ergebnis des EEA.*

Ergebnis: *die Zahl $|x|_m = \left|\frac{r}{s}\right|_m$.*

Bemerkung 2.24.

1. In Algorithmus 3 muss das Ergebnis x vor der Ausgabe explizit modulo m genommen werden.
2. Auf diese Art kann auch *schnell* die multiplikative Inverse von b modulo m bestimmt werden. Man bildet nämlich den Bruch $\frac{1}{b}$ ab, also wird der EEA mit der Matrix
$$\begin{pmatrix} m & 0 \\ b & 1 \end{pmatrix}$$
gestartet.
3. Es gibt das sogenannte „Zweischrittverfahren“, in dem man zuerst $b^{-1} (\mathrm{mod}\ m)$ bestimmt und anschließend: $|ab^{-1}|_m$. Es existiert aber kein Grund, das „Einschrittverfahren“, also genau den Algorithmus 3, nicht zu verwenden, vgl. [3].

Satz 2.25 (Algorithmus 3 funktioniert! [3, Theorem 6.30])**.** *Sei $\frac{r}{s} \in \mathbb{Q}$ mit $|s|_m \neq 0$ und sei $m \in \mathbb{Z}$ mit $m > 0$. Ist Startmatrix des Algorithmus*
$$\begin{pmatrix} m & 0 \\ s & r \end{pmatrix},$$
so terminiert der Algorithmus 3 nach $n+1$ Schritten mit $a_{n+1,1} = 0$. Dann gilt
$$\left|\frac{r}{s}\right|_m = |a_{n,2}|_m . \tag{2.7}$$

Der Beweis wird später im Text angegeben.

Definition 2.26. Zu einer Zahl m definiere

1. $\hat{\mathbb{Q}} := \left\{ \frac{a}{b} \in \mathbb{Q} : \mathrm{ggT}(b, m) = 1 \right\} \subset \mathbb{Q}$.
2. Die Mengen $\mathbb{Q}_0, \ldots, \mathbb{Q}_{m-1}$ mit $\frac{a}{b} \in \mathbb{Q}_k \Leftrightarrow \left|\frac{a}{b}\right|_m = k$, $k \in \{0, \ldots, m-1\}$.
3. Somit ist $\left|\frac{a}{b}\right|_m$ eine Abbildung von $\hat{\mathbb{Q}}$ nach $\hat{\mathbb{Z}}_m := \{\mathbb{Q}_0, \ldots, \mathbb{Q}_{m-1}\}$.

Die Bedingungen für die Eindeutigkeit der Abbildung EEA sowie seine Urbildmenge gibt der nächste Satz an [3, Theorem 5.14].

Satz 2.27 (Eindeutigkeit)**.** *Falls $\mathbb{Q}_k$ einen Farey–Bruch x der Ordnung N enthält, und falls*
$$2N^2 < m. \tag{2.8}$$
gilt, so ist x der einzige derartige Bruch in $\mathbb{Q}_k$.

Dieser Satz wird im noch folgenden Text bewiesen. Es ergibt sich die Abbildung $EEA : \mathbb{Q} \supset \mathbb{F}_N \to \hat{\mathbb{Z}}_m \subset \mathbb{Z}_m$. Beim Algorithmus 2 ist die Ausgabe des euklidischen Algorithmus nur die vorletzte Zeile, aber es gibt Fälle, in denen das ganze Ablauftableau des Algorithmus relevant ist. Der folgende Satz stammt aus [3, Theorem 6.39].

Satz 2.28 (rational reconstruction)**.** *Sei* $\frac{r}{s} \in \mathbb{F}_N$ *und* $x := \left|\frac{r}{s}\right|_m \in \hat{\mathbb{Z}}_m$*. In diesem Fall schiebt man das Vorzeichen des Bruchs in den Nenner, also* $\frac{r}{s} \in \mathbb{F}_N \Leftrightarrow 0 < r \leq N$ *und* $0 < |s| \leq N$*. So existiert ein Index* i*, der eine eindeutige Zeile* $(a_{1,i}, a_{2,i})$ *mit*

$$(r, s) = (a_{1,i}, a_{2,i}) \tag{2.9}$$

im Ausgabetableau $[(a_{1,j}, a_{2,j}) : j = 1, \ldots, n]$ *des mit der Startmatrix*

$$\begin{pmatrix} a_{1,1} & a_{2,1} \\ a_{2,1} & a_{2,2} \end{pmatrix} = \begin{pmatrix} m & 0 \\ x & 1 \end{pmatrix}$$

initiierten erweiterten euklidischen Algorithmus 2 liefert.

Die Existenz wird im Abschnitt 2.3.4 gezeigt, die Eindeutigkeit ist im Satz 2.27 mit einbegriffen. Der Name *rational reconstruction* stammt aus D. E. Knuth [6, Abschnitt 4.5.3, Übung 51], der seinerseits auf Veröffentlichungen von P. S. Wang bzw. P. Kornerup und R. T. Gregory, einen der Autoren von [3] verweist. Man kann darauf aufbauend den modifizierten Algorithmus angeben.

Algorithmus 4 (EEA mit Suche von $\mathbb{F}_N$)**.**
Eingabe: *Die Matrix*

$$\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix},$$

die Ordnung der Brüche N*.*

1. *Setze* $i \leftarrow 1$*.*
2. (a) *Falls* $a_{i+1,1} = 0$*, so brich ab.*
   (b) *Falls* $|a_{i,1}| < N$ und $|a_{i,2}| < N$*, so wiederhole die folgenden Schritte, ansonsten terminiere erfolgreich.*
3. *Setze* $q \leftarrow \left\lfloor \frac{a_{i,1}}{a_{i+1,1}} \right\rfloor$*.*
4. *Setze* $a_{i+2,1} \leftarrow a_{i,1} - q a_{i+1,1} = a_{i,1} \mod a_{i+1,1}$*.*
5. *Setze* $a_{i+2,2} \leftarrow a_{i,2} - q a_{i+1,2}$*.*
6. *Inkrementiere den Zähler:* $i \leftarrow i + 1$*.*
7. *Gehe zu dem Schritt 2.*

Ergebnis: *Das Paar*

$$(a_{i-1,1}, a_{i-1,2}),$$

welches in der Form $\frac{a_{i-1,1}}{a_{i-1,2}}$ *dem gesuchten Bruch entspricht, im Falle der erfolgreichen Terminierung und ein Fehler beim Abbruch.*

Bemerkung 2.29*.*

1. Der einzige Unterschied zwischen Algorithmus 2 und Algorithmus 4 liegt in dem Schritt 2.
2. Der Algorithmus 4 bricht entweder erfolgreich ab, sobald ein passender Bruch gefunden wurde, oder er läuft genauso lange, wie der „normale“ EEA und terminiert anschließend nicht erfolgreich. Es kann daraus gefolgert werden: Der Algorithmus 4 terminiert.

Als eine Nebenbemerkung wird (ohne Beweis!) noch eine weitere Aussage zu *rational reconstruction* angegeben. Ist $/\cdot/_m$ die Abbildung in den symmetrischen Restklassenringen – siehe Abschnitt 1.3 – so kann die folgende Methode [3, Theorem 6.40] verwendet werden.

Satz 2.30 (common denominator method)**.** *Sei $x = \frac{a}{b} \in \mathbb{F}_N$ und $k = |ab^{-1}|_m$. Man kann aus k den Wert von x wiederherstellen, falls es ein tb mit $t \in \mathbb{Z}$, $0 < t \leq N$ gibt, sodass*

$$ta = /tbk/_m \tag{2.10}$$

gilt. In diesem Fall ist

$$x = \frac{ta}{tb}. \tag{2.11}$$

2.2.6 Rationale Arithmetik in $\hat{\mathbb{Z}}_m$

Schaltet man vor $|\cdot|_m$ die Abbildung EEA, so kann man bestimmte rationale Zahlen – nämlich genau die Farey–Brüche $\mathbb{F}_N$ – modulo m abbilden. Mit $\frac{a}{b} \in \mathbb{F}_N$ schreibt man für die *Vorwärtsabbildung* $\left|EEA\left(\frac{a}{b}\right)\right|_m$ auch $\left|\frac{a}{b}\right|_m$. Die Operationen $\oplus, \ominus, \odot$ sind dieselben wie im ganzzahligen Fall. Die Division $\oslash$ stellt man wie gewohnt dar: Das Auffinden der multiplikativen Inversen modulo m – wiederum mit EEA – mit der anschließenden Multiplikation. Dafür schreibt man $(\hat{\mathbb{Z}}_m, \oplus, \odot)$. Im Gegensatz zu $\mathbb{Z}_m$ kann man das Ergebnis nicht direkt interpretieren: Die *Rückwärtsabbildung* ist genau $EEA^{-1} : \hat{\mathbb{Z}}_m \to \mathbb{F}_N$.

Bemerkung 2.31.

1. Eigentlich ist $\hat{\mathbb{Z}}_m = \{|EEA_m(x)|_m : x \in \mathbb{F}_N\}$ eine strikte Teilmenge von $\mathbb{Z}_m$. Das heißt, nicht alle verallgemeinerten Restklassen $\mathbb{Q}_k$ enthalten einen Farey–Bruch. Erhält man nach Rechnung in $\mathbb{Z}_m$ ein Ergebnis, das nicht in $\hat{\mathbb{Z}}_m$ ist, so kann EEA^{-1} keinen zugehörigen Farey–Bruch finden und bricht erfolglos ab. Dies wird nach Gregory und Krishnamurthy [3] als *Pseudoüberlauf* bezeichnet.

2. Im Folgenden wird oft $EEA : \mathbb{F}_N \ni \frac{a}{b} \to x \in \hat{\mathbb{Z}}_m$ für die durch den Algorithmus 3 definierte Abbildung und $EEA^{-1} : \hat{\mathbb{Z}}_m \ni x \to \frac{a}{b} \in \mathbb{F}_N$ für die entsprechende Abbildung aus dem Algorithmus 4 verwendet. Falls das m in $\hat{\mathbb{Z}}_m$ besonders hervorgehoben wird, so schreibt man entsprechend EEA_m und EEA_m^{-1} bzw. $EEA_{m,N}$ und $EEA_{m,N}^{-1}$, falls auch die Ordnung N der Farey–Brüche relevant ist.

Beispiel 2.32. *Sei $m = 19$, also $N = 3$.*

1. *Wegen $EEA_{19,3}\left(\frac{1}{3}\right) = |-6|_{19} = |13|_{19}$ und $EEA_{19,3}\left(\frac{1}{3}\right) = |12|_{19}$ ist $\left|\frac{1}{3}\right|_{19} \oplus \left|-\frac{2}{3}\right|_{19} = |13 + 12|_{19} = |6|_{19}$ und $EEA_{19,3}^{-1}(6) = -\frac{1}{3}$, was korrekt ist.*

2. *Wegen $EEA_{19,3}\left(\frac{1}{2}\right) = |-9|_{19} = |10|_{19}$ ist $\left|\frac{1}{2}\right|_{19} \oplus \left|-\frac{2}{3}\right|_{19} = |10 + 12|_{19} = |3|_{19}$. Aber $EEA_{19,3}^{-1}(3) = 3$, was das* falsche *Ergebnis 3 liefert, wobei die korrekte Antwort $-\frac{1}{6}$ ist.*

Bemerkung 2.33. Wie später die allgemeineren Betrachtungen verdeutlichen, können die Zwischenergebnisse auch nicht in $\mathbb{F}_N$ liegen. Solange das Endergebnis in $\mathbb{F}_N$ ist, wird das korrekte Ergebnis wiederhergestellt [3, Remark 5.15].

Satz 2.34 (Bedingung der Gleichheit der Brüche modulo m, [3, Theorem 5.6])**.** *Seien $x = a/b$ und $y = c/d$ und existieren die beiden Inversen $b^{-1} \pmod m$ und $d^{-1} \pmod m$. Dann ist*

$$|x|_m = |y|_m$$

genau dann, wenn

$$ad \equiv bc (\mathrm{mod}\ m).$$

Beweis. „⇐“ Multipliziere die beiden Seiten mit $b^{-1}d^{-1}$.

$$\begin{aligned} ad \equiv bc (\mathrm{mod}\ m) &\Rightarrow ab^{-1} \equiv cd^{-1} (\mathrm{mod}\ m) \\ &\Rightarrow |ab^{-1}|_m = |cd^{-1}|_m \Rightarrow |x|_m = |y|_m. \end{aligned}$$

„⇒“ Ist analog, der Faktor ist bd.

□

Zu beweisen ist eine entscheidende Aussage über die Beziehung zwischen der Ordnung der Farey–Brüche N und der Größe des Restklassenrings m, die mit dem EEA verkoppelt werden können.

Beweis des Satzes 2.27. Es wird das Gegenteil angenommen. Seien $x = \frac{a}{b}$ und $y = \frac{c}{d} \in \mathbb{F}_N$ mit $|x|_m$ und $|y|_m \in \mathbb{Q}_k$. Das bedeutet aber, dass $|x|_m = k = |y|_m$ und mit Satz 2.34 folgt $ad \equiv bc (\mathrm{mod}\ m)$, was man umschreiben kann als $|ad - bc|_m = 0$. Nun wegen (2.8) ist $0 \leq |ad - bc| \leq |a||d| + |b||c| \leq 2N^2 \leq m - 1$, also $ad - bc = 0$. Dies liefert aber $\frac{a}{b} = \frac{c}{d}$ bzw. $x = y$. □

2.3 Analyse

2.3.1 Kettenbrüche

> Alles wiederholt sich im Leben,
> ewig jung ist nur die Phantasie,
> Was sich nie und nirgends hat begeben,
> Das allein veraltet nie!
>
> Friedrich Schiller, „An die Freude“

Die Kettenbrüche stellen ein sehr altes Konzept dar [8, 13]. So waren sie schon den antiken Griechen bekannt. Der euklidische Algorithmus berechnet die Kettenbruchentwicklung „nebenbei“; aber auch andere Arbeiten, zum Beispiel die Berechnung der Schiefe der Ekliptik von Eratosthenes von Kyrene[2] bzw. Claudius Ptolemäus, könnten die Kettenbrüche verwendet haben. Nach von Fritz[3] hat Hippasos von Elis die Irrationalität des goldenen Schnittes mit Kettenbrüchen gezeigt. Laut Fowler waren die Kettenbrüche die übliche Form der Darstellung der Brüche in Platos Zeiten. Auch berühmte Diophantische Gleichungen lassen sich mithilfe der Kettenbrüche lösen.

Später wurden die Kettenbrüche von Madhava von Sangamagrama, E. Bombelli, P. A. Cataldi, J. Wallis, C. Huygens, J. H. Lambert, D. Rittenhouse, J.–L. Lagrange, C. Hermite, T. Stieltjes, A. Hurwitz, H. Padé, und L. O. Blumental untersucht oder verwendet. Der Begriff der Konvergente wurde von Leonard Euler eingeführt; so hat Euler im Jahre 1844 unter Verwendung der Kettenbrüche die Existenz der transzendenten Zahlen gezeigt. Carl Friedrich Gauß verwendete die Kettenbrüche,

[2] Eratosthenes von Kyrene. Der Mann mit dem Sieb. Jedoch außer dem Sieb von Eratosthenes, seiner wohl beständigsten Entdeckung, hat Eratosthenes auch den Durchmesser der Erdkugel (über die Messung des Einfallswinkels der Sonnenstrahlen in Syene und Alexandria), den Abstand zwischen Erde und Sonne und offenbar auch die Schiefe der Ekliptik, die er auf 11/83 von 180° einschätzte. Zu den weiteren Arbeiten von Eratosthenes zählen ein Kalender, geschichtliche und geographische Forschungen, sowie ein Sternenkatalog und mehrere Literaturwerke.

[3] Von Fritz und Fowler werden zitiert in I. Grattan–Guinness, *Numbers, ratios, and proportions in Euclid's Elements: How did he handle them?*, Historia Mathematica **23** (1996), 355–375, seinerseits zitiert in [8, The real numbers: Pythagoras to Stevin].

um die orthogonalen Polynome und Quadraturformeln zu konstruieren. Gauß verwendete dafür die Drei–Term–Rekursionsformel und den Satz von Bernoulli. Das klassische moderne Werk über Kettenbrüche ist „Die Lehre von den Kettenbrüchen“ von Oskar Peron, von dem in dieser Arbeit der erste Band [10] verwendet wird. Eine interessante Einführung in die Theorie der Kettenbrüche ist in [5] zu finden.

Definition 2.35. Ein endlicher *Kettenbruch* $[a_0; a_1 \ldots, a_n]$ wird rekursiv definiert als

- $[a] = a$
- $[a_0; a_1 \ldots, a_n] = a_0 + \frac{1}{[a_1; a_2, \ldots, a_n]}$.

Satz 2.36. *Jede rationale Zahl $x \in \mathbb{Q}$ ist als ein endlicher Kettenbruch darstellbar.*

Der Beweis sowie die Umkehrung dieser Aussage ist in [5, Theorem 14] dargestellt.

Definition 2.37 (Konvergenten)**.** Sei $[a_0; a_1, \ldots, a_n]$ eine Kettenbruchdarstellung einer Zahl $x \in \mathbb{Q}$. Dann ist $[a_0; a_1, \ldots, a_k]$ die *k-te Konvergente*, die Konvergente von der Ordnung k von x, im Zeichen: $K_k(x)$.

Bemerkung 2.38.

1. Es gibt mehrere Kettenbruchentwicklungen einer Zahl, z. B.

   $$[a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_n - 1, 1],$$

   aber man kann immer normieren, indem man für den letzen Eintrag des Kettenbruches ≥ 2 fordert. Somit kann man von „der“ Darstellung sprechen.

2. Die k-te Konvergente $K_k(x)$ ist für alle $k \in \mathbb{N}$ definiert. Da der ursprüngliche Kettenbruch nur bis a_n geht, setzt man $a_{n+1} = a_{n+2} = \cdots = 0$.

3. Die Konvergente eines endlichen Kettenbruchs ist ein endlicher Kettenbruch.

4. Der Zähler und der Nenner der Konvergente heißen *Kontinuanten* oder K–Polynome [6]. In diesem Abschnitt bezeichnen p_k und q_k (bzw. $p_k(x)$ und $q_k(x)$, bzw. $p_k(x_0, \ldots, x_k)$ und $p_{k-1}(x_1, \ldots, x_k)$) entsprechend den Zähler und Nenner der k–ten Konvergente.

Satz 2.39 (Konvergente Kettenbrüche sind Konvergenten)**.**

- *Für jede Konvergente $\frac{p}{q} = K_k(x)$ gilt*

  $$\left| x - \frac{p}{q} \right| < \frac{1}{q^2}. \tag{2.12}$$

- *Eine der zwei aufeinanderfolgenden Konvergenten $K_k(x)$ und $K_{k+1}(x)$, die als $\frac{p}{q}$ bezeichnet wird, erfüllt sogar*

  $$\left| x - \frac{p}{q} \right| < \frac{1}{2q^2}. \tag{2.13}$$

- *Jeder Bruch $\frac{p}{q}$, der (2.13) erfüllt, muss eine Konvergente $K_k(x)$ zu $k \in \mathbb{N}_0$ sein.*

Diesen Satz kann man in [10, Satz 2.11] finden, der Beweis ist auch in [5] als ein Teil des Beweises von Satz 2.36 (Theorem 14 in [5]) bzw. [5, Theorem 9] für endliche Kettenbrüche, angegeben.

Bemerkung 2.40 (Was ist a_i?). Die hier präsentierten Aussagen sind [5, Chapter 1] Eigenschaften der Kettenbrüche als einer Struktur.

2.3.2 Abfallgeschwindigkeit der euklidischen Funktion in verschiedenen Restklassenringen

Wählt man die Reste von $a \bmod b$ nicht als $\{0, \ldots, b-1\} = \mathbb{Z}_b$, sondern *symmetrisch*, so ist der EA schneller [11, Bemerkung 2.21]. Man muss aber die symmetrischen Reste $\mathbb{S}_m$ abweichend von Abschnitt 1.3 definieren.

Definition 2.41 (Symmetrische Reste)**.** Die Menge der *symmetrischen Reste* $\mathbb{S}_m$ modulo $m > 1$ wird folgendermaßen definiert:

$$\begin{aligned} \mathbb{S}_m &= \left\{ -\frac{m}{2} + 1, \ldots, \frac{m}{2} \right\} \text{ falls } m \text{ gerade} \\ \mathbb{S}_m &= \left\{ -\frac{m-1}{2}, \ldots, \frac{m-1}{2} \right\} \text{ falls } m \text{ ungerade.} \end{aligned} \tag{2.14}$$

Nun wählt man stets bei Algorithmen 1, 2 und 3 die Reste a_{i+1} nicht aus $\mathbb{Z}_{a_i}$, sondern aus $\mathbb{S}_{a_i}$, also

$$a_{i+1} \leftarrow /a_{i-1}/_{a_i}$$

in dem jeweiligen Schritt des Algorithmus.

Wählt man $|\cdot|$ als die euklidische Funktion d, so ist der *worst case* im „Standardrestklassenfall“ $d(a_{i+1}) = d(a_i) - 1$, aber im „symmetrischen“ Fall ist $d(a_{i+1}) = d(a_i)/2$. Somit ist die obere Schranke für die Laufzeit des EA nicht höchstens $d(a_0) - 1$, sondern nur $\log_2 d(a_0)$ Schritte [11]. In der Regel ist der Gewinn aber nicht so dramatisch: Nach [16, Exercise 3.30] benötigt man bei der Verwendung der symmetrischen Reste bei *ihrer worst case*–Eingabe etwa die Hälfte der Schritte des EA im Vergleich mit dem gewöhnlichen EA.

Beispiel 2.42 („Konvergenzgeschwindigkeit“ des EA)**.** *Man verfolgt den Ablauf des EA mit der Eingabe $a_0 = 55$, $a_1 = 34$. Links ist der „Standard–Fall“ $\mathbb{Z}_{a_i}$, rechts der „symmetrische Fall“ $\mathbb{S}_{a_i}$.*

i	q	a	q	a
1	–	55	–	55
2	1	34	2	34
3	1	21	3	−13
4	1	13	3	−5
5	1	8	2	−2
6	1	5	2	1
7	1	3		0
8	1	2		
9	2	1		
10		0		

Man kann nach [6, Abschnitt 4.5.3, Satz F] festhalten: „Am langsamsten“ ist der euklidische Algorithmus bei der Eingabe der zwei aufeinanderfolgenden Fibonacci–Zahlen F_n und F_{n-1}. Dazu bemerkt man, dass der *euklidische Schritt*, also der Schritt 2 des Algorithmus 1, auf zwei Fibonacci–Zahlen angewandt, wieder eine Fibonacci–Zahl F_{n-2} liefert. Die Kettenbruchdarstellung von F_n/F_{n-1}, die man aus den Zwischenergebnissen des euklidischen Algorithmus erhält, lautet $[1; 1, \ldots, 1]$. Die 2 im letzten Tabelleneintrag von q kommt von der Mehrdeutigkeit der Kettenbruchentwicklung, vgl. Bemerkung 2.38 (1),

$$\underbrace{[1; 1, \ldots, 1, 1, 1]}_{n \text{ Einträge}} = \underbrace{[1; 1, \ldots, 1, 2]}_{n-1 \text{ Einträge}}.$$

2.3.3 Analyse des Algorithmus 3

Die durch die Algorithmen 3 und 4 beschriebenen Abbildungen sind Homomorphismen, genaueres dazu folgt in Abschnitt 2.4. Folgendes wurde direkt (bis auf die Bezeichnungen und Indizes) aus [3, S. 38ff] übernommen.

Bemerkung 2.43 (Matrixform des EEA). Man kann den Schritt des Algorithmus 2 als

$$[a_{i+2,1}, a_{i+2,2}] = [1, -q_i] \begin{pmatrix} a_{i,1} & a_{i,2} \\ a_{i+1,1} & a_{i+1,2} \end{pmatrix} \tag{2.15}$$

schreiben.

