Grundlagen für Ingenieure & Informatiker

Zahlentheorie

Revision : 317

Inhaltsverzeichnis

Algorithmenverzeichnis

Symbolverzeichnis

$(G, \diamond)$	Verknüpfungsgebilde, bestehend aus Menge (Gruppe) G und Operation $\diamond$
$(R, +, \cdot)$	Ring
$(R[x], +, \cdot)$	Polynomring
$(R_m, +, \cdot)$	Restklassenring
$[M : K]$	Grad der Körpererweiterung M über K
$\lceil x \rceil$	Kleinste ganze Zahl größer als x (ceil)
$\left(\frac{a}{b}\right)$	L -Symbol: a über b, auch geschrieben als $(a \mid b)$
$\langle a \rangle$	Vom Element a erzeugte Untergruppe
$\lvert G : H \rvert$	Index der Gruppe G über H
$\lvert G \rvert$	Ordnung der Gruppe G
$\lfloor x \rfloor$	Größte ganze Zahl kleiner als x (floor)
$[r]_m$	Restklasse zum Modul m
$\{a_1, a_2, \ldots\}$	Menge von Elementen a_i
$\#G$	Ordnung der Gruppe G
$\diamond$	Operation $\diamond$
$\equiv$	Kongruenz
δ_{ij}	K -Symbol
$\phi(m)$	E 'sche Totient-Funktion
$\overline{a}$	Teilerfremder Teil von a bezüglich einer anderen Zahl
$a \perp b$	a und b sind teilerfremd
$a \mapsto b$	Abbildung von a auf b

$a \mid b$	a teilt b
$a \setminus b$	Differenzmenge a abzüglich b
$a \subseteq b$	a ist Teilmenge von b
$\mathbb{C}$	Körper der komplexen Zahlen
$\deg h(x)$	Grad des Polynoms $h(x)$
$\dim(V)$	Dimension des Vektorraums V
e	neutrales Element
$e_{(+)}$	Null-Element (neutrales Element der additiven Gruppe)
$e_{(\cdot)}$	Eins-Element (neutrales Element der multiplikativen Gruppe)
$\mathbb{F}_p^n$	Erweiterungskörper der Dimension n über p
$\mathbb{F}_q$	Endlicher Körper mit q Elementen
$\mathbb{F}_q^*$	Multiplikative Gruppe des endlichen Körpers $\mathbb{F}_q$
$\gcd(a,b)$	Größter gemeinsamer Teiler von a und b
GF(q)	G -Körper mit q Elementen
$\mathrm{htr}(\alpha)$	Halb-Trace von α
K	Körper
$\mathrm{lcm}(a,b)$	Kleinstes gemeinsames Vielfaches von a und b
mod	Modulo
$\mathbb{N}$	Menge der natürlichen Zahlen
$\mathrm{ord}(G)$	Ordnung der Gruppe G
$\mathbb{P}$	Menge der Primzahlen
p	Primzahl (bzw. Primelement)
$\mathbb{Q}$	Körper der rationalen Zahlen
$\mathbb{R}$	Körper der reellen Zahlen
$\mathrm{tr}(\alpha)$	Trace (Spur) von α
$\mathbf{v}$	Vektor v
V	Vektorraum

$\mathbb{Z}$	Menge der ganzen Zahlen
$\mathbb{Z}_m$	Restklassenring/Menge[1] der ganzen Zahlen $\bmod m$
$\mathbb{Z}_p$	Primzahlenkörper
$\mathbb{Z}_p[x]/m(x)$	Polynom-Restklassenring auf dem Grundkörper $\mathbb{Z}_p$

1 Algebraische Grundstrukturen

1.1 Gruppen

1.1.1 Allgemeine Definition

Eine Gruppe[2] $(G, \diamond)$ ist ein algebraisches System, daß aus einer nicht-leeren Menge von Elementen G besteht, für die eine Operation $\diamond$ mit folgenden Eigenschaften definiert ist [PW72, 2.1], [Kör88, 101 ff.]:

1. die Anwendung der Operation $\diamond$ auf zwei Elemente muß wieder zu einem Element in G führen (Abgeschlossenheit): $\diamond : G \times G \rightarrow G$;
2. für alle $a, b, c \in G$ muß gelten: $a \diamond (b \diamond c) = (a \diamond b) \diamond c$ (Assoziativität);
3. ein neutrales Element $e \in G$ muß existieren, so daß gilt: $a \diamond e = a$ mit $a \in G$ (Identität);
4. für jedes Element $a \in G$ muß ein inverses Element $b \in G$ existieren, so daß $a \diamond b = e$ gilt.

Die Menge G kann aus einer endlichen oder unendlichen Anzahl von Elementen a_i bestehen. Bei einer Aufzählung der Elemente werden diese in geschweifte Klammern eingeschlossen, z. B. so: $\{a_1, a_2, \ldots\}$. Ist die Zusatzbedingung $a \diamond b = b \diamond a$ erfüllt, wird die Gruppe kommutativ bzw. A 'sch genannt.