Lemma 2.44 ([3, Lemma 6.23])**.** *Ist* $a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \equiv 0 (\mathrm{mod}\ m)$*, so ist für* $i = 3, 4, \ldots$

$$a_{i,1}a_{i+1,2} - a_{i,2}a_{i+1,1} \equiv 0 (\mathrm{mod}\ m). \tag{2.16}$$

Bemerkung 2.45. Die Interpretation von (2.16) im Sinne von Bemerkung 2.43 ist

$$\det \begin{pmatrix} a_{i,1} & a_{i,2} \\ a_{i+1,1} & a_{i+1,2} \end{pmatrix} \equiv 0 (\mathrm{mod}\ m). \tag{2.17}$$

Beweis von Lemma 2.44. Setze $q = \lfloor a_{i-1,1}/a_{i,1} \rfloor$, dann ist $a_{i,1}a_{i+1,2} - a_{i,2}a_{i+1,1} = a_{i,1}(a_{i-1,2} - qa_{i,2})a_{i+1,2} - a_{i,2}(a_{i-1,1} - qa_{i,1}) = a_{i,1}a_{i-1,2} - a_{i,2}a_{i-1,1}$. Induktiv fortgesetzt: $a_{i,1}a_{i+1,2} - a_{i,2}a_{i+1,1} = (-1)^{i-1}(a_{1,1}a_{2,2} - a_{2,1}a_{1,2}) \equiv 0 (\mathrm{mod}\ m)$. □

Proposition 2.46 (Was geschieht in Algorithmus 3? [3, Lemma 6.27])**.** *Wähle die Startmatrix als*

$$\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} = \begin{pmatrix} m & 0 \\ c & d \end{pmatrix},$$

mit $0 < c < m$ *und* $\mathrm{ggT}(c, m) = 1$*. Dann liefert der Algorithmus 3 in jedem Schritt mit*[4] $i = 1, 2, \ldots, n$*:*

$$\left| \frac{a_{i+2,2}}{a_{i+2,1}} \right|_m = \left| \frac{a_{2,2}}{a_{2,1}} \right|_m \tag{2.18}$$

bzw. falls $0 < |d| < m$ *mit* $\mathrm{ggT}(d, m) = 1$

$$\left| \frac{a_{i+2,1}}{a_{i+2,2}} \right|_m = \left| \frac{a_{2,1}}{a_{2,2}} \right|_m . \tag{2.19}$$

Beweis. Da $a_{1,1} = m$ und $a_{1,2} = 0$ sind, ist $a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \equiv 0 (\mathrm{mod}\ m)$. Nach Lemma 2.44 folgt dann für $i = 2, 3, \ldots$ dass $a_{i,1}a_{i+1,2} - a_{i+1,1}a_{i,2} \equiv 0 (\mathrm{mod}\ m)$ gilt. Also nach Satz 2.34 ist

$$\left| \frac{a_{i,1}}{a_{i,2}} \right|_m = \left| \frac{a_{i+1,1}}{a_{i+1,2}} \right|_m .$$

□

Nun kann man leicht den Satz 2.25 beweisen und somit die Korrektheit des Algorithmus 3 zeigen.

[4] Das $i + 2$ kommt von $a_{i+2} = a_i - qa_{i+1}$. Es hat keinen tiefliegenden Grund.

Beweis von Satz 2.25. Da der Algorithmus 2 in n Schritten terminiert (siehe dazu Korollar 2.10) mit $a_{n,1} = \mathrm{ggT}(s,m) = 1$ ist nach Proposition 2.46

$$\left|\frac{a_{i,2}}{a_{i,1}}\right|_m = |a_{i,2}|_m = \left|\frac{a_{2,2}}{a_{2,1}}\right|_m = \left|\frac{r}{s}\right|_m .$$

□

2.3.4 Analyse des Algorithmus 4

Zunächst wird nochmal der in diesem Abschnitt zu beweisende Satz 2.28 zitiert.

Satz. *Sei $\frac{r}{s} \in \mathbb{F}_N$ und $k := \left|\frac{r}{s}\right|_m$ mit $0 < r \leq N$ und $0 < |s| \leq N$. So existiert ein Index i, der eine (eindeutige) Zeile im Ausgabetableau des erweiterten euklidischen Algorithmus 2 mit der Startmatrix*

$$\begin{pmatrix} m & 0 \\ k & 1 \end{pmatrix}$$

liefert, sodass

$$(r,s) = (a_{i,1}, a_{i,2})$$

gilt.

Die Proposition 2.46 liefert eine zusätzliche Aussage: Verwendet man bei der Eingabe des EEA

$$\begin{pmatrix} m & 0 \\ k & 1 \end{pmatrix}$$

mit $k = |r/s|_m$ einen Bruch $a_{i,1}/a_{i,2}$ für $i = 2, \ldots, n$, so ist k kongruent modulo m zu r/s. Da eigentlich die Brüche aus $\mathbb{Q}_k$ generiert werden, gilt: *falls* man einen Farey–Bruch findet, dann entsprechen sein Zähler und Nenner paarweise dem gesuchten Paar (r,s), da es nach Satz 2.27 höchstens einen Farey–Bruch in $\mathbb{Q}_k$ gibt. Für den genaueren Beweis dieser Aussage, muß man sich in das Reich der Kettenbrüche begeben. In den folgenden Betrachtungen wird [3, S. 43] und [11, S. 54ff] gefolgt. Betrachtet man den dreispaltigen EEA mit der Startmatrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{pmatrix} = \begin{pmatrix} m & 0 & -1 \\ k & 1 & 0 \end{pmatrix}, \tag{2.20}$$

so kann man die Tripel $\{a_{i,1}, a_{i,2}, a_{i,3}\}$ folgendermaßen rekursiv definieren: Für $i = 3, 4, \ldots, n$[5] und solange $a_{i-1,1} \neq 0$ setze

$$q_i = \left\lfloor \frac{a_{i-2,1}}{a_{i-1,1}} \right\rfloor \quad \text{und} \quad a_{i,j} = a_{i-2,j} - q_i a_{i-1,j} \quad \text{mit } j = 1,2,3. \tag{2.21}$$

Proposition 2.47 ([11, Lemma 3.23], ohne Beweis)**.** *Die nach* (2.21) *definierte Sequenz*

$$\left\{ \frac{a_{i,2}}{a_{i,3}} : i = 1, \ldots, n \right\} \tag{2.22}$$

ist die vollständige Folge der Konvergenten des Kettenbruchs für $\frac{m}{k}$ mit m und k aus (2.20)*.*

[5] Wiederum ist es nur die Sache der Indizes: man braucht $i-1$ und $i-2$, und i fängt mit 1 an. Würde man es genauso wie in [3] mit $i = -1$ machen, so hätte man „Für $i = 1, 2, \ldots$“.

Korollar 2.48. *Sind $a_{i,j}$ nach* (2.20) *und* (2.21) *gegeben, so ist*

$$K_l\left(\frac{m}{k}\right) = \frac{a_{l+2,2}}{a_{l+2,3}}.$$

Lemma 2.49 (Noch mehr über EEA)**.** *Hat der dreispaltige EEA folgendes Ablauftableau [3, Exercise 1.6.2]:*

	m	1	0
	k	0	1
q_1	r_1	s_1	t_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
q_n	r_n	s_n	t_n
q_{n+1}	0	s_{n+1}	t_{n+1}

so gilt

$$r_i = ms_i + kt_i \quad \text{für } i = 1, \ldots, n+1. \tag{2.23}$$

Beweis. Man sieht leicht, dass (2.23) auch für $i = -1$ und $i = 0$ also für $r_{-1} = m$, $s_{-1} = 1$, $t_{-1} = 0$ und $r_0 = k$, $s_0 = 0$, $t_0 = 1$ erfüllt ist. Nun kann man den Induktionsschritt machen, und zwar solange man den euklidischen Schritt machen kann: Für $i = 1, \ldots, n+1$ ist

$$\begin{aligned} r_i &= r_{i-2} - q_i r_{i-1} \\ &= ms_{i-2} + kt_{i-2} - q_i(ms_{i-1} + kt_{i-1}) \\ &= m(\underbrace{s_{i-2} - q_i s_{i-1}}_{=:s_i}) + k(\underbrace{t_{i-2} - q_i t_{i-1}}_{=:t_i}). \end{aligned}$$

□

Beweis von Satz 2.28. Der Beweis wird [3, S. 44–46] entnommen. Erweitere die Startmatrix

$$\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} = \begin{pmatrix} m & 0 \\ k & 1 \end{pmatrix}$$

des EEA mit passendem $[a_{1,3}, a_{2,3}]^T$ aus der Folge $[a_{i,3} : i = 1, \ldots, n]$, die (2.21) erfüllt. Mit der Proposition 2.47 ist (2.22) die vollständige Folge der Konvergenten des Kettenbruchs für $\frac{m}{k}$ (falls $k \neq 0$). Man kann die Annahme umschreiben: Sei t die eindeutige ganze Zahl, so dass

$$\left|\frac{r}{s}\right|_m = |k|_m \Leftrightarrow r \equiv ks (\mathrm{mod}\ m) \Leftrightarrow r = ks - mt$$

gilt.
Nun kann man schreiben

$$\left|\frac{k}{m} - \frac{t}{s}\right| = \left|\frac{ks - mt}{ms}\right| = \left|\frac{r}{ms}\right| \leq \frac{1}{s^2}\frac{|s|N}{2N^2+1} \leq \frac{1}{s^2}\frac{N^2}{2N^2+1} \leq \frac{1}{2s^2}.$$

Jetzt ist nach Satz 2.39, genauer gesagt nach (2.12), entweder $\frac{t}{s}$ oder $\frac{-t}{-s}$ eine Konvergente des Kettenbruchs für $\frac{k}{m}$. Da

$$\left\{\frac{a_{i,2}}{a_{i,3}} : i = 1, \ldots, n\right\} \tag{2.22}$$

die vollständige Folge der Konvergenten des Kettenbruchs für $\frac{m}{k}$ ist, schreibt man

$$\left\{\frac{0}{1}, \frac{a_{1,3}}{a_{1,2}}, \ldots, \frac{a_{n,3}}{a_{n,2}}\right\} \tag{2.24}$$

für die Folge der Konvergenten des Kettenbruchs für $\frac{k}{m}$. Dieses Ergebnis ist leicht zu rechtfertigen [11]: Vergleiche den ersten Schritt von EEA für $\frac{m}{k}$ und für $\frac{k}{m}$ samt ihrer Konvergenten. Beachte $k \in \mathbb{Z}_m$!

$a_{i,1}$	$a_{i,2}$	$a_{i,3}$
k	0	-1
m	1	0
k	0	-1

Folglich muss man nur die Spalten zwei und drei vertauschen, sowie die *beiden* Vorzeichen umkehren. Somit kann man in (2.24) einen Index i zwischen 1 und n finden, sodass $|a_{i,2}| = |s|$ und

$$\frac{t}{s} = \frac{a_{i,3}}{a_{i,2}}$$

gilt. Mit Lemma 2.49 folgt

$$\frac{a_{i,1}}{a_{i,2}} = k - m\frac{a_{i,3}}{a_{i,2}}$$

bzw. wegen $r = ks - mt$ folgt

$$\frac{r}{s} = k - m\frac{t}{s}.$$

Somit ist

$$\frac{a_{i,1}}{a_{i,2}} = \frac{r}{s}$$

bzw., da beides r und $a_{i,1} > 0$, paarweise $(r, s) = (a_{i,1}, a_{i,2})$. □

2.3.5 Interpretation der Ergebnisse oder „Was man alles gesehen hat?“

Der EEA liefert robuste Methoden für

- Berechnung des ggT und der Bézout–Koeffizienten,
- Berechnung der multiplikativen Inversen modulo m,
- Darstellung der Brüche modulo m als Restklassen $\hat{\mathbb{Z}}_m$,
- Berechnung der Abbildung EEA in $\hat{\mathbb{Z}}_m$,
- Berechnung der inversen Abbildung EEA^{-1}.

2.4 Homomorphismen

Definition 2.50 (Ringhomomorphismus)**.** Seien R, S Ringe mit Eins. Die Abbildung $\varphi : R \to S$ heißt ein *(Ring–)Homomorphismus*, falls für alle a, b in R gilt [17, Ring Homomorphism] bzw. [15, § 12, S. 40]:

1. $\varphi(a + b) = \varphi(a) + \varphi(b)$.

2. $\varphi(ab) = \varphi(a)\varphi(b)$.
3. $\varphi(0_R) = 0_S$.
4. $\varphi(1_R) = 1_S$.

Hierbei bezeichnet man mit 0_R usw. das entsprechende Element in dem jeweiligen Ring.

Lemma 2.51 (Homomorphismus von einem Homomorphismus ist ein Homomorphismus)**.** *Seien R, S, T Ringe mit Eins. Seien $\phi : R \to S$ und $\psi : S \to T$ Homomorphismen. Dann ist die Abbildung $\phi \circ \psi =: \chi$ mit $\chi : R \to T$ ein Homomorphismus.*

Beweis. Weise die entsprechenden Eigenschaften nach. Bezeichne mit $*_X$ die entsprechende Operation in dem Ring X. Seien $x, y \in R$ und $* \in \{+, \cdot\}$. Es gilt $\chi(x *_R y) = \psi(\phi(x *_R y)) = \psi(\phi(x) *_S \psi(y)) = \psi(\phi(x)) *_T \psi(\phi(y)) = \chi(x) *_T \chi(y)$. Bezeichne noch mit $e_X \in \{0_X, 1_X\}$ das entsprechende neutrale Element des Ringes X. Dann ist $\chi(e_R) = \psi(\phi(e_R)) = \psi(e_S) = e_T$. □

Definition 2.52 (Isomorphismus)**.** Ein *Isomorphismus* ist ein bijektiver Homomorphismus.

Korollar 2.53. *Ein Isomorphismus von Isomorphismus ein Isomorphismus.*

Proposition 2.54. *Die Abbildung $|\cdot|_m : \mathbb{Z} \to \mathbb{Z}_m$ mit $m > 1$ ist ein Homomorphismus.*

Beweis. Man weist einfach die Eigenschaften aus der Definition 2.50 nach, die explizite Behandlung von Restklassenringen und Abbildungen in diese findet sich in [15, § 15].

- 1 und 2 folgen aus dem Satz 1.10. Ist φ der Homomorphismus und ist $* \in \{+, \cdot\}$, sowie $\circledast \in \{\oplus, \odot\}$, so:

$$\varphi(a * b) = ||a|_m * |b|_m|_m = |a|_m \circledast |b|_m = \varphi(a) \circledast \varphi(b).$$

- zu 3: $\varphi(0) = |0|_m = 0 \in \mathbb{Z}_m$, zu 4: $\varphi(1) = |1|_m = 1 \in \mathbb{Z}_m$, daher die Forderung $m > 1$.

□

Proposition 2.55. *Die Abbildung $EEA : \mathbb{Q} \supset \mathbb{F}_N \to \hat{\mathbb{Z}}_m$, die durch den Algorithmus 3 mit $\frac{r}{s} \in \mathbb{F}_N$ und mit der Startmatrix*

$$\begin{pmatrix} m & 0 \\ s & r \end{pmatrix}$$

realisiert wird, ist ein Homomorphismus.

Beweis. Nach Satz 2.25 ist $EEA_m\left(\frac{a}{b}\right) = |ab^{-1}|_m \in \hat{\mathbb{Z}}_m$. Bleibt noch, die Homomorphismuseigenschaften zu zeigen. Da man in Brüchen modulo m kürzen kann, da $\mathbb{Z}_m$ ein kommutativer Ring ist und nach Satz 1.10 mit $\frac{a}{b}, \frac{c}{d} \in \mathbb{F}_N$ mit zu $m > 1$ und entsprechendem N folgt:

- $EEA_m\left(\frac{a}{b} + \frac{c}{d}\right) = EEA_m\left(\frac{ad+bc}{bd}\right) = |(ad+bc)(bd)^{-1}|_m = |ab^{-1}+cd^{-1}|_m = |ab^{-1}|_m \oplus |cd^{-1}|_m = EEA_m\left(\frac{a}{b}\right) \oplus EEA_m\left(\frac{c}{d}\right)$.
- $EEA_m\left(\frac{a}{b}\frac{c}{d}\right) = |ac(bd)^{-1}|_m = |ab^{-1}|_m \odot |cd^{-1}|_m = EEA_m\left(\frac{a}{b}\right) \odot EEA_m\left(\frac{c}{d}\right)$
- $EEA_m(0) = |0|_m = 0 \in \mathbb{Z}_m$,
- $EEA_m(1) = |1|_m = 1 \in \mathbb{Z}_m$.

□

Satz 2.56. *Die Abbildung* $EEA^{-1} : \hat{\mathbb{Z}}_m \ni x \to \frac{a}{b} \in \mathbb{F}_N$ *ist ein Homomorphismus.*

Der Beweis ist analog zu dem von Proposition 2.55, aber nicht mit Satz 2.25, sondern mit Satz 2.28.

Korollar 2.57. *Die Abbildung* $/\cdot/_m : \mathbb{Z} \to \mathbb{S}_m$ *für ungerade* $m > 1$ *ist ein Homomorphismus.*

Beweis. Es ist $/\cdot/_m = |\cdot|_m \circ \sigma$ mit σ aus dem Beweis von Satz 1.18. □

Korollar 2.58. *Die Abbildung* $\mathbb{F}_N \to \mathbb{S}_m$ *ist ein Homomorphismus.*

Beweis. Stelle die gesuchte Abbildung als $EEA \circ |\cdot|_m \circ \sigma$ dar. □

$$\mathbb{Q} \xrightarrow{\supset} \mathbb{F}_N \overset{EEA}{\longleftrightarrow} \hat{\mathbb{Z}}_M \xrightarrow{\supset} \mathbb{Z}_M \overset{\cong}{\longleftrightarrow} \mathbb{S}_M$$

Abbildung 2.1: Zusammenfassung der Kapitel 1 und 2.

Kapitel 3

Ganzzahliges multimodulares Rechnen

Come, let us hasten to a higher plane,
Where dyads tread the fairy fields of Venn,
Their indices bedecked from one to n,
Commingled in an endless Markov chain!

Stanislaw Lem, *Cyberiad*[a]

[a] Im Original ist natürlich kein Wort über „from one to n“:
Nieśbmiały cybernetyk potężcne ekstrema
Poznawał, kiedy grupy unimodularne
Cyberiady całkował w popołudnie parne,
Nie wiedząc, czy jest miłość, czy jeszcze jej nie ma?

Man geht „from one to n“ und betrachtet in diesem Kapitel die Restklassen modulo *mehrerer* Zahlen. Im Moment geht es noch um die ganzzahlige Darstellung und nicht ohne Grund: die Darstellung der Brüche mit mehreren Moduli hat eigene Schwierigkeiten und wird in Kapitel 4 betrachtet. Bevor aber das eigentliche Ziel dieser theoretischen Abhandlungen verfolgt werden kann, müssen viele Feinheiten der multimodularen Rechnung zuerst in der „harmloseren“ ganzzahligen Variante ausgearbeitet werden, und außerdem muss relativ viel neue Terminologie und etwas Methodologie eingeführt werden. Die Vorwärtsabbildung fällt leicht, aber in die Rückwärtsabbildung muss etwas mehr Arbeit investiert werden. Die nötige Richtung gibt der Chinesische Restsatz an, wobei die eigentliche Methode, die für die *Rekonstruktion* verwendet wird, auf H. L. Garner[1] zurückgeht. Diese verwendet die mixed–radix Darstellung, auch als die Darstellung mit gemischten Basen bekannt.

3.1 Allgemeines

Definition 3.1 (Standarddarstellung)**.**

1. Das n–Tupel
$$\beta = [m_1, \ldots, m_n]$$
mit $m_i \neq m_j$ und $\mathrm{ggT}(m_i, m_j) = 1$ für $i \neq j$ heißt *Modulvektor* des Restklassensystems.

[1] H. L. Garner, *IRE Trans.*, **EC–8** (1959), 140–147, zitiert in [6].

2. Zu jeder ganzen Zahl $s \in \mathbb{Z}$ bezeichne der Vektor

$$|s|_\beta = [|s|_{m_1}, \ldots, |s|_{m_n}] \tag{3.1}$$

die *Standarddarstellung* von s bezüglich des Modulvektors β.

3. Die einzelnen Restklassen $|s|_{m_i}$ heißen *Standardrestklassenziffern* von s bezüglich β.

4. Das *Standardrestklassenzahlensystem* ist $\mathbb{Z}_\beta := \{|s|_\beta : s \in \mathbb{Z}\}$.

Sind alle m_i in β ungerade, so kann man wieder zu den symmetrischen Restklassen umschalten. Als Elemente von β sind am besten Primzahlen zu wählen.

Definition 3.2 (symmetrische Darstellung)**.** Seien alle Elemente von β ungerade.

1. Analog zu der Definition 3.1 definiert man die *symmetrische Restklassendarstellung* von s bezüglich β mit

$$/s/_\beta = [/s/_{m_1}, \ldots, /s/_{m_n}], \tag{3.2}$$

wobei β ein Modulvektor ist.

2. Die einzelnen Restklassen $/s/_{m_i}$ heißen *symmetrische Restklassenziffern.*

3. Das *symmetrische Restklassenzahlensystem* ist $\mathbb{S}_\beta := \{/s/_\beta : s \in \mathbb{Z}\}$.

Bemerkung 3.3. Sei $\beta = [m_1, \ldots, m_n]$.

1. Das Produkt der Einträge von β wird als $M = m_1 \cdots m_n$ bezeichnet.

2. Zum gegebenen Modulvektor β seien die *unimodularen* Restklassenzahlensysteme $\mathbb{Z}_M$ und $\mathbb{S}_M$.

Beispiel 3.4.

$$\mathbb{S}_M = \left\{ -\frac{M-1}{2}, \ldots, -2, -1, 0, 1, 2, \ldots, \frac{M-1}{2} \right\}.$$

3.2 Der Chinesische Restsatz

Der Satz wird präsentiert nach [6, Abschnitt 4.3.2, Satz C].

Satz 3.5 (Chinesischer Restsatz)**.** *Seien $m_1, \ldots, m_n$ positive ganze koprime Zahlen und $M = m_1 \cdots m_n$. Seien noch α, $u_1, \ldots, u_n \in \mathbb{Z}$. Dann existiert genau eine ganze Zahl u, die die Bedingungen*

$$\alpha \leq u < \alpha + M \tag{3.3}$$

$$u \equiv u_i (\text{mod } m_i) \quad \text{für } i = 1, \ldots, n \tag{3.4}$$

erfüllt. Man wählt meist $\alpha = 0$.

Der folgende Beweis dieses Satzes findet sich in zahlreichen Büchern und wird auf sehr verschiedene Art und Weise ausgeführt. Der hier ausgeführte „Interpolationsbeweis“ stammt aus [11, Satz 2.43, Seite 41]. Dabei wird die Behauptung verallgemeinert:

Satz. *Seien R, $R_1, \ldots, R_n$ euklidische Ringe, sowie $m_1, \ldots, m_n \in R$ und M wie in der Behauptung. Dann gilt*

$$R/\langle M\rangle \cong R_1/\langle m_1\rangle \times \cdots \times R_n/\langle m_n\rangle \tag{3.5}$$
$$(R/\langle M\rangle)^* \cong (R_1/\langle m_1\rangle)^* \times \cdots \times (R_n/\langle m_n\rangle)^* . \tag{3.6}$$

Beweis. Sei $\mu_i = M/m_i = \prod_{j\neq i} m_j$ und

$$l_i = \left||\mu_i^{-1}|_{m_i}\mu_i\right|_M$$

für $i = 1, \ldots, n$. Für $q, q' \in R$ gilt $q \equiv q' (\mathrm{mod}\ M) \Leftrightarrow M \mid (q - q') \Leftrightarrow m_i \mid (q - q')$ für $i = 1, \ldots, n$. Also gilt für die lineare Abbildung

$$L : R/\langle M\rangle \to R_1/\langle m_1\rangle \times \cdots \times R_n/\langle m_n\rangle$$
$$L : q \mapsto [|q|_{m_1}, \ldots, |q|_{m_n}] =: [r_1, \ldots, r_n]$$

dass $\ker L = MR = \langle M\rangle$ gilt. L ist surjektiv, denn es kann zu jedem Satz $[r_1, \ldots, r_n]$ so ein q angegeben werden, dass $|q|_{m_i} \equiv r_i (\mathrm{mod}\ m_i)$, und zwar

$$q = \sum_{i=1}^{n} l_i r_i. \tag{3.7}$$

Somit gilt die Behauptung und L ist der erwünschte Homomorphismus. □

H. Cohen gibt zwei Algorithmen [1, Algorithmus 1.3.11] und [1, Algorithmus 1.3.12] zur Implementierung des Chinesischen Restsatzes an. Im Folgenden verfolgt man einen anderen Ansatz: Die im Abschnitt 3.5 dargestellte Vorgehensweise ist ein alternativer Beweis des Chinesischen Restsatzes.

3.3 Die Arithmetik

Bezeichne mit jeweils $\boxplus, \boxminus, \boxdot, \boxslash$ die entsprechende arithmetische Operation modulo β.

Satz 3.6 (Mehrmodulare Arithmetik in $\mathbb{Z}_\beta$, Teil 1)**.** *Seien a und b ganze Zahlen bzw. $x := |a|_\beta$ und $y := |b|_\beta$. Es gilt*

$$\begin{aligned} &x \boxplus y := |a + b|_\beta = [r_1, \ldots, r_n] \\ &\text{mit } r_i = |a|_{m_i} \oplus |b|_{m_i} \text{ für alle } i = 1, \ldots, n \end{aligned} \tag{3.8}$$

$$\begin{aligned} &x \boxminus y := |a - b|_\beta = [s_1, \ldots, s_n] \\ &\text{mit } s_i = |a|_{m_i} \ominus |b|_{m_i} \text{ für alle } i = 1, \ldots, n \end{aligned} \tag{3.9}$$

$$\begin{aligned} &x \boxdot y := |ab|_\beta = [t_1, \ldots, t_n] \\ &\text{mit } t_i = |a|_{m_i} \odot |b|_{m_i} \text{ für alle } i = 1, \ldots, n \end{aligned} \tag{3.10}$$

Beweis. Anwendung des Satzes 1.10 auf jede Komponente. □

Bemerkung 3.7. Mit Satz 3.6 ist die Addition, Subtraktion oder Multiplikation modulo β nur die Ausführung für $i = 1, \ldots, n$ der entsprechenden Grundrechenart für jedes $|a|_{m_i}$ und $|b|_{m_i}$ *wie normal* in $\mathbb{Z}$ mit anschließender Reduktion modulo m_i.

Definition 3.8 (Multiplikative Inverse)**.** Sei $a \in \mathbb{Z}$ und angenommen $a^{-1}(\mathrm{mod}\ m_i)$ existiert stets für alle $i = 1, \ldots, n$. So heißt das n–Tupel

$$a^{-1}(\mathrm{mod}\ \beta) = \left[a^{-1}(\mathrm{mod}\ m_1), \ldots, a^{-1}(\mathrm{mod}\ m_n)\right] \tag{3.11}$$

die *Standardrestklassendarstellung der multiplikativen Inversen* von a bezüglich des Modulvektors β. Es existiert genau dann, wenn alle seine Elemente existieren.

Lemma 3.9 (Die Inverse ist eine Inverse)**.** *Sei $|a|_\beta \in \mathbb{Z}_\beta$ und existiere die Inverse $|a^{-1}|_\beta$. Es gilt*

$$|a|_\beta \boxdot |a^{-1}|_\beta = |a^{-1}|_\beta \boxdot |a|_\beta = 1 \in \mathbb{Z}_\beta. \tag{3.12}$$

Beweis. Sei also $n = \#\beta$, $a \in \mathbb{Z}$ und $|a|_\beta \in \mathbb{Z}_\beta$ mit $\mathrm{ggT}(a, m_i) = 1$ für alle $i = 1, \ldots, n$. Nun ist

$$|a^{-1}|_\beta \boxdot |a|_\beta = \left[|a^{-1}|_{m_i} \odot |a|_{m_i} : i = 1, \ldots, n\right] = [|1|_{m_i} : i = 1, \ldots, n] = 1 \in \mathbb{Z}_\beta.$$

□

Satz 3.10 (Mehrmodulare Arithmetik in $\mathbb{Z}_\beta$, Teil 2)**.** *Seien a und b ganze Zahlen und existiere $b^{-1}(\mathrm{mod}\ \beta)$. Dann gilt*

$$\begin{aligned} |a/b|_\beta &= [c_1, \ldots, c_n] \\ \text{mit } c_i &= |a|_{m_i} \oslash |b|_{m_i} \\ &= |a|_{m_i} \odot \left|b^{-1}\right|_{m_i} \text{ für alle } i = 1, \ldots, n. \end{aligned} \tag{3.13}$$

Bemerkung 3.11.

1. Wie gewohnt, ist $\left|a/a\right|_\beta = [1, \ldots, 1]$.
2. Es ist die *ganzzahlige* Division, man kann nur mit *Einheiten* $b \in \mathbb{Z}_\beta^* = \{b \in \mathbb{Z}_\beta : b^{-1} \in \mathbb{Z}_\beta\}$ rechnen.

Beweisidee vom Satz 3.10. Man rechnet komponentenweise mit den multiplikativen Inversen von b modulo m_i. †

Beispiel 3.12 (Multimodulare Arithmetik)**.** *Sei $\beta = [3, 5, 7]$. Dann ist $M = 105$. Seien $a = 18$, $b = 26$. Dann ist $|a|_\beta = [0, 3, 4]$ und $|b|_\beta = [2, 1, 5]$.*

1. *$|a|_\beta \boxplus |b|_\beta = |a + b|_\beta = [|0 + 2|_3, |3 + 1|_5, |4 + 5|_7] = [2, 4, 2]$. Es ist auch $|44|_\beta = [2, 4, 2]$.*
2. *$|a|_\beta \boxminus |b|_\beta = |a - b|_\beta = [|0 - 2|_3, |3 - 1|_5, |4 - 5|_7] = [1, 2, 6]$. Es ist auch $|-8|_\beta = [1, 2, 6]$.*
3. *$|a|_\beta \boxdot |b|_\beta = |ab|_\beta = [|0 \cdot 2|_3, |3 \cdot 1|_5, |4 \cdot 5|_7] = [0, 3, 6]$. Es ist sowohl: $|468|_\beta = [0, 3, 6]$, als auch $|48|_\beta = [0, 3, 6]$. Allerdings ist $468 \equiv 48 (\mathrm{mod}\ 105)$.*
4. *Sei $a = 52$, b und β wie vorher. Es ist $|a|_\beta = [1, 2, 3]$. Also $|a|_\beta \boxslash |b|_\beta = \left||a|_\beta b^{-1}(\mathrm{mod}\ \beta)\right|_\beta = \left|[1, 2, 3] \cdot \left[2^{-1}(\mathrm{mod}\ 3), 1^{-1}(\mathrm{mod}\ 5), 5^{-1}(\mathrm{mod}\ 7)\right]\right|_\beta = |[1, 2, 3] \cdot [2, 1, 3]|_\beta = [2, 2, 2]$. Klar: $|2|_\beta = [2, 2, 2]$.*

3.4 Isomorphie

Proposition 3.13. *Zwei ganze Zahlen a und b haben die gleiche multimodulare Standarddarstellung $|a|_\beta = |b|_\beta$ bezüglich eines Modulvektors β genau dann, wenn $a \equiv b (\mathrm{mod}\ M)$ mit $M = m_1 \cdots m_n$.*

Beweis. $a \equiv b (\mathrm{mod}\ M) \Leftrightarrow M \mid (a-b)$. Wegen $m_i \mid M$, gilt $m_i \mid (a-b)$ für alle $i = 1, \ldots, n$, aus $m_i \mid (a-b)$ für alle $i = 1, \ldots, n$ folgt $M \mid (a-b)$. Also $M \mid (a-b) \Leftrightarrow$ für alle $i = 1, \ldots, n$ $a \equiv b (\mathrm{mod}\ m_i) \Leftrightarrow |a|_{m_i} = |b|_{m_i}$ für alle $i = 1, \ldots, n \Leftrightarrow |a|_\beta = |b|_\beta$. □

Korollar 3.14. *Aus $a \equiv b (\mathrm{mod}\ M)$ folgt $|a|_\beta = |b|_\beta$.*

Proposition 3.15. *Die Abbildung $|\cdot|_\beta$ ist ein Homomorphismus.*

Beweis. Alle arithmetischen Operationen werden unabhängig in jeder Komponente ausgeführt, also entspricht die Operation $\circledast \in \{\boxplus, \boxminus, \boxdot, \boxslash\}$ eingeschränkt auf die i–te Komponente der Operation $\circledast \in \{\oplus, \ominus, \odot, \oslash\}$ Da aber die Abbildung $|\cdot|_{m_i} : \mathbb{Z} \to \mathbb{Z}_{m_i}$ ein Homomorphismus für $i = 1, \ldots, n$ ist, ist auch $|\cdot|_\beta : \mathbb{Z} \to \mathbb{Z}_\beta$ mit $\beta = \{m_i : i = 1, \ldots, n\}$ ein Homomorphismus.

Man kann die Behauptung auch anders beweisen: Man wendet die Proposition 2.54 auf jede Komponente an. Dies wird später der Standardbeweisansatz im multimodularen Fall. □

Satz 3.16 ([3, S. 13])**.** *Die endlichen kommutativen Ringe mit Eins $(\mathbb{Z}_M, \oplus, \odot)$, $(\mathbb{Z}_\beta, \boxplus, \boxdot)$, $(\mathbb{S}_M, \oplus, \odot)$ und $(\mathbb{S}_\beta, \boxplus, \boxdot)$ sind isomorph zueinander.*

Beweisskizze.

1. Die o. g. Strukturen *sind* endliche kommutative Ringe mit Eins.
2. Sie sind isomorph zueinander.
   (a) $(\mathbb{Z}_M, \oplus, \odot) \cong (\mathbb{S}_M, \oplus, \odot)$ – bereits gezeigt in Satz 1.7.
   (b) $(\mathbb{Z}_M, \oplus, \odot) \cong (\mathbb{Z}_\beta, \boxplus, \boxdot)$ – bereits gezeigt mit dem Chinesischen Restsatz 3.5.
   (c) Alles andere folgt daraus.

†

Korollar 3.17. *Alle im Satz 3.16 aufgeführten Mengen haben M Elemente.*

3.5 Mixed–radix Darstellungen

Wie in Satz 3.16 dargestellt, sind $\mathbb{Z}_M$ und $\mathbb{Z}_\beta$ isomorph. Es ist aber noch nicht klar, wie die Standardmultimodulardarstellung in eine eindeutige ganze Zahl modulo M abgebildet wird. Einer der ältesten, vgl. [3, S. 17] bzw. [6, Abschnitt 4.3.2], Ansätze ist der *Chinesische Restsatz 3.5.* Gregory und Krishnamurthy [3] beschreiben ein anderes Verfahren, welches die *mixed–radix Darstellungen*[2] der ganzen Zahlen verwendet.

Definition 3.18 (Basenvektor)**.** Sei

$$\rho = [r_1, \ldots, r_n]$$

ein n–Tupel, seine Komponenten bezeichnet man als *Basen*, ρ selbst heißt *Basenvektor*. Sei noch $R := r_1 \cdots r_n$.

[2] Im Englischen heißt *radix* in diesem Fall die Basis, die mixed-radix Darstellung ist somit eine *Darstellung mit gemischten Basen*. In dieser Abhandlung wird „mixed-radix Darstellung“ verwendet.

Proposition 3.19. *Man kann jede ganze Zahl* $s \in \mathbb{Z}$ *mit* $0 \leq s < R$ eindeutig *als*

$$s = d_0 + d_1 r_1 + d_2 r_1 r_2 + \cdots + d_{n-1} r_1 r_2 \ldots r_{n-1}, \tag{3.14}$$

darstellen, wobei $d_0, \ldots, d_{n-1}$ *die* standard mixed–radix Ziffern *sind, die*

$$0 \leq d_i < r_{i+1} \text{ mit } i = 0, \ldots, n-1 \tag{3.15}$$

erfüllen.

Definition 3.20 (Mixed–radix Darstellung)**.** Die *Ziffernfolge* $d_0, \ldots, d_{n-1}$ nach (3.15) zur ganzen Zahl s und Basenvektor ρ bildet die *mixed–radix Darstellung* von $0 \leq s < R$ bezüglich ρ, in Zeichen:

$$\langle s \rangle_\rho = \langle d_0, \ldots, d_{n-1} \rangle. \tag{3.16}$$

Die Menge solcher Darstellungen wird mit $\mathbb{B}_\rho$ bezeichnet.

Bemerkung 3.21. Im Falle $r_1 = \cdots = r_n =: r$ ergibt sich die gewohnte r–adische Darstellung von s (zu Basis r).

Beweis der Proposition 3.19. Sei ρ gegeben und sei s so, dass $0 \leq s < R$ gilt. Definiere $R_i := r_1 \cdots r_i$ für $i = 1, \ldots, n$. Dann ist $r_{i-1} = R_i / r_i$. Nun folgt

$$\begin{aligned}
d_{n-1} &:= \left\lfloor \frac{s}{R_{n-1}} \right\rfloor, & s &\leftarrow s \pmod{R_{n-1}} \\
d_{n-2} &:= \left\lfloor \frac{s}{R_{n-2}} \right\rfloor, & s &\leftarrow s \pmod{R_{n-2}} \\
&\vdots & &\vdots \\
d_1 &:= \left\lfloor \frac{s}{R_1} \right\rfloor, & s &\leftarrow s \pmod{R_1} \\
d_0 &:= s
\end{aligned}$$

Solch eine Darstellung existiert stets, falls die Bedingung an s nicht verletzt ist, und zu festem ρ liefern verschiedene Werte von d_i, $i = 1, \ldots, n$ verschiedene Werte für s. Bleibt zu zeigen: (3.15) ist erfüllt. Klar: $0 \leq s < R_n = R$. Nun ist

$$d_{n-1} := \left\lfloor \frac{s}{R_{n-1}} \right\rfloor \leq r_n.$$

Jetzt $s \leftarrow s \mod R_{n-1}$, also $0 \leq s < R_{n-1}$. Man wiederhole diese Überlegung iterativ. □

Ein interessanterer Fall ist $\rho = \beta$. Damit hat eine ganze Zahl s zwei verschiedene Darstellungen: die „normale“ multimodulare Darstellung

$$|s|_\beta = [|s|_{m_1}, \ldots, |s|_{m_n}] \tag{3.1}$$

und die mixed–radix Darstellung

$$\langle s \rangle_\beta = \langle d_0, \ldots, d_{n-1} \rangle, \tag{3.16}$$

mit der man (hoffentlich) s wiederherstellen kann

$$s = d_0 + d_1 m_1 + d_2 m_1 m_2 + \cdots + d_{n-1} m_1 m_2 \cdots m_{n-1}. \tag{3.14}$$

Wie kann (3.1) in (3.16) abgebildet werden? Zu jedem $i = 1, \ldots, n$ ist $|s|_{m_i}$ gegeben, und man will daraus d_{i-1} erhalten. Dazu wird erst einmal festgestellt, dass $|s|_{m_i}$ und d_{i-1} im selben Intervall $[0, m_i - 1]$ liegen. Nun setze

$$\begin{aligned} t_1 &:= s \\ &= d_0 + m_1 \left(d_1 + d_2 m_2 + \cdots + d_{n-1} m_2 \cdots m_{n-1}\right) \\ &= d_0 + m_1 t_2, \end{aligned}$$

also

$$|t_1|_{m_1} = |d_0 + m_1 t_2|_{m_1} = |d_0|_m = |s|_{m_1}.$$

Setzt man dies fort, so erhält man

$$d_i = t_{i+1} = \left| \frac{t_i - d_{i-1}}{m_i} \right|_{m_{i+1}} \qquad i = 1, \ldots, n. \tag{3.17}$$

Das Problem ist, dass dies *nicht* in der gewöhnlichen Arithmetik berechnet werden kann, schließlich will man gerade s aus der mixed-radix Darstellung errechnen, man kennt es nicht! Aber man kennt $|s|_\beta$, und dieses Wissen genügt. Nochmals: *man rechnet hier in den Restklassenringen!*

Bemerkung 3.22. Mit β^i wird hier der *um i verkürzte* Modulvektor $[m_{i+1}, \ldots, m_n]$ notiert. Trivialerweise ist $\beta^0 = \beta$.

Algorithmus 5 (Berechnung von mixed-radix Darstellung)**.**
Gegeben: $|s|_\beta$.

1. *Setze* $t^{(1)} \leftarrow |s|_\beta$ *und* $i \leftarrow 1$. *Setze*

$$d_0 \leftarrow t_1^{(1)}, \tag{3.18}$$

2. *Wiederhole*

$$t^{(i+1)} \leftarrow \left| \frac{t^{(i)} - |d_{i-1}|_{\beta^i}}{m_i} \right|_{\beta^i} \tag{3.19}$$

$$\begin{aligned} d_i &\leftarrow t_{i+1}^{(i+1)} \\ i &\leftarrow i + 1 \end{aligned} \tag{3.20}$$

solange $i < n$.

Ergebnis: $\langle d_0, \ldots, d_{n-1} \rangle$.

Bemerkung 3.23 (Zu Algorithmus 5).

1. Der Algorithmus 5 ist eine effiziente Realisierung der Berechnung nach (3.17). Bei der Implementierung genügt es *einen* Vektor t zu benutzen, der in (3.19) stets überschrieben und verkürzt wird. Bei festem β können die Inversen von m_i in (3.19) vorberechnet werden.

2. Die Gleichung (3.18) gilt wegen

$$\left|t^{(1)}\right|_\beta = \left[t_1^{(1)}, t_2^{(1)}, \ldots, t_n^{(1)}\right] = \left[|d_0|_{m_1}, \left|t_2^{(1)}\right|_{m_2}, \ldots, \left|t_n^{(1)}\right|_{m_n}\right].$$

   Die Gleichung (3.20) setzt $|d_i|_{\beta^i} = \left[|d_i|_{m_{i+1}}, \ldots, |d_i|_{m_n}\right]$, welches im nächsten Schritt bei (3.19) verwendet wird. Allerdings braucht man im Endergebnis lediglich $|d_i|_{m_{i+1}}$ für $i = 0, \ldots, n-1$.

3. Die Gleichung (3.19) gilt wegen (3.17), da $|d_{i-1}|_{m_i} = t_i^{(i)}$ gilt, ist

$$\left|t^{(i)} - |d_{i-1}|_{\beta^{i-1}}\right|_{\beta^{i-1}} = \underbrace{\left[|0|_{m_{i-1}}, \left|t_i^{(i)} - d_{i-1}\right|_{m_i}, \ldots, \left|t_n^{(i)} - d_{i-1}\right|_{m_n}\right]}_{n-i+1 \text{ Elemente}}$$
$$\cong \underbrace{\left[\left|t_i^{(i)} - d_{i-1}\right|_{m_i}, \ldots, \left|t_n^{(i)} - d_{i-1}\right|_{m_n}\right]}_{n-i \text{ Elemente}} = \left|t^{(i)} - |d_{i-1}|_{\beta^i}\right|_{\beta^i},$$

sowie $m_i^{-1} (\mathrm{mod}\ \beta^i) = \left[m_i^{-1} (\mathrm{mod}\ m_{i+1}), \ldots, m_i^{-1} (\mathrm{mod}\ m_n)\right]$.

4. Aus dem Ergebnis $\langle s \rangle_\beta$ des Algorithmus 5 errechnet man leicht s (eigentlich: $|s|_M$, mit großem M), man kann (3.14) rekursiv umschreiben:

$$\begin{aligned} s_1 &= d_{n-1} \\ s_2 &= d_{n-2} + s_1 m_{n-1} \\ &\vdots \\ s_n &= d_0 + s_{n-1} m_1 \\ s &= s_n. \end{aligned} \tag{3.21}$$

Die Art des Algorithmus 5 wird im folgenden Beispiel verdeutlicht.

Beispiel 3.24. *Sei* $\beta = [5, 7, 11, 13]$ *und* $|s|_\beta = |-6|_\beta = [4, 1, 5, 7]$. *Berechne* $\langle s \rangle$ *und* $|s|_M$.

	$\lvert\cdot\rvert_5$	$\lvert\cdot\rvert_7$	$\lvert\cdot\rvert_{11}$	$\lvert\cdot\rvert_{13}$
$t = t^{(1)} = s$	**4**	1	5	7
$d_0 = 4$				
$\lvert d_0\rvert_{\beta^0}$	4	4	4	4
$\left\lvert t^{(1)} - \lvert d_0\rvert_{\beta^1}\right\rvert_{\beta^1}$		4	1	3
$\lvert m_1^{-1}\rvert_{\beta^1}$		3	9	8
$t = t^{(2)} = \left\lvert \frac{t^{(1)} - \lvert d_0\rvert_{\beta^1}}{m_1} \right\rvert_{\beta^1}$		**5**	9	11
$d_1 = 5$				
$\lvert d_1\rvert_{\beta^1}$		5	5	5
$\left\lvert t^{(2)} - \lvert d_1\rvert_{\beta^2}\right\rvert_{\beta^2}$			4	6
$\lvert m_2^{-1}\rvert_{\beta^2}$			8	2
$t = t^{(3)} = \left\lvert \frac{t^{(2)} - \lvert d_1\rvert_{\beta^2}}{m_2} \right\rvert_{\beta^2}$			**10**	12
$d_2 = 10$				
$\lvert d_2\rvert_{\beta^2}$			10	10
$\left\lvert t^{(3)} - \lvert d_2\rvert_{\beta^3}\right\rvert_{\beta^3}$				2
$\lvert m_3^{-1}\rvert_{\beta^3}$				6
$t = t^{(4)} = \left\lvert \frac{t^{(3)} - \lvert d_2\rvert_{\beta^3}}{m_3} \right\rvert_{\beta^3}$				**12**
$d_3 = 12$				

Somit ist $\langle s \rangle_\beta = \langle 4, 5, 10, 12 \rangle$. *Mit* $M = 5005$ *liefert es* $|s|_{5005} = 4 + 5 \cdot 5 + 10 \cdot 5 \cdot 7 + 12 \cdot 5 \cdot 7 \cdot 11 = 4999 = |-6|_{5005}$.

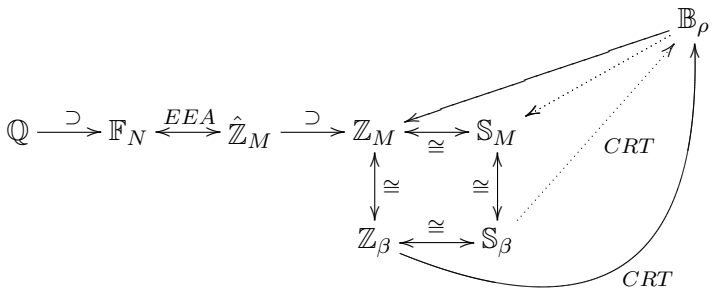


Abbildung 3.1: Zusammenfassung des Kapitels 3.

Kapitel 4

Rationales multimodulares Rechnen

So little time, so much to do.

Winston Churchill

Um sich in diesem Kapitel mit der *rationalen* Arithmetik zu bechäftigen, muss man zunächst eine Darstellung der Brüche angeben. Dazu kann der erweiterte euklidische Algorithmus 3 benutzt werden, der eine (via Algorithmus 4) invertierbare Darstellung einer Teilmenge von $\mathbb{Q}$ in $\mathbb{Z}_M$

$$\begin{aligned} \text{EEA} &: \mathbb{Q} \supset \mathbb{F}_N \to \mathbb{Z}_M, \\ \text{EEA} &: \frac{a}{b} \to |ab^{-1}|_M, \end{aligned}$$

liefert. Der zweite Teil der Kapitelüberschrift ist „*multimodulares* Rechnen“. Allerdings ist die naive Vorgehensweise

$$EEA \circ |\cdot|_\beta =: \phi^{(2)} \circ \phi^{(3)}$$

mit $n := \#\beta > 2$ – bei $n = 2$ geht es gerade noch, vgl. [3] – zum Scheitern verurteilt:

- man kann nur die Farey–Brüche bis zu einer bestimmten Ordnung richtig „wiederherstellen“, von daher sollte man bei der Wahl der Ordnung der „vorwärts“ abgebildeten Brüche beachten, dass das Endergebnis immer noch „zurückabbildbar“ ist. Allerdings wächst die Ordnung der Farey–Brüche mit der Anzahl der Operationen drastisch, vgl. Lemma 2.23.
- falls eine multimodulare Zahl modulo $\beta = [m_i : i = 1, \ldots, n]$ invertierbar sein sollte, soll diese auch modulo $M = m_1 \cdots m_n$ invertierbar sein. Nun ist aber M keine Primzahl.

Somit schränkt sich die Menge der Einheiten in der multimodularen Arithmetik drastisch ein, [11] gibt ein Beispiel für $\beta = [2, 5, 7, 11]$ an, wo es *keine* Einheiten außer ± 1 gibt.