Die Ordnung (auch Mächtigkeit oder Kardinalität) einer Gruppe ist die Anzahl der Elemente in G, geschrieben als $\mathrm{ord}(G)$, $|G|$ oder auch $\#G$. Ist U eine Teilmenge von G, d. h. $U \subseteq G$, dann wird $(U, \diamond)$ als Untergruppe von $(G, \diamond)$ bezeichnet, wenn mit derselben Operation $\diamond$ auch U alle Eigenschaften einer Gruppe erfüllt [Bos96, 1.], [PW72, 2.4]. Der Zusammenhang zwischen der Ordnung von U und der von G wird durch den Satz von L beschrieben.

$$|G| = i|U| \tag{1.1}$$

Die Ordnung $|U|$ ist danach ein Teiler von $|G|$ und i der so genannte Index von G über U (bzw. Index von U in G), welcher auch als $i = |G : U|$ geschrieben wird.[3]

[1] Die ganzen Zahlen von 0 bis $m-1$

[2] Der Begriff wurde zuerst von G benutzt.

[3] Die Bezeichnungsweise des Index in der Form $|G : U|$ ist dabei durch den Quotienten $\frac{|G|}{|U|}$ motiviert. Interessant ist in diesem Zusammenhang auch noch, daß die Ordnung der Untergruppe U höchstens halb so groß sein kann, wie die der Gruppe G.

1.1.2 Additive Gruppen

Entsprechend der allgemeinen Definition wird eine Menge G additive Gruppe $(G,+)$ genannt, wenn für sie

1. eine assoziative Operation $a+(b+c)=(a+b)+c$ mit $a,b,c \in G$ definiert ist;
2. ein neutrales (Null-) Element $e_{(+)}=0 \in G$ mit der Beziehung $a+0=0+a=a$ für alle $a \in G$ existiert;
3. und für sie außerdem ein inverses Element $-a \in G$ zu $a \in G$ mit der Relation $a+(-a)=0$ definiert ist.

Das bekannteste Beispiel einer solchen Gruppe ist die der ganzen Zahlen $(\mathbb{Z},+)$.

1.1.3 Multiplikative Gruppen

Für die multiplikative Gruppe $(G,\cdot)$ gilt äquivalent:

- eine assoziative Operation $a\cdot(b\cdot c)=(a\cdot b)\cdot c$ mit $a,b,c \in G$;
- ein neutrales (Eins-) Element $e_{(\cdot)}=1 \in G$ mit der Identität $a\cdot 1=1\cdot a=a$ für $a \in G$;
- und ein inverses Element $a^{-1} \in G$ zu jedem $a \in G$ mit der Relation $a\cdot a^{-1}=1$

sind definiert. Zur Multiplikation ist zu bemerken, daß nicht-negative Potenzen eines Elements induktiv definiert sind

$$a^0 = e_{(\cdot)} = 1, \qquad a^{n+1} = a\cdot a^n,$$

also durch n-malige Multiplikation.

$$a^n = \underbrace{a\cdot a\cdots a}_{n\text{-mal}}$$

Das Nullelement $e_{(+)}$ einer additiven Gruppe kann in einer multiplikativen Gruppe nicht enthalten sein, denn es ist grundsätzlich nicht invertierbar (siehe auch Abschnitt 1.3 zu den Körpern).

1.2 Ringe

Eine Menge R nennt man einen Ring $(R,+,\cdot)$, wenn auf ihr sowohl Addition als auch Multiplikation mit $R\times R \to R$ erklärt sind [PW72, 2.2]. Speziell müssen folgende Axiome gelten:

1. Bezüglich der Addition bildet R eine A ’sche Gruppe $(R,+)$.
2. $(R,\cdot)$ ist eine so genannte Halbgruppe,[4] d. h.

[4] Ein inverses Element wird dabei *nicht* gefordert, wenngleich einzelne Elemente von R ein Inverses besitzen können.

- $(R, \cdot)$ ist abgeschlossen $(\cdot : R \times R \to R)$
- und es gilt das Assoziativgesetz $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ für $a, b, c \in R$.

3. Außerdem muß für $a, b, c \in R$ das Distributivgesetz gelten $a \cdot (b + c) = ab + ac$.

Existiert zusätzlich noch die Identität $e_{(\cdot)} = 1 \in R$ mit $a \cdot 1 = 1 \cdot a = a$ für alle $a \in R$, dann wird $(R, +, \cdot)$ Monoid oder Ring mit *Eins* genannt. Die Menge der Elemente $r \in R$ in einem Monoid, welche mit $r \cdot r^{-1} = e_{(\cdot)}$ jeweils ein inverses Element besitzen, nennt man Einheitengruppe R^* des Ringes. Ein Ring wird außerdem als *kommutativ* bezeichnet, wenn auch für die Multiplikation das Kommutativgesetz $a \cdot b = b \cdot a$ gilt. Ein typisches Beispiel hierfür ist der Ring $(\mathbb{Z}, +, \cdot)$ der ganzen Zahlen, denn er erfüllt offensichtlich alle Axiome.

1.3 Körper

Ein Körper K wird durch ein System von Elementen (endlich oder unendlich) gebildet, die sich durch die definierten Operationen Addition, Subtraktion, Multiplikation und Division verknüpfen lassen[5] und für die das Distributivgesetz gilt [PW72, 2.3]. Präziser ausgedrückt:

1. $(K, +, \cdot)$ ist ein kommutativer Ring (zu den Eigenschaften vgl. Abschnitt 1.2);
2. $(K \setminus \{0\}, \cdot)$ ist eine kommutative Gruppe, d. h. alle Elemente ausgenommen 0 (auch geschrieben als $K^* = K \setminus \{0\}$) müssen ein inverses Element besitzen.[6]

Die bekanntesten Körper sind die der

- reellen Zahlen $(\mathbb{R}, +, \cdot)$,
- komplexen Zahlen $(\mathbb{C}, +, \cdot)$,
- rationalen Zahlen $(\mathbb{Q}, +, \cdot)$
- sowie der endliche Körper $\mathbb{Z}_2 := \{0, 1\}$.

Eine Menge von Funktionen über einem Körper K kann selbst auch wieder ein Körper sein, z. B. die Menge der rationalen Funktionen $\mathbb{R}(x)$ über $\mathbb{R}$, geschrieben als $(\mathbb{R}(x), +, \cdot)$.

1.4 Vektorraum

Eine Menge von Vektoren $V := \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots\}$, jeder bestehend aus einem geordneten n-Tupel von Elementen a aus einem Körper K, wird Vektorraum über dem Körper K genannt (vgl. [PW72, 2.5]), wenn:

1. die additive Gruppe $(V, +)$ existiert und vom A ’schen Typ ist $(+ : V \times V \to V)$;[7]

[5] Jede dieser Operationen muß wieder zu einem Element dieses Körpers führen (Abgeschlossenheit).

[6] Man sagt auch, daß K ein kommutativer Ring mit Eins sei, dessen Einheitengruppe aus allen Elementen außer der Null besteht.

[7] In Bezug auf den Begriff des Vektorraumes ist entscheidend (Abgrenzung zu einer Algebra, vgl. Abschnitt 1.5), daß nur die Addition zwischen Vektoren sowie die Multiplikation mit einem Skalar aus K definiert ist.

2. für jeden Vektor $\mathbf{v} \in V$ die Multiplikation mit einem Körperelement (Skalar) a definiert ist und wieder zu einem Vektor $\mathbf{u} = a\mathbf{v} \in V$ führt (Abgeschlossenheit der Abbildung $\cdot : K \times V \to V$);[8]
3. bezüglich zweier Skalare a, b und der Vektoren $\mathbf{u}, \mathbf{v}$ die folgenden Distributivgesetze gelten:

$$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$$
$$\mathbf{v}(a + b) = a\mathbf{v} + b\mathbf{v}$$

4. das Assoziativgesetz $(ab)\mathbf{v} = a(b\mathbf{v})$ gilt;
5. und es ein multiplikativ neutrales Element in Bezug auf die Vektoren in V gibt.

Wenn mit n die Anzahl der Tupel eines Vektors $\mathbf{v}$ bezeichnet wird, dann nennt man den zugehörigen Vektorraum (über dem Körper K) üblicherweise $V_n(K)$ oder kurz $V = K^n$, also z. B. $\mathbb{R}^3$ oder $\mathbb{Z}_2^n$. Entsprechend ist auch das neutrale Element, welches als $\mathbf{0} = (0, 0, \ldots, 0)$ geschrieben wird, ein Vektor mit n Elementen.

Kann man jeden Vektor von V als Linearkombination von unabhängigen Vektoren $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n} \in U \subseteq V$ darstellen, also als

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{u_i},$$

so nennt man $\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_n}$ die Basis, $\alpha_1, \alpha_2, \ldots, \alpha_n$ die Koeffizienten und $n = \dim(V) = |V|$ die Dimension von V.

1.5 Algebra (Verband)

Eine Algebra über einem Körper K ist eine Erweiterung des Begriffs Vektorraum in Bezug auf die Vektormultiplikation.[9] Speziell sind es folgende Axiome, die eine Algebra auszeichnen:

1. die Menge V ist ein Vektorraum über einem Körper K (vgl. Abschnitt 1.4);
2. für zwei Vektoren $\mathbf{u}, \mathbf{v} \in V$ ist ein Produkt $\mathbf{uv}$ definiert und in Bezug auf V abgeschlossen $(\cdot : V \times V \to V)$;
3. das Assoziativgesetz für Vektoren $(\mathbf{uv})\mathbf{w} = \mathbf{u}(\mathbf{vw})$ ist erfüllt;
4. auch das Distributivgesetz läßt sich auf Vektoren anwenden: $(\mathbf{u} + \mathbf{v})\mathbf{w} = \mathbf{uw} + \mathbf{vw}$.

1.6 Zusammenfassung

Tabelle 1.1 gibt einen Überblick zu den verschiedenen algebraischen Grundstrukturen und ihren Existenzbedingungen.

[8] Die mit $\cdot$ gekennzeichnete Operation nennt man auch die äußere,+ die innere Verknüpfung.

[9] Ohne für jeden Vektor ein multiplikativ inverses Element zu fordern. Man spricht deshalb auch manchmal von einem assoziativen Vektorring.