Es gibt aber einen Trick [3, S. 56–62], mit dem man die multimodulare rationale Arithmetik immer noch betreiben kann: Man trennt die Potenzen der Zahlen, modulo welche man rechnet, – nämlich m_i für $i = 1, \ldots, n$ – ab und schreibt sie separat nieder. Das macht die Abbildung $\phi^{(1)}$. Die Brüche, die teilerfremd zu m_i für $i = 1, \ldots, n$ sind, werden modulo M zu Restklassen konvertiert, wofür die Abbildung $\phi^{(2)}$ bzw. EEA verantwortlich ist. Anschließend bildet man modulo β ab, das ist $\phi^{(3)}$, die meistens zu $|\cdot|_\beta$ wird. Somit ergibt sich das Schema

$$\phi^{(1)} \circ \phi^{(2)} \circ \phi^{(3)}$$

für die Vorwärtsabbildung.

Die Rückwärtsabbildung ihrerseits besteht aus der (dieser oder jener) Behandlung der Potenzen von m_i für $i = 1, \ldots, n$. Weiter konvertiert man die multimodulare Darstellung in die unimodulare modulo M – dies geschieht mit dem im vorherigen Kapitel diskutierten mixed–radix–Algorithmus. Anschließend konvertiert man die Restklassen von $\hat{\mathbb{Z}}_M$ zurück in die Farey–Brüche mit dem erweiterten euklidischen Algorithmus. Am Ende müssen ggf. die entfernten Potenzen von m_i wieder in den Bruch hineinmultipliziert werden.

Die Vorwärts– und Rückwärtsabbildungen benötigen eine Langzahlarithmetik, die man in mehreren Fällen – z. B. überall bei dem euklidischen Algorithmus – in Restklassenringen der Größenordnung $M = m_1 \cdots m_n$ ausführen kann. Die Arithmetik dagegen bleibt in Restklassenringen $\mathbb{Z}_{m_i}$. Es erscheint sinnvoll $m_1, \ldots, m_n$ als Primzahlen, die gerade noch in ein Maschinenwort passen, zu wählen. Die Arithmetik ist unabhängig in jeder Komponente. Die Addition bereitet etwas Sorge mit mehreren Fallunterscheidungen, die sonstigen Grundrechenarten sind (fast) intuitiv.

Wie man es schon erahnen konnte, gibt es verschiedene Möglichkeiten der Behandlung der Potenzen von m_i, $i = 1, \ldots, n$. Es wird in diesem Text zunächst die eine Möglichkeit aufgegriffen. Anschließend schlägt man damit Brücken zu der anderen Vorgehensweise. Es ist zu betonen, dass die Unterschiede nur bei Vorwärts– und Rückwärtsabbildungen bestehen, die für den einen Fall eingeführte Arithmetik kann für den anderen übernommen werden.

4.1 Darstellung der $\mathbb{M}_\beta$

Definition 4.1 (multimodulare rationale Zahlen)**.** Sei $\mathcal{Z} \subseteq \mathbb{Z}$.

1. Das *n–Tupel der Paare* $[(u_i, v_i) : u_i \in \mathbb{Z}_{m_i}, v_i = \mathcal{Z}, i = 1, \ldots, n]$ ist eine *Darstellung der multimodularen rationalen Zahl* zu einem Modulvektor $\beta = [m_i : i = 1, \ldots, n]$.

2. Die Menge solcher Tupel zum Modulvektor β wird als $\mathbb{M}_\beta$ bezeichnet.

Die Einträge von u_i sind die aus der ganzzahligen multimodularen Arithmetik gewohnten Restklassen modulo m_i, während v_i die *Potenzen von* m_i darstellt. Die Menge $\mathcal{Z}$ beinhaltet die zulässige Werte für v_i. Man wird diese Menge üblicherweise mit den ganzen Zahlen $\mathbb{Z}$ identifizieren. Die Einträge von β werden grundsätzlich als Primzahlen gesehen, obwohl es im Text explizit angegeben ist, wann man diese Eigenschaft voraussetzt. Mit der Berechnung der Einträge von $\mathbb{M}_\beta$ beschäftigt sich der nächste Abschnitt.

4.2 Die Abbildungen

4.2.1 Vorwärtsabbildung

Die Vorwärtsabbildung $\phi : \mathbb{Q} \supset X \to \mathbb{M}_\beta$ wird für eine Teilmenge X von $\mathbb{Q}$ definiert, diese wird in dieser Arbeit nur implizit angegeben. Man ordnet zu ϕ stets einen Modulvektor $\beta = [m_1, \ldots, m_n]$.

Definition 4.2 (kanonische Vorwärtsabbildung)**.** Zu gegebenem Modulvektor $\beta = [m_1, \ldots, m_n]$ bildet man die Brüche aus X auf $\mathbb{M}_\beta$ folgendermaßen ab:

$$\begin{aligned} &\phi : X \to \mathbb{M}_\beta \\ &\phi : \frac{a}{b} \to [(u_i, v_i) : i = 1, \ldots, n] \end{aligned} \tag{4.1}$$

wobei

$$\begin{aligned}
\frac{a^{(1)}}{b^{(1)}} &= \frac{a}{b} m_1^{v_1}, && \operatorname{ggT}\big(a^{(1)}, b^{(1)}\big) = \operatorname{ggT}\big(a^{(1)}, m_1\big) = \operatorname{ggT}\big(b^{(1)}, m_1\big) = 1 \\
\frac{a^{(2)}}{b^{(2)}} &= \frac{a^{(1)}}{b^{(1)}} m_2^{v_2} = \frac{a}{b} m_1^{v_1} m_2^{v_2}, && \operatorname{ggT}\big(a^{(2)}, b^{(2)}\big) = \operatorname{ggT}\big(a^{(2)}, m_2\big) = \operatorname{ggT}\big(b^{(2)}, m_2\big) = 1 \\
&\vdots && \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
\frac{a^{(n)}}{b^{(n)}} &= \frac{a^{(n-1)}}{b^{(n-1)}} m_n^{v_n} = \frac{a}{b} m_1^{v_1} \dots m_n^{v_n}, && \operatorname{ggT}\big(a^{(n)}, b^{(n)}\big) = \operatorname{ggT}\big(a^{(n)}, m_n\big) = \operatorname{ggT}\big(b^{(n)}, m_n\big) = 1
\end{aligned} \tag{4.2}$$

$$u_i = \left| \frac{a^{(n)}}{b^{(n)}} \right|_{m_i} \qquad i = 1, \dots, n \tag{4.3}$$

gilt.[1] Im Folgenden wird $\phi\left(\frac{a}{b}\right)$ zum Modulvektor β oft als $\left|\frac{a}{b}\right|_\beta$ bezeichnet.

Bemerkung 4.3 (Wichtig!). Der jeweilige Faktor $m_i^{v_i}$ wird aus *allen* Modulen u_i ausmultipliziert. Dies wird hier als *kanonisch* bezeichnet.[2]

Nocheinmal formalisiert:

Algorithmus 6 (kanonische multimodulare rationale Vorwärtsabbildung)**.**
Eingabe: *Ein* gekürzter *Bruch* $x = \frac{a}{b} \in X \subset \mathbb{Q}$*, Modulvektor* β *mit* $\#\beta = n$*.*

1. *Für* $i = 1, \dots, n$ *bestimme ob* $m_i \mid a$ *oder* $m_i \mid b$*. Gilt es nicht, so gehe zu Schritt 2. Nun, da* $\frac{a}{b}$ *gekürzt ist, teilt* m_i *entweder* a *oder* b*, aber nicht beides. Setze* $v_i \leftarrow 0$*.*

   (a) *Solange* $m_i \mid a$ *gilt, setze* $v_i \leftarrow v_i + 1$*,* $a \leftarrow a/m_i$*.*
   (b) *Solange* $m_i \mid b$ *gilt, setze* $v_i \leftarrow v_i - 1$*,* $b \leftarrow b/m_i$*.*

   Anschließend ist für $i = 1, \dots, n$*:* $m_i \nmid a$ *und* $m_i \nmid b$*.*

2. *Berechne für alle* $i = 1, \dots, n$ *den Wert für* u_i *modulo* m_i

   $$u_i \leftarrow a \oslash b$$

   mit dem erweiterten euklidischen Algorithmus 3. Die Startmatrix dafür ist

   $$\begin{pmatrix} m_i & 0 \\ b & a \end{pmatrix}.$$

Ergebnis: $|x|_\beta = [(u_i, v_i) : i = 1, \dots, n] \in \mathbb{M}_\beta$*.*

Da der Bruch $a^{(n)}/b^{(n)}$ durch den Algoritmus 3 verarbeitet wird, existiert nach Satz 2.27 ein eindeutiger Farey–Bruch der Ordnung N mit $2N^2 < M = m_1 \cdots m_n$. Somit ist die Menge X der eindeutig abbildbaren Brüche auf die Produkte der Farey–Brüche der Ordnung bis zu N und der Potenzen von m_i mit $i = 1, \dots, n$ beschränkt. Die Bedingung an N wird in Korollar 4.5 verdeutlicht.

Es ist zu beachten, dass man in den theoretischen Überlegungen $\phi^{(2)} = EEA_{M,N}$ und $\phi^{(3)} = |\cdot|_\beta$ verwendet, aber der Algorithmus 6 im Schritt 2 berechnet $EEA_{m_i,N} \circ |\cdot|_{m_i}$ für $i = 1, \dots, n$. Das folgende Lemma zeigt, dass es keinen Unterschied ausmacht.

[1] Also ist (4.2): $\frac{a^{(0)}}{b^{(0)}} := \frac{a}{b}$ und $\frac{a^{(i)}}{b^{(i)}} = \frac{a^{(i-1)}}{b^{(i-1)}} m_i^{v_i}$ für $i = 1, \dots, n$.

[2] Und es stimmt nicht wirklich damit überein, was [3] als kanonisch bezeichnet, sieht aber genauso aus. Jedoch dasjenige, was später als *nichtkanonisch* bezeichnet wird, stimmt gar nicht mit dem überein, was in [3] nichtkanonisch heißt. Vgl. Abbildung 4.5, Seite 47.

$$\mathbb{Q} \supset X \xrightarrow{\phi^{(1)}} \mathbb{F}_N \times \mathcal{Z}^n \xrightarrow[EEA]{\phi^{(2)}} \mathbb{Z}_M \times \mathcal{Z}^n \xrightarrow[|\cdot|_\beta]{\phi^{(3)}} \mathbb{Z}_\beta \times \mathcal{Z}^n = \mathbb{M}_\beta$$

Abbildung 4.1: Struktur der Vorwärtsabbildung.

Lemma 4.4. *Seien β ein Modulvektor mit $\beta = [m_1, \ldots, m_n]$, $M = m_1 \cdots m_n$, $2N^2 < M$ und $x \in \mathbb{F}_N$. Seien noch $[s_i : i = 1, \ldots, n]$, $[t_i : i = 1, \ldots, n] \in \mathbb{Z}_\beta$ mit $[s_1, \ldots, s_n] = |EEA_{M,N}(x)|_\beta$ und $t_i = |EEA_{m_i,N}(x)|_{m_i}$. Es gilt*

$$[s_1, \ldots, s_n] = [t_1, \ldots, t_n].$$

Beweis. Sei $x = a/b \in \mathbb{F}_N$. Dann ist $|EEA_{M,N}(a/b)|_M = |ab^{-1}|_M =: S$ und $|EEA_{m_i,N}(a/b)|_{m_i} = |ab^{-1}|_{m_i} = t_i$ für $i = 1, \ldots, n$. Man kann mit dem Chinesischen Restsatz 3.5 aus $t_1, \ldots, t_n$ eindeutig S rekonstruieren (mit $\alpha = 0$). Andererseits ist $[s_1, \ldots, s_n] = |S|_\beta$. □

Dabei soll ϕ eine schöne Eigenschaft besitzen: die Abbildung ϕ soll ein Homomorphismus sein. Jedoch wird es in diesem Text nicht bewiesen. Im Folgenden wird aber eine ähnliche Aussage formuliert und bewiesen: Satz 4.31.
Man erinnere sich aber an die Darstellung $\phi = \phi^{(1)} \circ \phi^{(2)} \circ \phi^{(3)}$. Die folgende Aussage folgt aus dem unimodularen Fall (Satz 2.27), dieser trifft ein bei $\phi^{(2)}$ *vor* der eigentlichen „Multimodularität“ in $\phi^{(3)}$.

Korollar 4.5. *Sei N die Ordnung der Farey–Brüche und sei $M = m_1 \cdots m_n$. Falls $\mathbb{Q}_x$ einen Farey–Bruch der Ordnung N enthält und*

$$2N^2 < M. \tag{4.4}$$

gilt, so enthält $\mathbb{Q}_x$ genau einen Farey–Bruch der Ordnung N.

4.2.2 Normalisierung

Die Darstellung von $x = [(u_i, v_i) : i = 1, \ldots, n] \in \mathbb{M}_\beta$ heißt die *kanonische Normalform*, falls

- x die Bemerkung 4.3 erfüllt und
- x die *höchstmögliche* Anzahl der $v_i \neq 0$ hat.

Die Darstellung von x heißt *ausmultipliziert*, falls *alle* $v_i = 0$ sind. Nach Algorithmus 6 befinden sich die multimodulare Zahlen in der kanonischen Normalform.

Beispiel 4.6. *Sei $\beta = [3, 5, 7]$. Es ist $x =: \phi(30) = \phi((3 \cdot 5) \cdot 2) = [(2, 1), (2, 1), (2, 0)]$, man bildet eigentlich nur 2 ab, die Faktoren 3 und 5 werden in v_1 und v_2 gespeichert. Das ist die kanonische Normalform von x. Man kann aber x auch als $[(0, 0), (1, 1), (6, 0)]$ schreiben. In diesem Fall hat man mit Faktor 3 multipliziert. Jedoch sind noch nicht alle v_i Null. Mit $[(0, 0), (0, 0), (2, 0)]$, erhalten durch multiplizieren mit Faktor 5, hat man die ausmultiplizierte Darstellung erreicht.*

Man kann $\theta : \mathbb{M}_\beta \to \mathbb{M}_\beta$ definieren. Die Abbildung θ liefert die ausmultiplizierte Form, wobei die Bildmenge von θ als $\mathbb{M}_\beta^\circ = \{[(u_i, v_i) : i = 1, \ldots, n] \in \mathbb{M}_\beta : v_i = 0 \text{ für alle } i = 1, \ldots, n\} \subset \mathbb{M}_\beta$ bezeichnet werden kann.

Algorithmus 7 (Berechnung von $\theta(\mathbb{M}_\beta)$)**.**
Eingabe: $[(u_i, v_i) : i = 1, \ldots, n]$.
Solange es ein i mit $v_i \neq 0$ gibt, wiederhole:

1. *Für alle $j = 1, \ldots, n$ mit $j \neq i$ setze*

$$u_j \leftarrow u_j \odot m_i^{v_i}.$$

2. *Setze $(u_i, v_i) \leftarrow (0, 0)$.*

Ergebnis: $[(u_i, 0) : i = 1, \ldots, n]$.

Bemerkung 4.7. $\mathbb{Z}_\beta$ kann als ein ganzzahliges Analogon der ausmultiplizierten $\mathbb{M}_\beta$ betrachtet werden. Für $n \leq 2$ ist $\hat{\mathbb{Z}}_\beta$ völlig equivalent zu ausmultiplizierten $\mathbb{M}_\beta$.

$$\xrightarrow{\phi} \mathbb{M}_\beta \xrightarrow{\psi}$$
$$\theta \downarrow$$
$$\mathbb{M}_\beta^\circ$$

Abbildung 4.2: Abbildungen auf $\mathbb{M}_\beta$.

4.2.3 Rückwärtsabbildung

Definiere eine Abbildung $\psi : \mathbb{M}_\beta \to Y \subset \mathbb{Q}$. Gregory und Krischnamurthy [3, S. 58–60] beschreiben ψ wie folgt.

Definition 4.8 (Produkte von m_i)**.**

$$I_+ := \{i : v_i > 0\}, \qquad I_0 := \{i : v_i = 0\}, \qquad I_- := \{i : v_i < 0\}, \tag{4.5}$$

$$M_+ := \prod_{i \in I_+} m_i, \qquad M_0 := \prod_{i \in I_0} m_i, \qquad M_- := \prod_{i \in I_-} m_i. \tag{4.6}$$

Das leere Produkt ist per Konvention Eins.

Klar: $M = M_+ M_0 M_-$.

Algorithmus 8 (kanonische multimodulare rationale Rückwärtsabbildung nach [3])**.**
Eingabe: $|x|_\beta = [(u_i, v_i) : i = 1, \ldots, n] \in \mathbb{M}_\beta$*, der zugehörige Modulvektor β.*

1. *Berechne M_+, M_0 und M_-.*

2. *Finde ein q, so dass für alle $i \in I_0$*

$$\phi_{m_i}(q) = (u_i, 0). \tag{4.7}$$

   Dies wird durch den mixed-radix Algorithmus 5 geleistet.

3. *Setze $q' \leftarrow \left| qM_- \left(M_+^{-1} (\mathrm{mod}\ M_0) \right) \right|_{M_0}$.*

4. Berechne die Ordnung N der Farey-Brüche $\mathbb{F}_N$ als

$$N = \sqrt{\frac{1}{2}\prod_{i=1}^{n} m_i} = \sqrt{M/2}. \tag{4.8}$$

5. Suche mit dem erweiterten euklidischen Algorithmus 4 einen passenden Farey–Bruch. Verwende dafür die Startmatrix

$$\begin{pmatrix} M_+M_0 & 0 \\ M_+q' & M_- \end{pmatrix}.$$

Falls der Algorithmus 4 erfolgreich den Bruch $\frac{a}{b}$ findet, so terminiere erfolgreich, ansonsten brich ab.

Ergebnis: *im Erfolgsfall der Farey-Bruch $\frac{a}{b}$ der Ordnung N, andernfalls ein Fehler.*

Bemerkung 4.9. Die Berechnung von Schritt 3 und Schritt 5 soll mit der genauen Langzahlarithmetik durchgeführt werden. Allerdings finden diese Berechnungen in viel größeren Restklassen als sonst bei $\mathbb{M}_\beta$, aber immer noch in Restklassenringen statt.

Bei der Implementierung des Algorithmus 8 habe ich diesen im Sinne von [11, Bemerkung 3.35.3] umgeändert.

Algorithmus 9 (modifizierte kanonische multimodulare rationale Rückwärtsabbildung)**.**
Eingabe: $|x|_\beta = \left|\frac{a}{b}\right|_\beta = [(u_i, v_i) : i = 1, \ldots, n] \in \mathbb{M}_\beta$*, der zugehörige Modulvektor* β*.*

1. „Herausziehen“ der Faktoren

$$a' \leftarrow \prod_{i\in I_+} m_i^{v_i} \quad b' \leftarrow \prod_{i\in I_-} m_i^{v_i} \tag{4.9}$$

$$(u_i, v_i) \leftarrow (u_i, 0) \quad \text{für } i \in (I_+ \cup I_-) \tag{4.10}$$

2. Berechne

$$M \leftarrow \prod_{i=1}^{n} m_i \tag{*}$$

3. Finde ein q, so dass für alle $i = 1, \ldots, n$

$$\phi_{m_i}(q) = (u_i, 0). \tag{4.11}$$

Dies wird durch den mixed-radix Algorithmus 5 geleistet.

4. Berechne die Ordnung N der Farey-Brüche $\mathbb{F}_N$ als

$$N = \sqrt{\frac{1}{2}\prod_{i=1}^{n} m_i} = \sqrt{M/2}. \tag{**}$$

5. Suche mit dem erweiterten euklidischen Algorithmus 4 einen passenden Farey–Bruch. Verwende dafür die Startmatrix

$$\begin{pmatrix} M & 0 \\ q & 1 \end{pmatrix}.$$

Falls der Algorithmus 4 keinen passenden Bruch findet, so brich ab.

6. „Wiedergutmachen vom Schritt 1.“ Setze

$$\frac{a}{b} \leftarrow \frac{a}{b}\frac{a'}{b'}$$

und terminiere erfolgreich.

Ergebnis: *im Erfolgsfall der Bruch $\frac{a}{b}$, andernfalls ein Fehler.*

Bemerkung 4.10. Die Nachteile von Algorithmus 8 sind auch hier nicht vermeidbar, die Schritte 1 und 6 müssen mit der „normalen“ genauen rationalen Arithmetik und der Schritt 5 muss in der langzahligen Restklassenarithmetik durchgeführt werden, die aber „nicht so langzahlig“ sein müssen wie im Algorithmus 8. Man kann auch die mit (*) und (**) markierten Werte im voraus berechnen.[3] Außerdem kann man eine korrekte Abbildung in einigen Problemfällen erwarten [11, Bemerkung 3.35.2].

Bemerkung 4.11 (Güte des Ergebnisses des Algorithmus 9). Eigentlich soll das Ergebnis $\frac{a}{b}$ vom Algorithmus 9 gleich x sein, also $\psi(\phi(x)) = x$, es gilt aber nicht für alle $x \in \mathbb{Q}$. Man kann $\mathbb{Q} \ni x = \hat{x} m_1^{v_1} \cdots m_n^{v_n}$ schreiben. Also spaltet man aus dem Nenner und dem Zähler der Eingabe x so viele Potenzen von m_i, $i = 1, \ldots, n$, wie möglich. Falls beides, Nenner und Zähler von $\hat{x}$

- zu allen m_i, $i = 1, \ldots, n$, teilerfremd sind,
- betraglich kleiner oder gleich N sind,

so soll ψ korrekt den Wert von x aus $\mathbb{W}_\beta$ zurück in $\mathbb{Q}$ abbilden können. Die erste Bedingung an $\hat{x}$ ist per Konstruktion von ϕ gegeben, und die zweite heißt genausoviel wie $\hat{x} \in \mathbb{F}_N$. Weitere Informationen zu diesem Thema sind im Abschnitt 4.3 zu finden. Man führt trotzdem die Mengen X und Y ein. Es gilt die implizite Definition:

$$X := \{x \in \mathbb{Q} : \phi(x) \text{ eindeutig}\}, \quad \phi(X) = \mathbb{M}_\beta \tag{4.12}$$

$$Y := \text{bild } \psi, \quad \psi(\mathbb{M}_\beta) = Y. \tag{4.13}$$

Nach kurzen Überlegungen kann man feststellen, dass beides (4.12) und die Definition 4.1 die Menge $\mathbb{M}_\beta$ wohldefinieren. Die Mengen X und Y haben eine komplexe Struktur.

4.2.4 Arithmetik

Alle arithmetischen Operationen werden grundsätzlich *komponentenweise* ausgeführt.

Addition

Die Addition ist die komplizierteste Operation auf $\mathbb{M}_\beta$. Addiert man die i-ten Komponenten (u_i, v_i) und (μ_i, ν_i), so lautet die Faustregel

- Falls $v_i = \nu_i$, addiere $u_i + \mu_i$, das Ergebnis ist $(u_i \oplus \mu_i, v_i)$.
- Falls $v_i < \nu_i$, übernimm das Paar mit der *kleineren* zweiten Komponente: (u_i, v_i).

Präziser kann dies in einer Tabelle zusammengefasst werden. Es gilt

[3] Das ist: bei Algorithmus 9, bei Algorithmus 8 kann man zwar die Berechnung der Ordnung von $\mathbb{F}_N$, also (**), vorziehen, aber man muss jedes Mal M_ξ, $\xi \in \{-, 0, +\}$ neu berechnen, oder *alle* mögliche Kombinationen zwischenspeichern, was sicherlich nicht speichereffizient ist. Die Kenntnis von I_+ und I_- ist für die Berechnung von (4.9) nicht erforderlich, man kann $a'/b' = m_1^{v_1} \cdots m_n^{v_n}$ in einer `for`–Schleife berechnen.

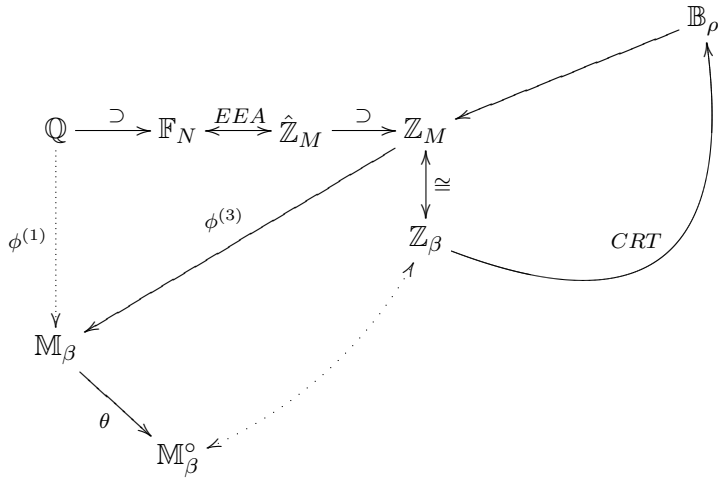


Abbildung 4.3: Zusammenfassung der bisherigen Ergebnisse. Die hier unwesentlichen symmetrischen Restklassen $\mathbb{S}_M$ und $\mathbb{S}_\beta$ sind nicht eingezeichnet.