Tabelle 1.1: Eigenschaften algebraischer Strukturen

	Addition				Multiplikation				
	assoziativ	kommutativ	neutral	invers	assoziativ	kommutativ	neutral	invers	distributiv
Additive Gruppe	×		×	×					
(Multiplikative) Halbgruppe					×				
Monoid					×		×		
Multiplikative Gruppe					×		×	×	
Multiplikative A ’sche Gruppe					×	×	×	×	
Ring	×	×	×	×	×		×		×
Schiefkörper	×	×	×	×	×		×	×	×
Körper	×	×	×	×	×	×	×	×	×

2 Endliche Strukturen

2.1 Multiplikative A ’sche Gruppen

Entsprechend Abschnitt 1.1.2 handelt es sich bei endlichen kommutativ-multiplikativen Gruppen um algebraische Strukturen $G^* := (G, \cdot)$, welche bzgl. der Multiplikation

- assoziativ;
- mit einem neutralen Element $e_{(\cdot)} = 1$ ausgestattet;
- eindeutig invertierbar;
- und kommutativ

sind.[10]

2.1.1 Ordnung von Elementen

Endliche multiplikative Gruppen sind insbesondere wegen ihrer Eigenschaften beim Potenzieren von Gruppenelementen sehr interessant. Betrachten wir dazu die Folge der Potenzen $r^1, r^2, r^3, r^4, \ldots$ irgendeines Elements $r \in G^*$. Nach dem Prinzip der Abgeschlossenheit (Gruppenaxiom) wird auch jede Potenz von r wieder in G^* liegen. Wegen der endlichen Zahl $|G^*|$ von Elementen, muß sich ab irgendeiner Potenz $l \leq |G^*|$ die Folge wiederholen. Eine solche Wiederholung läßt den Ansatz $r^i = r^l$ zu. Multiplikation mit dem inversen Element von r^i ergibt $1 = r^{l-i}$, was wegen $l > i$ wiederum bedeutet, das es immer ein Element mit

[10] Mit diesen Eigenschaften können sie als Einheitengruppe eines kommutativen Ringes mit Eins (vgl. Abschnitt 1.2) oder als multiplikative Gruppe eines Körpers (vgl. Abschnitt 1.3) aufgefaßt werden.

$$r^k = e_{(\cdot)} = 1, \qquad r \in G^* \tag{2.1}$$

gibt. Die Folge $\{r, r^2, r^3, \ldots, r^{k-1}, r^k = 1 = r^0\}$ nennt man die vom Element r erzeugte zyklische Untergruppe und kennzeichnet sie mit

$$\langle r \rangle = \{ r^n \,\big|\, 1 \le n \le k \} \subseteq G^* \; . \tag{2.2}$$

Unter Zuhilfenahme von $r^k = r^0$ läßt sich die Menge der Elemente auch so definieren:

$$\langle r \rangle = \{ r^{n \bmod k} \,\big|\, n \in \mathbb{N} \} \; . \tag{2.3}$$

Abbildung 2.1 stellt die Periodizität der Folge am Beispiel $k = 12$ als Kreisteilung dar.

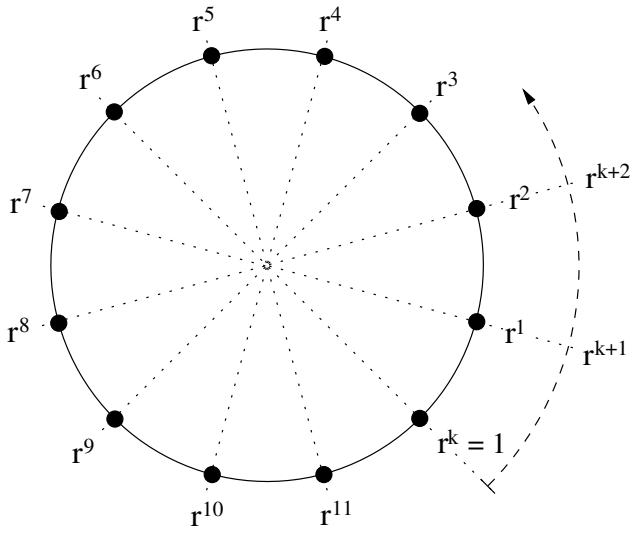


Abbildung 2.1: Zyklische Untergruppe $\langle r \rangle$

Die Ordnung (Anzahl der Elemente) der Untergruppe $\langle r \rangle$ ist $|\langle r \rangle| = k$. Sie ist gleichzeitig die kleinste Potenz k, die zu $r^k = 1$ führt. Man nennt sie auch Ordnung des Elements r und schreibt statt $\mathrm{ord}(r) = \#r = |\langle r \rangle|$ einfach nur $|r|$. Die Ordnung des neutralen Elements $e_{(\cdot)}$ ist 1, denn der kleinste Exponent k der zu $e^k_{(\cdot)} = e_{(\cdot)}$ führt, ist 1.[11]

Formel 2.1 gibt uns, wenn man sie mit r^{-1} multipliziert, eine Berechnungsvorschrift für das inverse Element an die Hand:

[11] Und dies ist das einzige Element mit der Eigenschaft $|r| = 1$.

$$r^{-1} = r^{k-1} \,. \tag{2.4}$$

2.1.2 Potenzen eines Elements

Betrachten wir jetzt die n-te Potenz irgendeines Elements r und stellen die Frage nach der Ordnung des so erzeugten Elements $s = r^n$.

Bezeichnet man dazu mit $k = |r|$ und $l = |s|$ die jeweilige Ordnung, so gilt nach Beziehung 2.1:

$$r^k = r^{|r|} = 1 \qquad\qquad s^l = s^{|s|} = 1$$

und für weitere Potenzen von r^k:

$$r^{ik} = (r^k)^i = 1^i = \underbrace{1 \cdot 1 \cdot 1 \cdots 1 \cdot 1}_{i\text{-mal}} = 1 \,. \tag{2.5}$$

Unter diesen Voraussetzungen kann man

$$s^l = (r^n)^l = r^{nl} = 1 = r^k = r^{ki}$$

formulieren und so $nl = ki$ schlußfolgern (Exponentenvergleich). Da sich unser Interesse auf den kleinsten Exponenten l beschränkt, haben wir es hierbei mit der Frage nach dem kleinsten gemeinsamen Vielfachen von n und k zu tun. Ein Beispiel für $k = 6$ und $n = 4$, also $\mathrm{lcm}(k,n) = 12 = 4 \cdot 3 = 6 \cdot 2$ zeigt Abbildung 2.2.

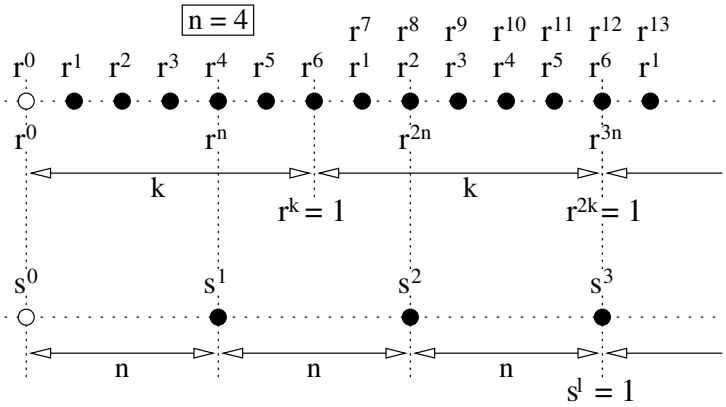


Abbildung 2.2: Potenzen eines Elements

Obwohl mit $nk = \text{lcm}(n,k)\gcd(n,k)$ auch eine Berechnungsvorschrift zur Verfügung steht, wollen wir aus Verständnis-gründen den ausführlichen Weg beschreiten. Dazu wird unter Zuhilfenahme der Abkürzung $d = \gcd(k,n)$ und mittels der Produktdarstellungen $k = \overline{k}d$ und $n = \overline{n}d$ zuerst der gemeinsame Teiler d eliminiert.

$$\overline{n}l = \overline{k}i \tag{2.6}$$

Wegen der Teilerfremdheit von $\overline{n}$ und $\overline{k}$ kann nur die Multiplikation mit der jeweils anderen Größe zum kleinsten gemeinsamen Vielfachen $\text{lcm}(\overline{n},\overline{k}) = \overline{n}l = \overline{k}i$ führen.

$$i = \overline{n} = \frac{n}{\gcd(k,n)} \qquad\qquad l = \overline{k} = \frac{k}{\gcd(k,n)} \tag{2.7}$$

Die Konsequenzen aus dem Ergebnis

$$l = \left|r^n\right| = \frac{|r|}{\gcd(|r|,n)} \tag{2.8}$$

sind recht interessant:

1. Die Ordnung eines durch Potenzieren erzeugten Elements $s = r^n$ ist immer kleiner/gleich der Ordnung des Ausgangselements r. Für den Fall $\gcd(|r|,n) = 1$ ist sie maximal (siehe auch Punkt 4).
2. Als Bestätigung für den Satz von L (vgl. auch Abschnitt 1.1.1) ist festzustellen, daß die Ordnung von $s = r^n$ genau die Ordnung von r teilt.[12] Im Sinne der Definition des Index einer Gruppe über deren Untergruppe, hier von $\langle r\rangle$ über $\langle s\rangle$, gilt deshalb:

   $$|\langle r\rangle : \langle s\rangle| = \gcd(|r|,n) \qquad (s = r^n)\,.$$

   Die Ordnung $|s|$ dabei (entsprechend der Bedeutung eines größten gemeinsamen Teilers) der bezüglich n teilerfremde Anteil in $|r|$.
3. Bei Kenntnis der Ordnung $k = |r|$ ist automatisch die Ordnung jedes Elements in der zyklischen Untergruppe $\langle s\rangle$ bekannt (vgl. Formel 2.8 mit Mengendefinition 2.2).
4. Zwei Elemente $r \neq s$ mit derselben Ordnung sind nach Formel 2.8 dadurch gekennzeichnet, daß $\gcd(k,n) = 1$ gilt. Die Anzahl der zur Ordnung k teilerfremden Zahlen (die kleiner als k sind) entspricht damit der Anzahl von möglichen Potenzen n, für die $k = |r^n| = |s|$ wird. Aus diesem Grund wird in einer multiplikativen Gruppe die Anzahl der Elemente mit jeweils gleicher Ordnung k genau durch E 's Totient-Funktion[13] $\phi(k)$ bestimmt.

[12] Das ist allerdings keine neue Erkenntnis, sondern eigentlich der Ausgangspunkt nach Formel 2.5.

[13] E 's Totient-Funktion $\phi(k)$ liefert eine Aussage über die Anzahl der zu k teilerfremden Zahlen, welche kleiner als k sind (siehe auch Abschnitt 3.3.2).

5. Bei Kombination der Formeln 2.8 und 2.4 stellt man fest, daß sich die Ordnung eines Elements r beim Übergang zu dessen Inversen r^{-1} nicht ändert: $|r^{-1}| = |r^{k-1}| = k/\gcd(k, k-1) = k$.