- $u := u_i$ usw. – die Indizes sind in der Tabelle ausgelassen. Es wird konsequent mit den i–ten Einträgen gerechnet.
- $u \oplus \mu = |u + \mu|_m$.
- alle $v, \nu > 0$, ansonsten lautet der Tabelleneintrag $-v$ bzw. $-\nu$.
- $[(0, z_i) : i = 1, \ldots, n]$ ist das neutrale Element, es wird auf der Seite 44 ausführlich beschrieben. Die Einträge z_i, ζ_i sind in $\mathcal{Z}$ für $i = 1, \ldots, n$.

$+$	$(0, z)$	(u, v)	$(u, 0)$	$(u, -v)$
$(0, \zeta)$	$(0, 0)$	(u, v)	$(u, 0)$	$(u, -v)$
(μ, ν)	(μ, ν)	$*$	$(u, 0)$	$(u, -v)$
$(\mu, 0)$	$(\mu, 0)$	$(\mu, 0)$	$(u \oplus v, 0)$	$(u, -v)$
$(\mu, -\nu)$	$(\mu, -\nu)$	$(\mu, -\nu)$	$(\mu, -\nu)$	$**$

Die zwei markierten Fälle sind etwas komplizierter:

$$(u, v) + (\mu, \nu) = \begin{cases} (u, v) & \text{falls } v < \nu \\ (u \oplus \mu, v) & \text{falls } v = \nu \\ (\mu, \nu) & \text{falls } v > \nu \end{cases} \tag{*}$$

$$(u, -v) + (\mu, -\nu) = \begin{cases} (u, -v) & \text{falls } -v < -\nu \\ (u \oplus \mu, -v) & \text{falls } v = \nu \\ (\mu, -\nu) & \text{falls } -v > -\nu \end{cases} \tag{**}$$

Multiplikation

Definition 4.12 (Rationale multimodulare Multiplikation)**.** Für alle $i = 1, \ldots, n$ gilt

$$(u_i, v_i) \cdot (\mu_i, \nu_i) = (u_i \odot \mu_i, v_i + \nu_i). \tag{4.14}$$

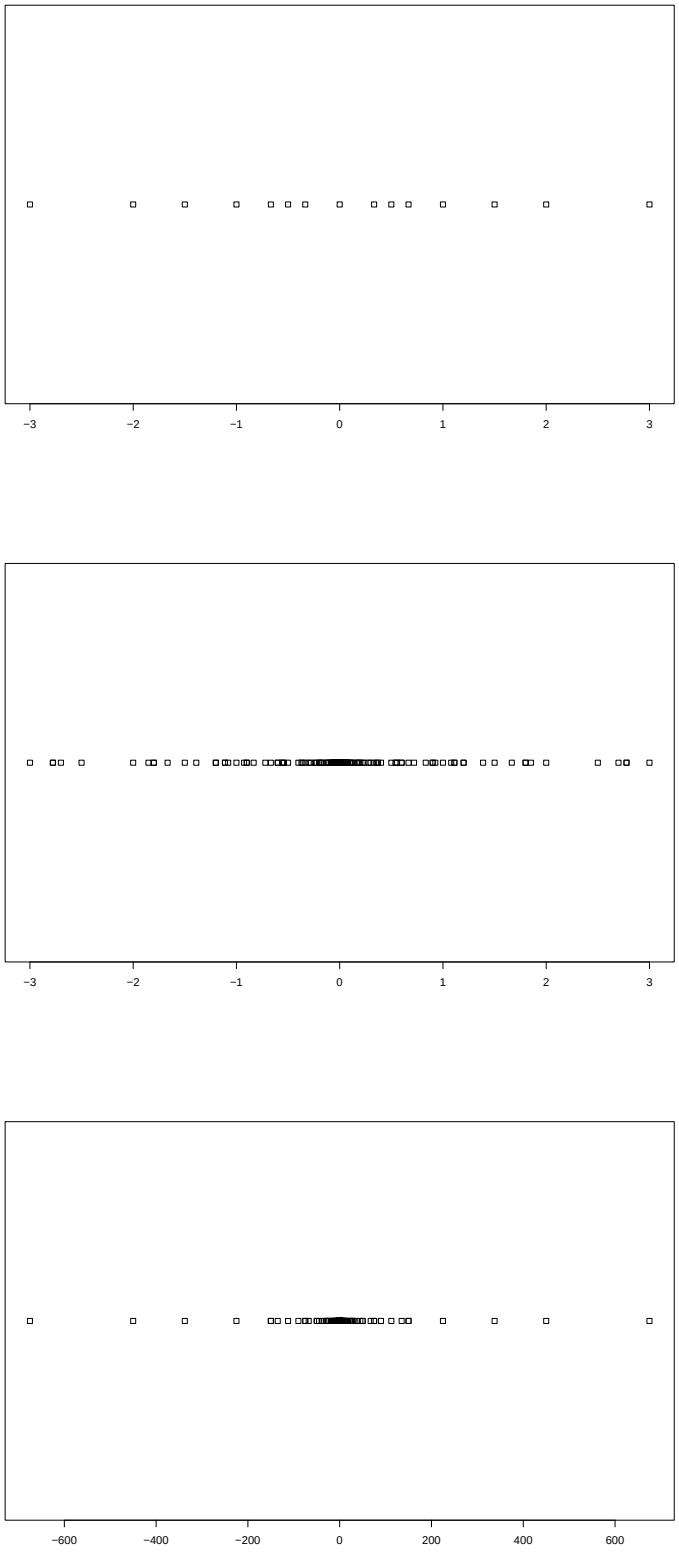


Abbildung 4.4: Die Farey–Brüche in $\mathbb{F}_3$ (oben) und die Punkte in X zu $\mathbb{F}_3$, $\beta = [3, 5]$ und für die Menge $\mathcal{Z} = \{-2, -1, 0, 1, 2\}$ der zulässigen Werte für v_i. Die Punkte von X, die beträglich kleiner oder gleich 3 sind, sieht man in der Mitte. Unten ist die ganze Menge X. Man beachte das Maßstab.

Darstellung von Null und Eins

Wie implizit in der Definition der Addition im Abschnitt 4.2.4 verwendet, ist $[(0, z_i) : i = 1, \ldots, n]$ die Darstellung von Null, unabhängig von den Werten von $z_i \in \mathcal{Z}$, $i = 1, \ldots, n$.
Die Eins wird natürlich als $[(1, 0) : i = 1, \ldots, n]$ dargestellt.

Additive Inverse

Man nimmt leicht an, dass

$$\phi(-x) = [(-u_i, v_i) : i = 1, \ldots, n] \tag{4.15}$$

die Inverse von $\phi(x) = [(u_i, v_i) : i = 1, \ldots, n]$ ist. Das ist auch der Fall, wenn die Werte $v_1, \ldots, v_n$ in $x = u\, m_1^{v_1} \cdots m_n^{v_n}$ und in $-x = -u\, m_1^{v_1} \cdots m_n^{v_n}$ identisch sind, und der Unterschied zwischen u und $-u$ nur in dem Vorzeichen liegt. Nach Lemma 2.20 ist $-u_i$ genau die additive Inverse[4] von u_i. Nun, für alle $i = 1, \ldots, n$ ist $(u_i, v_i) + (-u_i, v_i) = (u_i \ominus u_i, v_i) = (0, v_i) = (0, 0)$. Die letzte Gleichheit gilt wegen der Darstellung der Null.

Multiplikative Inverse

Wieder mit Lemma 2.20 und nach der Grundvorraussetzung, dass alle m_i prim sind (genauer: für $a/b \in X$ ist $\mathrm{ggT}(b, m_i) = 1 = \mathrm{ggT}(a, m_i)$, $i = 1, \ldots, n$), kann man zu gegebenem Modulvektor β die Zahl

$$\phi(x) = [(u_i, v_i) : i = 1, \ldots, n]$$

invertieren, indem man u_i invertiert. Das ist der Fall, wenn alle $v_i = 0$ sind (also: x ist ausmultipliziert) und falls die Grundvorraussetzung erfüllt ist. Aber der Fall $v_i \neq 0$ schadet nicht, denn falls

$$x = \frac{a m_i^{v_i}}{b}$$

war, so ist – wieder mit der Grundvorraussetzung über m_i, $i = 1, \ldots, n$ –

$$x^{-1} = \frac{b}{a m_i^{v_i}} = \frac{b m_i^{-v_i}}{a},$$

und umgekehrt. Somit ist

$$\phi(x^{-1}) = \left[\left(u^{-1} (\mathrm{mod}\ m_i), -v_i\right) : i = 1, \ldots, n\right]. \tag{4.16}$$

Ein kurzer Beweis, dass es wirklich eine Inverse ist: für alle $i = 1, \ldots, n$ ist

$$(u_i, v_i) \cdot \left(u^{-1} (\mathrm{mod}\ m_i), -v_i\right) = \left(u_i \odot u_i^{-1}, v_i - v_i\right) = (1, 0).$$

Zwischenergebnisse

Bemerkung 4.13 (Überlauf ist nicht schlimm). Die Zwischenergebnisse der Berechnungen in $\mathbb{M}_\beta$ können auch *nicht* in dem durch ψ darstellbaren Bereich Y liegen. Formalisiert hat man die folgende Situation für a_i, $b_i \in \mathbb{M}_\beta$ und $* \in \{+, -, \cdot, /\}$:

1. Ist $a_0 * b_0 =: a_1$ und existiert $\psi(a_1) \in Y$, so ist das Ergebnis korrekt (modulo M). Existiert es nicht, so darf a_1 immer noch ein Zwischenergebnis sein.

[4] was genau $-u_i$ ist, hängt sehr stark von der gewählten Darstellung ab, $|m_i - u_i|_{m_i} \in \mathbb{Z}_{m_i}$ bzw. $-/u_i/_{m_i} \in \mathbb{S}_{m_i}$.

2. Ist das Ergebnis a_n der letzten Operation zurückabbildbar, so interessiert man sich nicht für die Zwischenergebnisse. Diese können auch nicht zurückabbildbar sein. Das Endergebnis soll immer noch stimmen.

3. Ist das Endergebnis nicht zurückabbildbar, so scheitert ψ bei der Wiederherstellung der Brüche: entweder findet der EEA einen falschen Bruch oder gar keinen. Der falsche Bruch und der korrekte Bruch sind aber kongruent modulo M_0 (Algorithmus 8) bzw. modulo M (Algorithmus 9). Wurde kein Bruch gefunden, so hat man bloß eine Zahl in $\hat{\mathbb{Z}}_{M_0}$ bzw. $\hat{\mathbb{Z}}_M$, kongruent dem korrekten Bruch. In beiden Fällen hat man die extrahierten Potenzen von m_i – also genau v_i – mit $i = 1, \ldots, n$. Das heißt, auch im Falle eines „hoffnungslosen“ Überlaufs kann man den korrekten Wert mit Zusatzinformation (z. B. mit derselben Berechnung mit anderen Werten von M) wiederherstellen. Vergleiche dazu den Abschnitt 5.4.1.

Beispiel 4.14 („Rückgängig gemachter“ Überlauf)**.** *Man kann modulo $M = 5005$, also zu $N = 50$, das Zwischenergebnis*

$$\frac{1}{34} \cdot \frac{1}{2} = \frac{1}{68}$$

nicht zurückabbilden. Das Endergebnis

$$\left(\frac{1}{34} \cdot \frac{1}{2}\right) \cdot 4 = \frac{1}{17}$$

wird korrekt berechnet und zurückabgebildet.

4.3 Ein Gegenbeispiel

Betrachte man ein längeres Beispiel. Sei

$$\beta = [5, 7, 11, 13],$$

es soll die Summe von $a = \frac{1}{21}$ und $b = \frac{1}{3}$ berechnet werden. Beide Brüche und ihre Summe sind in $\mathbb{F}_{50}$, denn mit $M = 5 \cdot 7 \cdot 11 \cdot 13 = 5005$ ist $50 = \sqrt{\frac{1}{2}M}$. D. h. es *sollte* keine Probleme geben.

Man bildet die Werte a und b mit dem Algorithmus 6 in $\mathbb{M}_\beta$ ab:

$$|a|_\beta = [(2,0), (5,0), (4,0), (9,0)]\,, \tag{4.17}$$

$$|b|_\beta = [(2,0), (5,-1), (4,0), (9,0)]\,. \tag{4.18}$$

Es folgt

$$|c|_\beta := a + b = [(4,0), (5,-1), (8,0), (5,0)]\,. \tag{4.19}$$

Bildet man jetzt

$$\frac{1}{7}\underbrace{[(4,0), (5,0), (8,0), (5,0)]}_{=:\hat{c}}$$

zurück in $\mathbb{F}_N$, so ergibt sich mit mixed radix Algorithmus 5

β	5	7	11	13
$\lvert t^1\rvert_{\beta^0} = \lvert\hat{c}\rvert_\beta$	4	5	8	5
$\lvert d_0\rvert_{\beta^0}$	$\boxed{4}$	4	4	4
$\lvert t^1 - d_0\rvert_{\beta^1}$		1	4	1
$m_1^{-1} (\mathrm{mod}\ \beta_1)$		3	9	8
$\lvert t^2\rvert_{\beta^1}$		3	3	8
$\lvert d_1\rvert_{\beta^1}$		$\boxed{3}$	3	3
$\lvert t^2 - d_0\rvert_{\beta^2}$			0	5
$m_2^{-1} (\mathrm{mod}\ \beta_2)$			8	2
$\lvert t^3\rvert_{\beta^2}$			0	10
$\lvert d_2\rvert_{\beta^2}$			$\boxed{0}$	0
$\lvert t^3 - d_2\rvert_{\beta^3}$				10
$m_3^{-1} (\mathrm{mod}\ \beta_3)$				6
$\lvert t_4\rvert_\beta = \lvert d_3\rvert_{\beta^3}$				$\boxed{8}$

die Darstellung von $\langle\hat{c}\rangle = \langle 4, 3, 0, 8\rangle$. Somit ist $\lvert\hat{c}\rvert_M = 4 + 3 \cdot 5 + 0 \cdot 5 \cdot 7 + 8 \cdot 5 \cdot 7 \cdot 11 = 3099$. Mit dem erweiterten euklidischen Algorithmus 4 folgt

5005	0
3099	1
1906	−1
1193	2
713	−3
480	5
233	−8
14	**21**
9	−344
5	365
4	−709
1	1074
0	−5005

Der einzige Farey-Bruch der Ordnung kleiner oder gleich 50 in dieser Liste ist in der markierten Zeile. Dies liefert den Bruch $\hat{c} = \frac{14}{21} = \frac{2}{3}$, somit

$$c = \frac{1}{7}\frac{2}{3} = \frac{2}{21} \neq \frac{8}{21} = \frac{1}{21} + \frac{1}{3} = a + b.$$

Man kann auch andere Gegenbeispiele angeben. Sei nachwievor $\beta = [5, 7, 11, 13]$, seien $d = \phi(5)$ und $e = \phi(7)$. Sind d und e in Normalform, so wird die Summe $d + e$ zwar berechnet, kann aber nicht zurückabgebildet werden. Der EEA schlägt fehl mit der Startmatrix

$$\begin{pmatrix} 5005 & 0 \\ 3291 & 1 \end{pmatrix},$$

es kann kein Farey–Bruch der Ordnung 50 oder kleiner gefunden werden. Rechnet man mit $\theta(d)$ und $\theta(e)$, so ist das Ergebnis korrekt.

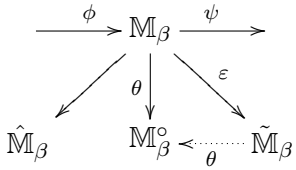


Abbildung 4.5: Mögliche Klassifikation der Teilmengen von $\mathbb{M}_\beta$: $\hat{\mathbb{M}}_\beta$ ist nicht rückwärtsabbildbar, dessen Elemente können nicht mit Elementen aus $\mathbb{Q}$ identifiziert werden. Die Menge $\mathbb{M}_\beta^\circ \subsetneq \mathbb{M}_\beta$ ist ausmultipliziert, alle v_i sind 0. Die Menge $\tilde{\mathbb{M}}_\beta$ enthält die nach [3] „nichtkanonischene Darstellungen“, denn die Darstellung der Summe kann sich von der „normalen“ Darstellung desselben Wertes unterscheiden. Solcher Übergang wird im Bild mit der „Pseudoauslöschung“ ε bezeichnet.
Achtung: $\tilde{\mathbb{M}}_\beta$ ist nicht $\mathbb{W}_\beta$ aus dem nächsten Abschnitt!

Es scheint, dass die Summe im ersten Gegenbeispiel gar nicht im „richtigen“ Bereich liegt. Bezeichnet man die Vorwärtsabbildung als ϕ und die Rückwärtsabbildung als ψ, so ist nicht nur $\phi(a) + \phi(b) \neq \phi(a+b)$, sondern auch $\phi\left(\psi\left(\phi(a+b)\right)\right) \neq \phi(a+b)$. Das wird hier als „Pseudoauslöschung“ bezeichnet. Außerdem ist $\mathbb{M}_\beta$ strikt in seinem Abschluss enthalten. Das heißt, dass beispielsweise bei der Addition zweier Elemente von $\mathbb{M}_\beta$ das Ergebnis auch *nicht* in $\mathbb{M}_\beta$ liegen kann. Man könnte versuchen, eine „nicht rückwärtsabbildbare“ Teilmenge $\hat{\mathbb{M}}_\beta$ zu definieren, aber in dieser Arbeit wird anders vorgegangen.

4.4 Modifizierte Abbildungen

And now for something completely different.

Monty Python's Flying Circus

Schreibt man bei der Definition der kanonischen Vorwärtsabbildung

$$\begin{aligned} \phi &: X \to \mathbb{M}_\beta \\ \phi &: \frac{a}{b} \to [(u_i, v_i) : i = 1, \ldots, n] \end{aligned} \tag{4.1}$$

$$\begin{aligned} \frac{a^{(1)}}{b^{(1)}} &= \frac{a}{b} m_1^{v_1}, & \operatorname{ggT}\left(a^{(1)}, m_1\right) &= \operatorname{ggT}\left(b^{(1)}, m_1\right) = 1 \\ \frac{a^{(2)}}{b^{(2)}} &= \frac{a^{(1)}}{b^{(1)}} m_2^{v_2} = \frac{a}{b} m_1^{v_1} m_2^{v_2}, & \operatorname{ggT}\left(a^{(2)}, m_2\right) &= \operatorname{ggT}\left(b^{(2)}, m_2\right) = 1 \\ &\vdots & &\vdots \\ \frac{a^{(n)}}{b^{(n)}} &= \frac{a^{(n-1)}}{b^{(n-1)}} m_n^{v_n} = \frac{a}{b} m_1^{v_1} \ldots m_n^{v_n}, & \operatorname{ggT}\left(a^{(n)}, m_n\right) &= \operatorname{ggT}\left(b^{(n)}, m_n\right) = 1 \end{aligned} \tag{4.2}$$

$$u_i = \left| \frac{a^{(n)}}{b^{(n)}} \right|_{m_i} \quad i = 1, \ldots, n \tag{4.3}$$

die Gleichungen (4.2) etwas knapper, so sehen sie folgendermaßen aus:

$$\text{Für } i = 1, \ldots, n \quad \frac{a'}{b'} = \frac{a}{b} \prod_{i=0}^{n} m_i^{v_i}, \quad \operatorname{ggT}(a', m_i) = \operatorname{ggT}(b', m_i) = 1, \quad u_i = \left| \frac{a'}{b'} \right|_{m_i}. \tag{4.20}$$

Dies macht den Raum offen für eine andere Interpretation der Vorwärtsabbildung als in [3]. Anstelle von $(\mathbb{M}_\beta, +, \cdot)$ wird die Bildmenge der *modifizierten* Vorwärtsabbildung mit $(\mathbb{W}_\beta, +, \cdot)$ bezeichnet. Die Arithmetik kann jedoch, wie man später sehen wird, von $\mathbb{M}_\beta$ übernommen werden.

4.4.1 Vorwärtsabbildung

Definition 4.15 (modifizierte Vorwärtsabbildung)**.** Zu gegebenem Modulvektor $\beta = [m_1, \ldots, m_n]$ bildet man die Brüche aus $\mathbb{F}_N$ auf $\mathbb{W}_\beta$ folgendermaßen ab:

$$\begin{aligned} &\varphi : X \to \mathbb{W}_\beta \\ &\varphi : \frac{a}{b} \to [(u_i, v_i) : i = 1, \ldots, n] \end{aligned} \tag{4.21}$$

wobei $\frac{a}{b}$ gekürzt ist, es folgt mit $a^{(i)} = am_i^{s_i}$, $b^{(i)} = bm_i^{t_i}$ und $\mathrm{ggT}\big(a^{(i)}, m_i\big) = \mathrm{ggT}\big(b^{(i)}, m_i\big) = 1$

$$v_i = s_i - t_i \tag{4.22}$$

$$u_i = \left| \frac{a^{(i)}}{b^{(i)}} \right|_{m_i} \tag{4.23}$$

für $i = 1, \ldots, n$. Im Folgenden wird $\varphi_\beta\left(\frac{a}{b}\right)$ auch als $\left\|\frac{a}{b}\right\|_\beta$ bezeichnet. Es ist anzumerken, dass für alle $i = 1, \ldots, n$ entweder s_i oder t_i oder beides gleich Null ist.

Bemerkung 4.16 (Schreibweise und Unterschied zu $\mathbb{M}_\beta$). Man könnte die Gleichungen (4.22), (4.23) auch so schreiben:

$$\begin{aligned} \frac{a^{(1)}}{b^{(1)}} &= \frac{a}{b} m_1^{v_1} \\ \frac{a^{(2)}}{b^{(2)}} &= \frac{a}{b} m_2^{v_2} \\ &\vdots \\ \frac{a^{(n)}}{b^{(n)}} &= \frac{a}{b} m_n^{v_n} \\ u_i &= \left| \frac{a^{(i)}}{b^{(i)}} \right|_{m_i} \end{aligned}$$

für $i = 1, \ldots, n$. Der jeweilige Faktor $m_i^{v_i}$ wird nur aus *einem* Eintrag u_i ausmultipliziert. Dies wird als *nichtkanonisch* bezeichnet.

Algorithmus 10 (nichtkanonische multimodulare rationale Vorwärtsabbildung)**.**
Eingabe: *Ein* gekürzter *Bruch* $x = \frac{a}{b} \in X \subset \mathbb{Q}$*, Modulvektor* β *mit* $\#\beta = n$*.*

1. *Für* $i = 1, \ldots, n$ *setze setze* $v_i \leftarrow 0$*,* $a^{(i)} \leftarrow a$ *und* $b^{(i)} \leftarrow b$*.*

   (a) *Solange* $m_i \mid a^{(i)}$*:*

$$\begin{aligned} v_i &\leftarrow v_i + 1 \\ a^{(i)} &\leftarrow a^{(i)}/m_i. \end{aligned}$$

(b) Solange $m_i \mid b^{(i)}$*:*

$$v_i \leftarrow v_i - 1$$
$$b^{(i)} \leftarrow b^{(i)}/m_i$$

Da $\frac{a}{b}$ *gekürzt ist, wird für jedes* i *entweder 1a oder 1b oder keins von beiden ausgeführt.*

2. *Berechne für alle* $i = 1, \ldots, n$ *die Werte* $b^{-1} (\mathrm{mod}\ m_i)$ *und setze modulo* m_i

$$u_i \leftarrow a^{(i)} \oslash b^{(i)}$$

mit dem erweiterten euklidischen Algorithmus 3. Die Startmatrix des EEA ist in diesem Fall

$$\begin{pmatrix} m_i & 0 \\ b^{(i)} & a^{(i)} \end{pmatrix}.$$

Ergebnis: $\|x\|_\beta = [(u_i, v_i) : i = 1, \ldots, n] \in \mathbb{W}_\beta$.

Beispiel 4.17.