2.1.3 Beziehung zur Gruppenordnung

Variante 1 Die Ordnung der von r erzeugten zyklischen Untergruppe $\langle r \rangle$ ist nach dem Satz von L (siehe Abschnitt 1.1.1) ein Teiler der Gruppenordnung $|G^*|$. Man kann die Ordnung der multiplikativen Gruppe deshalb auch folgendermaßen ausdrücken [Bos96, 1.2, Satz 3]:

$$|G^*| = ik \,. \tag{2.9}$$

Daß die Ordnung von $\langle r \rangle$ ein Teiler von $|G^*|$ ist, führt in Verbindung mit Gleichung 2.1 zu:

$$r^{|G^*|} = r^{ik} = (r^k)^i = 1 \,. \tag{2.10}$$

Anschaulich (siehe auch Abbildung 2.2) bedeutet dies, daß sich die Potenzen r^n nach k Elementen periodisch wiederholen .

$$\underbrace{\underbrace{r^1, r^2, r^3, \ldots, r^k}_{k \text{ Elemente}}, \underbrace{r^1, r^2, r^3, \ldots, r^k}_{k \text{ Elemente}}, \ldots, \underbrace{r^1, r^2, r^3, \ldots, r^k}_{k \text{ Elemente}}}_{|G^*| \text{ Elemente } (i \text{ Perioden})}$$

Variante 2 Möchte man einen Rückgriff auf den Satz von L vermeiden, so kann für Beziehung 2.9 auch der folgende Beweis angeführt werden. Multipliziere jedes der $|G^*|$ Elemente aus $G^* = \{r_1, r_2, r_3, \ldots, r_{|G^*|}\}$ mit dem Element r, welches ebenfalls G^* entstammt (r ist eines der r_n, mit $n = 1, 2, \ldots, |G^*|$). Nach dem Prinzip der Abgeschlossenheit muß auch jedes Produkt rr_n wieder in G^* liegen. Außerdem müssen die Produkte paarweise verschieden sein, sonst würde die Multiplikation mit dem inversen Element r^{-1} zu zwei gleichen Elementen führen (Bed. $rr_\nu \neq rr_\mu$). Abgesehen von der Reihenfolge kann deshalb folgende eindeutige Mengenabbildung (Bijektion) angegeben werden: $\{rr_1, rr_2, rr_3, \ldots, rr_{|G^*|}\} \mapsto \{r_1, r_2, r_3, \ldots, r_{|G^*|}\}$. Bildet man jetzt das Produkt aller so erzeugten Elemente

$$\begin{aligned} rr_1 \cdot rr_2 \cdot rr_3 \cdots rr_{|G^*|} &= r_1 r_2 r_3 \cdots r_{|G^*|} \\ r^{|G^*|} r_1 r_2 r_3 \cdots r_{|G^*|} &= r_1 r_2 r_3 \cdots r_{|G^*|} \end{aligned}$$

und multipliziert noch mit den Inversen r_n^{-1}, dann bestätigt sich $r^{|G^*|} = 1$. Einzig mögliche Schlußfolgerung aus $r^{|G^*|} = 1$ und $r^k = 1$ kann aber (konform zu Beziehung 2.9) nur die sein, daß die Elementeordnung k genau die Gruppenordnung $|G^*|$ teilt.

2.1.4 Generatorelemente

Allgemein nennt man jede, von mindestens einem Element g durch Potenzierung erzeugte multiplikative Gruppe, eine zyklische Gruppe (siehe Abschnitt 2.1.1). Umfaßt die von g erzeugte Untergruppe $\langle g \rangle$ alle Elemente der multiplikativen Gruppe G^* (in irgendeiner Abfolge), dann bezeichnet man g als Generatorelement oder primitives Element.

$$\langle g \rangle := \left\{ g^1, g^2, g^3, \ldots, g^{|G^*|-1}, g^{|G^*|} = g^0 = 1 \right\} = G^* \tag{2.11}$$

Die Ordnung eines solchen Elements muß in diesem Sinne der Gruppenordnung entsprechen.[14]

$$|g| = |G^*|$$

Wie alle anderen Elemente muß auch ein Generatorelement Gleichung 2.1 erfüllen - aber eben nur für Generatorelemente ist $|G^*|$ der kleinste Exponent, welcher zu $g^{|G^*|} = 1$ führt. Bei Kenntnis eines Generatorelements aus G^* sind so nicht nur alle Elemente der multiplikativen Gruppe bekannt, sondern nach Formel 2.8 auch deren Ordnung.

Um die Frage zu beantworten, ob es immer mindestens ein Generatorelement gibt, rekapitulieren wir folgende Fakten:

1. Die Ordnung $k = |r|$ eines jeden Elements teilt die Gruppenordnung (Formel 2.9): $|G^*| = ik$.
2. Jedes Element r kann durch Potenzieren aus g erzeugt werden: $r = g^n$.
3. Die Ordnung $|g|$ eines Generatorelements erfüllt (wie die eines jeden anderen Elements) Formel 2.8: $|g| = |g^n| \gcd(|g|, n)$.
4. Die Ordnung eines Generatorelements muß der Gruppenordnung entsprechen: $|g| = |G^*|$.