1. *Sei* $\beta = [2, 3]$*. Es gilt:* $\phi(5) = [(1, 0), (2, 0)] = \varphi(5)$.
2. *Sei* $\beta = [2, 3, 5]$*. Es gilt* $\varphi(3) = [(1, 0), (1, 1), (3, 0)]$*, im Gegensatz zu dem kanonischen Ergebnis* $\phi(3) = [(1, 0), (1, 1), (1, 0)]$.
3. *Sei* $\beta = [5, 7, 11, 13]$*. Es folgt:*

$$\varphi\left(\frac{5}{8}\right) = [(2, 1), (5, 0), (2, 0), (12, 0)]$$
$$\phi\left(\frac{5}{8}\right) = [(2, 1), (1, 0), (7, 0), (5, 0)].$$

$$\mathbb{Q} \supset X \xrightarrow{\varphi^{(1)}} \mathbb{F}_N \times \mathcal{Z}^n \xrightarrow[EEA]{\varphi^{(2)}} \mathbb{Z}_M \times \mathcal{Z}^n \xrightarrow[|\cdot|_\beta]{\varphi^{(3)}} \mathbb{Z}_\beta \times \mathcal{Z}^n = \mathbb{W}_\beta$$

Abbildung 4.6: Struktur der Vorwärtsabbildung. Keinen wesentlichen Unterschied zu ϕ ist zu sehen, alles ist in $\varphi^{(1)}$ verborgen.

$$\xrightarrow{\varphi} \mathbb{W}_\beta \xrightarrow{\psi}$$
$$\Theta \downarrow$$
$$\mathbb{W}_\beta^\circ$$

Abbildung 4.7: Abbildungen auf $\mathbb{W}_\beta$.

4.4.2 Normalisierung

Analog zum Abschnitt 4.2.2 definiert man eine *nichtkanonische Normalform* von $x = [(u_i, v_i) : i = 1, \ldots, n] \in \mathbb{W}_\beta$ falls

- x die Bemerkung 4.16 erfüllt.
- die größtmögliche Anzahl von v_i verschieden ist von Null.

Die Definition der *ausmultiplizierten* Darstellung für $\mathbb{W}_\beta$ stimmt mit der für $\mathbb{M}_\beta$ überein. Nach dem Algorithmus 10 befinden sich die multimodularen Zahlen in der nichtkanonischen Normalform. Man kann $\Theta : \mathbb{W}_\beta \to \mathbb{W}_\beta$ definieren, das Bild von $\Theta(\mathbb{W}_\beta) =: \mathbb{W}_\beta^\circ$ soll alle $v_i = 0$ haben.

Algorithmus 11 (Berechnung von $\Theta(\mathbb{W}_\beta)$)**.**
Eingabe: $x = [(u_i, v_i) : i = 1, \ldots, n] \in \mathbb{W}_\beta$.
Für alle i, für die $v_i \neq 0$ gilt, setze

$$(u_i, v_i) \leftarrow (0, 0).$$

Ergebnis: $x \in \mathbb{W}_\beta^\circ$.

Bemerkung 4.18.

1. Wie das Beispiel $\Theta([(1,1),(1,1)]) = [(0,0),(0,0)]$ zeigt, ist $\mathbb{W}_\beta^\circ \subset \mathbb{W}_\beta$.
2. Die ausmultiplizierte Form $\Theta(x)$ von $x \in \mathbb{W}_\beta$ ist sowohl ein Element von $\mathbb{W}_\beta$ als auch ein Element von $\mathbb{M}_\beta$. Es gilt sogar mehr: Sei $\xi \in X$, sowie $x = \varphi(\xi) \in \mathbb{W}_\beta$ und $y = \phi(\xi) \in \mathbb{M}_\beta$. Es ist $\Theta(x) = \theta(y)$. Jedoch ist klar: x und y sind elementweise verschieden.

In der weiteren Entwicklung des Algorithmus zur Rückwärtsabbildung wird die erweiterte Abbildung in $\mathbb{W}_\beta^\circ$ benötigt, und zwar die *erweiterte* nichtkanonische Abbildung $\hat{\Theta} : \mathbb{W}_\beta \to (\mathbb{W}_\beta \times \mathbb{Q})$. Im Gegensatz zu Θ wird bei

$$\hat{\Theta}(x) = \left(\hat{x}, \frac{\hat{a}}{\hat{b}}\right)$$

der ursprüngliche Wert von x um $\frac{\hat{a}}{\hat{b}}$ geändert, genauer gesagt

$$x = \hat{x}\frac{\hat{a}}{\hat{b}}. \tag{4.24}$$

Dabei ist $\hat{x} = [(u_i, v_i) : i = 1, \ldots, n]$ eigentlich in $\mathbb{W}_\beta^\circ$, analog zu Θ soll $v_i = 0$ für alle $i = 1, \ldots, n$ gelten.

Algorithmus 12 (Berechnung von $\hat{\Theta}(\mathbb{W}_\beta)$)**.**
Eingabe: $x = [(u_i, v_i) : i = 1, \ldots, n] \in \mathbb{W}_\beta$, *zugehöriges β mit $\#\beta = n$.*

1. *Setze*

$$\frac{\hat{a}}{\hat{b}} \leftarrow \prod_{i=1}^{n} m_i^{v_i}. \tag{4.25}$$

2. *Für alle $i \in I_+ \cup I_-$*

   (a) *Für alle $j = 1, \ldots, n$, $j \neq i$ setze*

$$u_j \leftarrow u_j \odot \left|m_i^{-v_i}\right|_{m_j}. \tag{4.26}$$

   (b) *Setze $(u_i, v_i) \leftarrow (u_i, 0)$.*

Ergebnis: $\hat{x} = [(u_i, v_i) : i = 1, \ldots, n] \in \mathbb{W}_\beta^\circ$, *Bruch* $\frac{\hat{a}}{\hat{b}}$.

Proposition 4.19 (Algorithmus 12 funktioniert)**.** *Bezeichnet man mit x die Eingabe des Algorithmus 12 und mit $(\hat{x}, \frac{\hat{a}}{\hat{b}})$ die Ausgabe, so gilt*

$$x = \hat{x}\frac{\hat{a}}{\hat{b}}. \tag{4.24}$$

Beweis. Der Beweis erfolgt mit der Induktion nach der Anzahl $\#(I_+ \cup I_-) =: A$ der Einträge (u_i, v_i) mit $v_i \neq 0$. Sei $\xi \in X$ ein gekürzter Bruch mit $x = \varphi(\xi)$. Die entfernten Faktoren werden abhängig von dem Vorzeichen von v_i in $\hat{a}$ und $\hat{b}$ gespreichert,

$$\hat{a} = \prod_{i \in I_+} m_i^{v_i}, \quad \hat{b} = \prod_{i \in I_-} m_i^{v_i}.$$

Mit $z(q)$ und $n(q)$ bezeichnet man in dem Beweis den Zähler und den Nenner von $q \in \mathbb{Q}$. Nun schaut man sich x genauer an.

- $A = 0$. In diesem Fall befindet sich x bereits in der ausmultiplizierten Form, setzte $\hat{a} = \hat{b} = 1$. Offensichtlich gilt (4.24).
- $A = 1$. Sei $v_i \neq 0$. Man berechnet $\hat{a}/\hat{b} = m_i^{v_i}$ nach (4.25). Andererseits lässt sich ξ schreiben als $\xi = \xi' m_i^{v_i}$ (d. h. für $v_i \neq 0$ ist entweder $v_i > 0$ und $m_i^{v_i} \mid z(\xi)$ oder $v_i < 0$ und $m_i^{-v_i} \mid n(\xi)$) und $u_i = |\xi'|_{m_i}$, aber $u_j = |\xi|_{m_j} = |\xi' m_i^{v_i}|_{m_j}$ für $j = 1, \ldots, n$, $j \neq i$. In zweitem Schritt des Algorithmus 12 wird für alle $j = 1, \ldots, n$, $j \neq i$,

  $$u_j \leftarrow u_j \odot \left|m_i^{-v_i}\right|_{m_j} = \left|\xi' m_i^{v_i} m_i^{-v_i}\right|_{m_j} = \left|\xi'\right|_{m_j}$$

  gerechnet. Mit anschließendem $(u_i, v_i) \leftarrow (u_i, 0)$ ist nun $[(u_j, v_j) : j \in 1, \ldots, n] = \varphi(\xi')$ und (4.24) gilt.
- Sei jetzt $A \geq 2$. Der Wert für $\hat{a}/\hat{b}$ wird nach (4.25) berechnet und ist korrekt, man muss also noch zeigen, dass $\hat{x}$ den richtigen Wert hat.[5] Man nimmt an, die Aussage der Proposition wäre für $A-1$ bereits bewiesen. Man wählt ein kleinstmögliches i mit $v_i \neq 0$ aus der Menge $(I_+ \cup I_-)$. Man betrachtet die Situation nur für dieses i, alles andere ergibt sich per Induktion. Für $\xi \in X$ gekürzt mit $\varphi(\xi) = x$ lässt sich $u_i = |\xi'|_{m_i}$ schreiben, mit $\xi = \xi' m_i^{v_i}$. Klar: $m_i^{-v_i} \mid n(\xi)$ oder $m_i^{v_i} \mid z(\xi)$. Der Eintrag u_j hat die Form $\left|\xi^{(j)}\right|_{m_j}$ mit $\xi = \xi^{(j)} m_j^{v_j}$ für $j = 1, \ldots, n$, $j \neq i$. Es gilt für $m_1, \ldots, m_n$ koprim und für $j \in (I_+ \cup I_-)$, $j \neq i$, (d. h. so ein j, dass $v_j \neq 0$ und $j \neq i$ gilt) entweder $v_i > 0$ und $m_i^{v_i} \mid z\left(\xi^{(j)}\right)$ oder $v_i < 0$ und $m_i^{-v_i} \mid n\left(\xi^{(j)}\right)$, wegen $m_i^{v_i} \mid z(\xi)$ oder $m_i^{-v_i} \mid n(\xi)$ für $v_i > 0$ und $v_i < 0$ entsprechend.

  Nun berechnet man für $j = 1, \ldots, n$, $j \neq i$

  $$u_j \leftarrow u_j \odot \left|m_i^{-v_i}\right|_{m_j} = \left|\xi^{(j)} / m_i^{v_i}\right|_{m_j}$$

  (also $z(\xi) \leftarrow z(\xi) m_i^{-v_i}$ für $v_i > 0$ oder $n(\xi) \leftarrow n(\xi) m_i^{v_i}$ für $v_i < 0$) und bei $(u_i, v_i) = (|\xi'|_{m_i}, v_i)$ wird $v_i \leftarrow 0$ gesetzt. Somit ist nach einem Durchlauf des Schleifenrumpfes in zweitem Schritt des Algorithmus 12 $x' = [(u_1, v_1), \ldots, (u_n, v_n)]$. Die Anzahl A der Einträge von x mit $v_i \neq 0$ ist jetzt um Eins geringer. Es gilt $x' = \varphi(\xi')$.

Wiederholt man den letzten Schritt iterativ, so erreicht x' die in (4.24) für $\hat{x}$ beschriebene Form, mit $\#(I_+ \cup I_-) = 0$. Im Falle $A = n$ haben alle Einträge u_j die Form $\left|\xi^{(j)}\right|_{m_j}$, $j = 1, \ldots, n$. □

[5] Es wird hier zwischen x' und $\hat{x}$ unterschieden. Die Darstellung x hat A Einträge mit $v_i \neq 0$, es ist $2 \leq A \leq n$. Ein Durchlauf reduziert $\#(I_+ \cup I_-)$ auf $A - 1$, x wird hierdurch zu x'. Im Gegensatz dazu ist $\hat{x}$ die Darstellung nach A Durchläufen, es ist $\#(I_+ \cup I_-) = 0$. Es wird hier nur ein Durchlauf betrachtet, alles andere ergibt sich aus dem Induktionansatz.

4.4.3 Rückwärtsabbildung

Algorithmus 13 (nichtkanonische multimodulare rationale Rückwärtsabbildung à la [3])**.**
Eingabe: $\|x\|_\beta = [(u_i, v_i) : i = 1, \ldots, n] \in \mathbb{W}_\beta$*, der zugehörige Modulvektor* β*.*

1. *Berechne* M_+*,* M_0 *und* M_-*.*
2. *Finde ein* q*, so dass für alle* $i \in M_0$

$$\varphi_{m_i}(q) = (u_i, 0). \tag{4.27}$$

   Dies wird durch den mixed–radix Algorithmus 5 geleistet.

3. *Setze* $q' \leftarrow \left| qM_-M_+^{-1} (\mathrm{mod}\ M_0) \right|_{M_0}$*.*
4. *Berechne die* Ordnung N *der Farey–Brüche* $\mathbb{F}_N$ *als*

$$N = \sqrt{\frac{1}{2} \prod_{i=1}^{n} m_i} = \sqrt{M/2}. \tag{4.28}$$

5. *Suche mit dem erweiterten Euklidischen Algorithmus 4 einen passenden Farey–Bruch. Verwende dafür die Startmatrix*

$$\begin{pmatrix} M_+M_0 & 0 \\ M_+q' & M_- \end{pmatrix}.$$

   Falls der Algorithmus 4 erfolgreich den Bruch $\frac{a}{b}$ *findet, so terminiere erfolgreich, ansonsten brich ab.*

Ergebnis: *im Erfolgsfall der Bruch* $\frac{a}{b}$*, andernfalls ein Fehler.*

Man kann Algorithmus 13 analog zu Algorithmus 9 modifizieren. Die angeführte Version soll den wiederherstellbaren Bereich von $\mathbb{W}_\beta$ erweitern, wird aber in dieser Arbeit nicht weiter diskutiert.

Algorithmus 14 (modifizierte nichtkanonische multimodulare rationale Rückwärtsabbildung)**.**
Eingabe: $x \in \mathbb{W}_\beta$*, der zugehörige Modulvektor* β *mit* $\#\beta = n$*.*

1. *„Herausziehen“ der Faktoren. Führe Algorithmus 12 aus, erhalte* $\hat{x}$ *und* $\frac{a'}{b'}$*.*
2. *Berechne*

$$M \leftarrow \prod_{i=1}^{n} m_i \tag{*}$$

3. *Finde mit dem mixed-radix Algorithmus 5 ein* q*, so dass für alle* $i = 1, \ldots, n$

$$\varphi_{m_i}(q) = (u_i, 0).$$

4. *Berechne die Ordnung* N *der Farey–Brüche* $\mathbb{F}_N$ *als*

$$N = \sqrt{\frac{1}{2} \prod_{i=1}^{n} m_i} = \sqrt{M/2}. \tag{**}$$

5. *Suche mit dem erweiterten Euklidischen Algorithmus 4 einen passenden Farey–Bruch. Verwende dafür die Startmatrix*

$$\begin{pmatrix} M & 0 \\ q & 1 \end{pmatrix}.$$

Falls der Algorithmus 4 erfolgreich den Bruch $\frac{a}{b}$ findet, so terminiere erfolgreich, ansonsten brich ab.

Ergebnis: $\frac{a}{b}\frac{a'}{b'}$.

Bemerkung 4.20.

1. Die Bemerkung 4.13 gilt auch für $\mathbb{W}_\beta$.

2. Die Berechnungen in (*) und (**) kann man nach wie vor in der Vorbereitungsphase ausführen. Es gilt noch mehr: Das Produkt $M = m_1 \cdots m_n$ bzw. die Ordnung der Farey–Brüche N sind miteinander durch (2.8) verbunden, sie bilden einen globalen Parameter. Dieser kann *a priori* bestimmt werden, falls man die „Größe“ von dem Endergebnis kennt. In diesem Fall sind (*) und (**) *bereits* vorberechnet vor dem Beginn der eigentlichen Rechnung, denn sie sind für die effiziente Dimensionierung der Arithmetik wichtig: Man muss „noch“ korrekt das Endergebnis finden können, andererseits will man keinen Mehraufwand. Mehrere angewandte Beispiele sind in Kapitel 5 zu finden.

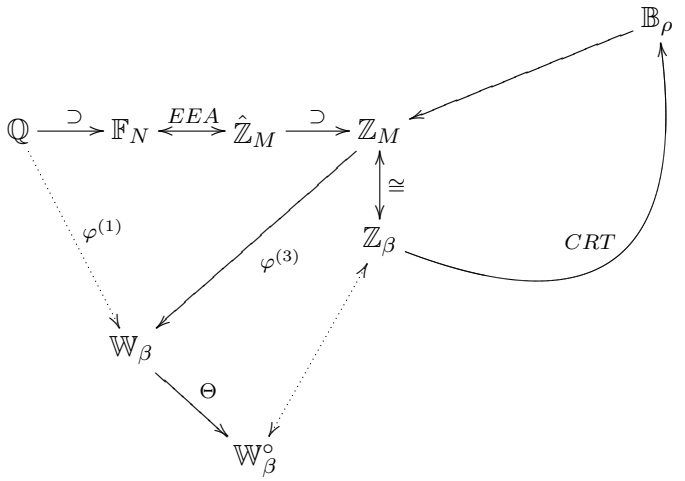


Abbildung 4.8: Zusammenfassung der bisherigen Ergebnisse. Das Graphik entspricht jetzt der Version am Anfang der Arbeit. Die Mengen $\mathbb{M}_\beta$, $\mathbb{M}_\beta^\circ$, $\mathbb{S}_M$ und $\mathbb{S}_\beta$ wurden übersichtlichkeitshalber entfernt. Der Bezug zwischen $\mathbb{W}_\beta$ und $\mathbb{M}_\beta$ wird in der nächsten Abbildung verdeutlicht.

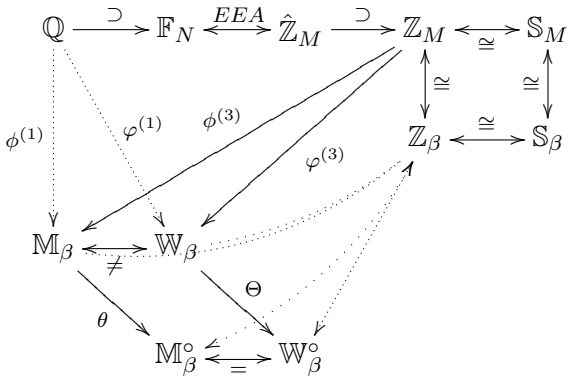


Abbildung 4.9: Die Ähnlichkeiten von $\mathbb{M}_\beta$ und $\mathbb{W}_\beta$. Bis auf $\mathbb{B}_\rho$ ist das Bild vollständig.

4.5 Analyse der Abbildungen

Die folgenden Abschnitte werden zum Beweis der Aussage des Satzes 4.31: „φ ist ein Homomorphismus“ verwendet, sind aber auch losgelöst davon sinnvoll.

4.5.1 Hilfsaussagen

Hier bedeutet $-a$ die additive Inverse von a. Die Darstellung von $|-a|_m$ bzw. $/-a/_m$ ist der Implementierung überlassen.

Lemma 4.21. *Seien a, $m \in \mathbb{Z}$, $m > 1$. Es gilt $-|a|_m = |-a|_m$.*

Beweis. $\mathbb{Z}_m \ni -|a|_m =: |b|_m$ mit $b \in \mathbb{Z}$ und $|a|_m \oplus |b|_m \equiv 0 (\mathrm{mod}\ m)$. Sei $c := -a$, also $a+c = 0 \in \mathbb{Z}$. Man kann $|c|_m$ bilden, und $|a|_m \oplus |c|_m = |a+c|_m = |0|_m$. Somit ist $-|a|_m = |b|_m = |c|_m = |-a|_m$. □

Lemma 4.22. *Sei G eine multiplikative kommutative Gruppe und $a \in G$. Dann ist $(a^{-1})^{-1} = a$.*

Beweis. Die Inverse a^{-1} ist definiert durch $aa^{-1} = 1 \in G$. Im Falle „$(a^{-1})^{-1}$“, existiert zu a^{-1} so ein $(a^{-1})^{-1} =: b$ mit $a^{-1}b = ba^{-1} = 1 \in G$. Dann ist aber a^{-1} die Inverse von b und wegen der Eindeutigkeit der Inversenbildung[6] ist $a = b$. □

Lemma 4.23. *Seien $a, b \in \mathbb{Q}$ gekürzt und $m \in \mathbb{Z}$. Teilt m entweder Nenner oder Zähler von a und entweder Nenner oder Zähler von b, so teilt m entweder den Nenner oder den Zähler der Summe $a+b$ in der gekürzten Darstellung.*

Beweis. Seien $a = p/q$ und $b = r/s$ gekürzt, sowie $m \in \mathbb{Z}$. Die Summe $a + b$ sieht folgendermaßen aus:

$$\frac{p}{q} + \frac{r}{s} = \frac{ps + qr}{qs}.$$

Es gilt entweder $m \mid p$ oder $m \mid q$ und gleichzeitig entweder $m \mid r$ oder $m \mid q$. Entsprechend gibt es vier Fälle zu unterscheiden. Es werden mehrmals die Aussagen aus der Lemma 1.4 verwendet: $m \mid x$, $m \mid y \Rightarrow m \mid (x + y)$ und $m \mid x \Rightarrow m \mid xy$ für $m, x, y \in \mathbb{Z}$. Da der Bruch p/q gekürzt sind, aus $m \mid p$ folgt unmittelbar $m \nmid q$ und umgekehrt, analog für r/s.

[6] Angenommen [4], es gäbe zwei verschiedene Inverse y und z zu einem Element x der multiplikativen Gruppe G. Dann ist aber $y = y1 = y(xz) = (yx)z = 1z = z$.

- „$m \mid p$, $m \mid r$“. Es gilt $m \mid ps$, $\mid qr$, $m \nmid qs$, also $m \mid (ps + qr)$.
- „$m \mid p$, $m \mid s$“. Es folgt $m \mid ps$, $m \nmid qr \Rightarrow m \nmid (ps + qr)$. Aber $m \mid qs$.
- „$m \mid q$, $m \mid r$“ entspricht dem vorherigen Fall mit vertauschten Rollen, es folgt $m \nmid (ps + qr)$ und $m \mid qs$.
- „$m \mid q$, $m \mid s$“. Dieser Fall ist interessanter. Es gilt $m \nmid p$, $m \nmid r$, sowie $m \mid ps$, $m \mid qr$, also $m \mid (ps+qr)$ und $m \mid qs$. Also ist die Summe nicht gekürzt. Angenommen, man kann $\frac{p}{q} = \frac{p}{q'}m^v$ und $\frac{r}{s} = \frac{r}{s'}m^\nu$ schreiben, mit $v, \nu < 0$ und $m \nmid q'$, $m \nmid s'$. Dann ist

$$\frac{p}{q} + \frac{r}{s} = \frac{ps'm^{-\nu} + q'rm^{-v}}{q's'm^{-v}m^{-\nu}} = \frac{ps'm^{v-\nu} + q'r}{q's'm^{-\nu}}.$$

  Die letzte Gleichheit folgt o. B. d. A. mit $|v| < |\nu|$ (sonst einfach die Rollen vertauschen). Dann kann man den Bruch auf der linken Seite mit m^{-v} kürzen und es ergibt sich die rechte Seite. Gegebenenfalls wird man den Bruch noch weiterhin kürzen können, aber nicht mehr mit einem Faktor von m. Es ist offensichtlich, dass m den Nenner teilt, aber nicht den Zähler.

□

4.5.2 Die Addition

Satz 4.24 (Die Addition funktioniert)**.** *Für a, b in X ist*

$$\varphi(a + b) = \varphi(a) + \varphi(b) \quad (\in \mathbb{W}_\beta). \tag{4.29}$$

Bemerkung 4.25. Im folgenden Abschnitt bezeichnet $\varphi_i(x)$ bzw. $(\varphi(x))_i$ die i–te Komponente von $\varphi(x) = [(u_i, v_i) : i = 1, \ldots, n] \in \mathbb{W}_\beta$, also genau das Paar (u_i, v_i).
Der Beweis wird in zwei Lemmata aufgespalten.

Lemma 4.26. *Seien a, b in X und $(u, v) := \varphi_i(a)$ sowie $(\mu, \nu) := \varphi_i(b)$, mit $v = \nu$, so ist*

$$(u, v) + (\mu, v) = (u \oplus \mu, v). \tag{4.30}$$

Beweis. Setze $m := m_i$, seien $a = a'm^v$, $b = b'm^v$. Es gilt $a + b = a'm^v + b'm^v = (a' + b')m^v$. Mit $u = |a'|_m$ und $\mu = |b'|_m$ folgt (4.30). □

Lemma 4.27. *Für die i–ten Einträge (u, v), (μ, ν) zweier Zahlen aus $\mathbb{W}_\beta$, für die $v > \nu$ gilt, folgt, dass der i–te Eintrag der Summe dieser Zahlen (μ, ν) lautet.*

Beweisidee. Angenommen $v > \nu$, man schreibt $a'm^v + b'm^\nu = (a'm^{v-\nu} + b')m^\nu$, „merkt“ sich m^ν und reduziert den restlichen Ausdruck modulo m, so ergibt sich $|b'|_m$ und das vorgemerkte m^ν. Man übersieht aber gerne bei solchem Ansatz, dass a' und b' gekürzte Brüche sind, und dass v und ν auch negativ sein können. Das Ergebnis stimmt immer noch, aber um es sauber zu zeigen, braucht man etwas mehr Arbeit.