Kombination der formelmäßigen Voraussetzungen ergibt:

$$\begin{aligned} |G^*| &= k \gcd(|G^*|, n) \\ i &= \gcd(ik, n) \end{aligned} \tag{2.12}$$

und diese Bedingung muß für irgendeine Potenz $n = 1, 2, \ldots, |G^*|$ zu gewährleisten sein (und zwar eindeutig für jedes Element r). Äquivalenz 2.12 kann aber nur erfüllt werden, wenn man $n = i\overline{n}$, mit $\gcd(k, \overline{n}) = 1$ annimmt. Die Anzahl der teilerfremden Zahlen $\overline{n}$ kleiner als k liefert mit $\phi(k)$ E ’s Totient-Funktion (vgl. Abschnitt 3.3.2). Sie ist immer größer als 0, weshalb stets *ein* primitives Element existiert. Aus diesem Grund ist jede multiplikative A ’sche Gruppe zyklisch, mit $\phi(|G^*|)$ Generatorelementen.[15]

[14] Für den Spezialfall, daß es sich bei der Gruppenordnung $|G^*|$ um eine Primzahl handelt, sind alle Elemente primitiv.
[15] Zur Anzahl von Elementen mit derselben Ordnung siehe auch Bemerkung 4 auf Seite 12.

2.1.5 Elemente als Nullstellen

Wenden wir uns jetzt einer etwas anderen Sichtweise auf die Elemente der multiplikativen Gruppe G^* zu, nämlich der Betrachtung über Nullstellen. Dazu gehen wir von dem Polynom $\varphi(x) = x^{|G^*|} - 1$ mit $x, \varphi(x) \in G^*$ aus, welches die Nullstellen bzw. Einheitswurzeln α haben soll.

$$\alpha^{|G^*|} - 1 = 0 \qquad \alpha^{|G^*|} = 1 \qquad (\alpha \in G^*) \tag{2.13}$$

Wir stellen nun fest, daß entsprechend des Fundamentalsatzes der Algebra $\varphi(x)$ genau $|G^*|$ Nullstellen haben muß. Nach Formel 2.9 wird diese Bedingung aber auch durch jedes Element aus G^* erfüllt. Da deren Anzahl genau mit der Anzahl der Nullstellen übereinstimmt, kann es sich bei den Elementen $r \in G^*$ nur um die $|G^*|$ Nullstellen von $\varphi(x)$ handeln [PW72, Satz 6.18]. Das mehrfache Nullstellen nicht vorhanden sind, kann man (in Verbindung mit Gleichung 2.13) durch Ableitung von $\varphi(x)$ an den Stellen α nachprüfen.

$$\left.\frac{\mathrm{d}\varphi(x)}{\mathrm{d}x}\right|_{x=\alpha} = |G^*|\, x^{|G^*|-1}\Big|_{x=\alpha} = |G^*|\, \alpha^{-1}\alpha^{|G^*|} = |G^*|\, \alpha^{-1} \neq 0$$

Mit diesen Erkenntnissen läßt sich für $\varphi(x)$ eine Linearfaktordarstellung auf Basis der Elemente r angeben.

$$\varphi(x) = x^{|G^*|} - 1 = \prod_{r \in G^*} (x - r) \tag{2.14}$$

Ausmultiplizieren der rechten Seite führt mit $\varphi(0) = -1$ noch zu der interessanten Äquivalenz:

$$1 + \prod_{r \in G^*} r = 0\,. \tag{2.15}$$

Da alle Elemente r der multiplikativen Gruppe G^* als Potenzen eines Generatorelements g darstellbar sind, kann man die Linearfaktordarstellung 2.14 auch folgendermaßen schreiben:

$$\varphi(x) = \prod_{n=1}^{|G^*|} (x - g^n) = (x-g)(x-g^2)(x-g^3)\cdots(x-g^{|G^*|-1})(x-1)\,.$$

2.2 Endliche Körper

2.2.1 Definition

Jeder endliche Körper ist durch eine beschränkte Anzahl von Elementen gekennzeichnet, auf welche die Körperaxiome von Abschnitt 1.3 zutreffen. Man nennt solche algebraischen Strukturen auch G -Körper und bezeichnet sie mit $\mathbb{F}_q$ oder GF(q), wobei q die Ordnung (Anzahl der Elemente) des Körpers angibt.[16] Bei der Konstruktion eines endlichen Körpers ist von allergrößter Bedeutung, daß zu den Eigenschaften eines Ringes noch die der multiplikativen Gruppe kommen. Zusätzliches Kriterium ist danach die Existenz des multiplikativ inversen Elements r^{-1} zu jedem $r \in \mathbb{F}_q^*$.

2.2.2 Ordnung

Definiert q die Anzahl der Elemente im Körper, d. h. inklusive Nullelement $e_{(+)}$, dann muß für die Ordnung der multiplikativen Gruppe $\mathbb{F}_q^* := \mathbb{F}_q \setminus \{0\}$ gelten:

$$|\mathbb{F}_q^*| = q - 1 \,. \tag{2.16}$$

Entsprechend Abschnitt 2.1.3 muß die Ordnung der multiplikativen Gruppe ein Vielfaches der Elementeordnung $k = |r|$ sein: $|\mathbb{F}_q^*| = ik$ (Satz von L). Ein G -Körper kann deshalb nicht jede beliebige Ordnung annehmen – wegen vorgenannter Bedingung müssen $q = |\mathbb{F}_q^*| + 1$ und die Ordnung k eines jeden Elements $r \in \mathbb{F}_q^*$ teilerfremd sein.[17]

$$\gcd(q, k) = 1 \tag{2.17}$$

Diese Erkenntnis läßt sich (einerseits logisch, aber auch rein analytisch) aus

$$1 = q - ik \,. \tag{2.18}$$

mit Hilfe des Satzes von B´ ableiten. Dazu vergleicht man Beziehung 16 mit Formel 4.3 aus Abschnitt 4.1.1 und stellt fest:

[16] Unerheblich ist dabei, ob die Menge der Körperelemente r durch eine Restklassenoperation oder irgendeine andere Methode definiert wird. Bei gleicher Anzahl von Elementen kann man solche Mengen nämlich immer (strukturerhaltend, eineindeutig) aufeinander abbilden [PW72, 6.5].