Beweis. Seien $a, b \in X$, man betrachtet die i–te Kompomnente von $\varphi(\cdot)$ und bezeichnet $m := m_i$. Seien $\varphi_i(a) = (u, v) = (|a'|_m, v)$ und $\varphi_i(b) = (\mu, \nu) = (|b'|_m, \nu)$ mit $a = a'm^v = \frac{p}{q}m^v$ und $b = b'm^v = \frac{r}{s}m^\nu$. Man betrachtet stets

$$\frac{p}{q}m^v + \frac{r}{s}m^\nu.$$

Der Fall $v = \nu$ ist ausgeschlossen. Es gibt mehrere Fälle, von dennen es wegen der Einschränkung $v > \nu$ genügt drei zu betrachten.

- Seien $v > \nu > 0$. Es gilt

$$\frac{pm^v}{q} + \frac{rm^\nu}{s} = \frac{psm^v + qrm^\nu}{qs} = \frac{psm^{v-\nu} + qr}{qs} m^\nu.$$

  Insgesamt kann man m^ν abspalten, es ist $\left|(psm^{v-\nu} + qr)(qs)^{-1}\right|_m = \left|qr(qs)^{-1}\right|_m = \left|rs^{-1}\right|_m$. Mit dem abgespaltenen Faktor m^ν ergibt sich genau (μ, ν).

- Seien $v > 0$, $\nu < 0$. Es gilt

$$\frac{pm^v}{q} + \frac{r}{sm^{-\nu}} = \frac{psm^{v-\nu} + rq}{qsm^{-\nu}} = \frac{psm^{v-\nu} + rq}{qs} m^\nu.$$

  Nun kann man m^ν abspalten und sich den restlichen Ausdruck modulo m anschauen. Es gilt $\left|(psm^{v-\nu} + rq)(qs)^{-1}\right|_m = \left|rq(qs)^{-1}\right|_m = \left|rs^{-1}\right|_m$ und der „vorgemerkte“ Faktor ist m^ν, also genau (μ, ν).

- Seien $\nu < v < 0$. Dann ist

$$\frac{p}{qm^{-v}} + \frac{r}{sm^{-\nu}} = \frac{psm^{-\nu} + rqm^{-v}}{qsm^{-v-\nu}} = \frac{psm^{-\nu+v} + rq}{sqm^{-\nu}} = \frac{psm^{-\nu+v} + rq}{sq} m^\nu,$$

  dabei kürzt man mit m^v. Der Faktor m^ν wird abgesplitten, was den restlichen Ausdruck angeht, so ist $\left|(psm^{-\nu+v} + rq)(sq)^{-1}\right|_m = \left|rq(sq)^{-1}\right|_m = \left|rs^{-1}\right|_m$. Zusammen mit dem „vorgemerkten“ Faktor m^ν ergibt sich genau (μ, ν).

□

Beweis von Satz 4.24.

1. Die Fälle „irgendetwas+0“ ergeben sich nach der Definition der multimodularen Addition.[7]
2. Die Falle mit $v_i = \nu_i$ sind korrekt nach Lemma 4.26.
3. Die Fälle mit $v_i > \nu_i$ sind korrekt nach Lemma 4.27.
4. Die Fälle mit $\nu_i > v_i$ sind ebenso korrekt nach Lemma 4.27.

Damit sind alle bei der Addition möglichen Fälle abgedeckt. In der folgenden Tabelle sind $u_i, \mu_i \in \mathbb{Z}_{m_i}$ und $z_i, \zeta_i, v_i, \nu_i \in \mathcal{Z}$ mit $v_i, \nu_i > 0$ für $i = 1, \ldots, n$. Alle anderen für v_i, ν_i möglichen Werte sind mit Vorzeichen bzw. 0 verdeutlicht.

Fall	$(0, z_i)$	(u_i, v_i)	$(u_i, 0)$	$(u_i, -v_i)$
$(0, \zeta_i)$	1.	1.	1.	1.
(μ_i, ν_i)	1.	(A)	4.	4.
$(\mu_i, 0)$	1.	3.	2.	4.
$(\mu_i, -\nu_i)$	1.	3.	3.	(B)

$$(A)\begin{cases} v_i > \nu_i & 3. \\ v_i = \nu_i & 2. \\ v_i < \nu_i & 4. \end{cases} \qquad (B)\begin{cases} v_i < \nu_i & 4. \\ v_i = \nu_i & 2. \\ v_i > \nu_i & 3. \end{cases}$$

□

[7] Falls man es genauer betrachten will: Ist bei $(u, v) + (\mu, \nu)$ der Wert $u = 0$, so übernimmt man (μ, ν) als Ergebnis. Der Wert von v wird dabei gar nicht Rücksicht genommen. Es kann allerdings vorteilhaft sein die Ausdrücke der Form $[(0, z_i) : i = 1, \ldots, n]$ mit $z_i \in \mathcal{Z}$ für $i = 1, \ldots, n$ stets auf $[(0, 0) : i = 1, \ldots, n]$ zu normieren.

4.5.3 Die Multiplikation

Satz 4.28 (Die Multiplikation funktioniert)**.** *Seien a, b in X. Es gilt*

$$\varphi(ab) = \varphi(a)\varphi(b) \quad (\in \mathbb{W}_\beta). \tag{4.31}$$

Bezeichnet man mit $n(x)$ und $z(x)$ den Nenner und den Zähler des Bruches $x \in \mathbb{Q}$, so kann das folgende Lemma formuliert werden.

Lemma 4.29. *Seien a, b in X. Seien v und ν die Faktoren vom i–ten Element von $\varphi(a)$ und $\varphi(b)$ entsprechend, also ist $a = a'm_i^v$ und $b = b'm_i^\nu$. Dann gilt für den Faktor c'*

$$c' = a'b', \tag{4.32}$$

für $c = ab = c'm_i^{v+\nu}$ mit $m_i \nmid n(c')$ und $m_i \nmid z(c')$.

Beweis. Sei $c = ab$. Dann ist $c = ab = a'm_i^v b'm_i^\nu = \underbrace{a'b'}_{=:c'} m_i^{v+\nu}$. □

Beweis vom Satz 4.28. Da die Multiplikation komponentenweise ausgeführt wird, genügt es die i–te Komponente zu betrachten. Seien $c = ab$, $c' = a'b'$ und v, ν im Sinne von Lemma 4.29. Setze $(u, v) := \varphi_i(a)$ und $(\mu, \nu) := \varphi_i(b)$. Wegen der Definition 4.15 ist klar: $u = |a'|_{m_i}$ und $\mu = |b'|_{m_i}$. Die Multiplikation nach der Definition 4.12 in $\mathbb{W}_\beta$ liefert

$$(\varphi(a)\varphi(b))_i = (u \odot \mu, v + \nu) = (|u\mu|_{m_i}, v + \nu)\,.$$

Aus dem Beweis von Lemma 4.29 und der Bemerkung 4.16 zu der Definition 4.15 von φ ist ersichtlich, dass

$$\varphi_i(ab) = \varphi_i(c) = \left(|c'|_{m_i}, v + \nu\right).$$

Bleibt zu zeigen:

$$|c'|_{m_i} = u \odot \mu. \tag{4.33}$$

Es ist einerseits $u \odot \mu = |u\mu|_{m_i}$, und andererseits ist nach Lemma 4.29

$$|c'|_{m_i} = |a'b'|_{m_i} = \left||a'|_{m_i}|b'|_{m_i}\right|_{m_i}.$$

Nun folgt (4.33) aus dem Satz 1.10. □

Bemerkung 4.30 (Wichtig!). Der entscheidende Nachteil des gesamten Konzepts ist die Tatsache, dass a und b Brüche aus $\mathbb{F}_N$ enthalten. Die obigen Sätze versichern, dass die jeweilige Operation korrekt ausgeführt wird, falls das Ergebnis wieder in X ist, also Brüche aus $\mathbb{F}_N$, aber nicht $\mathbb{F}_{\hat{N}}$, $\hat{N} > N$ in $\mathbb{N}$, enthält. Das ist aber nicht mit Farey–Brüchen garantiert, siehe Lemma 2.23.

4.5.4 φ ist ein Homomorphismus

Satz 4.31. *Die Abbildung $\varphi : X \to \mathbb{W}_\beta$ ist ein Homomorphismus.*

Beweis. Die Vorwärtsabbildung φ erhält die Addition (Satz 4.24) und die Multiplikation (Satz 4.28). Es muss noch gezeigt werden, dass die neutralen Elemente erhalten bleiben.
Nun ist $0 := \varphi(0) = [(0, z_i) : i = 1, \ldots, n]$ mit z_i beliebig für alle i. Nach der Definition der Addition in $\mathbb{W}_\beta$ (Abschnitt 4.2.4) ist $x + 0 = x$ für alle $x \in \mathbb{W}_\beta$.
Die Eins ist $1 := \varphi(1) = [(1, 0) : i = 1, \ldots, n]$ und, da die Multiplikation komponentenweise ausgeführt wird, so wegen $|1|_{m_i} \odot x = x$ und $m_i^v m_i^0 = m_i^v$ für alle $x \in \mathbb{Z}_{m_i}$ und alle $v \in \mathcal{Z}$, ist $1 \in \mathbb{W}_\beta$ das multiplikativ neutrale Element. □

4.5.5 Inverse

Bemerkung 4.32. Mit $-x$ wird die *additive Inverse* von x bezeichnet.

Satz 4.33. *Für $a \in X$ gilt es:*

$$\varphi(-a) = -\varphi(a) \tag{4.34}$$

und

$$\varphi(a^{-1}) = (\varphi(a))^{-1}. \tag{4.35}$$

Beweis.

1. Zu (4.34): Man betrachtet die i–te Komponente von $\mathbb{W}_\beta$: Setze $m := m_i$. Es ist $\varphi_i(a) = (|a'|_m, v)$ und $a = a'm^v$. Ebenso ist $(-\varphi(a))_i = (-|a'|_m, v)$ und $(|a'|_m \oplus -|a'|_m, v) = (0, v) = \varphi_i(0)$. Nun ist $\varphi_i(-a) = (|c|_m, \nu)$ mit $-a = cm^\nu$. Nach der Definition von v_i für $i = 1, \ldots, n$ in (4.22) ist das Vorzeichen von a bei der Bildung der Faktoren $m_i^{v_i}$ für $i = 1, \ldots, n$ nicht relevant. Somit ist $v = \nu$. Aus $a = a'm^v$ folgt $-a'm^v = -a = cm^v$, also ist auch $|c|_m = |-a'|_m$. Nun ist $|c|_m \oplus |a'|_m = 0 (\mathrm{mod}\ m)$, das heißt $(|c|_m, \nu) = \varphi_i(-a) = (-\varphi(a))_i$.

2. Zu (4.35): Es wird wieder komponentenweise vorgegangen, es ist *zu zeigen:* $\varphi_i\left(\frac{s}{t}\right)^{-1} = \varphi_i\left(\frac{t}{s}\right)$. Es ist wieder $m := m_i$. Sei $\frac{s}{t} = \frac{s'}{t'}m^v$ und $u = |s't'^{-1}|_m$, also $\varphi_i\left(\frac{s}{t}\right) = (u, v)$. Nun ist mit Lemma 4.22: $(u, v)^{-1} = \left(|u|_m^{-1}, -v\right) = \left(|s'^{-1}(t'^{-1})^{-1}|, -v\right) = \left(|s'^{-1}t'|_m, -v\right)$. Andererseits ist $\varphi_i\left(\frac{t}{s}\right) = (\mu, \nu)$ mit $\frac{t}{s} = \frac{t'}{s'}m^\nu$ und $\mu = |t's'^{-1}|_m$. Man sieht: $|u|_m^{-1} = \mu$.
   Noch zu zeigen: $v = -\nu$. Der Fall $v = 0 = \nu$ ist trivial. Sei also $v > 0$. Dann ist $\frac{s}{t} = \frac{s'm^v}{t'} \Rightarrow \nu < 0$ mit $\frac{t}{s} = \frac{t'}{s'm^\nu}$. Der Fall $v < 0$ geht analog.

□

Kapitel 5

Berechnung der Determinante der ganzzahligen Matrix

5.1 Motivation

But... what is it good for?

Engineer at the Advanced Computing Systems
Division of IBM, 1968, commenting
on the microchip.

Es gibt bestimmte Klassen von Matrizen mit kleinen Determinanten. Beispielsweise haben die ganzzahligen *unimodularen* Matrizen die Determinante ± 1 (genauer gesagt: die Matrix $\mathbf{A}$ über einem Ring R ist unimodular genau dann, wenn $\det \mathbf{A} \in R^*$ ist). Nun ist die Arithmetik in $\mathbb{W}_\beta$ bestens geeignet, um die Matrizen auf die *Nicht*unimodularität zu testen. Die Zwischenergebnisse bei der Rechnung in $\mathbb{W}_\beta$ können und *werden* falsch, vgl. Bemerkung 4.13 (bzw. die Bemerkung 4.20), aber solange das Endergebnis in Y ist, so ist es korrekt. Da man Primzahlen für m_i, $i = 1, \ldots, n$ wählt, kann man in jedem Restklassenring $\mathbb{Z}_{m_i}$ die Eins darstellen, es geht mit *allen* Primzahlen. Das heißt, die Primzahlen in β und deren Anzahl $\#\beta$ müssen nicht so groß sein. Extremfall ist $\beta = [2]$. Und entweder ist das Ergebnis 0 und die Eingabematrix ist *sicherlich nicht* unimodular, oder das Ergebnis ist 1, und man weiss es nicht: die Matrix *könnte* unimodular sein. In diesem extremen Fall bekommt man eine korrekte Aussage trotz der Tatsache, dass *alle* Zwischenergebnisse falsch sind.

Beispiel 5.1 (Das Beispiel *in extremo*)**.** *Die Matrix*

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

ist nicht *unimodular. Die Darstellung von* $\mathbf{A}$ *modulo* 2 *ist*

$$|\mathbf{A}|_2 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix},$$

die Determinante dieser Matrix ist 0*. Man sieht:* $\mathbf{A}$ *ist* sicherlich *nicht unimodular.*
Die Matrix

$$\mathbf{B} = \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}$$

ist nicht *unimodular. Die Darstellung von* $\mathbf{B}$ *modulo* 2 *ist*

$$|\mathbf{B}|_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

die Determinante dieser Matrix ist 1*. Testergebnis: die Matrix* $\mathbf{B}$ könnte *unimodular sein. Die Matrix*

$$\mathbf{C} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

ist *unimodular. Die Darstellung von* $\mathbf{C}$ *modulo* 2 *stimmt mit* $\mathbf{C}$ *überein. Die Determinante von* $\mathbf{C}$ *ist* 1*, die Matrix* könnte *unimodular sein.*

Im Folgenden wird ein allgemeines Verfahren zur Determinantenberechnung präsentiert. Das beschränkte Wachstum der Matrixeinträge im Laufe der Berechnung und – bei korrekt gesetztem Parameter β – korrektes genaueres Ergebnis bilden den Reiz, die multimodulare Arithmetik für diese Problemstellung zu verwenden. Die Wahl der ganzzahligen Matrizen liefert eine ganzzahlige Determinante, obwohl die Zwischenergebnisse i. A. nur rational sind. Andererseits bieten große Einträge und relativ viele Rechenoperationen eine gute Möglichkeit, die multimodulare Arithmetik an einem praxisnahen Beispiel zu testen.

5.2 Naiver Ansatz

Man definiert zunächst präzise den Gegenstand der Berechnung und einige Eigenschaften davon.

Definition 5.2. Die *Determinante* der $r \times r$–Matrix $\mathbf{A}$ ist in Zeichen:

$$\det \mathbf{A} = \det \begin{pmatrix} a_{1,1} & \dots & a_{1,r} \\ \vdots & \ddots & \vdots \\ a_{r,1} & \dots & a_{r,r} \end{pmatrix} = \begin{vmatrix} a_{1,1} & \dots & a_{1,r} \\ \vdots & \ddots & \vdots \\ a_{r,1} & \dots & a_{r,r} \end{vmatrix}. \tag{5.1}$$

Definition 5.3. Sei $\mathbf{A}$ eine $r \times r$–Matrix. Mit $\mathbf{A}^{(i,j)}$ wird eine $(r-1) \times (r-1)$–Matrix bezeichnet, die die Matrix $\mathbf{A}$ mit *gestrichener* i–ten Zeile und j–ter Spalte ist.

Proposition 5.4 (Nach [7, S. 144–150, 162], ohne Beweis)**.** *Betrachte eine* $r \times r$*–Matrix* $\mathbf{A}$*.*

1. *Die Determinante ist eine alternierende multilineare Form.*
2. *Die Determinante der Einheitsmatrix* $\mathbf{I}$ *ist Eins.*
3. *Die Leibnizsche Entwicklungsformel lautet*

$$\det \mathbf{A} = \sum_{i=1}^{r} (-1)^{i+j} a_{i,j} \det \mathbf{A}^{(i,j)}. \tag{5.2}$$

4. *Die Determinante der transponierten Matrix ist gleich der Determinante der Matrix selbst.*
5. *Vertauscht man zwei Zeilen der Matrix, so verändert sich nur das Vorzeichen der Determinante.*
6. *Es ist äquivalent*
   (a) $\mathbf{A}$ *ist invertierbar. (Also:* $\mathbf{A}$ *ist* nicht singulär*.)*
   (b) *Die Spalten von* $\mathbf{A}$ *sind linear unabhängig.*
   (c) $\det \mathbf{A} \neq 0$.

Die Verwendung der Leibnizschen Entwickungsformel (5.2) ist bei größeren Matrizen zum Scheitern verurteilt, da sie eine exponentielle Laufzeit hat: Berechnet man mit (5.2) die Determinante einer $r \times r$–Matrix $\mathbf{A}$, so müssen n Determinanten von $(r-1) \times (r-1)$–Matrizen berechnet werden, für die wiederrum $r(r-1)$ Determinanten von $(r-2) \times (r-2)$–Matrizen benötigt werden, usw.

5.3 Die LU–Zerlegung

Das Mittel der Wahl bei der numerischen Determinantenberechnung ist die **LU**–Zerlegung, die auf Gauß zurückgeht. Durch geschicktes sukzessives Multiplizieren von geeigneten Matrizen $\mathbf{Y}_i$ an die quadratische Ausgangsmatrix $\mathbf{A}$ erhält man die *obere Dreiecksmatrix* $\mathbf{U}$, und das Produkt der $\mathbf{Y}_i$ ist die *untere Dreiecksmatrix* $\mathbf{L}$, wobei

$$\mathbf{A} = \mathbf{LU}.$$

Man kann die Dreiecksmatrizen so wählen, dass die *Diagonale* der Matrix $\mathbf{L}$ nur Einsen enthält:

$$l_{i,j} = \begin{cases} * & i > j \\ 1 & i = j \\ 0 & i < j. \end{cases}$$

Solch eine Matrix wird nach [2] die *Einheitsdreiecksmatrix* genannt. In diesem Fall ist die Determinante von $\mathbf{A}$ das Produkt der Diagonalelemente von $\mathbf{U}$.

Lemma 5.5. *Die Determinante der Dreiecksmatrix* $\mathbf{U}$ *ist das Produkt ihrer Diagonalelemente.*

Beweis. Man geht induktiv über Länge r der Matrix $\mathbf{U}$ vor. Der Fall $r = 1$ ist trivial. Sei nun die Behauptung für den Fall $r-1$ bewiesen. So ist die Determinante der $(r-1)\times(r-1)$–Teilmatrix $\mathbf{U}^{(r,r)}$ das Produkt ihrer Diagonalelemente. Betrachtet man die Leibnizsche Entwicklung (5.2) nach dem Element $u_{r,r}$, das *nicht* in der Teilmatrix $\mathbf{U}^{(r,r)}$ ist, so ergibt sich die Behauptung. □

Lemma 5.6 (Permutationsmatrizen)**.** *Die Determinante einer* Permutationsmatrix $\mathbf{P}$*, also solch einer Matrix, dass in jeder Zeile oder Spalte jeweils nur eine Eins steht und sonst Nullen, ist* ± 1*. Das Vorzeichen kann durch Zählen der Permutationen bestimmt werden.*

Beweis. Durch sukzessives Vertauschen der Zeilen der Permutationsmatrix erhählt man die Einheitsmatrix $\mathbf{I}$. Anhängig von der Anzahl der Vertauschungen ist $\det \mathbf{P} = 1$ oder $\det \mathbf{P} = -1$. □

Also ist, falls man die Pivotsuche verwendet, $\mathbf{A} = \mathbf{PLU}$ mit der Permutationsmatrix $\mathbf{P}$. In diesem Fall wird die Determinante von $\mathbf{A}$ bis auf das Vorzeichen durch $u_{1,1} \cdots u_{r,r}$ bestimmt. Die folgende Aussagen stammen aus [2, Section 3.1.8], hier werden sie ohne Beweis angegeben.

Proposition 5.7 (Eigenschaften der Dreiecksmatrizen)**.**

1. *Die* Inverse *der oberen (unteren) Dreiecksmatrix ist eine obere (untere) Dreiecksmatrix.*
2. *Das* Produkt *der oberen (unteren) Dreiecksmatrix ist eine obere (untere) Dreiecksmatrix.*
3. *Die* Inverse *der oberen (unteren) Einheitsdreiecksmatrix ist eine obere (untere) Einheitsdreiecksmatrix.*
4. *Das* Produkt *der oberen (unteren) Einheitsdreiecksmatrix ist eine obere (untere) Einheitsdreiecksmatrix.*

5.3.1 Das Verfahren

Man bezweckt die Zerlegung der Matrix $\mathbf{A}$ in $\mathbf{A} = \mathbf{LU}$. Dieser Abschnitt ist präsentiert nach [12, S. 38ff.] und [2, Chapter 3].

Definition 5.8 (Teilmatrizen)**.** Sei $\mathbf{A}$ eine $r \times r$–Matrix. Dann bezeichne mit $\mathbf{A}_k$ die $k \times k$–*Teilmatrix* von $\mathbf{A}$:

$$\mathbf{A}_k = \begin{pmatrix} a_{1,1} & \dots & a_{1,k} \\ \vdots & \ddots & \vdots \\ a_{k,1} & \dots & a_{k,k} \end{pmatrix}. \tag{5.3}$$

Satz 5.9 ([2, Theorem 3.2.1], ohne Beweis)**.** *Sei* $\mathbf{A}$ *eine* $r \times r$*–Matrix.*

1. *Falls alle Teilmatrizen* $\mathbf{A}_k$ *von* $\mathbf{A}$ *mit* $k = 1, \dots, r$ *von Null verschiedenen Determinanten haben, so existiert die* $\mathbf{LU}$*–Zerlegung von* $\mathbf{A}$*.*
2. *Existiert die* $\mathbf{LU}$*–Zerlegung von* $\mathbf{A}$ *und ist* $\mathbf{A}$ *nicht singulär, dann gilt*
   (a) *Die* $\mathbf{LU}$*–Zerlegung von* $\mathbf{A}$ *ist eindeutig.*
   (b) *Falls* $\mathbf{L}$ *als Einheitsdreiecksmatrix konstruiert ist, so ist* $\det \mathbf{A} = u_{1,1} \cdots u_{r,r}$*.*

Einige Details der Implementierung folgen.

Definition 5.10 (Gauß–Transformationen)**.** Die Matrizen $\mathbf{Y}_i$ heißen *Gauß–Matrizen* bzw. *Gauß–Transformationen* und haben die Form

$$\mathbf{Y}_i = \mathbf{I} - \mathbf{y}\mathbf{1}_i^T = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & -y_{i+1} & 1 & \\ & & \vdots & & \ddots \\ & & -y_r & & 1 \end{pmatrix}. \tag{5.4}$$

Achtung: $y_1, \dots, y_i = 0$.

Lemma 5.11 (Inverse der Gauß–Transformation)**.**

$$(\mathbf{I} - \mathbf{y}\mathbf{1}_i^T)^{-1} = \mathbf{I} + \mathbf{y}\mathbf{1}_i^T. \tag{5.5}$$

Beweis. $(\mathbf{I} - \mathbf{y}\mathbf{1}_i^T)(\mathbf{I} + \mathbf{y}\mathbf{1}_i^T) = \mathbf{I} - \mathbf{y}\mathbf{1}_i^T + \mathbf{y}\mathbf{1}_i^T - \mathbf{y}\mathbf{1}_i^T\mathbf{y}\mathbf{1}_i^T = \mathbf{I}$, da $\mathbf{1}_i^T\mathbf{y} = y_i = 0$. □

Algorithmus 15 (Einfache Gauß–Elimination)**.**
Eingabe: *$r \times r$–Matrix* $\mathbf{A}$*.*
Für $k = 1, \dots, r$ *und für* $j = k + 1, \dots, r$ *wiederhole*

1. $y \leftarrow \frac{a_{j,k}}{a_{k,k}}$, $a_{j,k} \leftarrow y$.
2. *Für* $l = k + 1, \dots, r$ *wiederhole*
$$a_{j,l} \leftarrow a_{j,l} - y a_{k,l}. \tag{5.6}$$

Ausgabe: *Umgeformte Matrix* $\mathbf{A} = \mathbf{LU}$*.*

Algorithmus 16 (Schnelle Berechnung der Determinante)**.**
Eingabe: $r \times r$*–Matrix* $\mathbf{A}$*.*
Bestimme mit Algorithmus 15 die untere Einheitsdreiecksmatrix $\mathbf{L}$ *und obere Dreiecksmatrix* $\mathbf{U}$ *mit* $\mathbf{A} = \mathbf{LU}$*. Berechne* $\det \mathbf{A} = u_{1,1} \cdots u_{r,r}$*.*
Ausgabe: $\det \mathbf{A}$*.*

Bemerkung 5.12 (Wichtig!)*.* Der Algorithmus 15 (und entsprechend auch der Algorithmus 16) funktioniert nicht für jede Matrix, sondern *nur* für diejenigen Matrizen, die die Voraussetzungen des Satzes 5.9 erfüllen. Zum Beispiel scheitert bei der Matrix

$$\begin{pmatrix} 0 & 3 \\ 4 & 0 \end{pmatrix}$$

die **LU**–Zerlegung.

Zum Robustheitstest der rationalen Arithmetik in $\mathbb{W}_\beta$ stellt die Pivot–Suche keine Abhilfe, sondern eher einen Störungsfaktor dar. Dabei verändert sich die Theorie kaum, da die Determinante der Pivot–Matrizen ± 1 ist, vgl. Lemma 5.6. Der einzige Fall, in dem die Pivotsuche von Nutzen ist, ist der Fall der singulären Teilmatrix $\mathbf{A}_k$. Da in diesem Fall die Vorrausetzungen des Satzes 5.9 nicht erfüllt sind, muss man eine passende Permutation anwenden. In dem obigen Beispiel wäre die Permutationsmatrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

eine mögliche Lösung.

5.3.2 Maße und Abschätzungen

Um die geeigneten Grenzen anzugeben, braucht man einige Abschätzungen, sowohl für das Endergebnis als auch für die bei der Berechnung vorkommenden Werte.

Größe der Zwischenergebnisse

Da die Arithmetik in $(\mathbb{W}_\beta, +, \cdot)$ beschränkt ist, braucht man ein Verfahren zur Bestimmung des größten bei der Berechnung vorkommenden Elementes. Genauer gesagt zählt nicht die *absolute* Größe, sondern die Größe von Zähler und Nenner, denn die Dimensionierung der Arithmetik geschieht mit (2.8), was Farey–Brüche involviert.

Definition 5.13.

1. Definiere die „Farey–Größe“ γ eines Bruches $\frac{a}{b} \in \mathbb{Q}$ als

$$\gamma\left(\frac{a}{b}\right) = \max\{|a|, |b|\}. \tag{5.7}$$

2. Definiere das *maximale Element der gaußschen* $\mathbf{LU}$*-Zerlegung der Matrix* $\mathbf{A}$*,* $\Gamma(\mathbf{A})$ bezüglich γ als

$$\Gamma(\mathbf{A}) = \max_{1 \le i,j,k \le r} \gamma\left(a_{j,k}^{(i)}\right). \tag{5.8}$$

   Hierbei bezeichnet $a_{j,k}^{(i)}$ den Eintrag auf der j–ten Zeile und k–ten Spalte des Ergebnisses des i–ten Schrittes der $\mathbf{LU}$–Zerlegung.

Größe der Determinante

Man kann die Größe der Determinante aus der Größe der Einträge der Matrix $\mathbf{A}$ abschätzen. Die Hadamard–Abschätzung für reellwertige Matrizen gibt die obere Schranke an [17, Hadamard's Inequality].

Lemma 5.14 (Hadamard–Abschätzung)**.** *Sei* $\mathbf{A}$ *eine* $r \times r$*–Matrix. Es gilt*

$$\det \mathbf{A} \leq \sqrt{\prod_{i=1}^{r} \sum_{j=1}^{r} a_{i,j}^2} =: H(\mathbf{A}).$$

Beweisidee. Betrachte die Determinante von $\mathbf{A}\mathbf{A}^T$, verwende die Gram–Schmidt Orthonormalisierung. Vgl. [6, Abschnitt 4.6.1, Übung 15]. †

Diese fundamentale Abschätzung erlaubt, die Dimensionierung der Arithmetik vor der eigentlichen Berechnung vorzunehmen. Die Ordnung der Farey–Brüche N soll $N > H(\mathbf{A})$ erfüllen, also

$$M > 2H(\mathbf{A})^2. \tag{5.9}$$

Der Aufwand

Der Aufwand der Gauß–Elimination beträgt nach [12, S. 45] für eine $r \times r$–Matrix

$$\frac{3}{2}r^3 - \frac{1}{2}r^2 - \frac{1}{6}r \text{ } \textit{flops}.$$

Allerdings ist es die Anzahl der ***fl****oating point* ***op****erations*; der Aufwand der *exakten rationalen* Berechnung hängt von der Größe der Zähler und Nenner, also genau von $\gamma(\cdot)$ aller Einträge aller Zwischenergebnisse ab. Die einfache Abschätzung für die obere Grenze gibt exponentielles Wachstum an, allerdings kann man zeigen, dass das Wachstum der Zwischenergebnisse und die Anzahl der Wortoperationen in $\mathbb{Q}$ polynomial beschränkt bleibt [16]. Man verwendet hier $\mathcal{O}(\cdot)$, das Landau–Symbol, vgl. [17, Landau Symbols]. Rechnet man in $\mathbb{Z}_M$, so ergibt sich der Aufwand von

$$\mathcal{O}\big(r^3 \cdot r^2(\log r + \log B)^2\big),$$

wobei $r^{r/2}B^r = H(\mathbf{A})$ gilt. Falls man multimodular rechnet, und falls der Aufwand für eine Operation modulo m_i für $i = 1, \ldots, n$ konstant ist, so ist der Aufwand der Determinantenberechnung sogar[1]

$$\mathcal{O}(r^3 n).$$

Man muss noch Aufwand für die Vorwärts– und die Rückwärtsabbildung miteinbeziehen, aber es ist zu betonen, dass die Berechnung so ausgelegt werden soll, dass die Vorwärts– und die Rückwärtsabbildung nur ein Bruchteil des eigentlichen Aufwandes sind. Man wird im nächsten Abschnitt sehen, dass die Determinantenberechnung ein solcher Fall ist.

[1] Natürlich müssen die Werte von n und $[m_i : i = 1, \ldots, n]$ so ausgelegt sein, dass

- n möglichst klein ist,
- m_i für alle $1, \ldots, n$ möglichst große verschiedene Primzahlen,
- m_i für alle $1, \ldots, n$ noch unter der Schranke, die den konstanten Aufwand für die Berechnungen in $\mathbb{Z}_{m_i}$ liefert,
- $M = m_1 \cdots m_n$ groß genug ist, um das korrekte Endergebnis in $\mathbb{Z}_M$ zu beinhalten.

Der Wechsel zu der multimodularen Restklassenarithmetik bringt im Vergleich zu dem unimodularen Fall den Aufwand (hier: ohne die logarithmische Terme) von $\mathcal{O}(r^5 \dots)$ auf $\mathcal{O}(r^3 n)$. Die in Kapitel 4 dargestellte Arithmetik in $(\mathbb{W}_\beta, +, \cdot)$ liefert sogar eine rationale endliche Arithmetik, d. h. man kann mit geringfügig größerem Aufwand bei der Vorwärts– und Rückwärtsabbildung die Berechnungen in der bestimmten Teilmenge der rationalen Zahlen vornehmen. Wie aus den praktischen Tests ersichtlich, reduziert die Berechnung in $\mathbb{W}_\beta$ mit passend gewählten β deutlich die Berechnungszeit, bei größeren Beispielen war der Unterschied zu $\mathbb{Q}$ mehrfach bis zu Faktor 10.

5.4 Die Berechnung

> The purpose of computing is insight, not numbers.
>
> R.W. Hamming

Um die verschiedene Möglichkeiten zu Determinantenberechnung zu vereinheitlichen, definiere zu einer quadratischen Matrix $\mathbf{A}$

- $\delta(\mathbf{A})$ – die Determinante von $\mathbf{A}$, berechnet mit rationalen Langzahlarithmetik.
- $d(\mathbf{A})$ – die Determinante von $\mathbf{A}$, beides die $\mathbf{LU}$–Zerlegung und das Produkt der Diagonale von $\mathbf{U}$ wird in $\mathbb{W}_\beta$ berechnet.
- $\Delta(\mathbf{A})$ – die Determinante von $\mathbf{A}$, bekannt aus der Literatur.

Anzahl der Operationen

Für die Bestimmung von $d(\mathbf{A})$ braucht man r^2 Vorwärtsabbildungen, $\mathcal{O}(r^3)$ Rechenoperationen in $\mathbb{W}_\beta$ und *eine* Rückwärtsabbildung.

Generell, da die Rückwärtsabbildung relativ „teuer“ ist, können die Vorteile der multimodularen Arithmetik sich vollständig in genau solchen Aufgaben, wie die Berechnung von d entfalten. Es wird ein Satz Daten in $\mathbb{W}_\beta$ abgebildet, *lange* dort gerechnet, und es werden vergleichmäßig wenige Ergebnisse zurück in $\mathbb{Q}$ abgebildet.

5.4.1 Die Hilbert–Matrizen

Die r–te Hilbert–Matrix $\mathbf{H} = \mathbf{H}_r$ wird definiert als $h_{i,j} = (i + j - 1)^{-1}$ für $1 \le i, j \le r$ [17, Hilbert Matrix]. Für $\mathbf{H}_{10}$ konnte mit $n = 35$ und $m_i \approx 1000$, also mit

$$N = 153717642103548757372439836543709250071816906093808599,$$

das Ergebnis

$$d(\mathbf{H}_{10}) = \det \mathbf{H_{10}} = \frac{1}{46206893947914691316295628839036278726983680000000000}$$

errechnet werden. Dieser Wert der Determinante ist korrekt.[2] Dasselbe Ergebnis konnte bei $n = 12$ mit $m_i \approx 2^{31}$ also bei

$$N = 693528087918401104528842234620872737241684748400310695 83$$

[2] Vgl. die Sequenz A005249 in N. J. A. Sloane, *The On–Line Encyclopedia of Integer Sequences,* `http://www.research.att.com/~njas/sequences/A005249`.

erreicht werden.[3] Im letzten Falle ist

$$(198403995, 0), (1736491089, 0), (245747043, 0), (2059738299, 0), (1599372216, 0), (58231925, 0),$$
$$(2061095452, 0), (815462011, 0), (1489332235, 0), (231928716, 0), (1655346919, 0), (63356032, 0)$$

der Wert der Determinante in $\mathbb{W}_\beta$. Verringert man n auf 4, so ist

$$M = 212676464470306383125965308282830336 99$$

nicht ausreichend, um das korrekte Ergebnis darzustellen. Es wird

$$d(\mathbf{H}_{10}) = \frac{1259068629079026274}{2644785098613885589}$$

berechnet. Allerdings gilt

$$\begin{aligned} |d(\mathbf{H}_{10})|_M &= \left|1259068629079026274 \cdot 2644785098613885589^{-1}\right|_M \\ &= 230371801179512067341225908497011413 \\ &\cong \left|46206893947914691316295628839036278726983680000000000^{-1}\right|_M \\ &= |\delta(\mathbf{H}_{10})|_M \,. \end{aligned}$$

Die Hilbert–Matrizen sind nicht ganzzahlig, aber ihre Inverse schon. Die *Inverse der r–ten Hilbert–Matrix* $\mathbf{H}_r^{-1}$ wird nach [17] komponentenweise definiert als[4]

$$h_{i,j}^{-1} = (-1)^{i+j}(i+j-1)\binom{r+i-1}{r-j}\binom{r+j-1}{r-i}\binom{i+j-2}{i-1}^2 .$$

Mit demselben β wie in vorherigem Beispiel und mit $n = 12$ konnte der korrekte Wert für $d(\mathbf{H}_{10}^{-1})$ bestimmt werden, nämlich $1/\Delta(\mathbf{H}_{10})$. Dieser Wert in $\mathbb{W}_\beta$ wird repräsentiert durch

$$(1590881533, 0), (549013557, 0), (2059466049, 0), (2143063689, 0), (1902246556, 0), (1622744607, 0),$$
$$(834765503, 0), (961198508, 0), (1414920503, 0), (1285403005, 0), (1930927081, 0), (1345675870, 0).$$

5.4.2 Die Pascal–Matrizen

Definiere die $r \times r$–Pascal–Matrix $\mathbf{B}$ als[5]

$$b_{i,j} = \binom{i+j-2}{i-1}$$

für $1 \le i, j \le r$. Diese Matrix ist unimodular, vgl. [17, Pascal Matrix]. Mit der $\mathbf{LU}$–Zerlegung ist es möglich die Determinante von $\mathbf{B}$ zu berechnen; wegen der besonderen Eigenschaften der Matrix ist die Zerlegung in diesem Fall ganzzahlig. Da dieser Fall weniger interessant ist, werden die Zeilen

[3] Mit

$$\beta = [2147483399, 2147483423, 2147483477, 2147483489, 2147483497, 2147483543,$$
$$2147483549, 2147483563, 2147483579, 2147483587, 2147483629, 2147483647].$$

[4] Dabei ist $\binom{n}{k}$ der Binomialkoeffizient, es ist $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

[5] $\mathbf{B}$ für „Blaise“, denn $\mathbf{P}$ steht für die Permutationsmatrix.

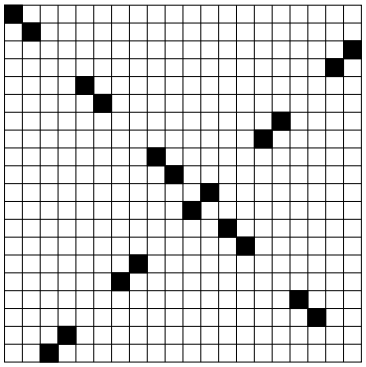


Abbildung 5.1: Die Permutationsmatrix $\mathbf{P}$ für $r = 20$, $\det \mathbf{P} = -1$. Ein schwarz eingefärbter Punkt bedeutet eine Eins, ein weißer eine Null.

von $\mathbf{B}$ vor der Berechnung permutiert. Dies liefert Brüche in der $\mathbf{LU}$–Zerlegung und vergrößert somit den Wert dieses Beispiels. Die hier verwendete Permutationsmatrix $\mathbf{P}$ wird definiert durch

$$p_{i,j} = \begin{cases} 1 & \text{falls } i = j \text{ und } 4 \nmid i,\ 4 \nmid (i+1) \\ 1 & \text{falls } i = 4k-1, k \in \mathbb{Z} \text{ und } j = r - i + 3 \\ 0 & \text{sonst,} \end{cases}$$

mit $1 \leq i, j \leq r$. In der $\mathbf{LU}$–Zerlegung der $\mathbf{PB}$ ist die Diagonale von $\mathbf{U}$

$$\begin{aligned} \operatorname{diag}\ \mathbf{U} = \Big[& 1, 1, 171, -51, 105, \frac{91}{3}, 385, -65, 77, \frac{75}{7}, \frac{77}{15}, \\ & -\frac{15}{77}, \frac{7}{75}, \frac{1}{77}, \frac{1}{65}, -\frac{1}{385}, \frac{3}{91}, \frac{1}{105}, \frac{1}{51}, -\frac{1}{171} \Big] \end{aligned}$$

somit ist auch $d(\mathbf{PB}) = \det(\mathbf{PB}) = -1$. Insgesamt waren in der zerlegten Matrix ca. 66% der Einträge nicht ganzzahlig. Die Berechnung wurde bei $r = 20$, $n = \#\beta = 4$, $\beta = [1009, 1013, 1019, 1021]$, also bei $N = 729180$, ausgeführt. Der größte in der Berechnung vorkommende Wert $\Gamma(\mathbf{PB})$ war allerdings 160291041068956. Man kann dieses Ergebnis vertiefen: Bei 50×50–Matrix sind 2222 Einträge in $\mathbb{Q} \setminus \mathbb{Z}$, $\Gamma(\mathbf{PB}) = 938380971435091317709267448795262806502 5$. Die Parameter der Arithmetik $(\mathbb{W}_\beta, +, \cdot)$ blieben dieselben, die Determinante wurde korrekt berechnet, obwohl sogar in der Diagonale von $\mathbf{U}$ einige Werte nicht zurückabgebildet werden konnten. Weitere Ergebnisse werden in einer Tabelle zusammengefasst.

r	N	d	$\Gamma(\mathbf{PB}) \approx$	Anzahl der Einträge von $\mathbf{LU}$			Zeit
				insgesamt	nicht in $\mathbb{Z}$ (%)	nicht in Y (%)	
10	729180	1	$2 \cdot 10^5$	100	50 (50%)	0 (0%)	0,03
20	729180	−1	$1,6 \cdot 10^{14}$	400	265 (66%)	36 (9%)	0,24
25	729180	1	$1,6 \cdot 10^{17}$	625	492 (78%)	79 (13%)	0,45
30	729180	−1	$3 \cdot 10^{23}$	900	722 (80%)	137 (15%)	0,82
40	729180	1	$1 \cdot 10^{31}$	1600	1333 (83%)	303 (19%)	1,94
50	729180	1	$9 \cdot 10^{39}$	2500	2222 (88%)	687 (27%)	3,75
10	10	1	$2 \cdot 10^5$	100	50 (50%)	44 (44%)	0,03

Der Fall $r = 15$ wurde in der Tabelle nicht ausgeführt, da es ein Sonderfall ist. Es gilt $\mathbf{P} = \mathbf{I}$, die Matrizen $\mathbf{L}$ und $\mathbf{U}$ sind in diesem Fall ganzzahlig, was die Berechnung weniger interessant macht.

Diese Beispiele zeigen alle Vorteile der Arithmetik in $\mathbb{W}_\beta$ für diese Art der Berechnungen. Das Endergebnis wurde genau und korrekt bestimmt, es wurde nur eine relativ kostspielige Rückwärtsabbildung verwendet, und die Berechnung war mehrfach schneller im Vergleich zu der rationalen Langzahlarithmetik. Zusätzlich ist anzumerken, dass in dem Beispiel mit $r = 50$ eine direkte Lösung in `octave` getestet wurde, und falsche Ergebnisse lieferte, da während der Berechnung für die 32–bit *floating point*–Arithmetik zu große Werte vorgekommen sind.

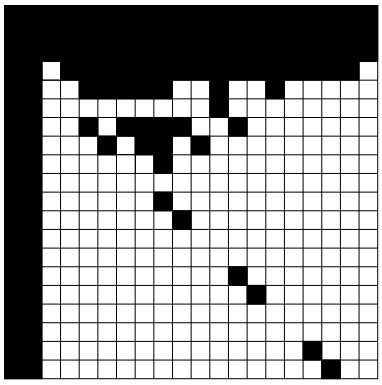


Abbildung 5.2: Das Ergebnis $\mathbf{L} + \mathbf{U} - \mathbf{I}$ der $\mathbf{LU}$–Zerlegung von $\mathbf{PB}$ mit $r = 20$. Ein schwarz eingefärbter Punkt bedeutet eine Zahl in $\mathbb{Z}$, ein weißer eine in $\mathbb{Q} \setminus \mathbb{Z}$.

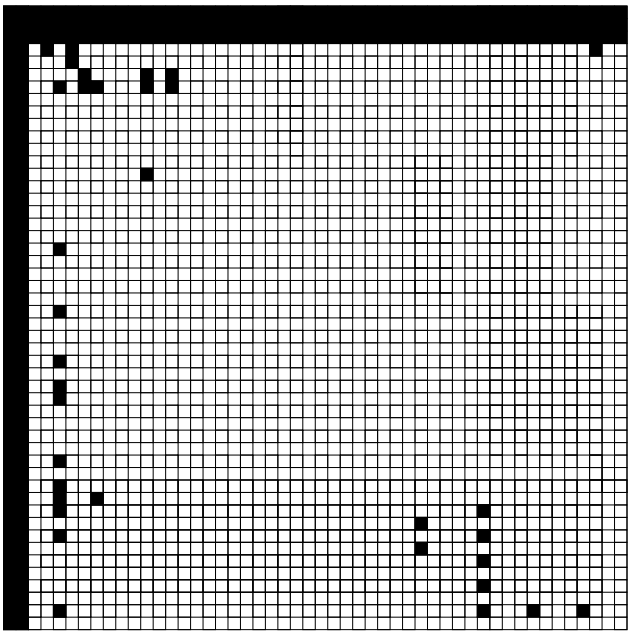


Abbildung 5.3: Dasselbe, wie in der Abbildung 5.2, nur mit $r = 50$. Nachwievor bedeutet ein eingefärbter Punkt eine Zahl in $\mathbb{Z}$, ein weißer eine Zahl in $\mathbb{Q} \setminus \mathbb{Z}$.

Anhang

Eine lauffähige Implementierung sowohl von $\mathbb{M}_\beta$ als auch von $\mathbb{W}_\beta$ findet man auf der beigelegten CD. Für eine genauere Dokumentation sei man auf die `README`–Datei, Beispielanwendungen und den Quelltext selbst hingewiesen. Die Implementierungssprache ist `C++`, es wurden die Langzahlbibliothek `cln` und die *smart pointers* aus `boost` verwendet.

Literaturverzeichnis

[1] Henri Cohen. *A Course in Computational Algebraic Number Theory*. Springer, 1995.

[2] Gene H. Golub and Charles F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, London, second edition, 1989.

[3] R. T. Gregory and E. V. Krishnamurthy. *Methods and Applications of Error–Free Computation*. Springer, 1984.

[4] Larry C. Grove. *Algebra*. Dover Publications Inc., Mineola, New–York, 2004. Republication of edition by Academic Press, 1983.

[5] A. Ya. Khinchin. *Continued Fractions*. Dover Publications, Inc., Mineola, New York, 1997. Originally published by University of Chicago Press, 1964.

[6] Donald E. Knuth. *The Art of Computer Programming*, volume 2. Addison–Wesley, third edition, 1998.

[7] Serge Lang. *Linear Algebra*. Springer, third edition, 1987.

[8] MacTutor. In *History of Mathematics Archive*. University of St. Andrews, Scotland, `http://www-history.mcs.st-andrews.ac.uk/`, 2006.

[9] P. Montgomery. Modular multiplication without trial division. *Math. Comp.*, 44:519–521, 1985.

[10] Oskar Perron. *Die Lehre von den Kettenbrüchen*, volume I. B. G. Teubner Verlagungsgesellschaft, Stuttgart, dritte edition, 1954.

[11] Tomas Sauer. Computeralgebra. Vorlesungsskript, Justus–Liebig–Universität Gießen, 2002.

[12] Tomas Sauer. Numerische Mathematik I. Vorlesungsskript, Justus–Liebig–Universität Gießen, 2002.

[13] Tomas Sauer. Kettenbrüche. Vorlesungsskript, Justus–Liebig–Universität Gießen, 2005.

[14] B. L. van der Waerden. *Algebra*, volume II. Springer, 1959.

[15] B. L. van der Waerden. *Algebra*, volume I. Springer, 1971.

[16] Joachim von zur Gathen and Jürgen Gerhard. *Modern Computer Algebra*. Cambridge University Press, 1999.

[17] Eric W. Weisstein. In *MathWorld*. Wolfram Web Resource, `http://mathworld.wolfram.com/`, 2006.

Selbstständigkeitserklärung

Hiermit versichere ich, die vorliegende Diplomarbeit im Fach Mathematik selbstständig verfasst und nur die angegebene Hilfsmittel verwendet zu haben.

Gießen, den 14. März 2007 .....