[17] Im Umkehrschluß können nur solche Elemente r der multiplikativen Gruppe $\mathbb{F}_q^*$ angehören, deren Ordnung k keinen Teiler mit $|\mathbb{F}_q|$ hat.

$$\gcd(q,k) = 1, \quad \text{mit} \quad \alpha = 1, \beta = -i\,.$$

Die einfachste Möglichkeit Teilerfremdheit zu gewährleisten, ist die Wahl von q als Primzahl oder als Potenz einer Primzahl. Im ersten Fall nennt man $\mathbb{F}_p$ einen Primzahlenkörper, für $q = p^n$ einen Erweiterungs- oder Binärkörper (weitere Ausführungen zu $\mathbb{F}_{p^n}$ in Abschnitt 3.4).

2.2.3 Elemente als Nullstellen

Aus Abschnitt 2.1.5 (Gleichung 2.14) ist bekannt, daß man die Elemente der multiplikativen Gruppe $\mathbb{F}_q^*$ als Nullstellen des Polynoms $\varphi(x) = x^{q-1} - 1$ auffassen kann. Nimmt man jetzt noch das Nullelement $e_{(+)} = 0$ hinzu, dehnt also die Betrachtung von der multiplikativen Gruppe auf alle Elemente des Körpers aus, dann kann man als zusätzlichen Linearfaktor $x - 0$ einbeziehen:

$$\psi(x) = x\varphi(x) = x^q - x = \prod_{r \in \mathbb{F}_q} (x - r)\,. \tag{2.19}$$

Deshalb bilden die Nullstellen von $x^q - x = x(x^{q-1} - 1)$ einen endlichen Körper $\mathbb{F}_q$.

3 Restklassen

3.1 Definition

Restklassen sind Kongruenzen von Elementen (einer algebraischen Struktur) modulo eines fixen Elements m. Das Rechnen mit solchen Elementen wird als modulare oder Modulo-Arithmetik und m als das Modul bezeichnet. In diesem Sinne definiert man

$$r \equiv a \bmod m \tag{3.1}$$

als den Rest r, der bei der „Division“ a/m entsteht und sagt: r ist kongruent a modulo m. Im Umkehrschluß sind die Restklassenelemente durch die Relation

$$a = qm + r, \qquad 0 \le r < m \tag{3.2}$$

bestimmt. Die Äquivalenzrelation $r \equiv a \bmod m$ kann man z. B. im Ring der ganzen Zahlen $\mathbb{Z}$ definieren, ist aber nicht auf diesen beschränkt. Im allgemeinen Fall $a \in R$ schreibt man für die Restklasse

$$[r]_m = \{a | a \in R, r \equiv a \bmod m\},$$

d. h. die Restklasse $[r]_m$ ist die Menge aller Elemente $a \in R$, die bei der Division modulo m genau den Rest r ergeben. Ein Beispiel für $a \in \mathbb{Z}$ mit einem Modul von $m = 4$ soll das veranschaulichen.

$$[0]_4 = \{\ldots, -12, -8, -4, 0, 4, 8, \ldots\} \qquad [2]_4 = \{\ldots, -10, -6, -2, 2, 6, 10, \ldots\}$$
$$[1]_4 = \{\ldots, -11, -7, -3, 1, 5, 9, \ldots\} \qquad [3]_4 = \{\ldots, -9, -5, -1, 3, 7, 11, \ldots\}$$

Üblicherweise bezeichnet man die Menge aller Restklassen mit

$$R_m = \{[r_0]_m, [r_1]_m, \ldots, [r_{n-1}]_m\},$$

also z. B. für die ganzen Zahlen $a \in \mathbb{Z}$ modulo m:

$$\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z} = \{[0]_m, [1]_m, \ldots, [m-1]_m\}\ .$$

Im Allgemeinen werden Verknüpfungsoperationen (Addition, Multiplikation usw.) zwischen zwei Restklassen dadurch definiert, daß man jeweils einen Vertreter aus den Restklassen auswählt und dann die Operation mit diesem Repräsentanten durchführt.

3.2 Restklassenringe

Um von einem Restklassenring $(R_m, +, \cdot)$ sprechen zu können ist der Nachweis aller Ringaxiome von Abschnitt 1.2 in Bezug auf die Menge der Restklassen R_m notwendig [Wel01, 5.]. Dazu geht man von der Modulo-Arithmetik nach Gleichung 3.1 und 3.2 aus und prüft jeden Punkt durch Einzelbetrachtung:

1. Bezüglich der Addition muß $(R_m, +, \cdot)$ eine A 'sche Gruppe bilden, d. h.

   a) die Addition existiert und ist abgeschlossen;[18]

   $$\begin{aligned}[r]_m = [r_1]_m + [r_2]_m &= (a_1 - mq_1 + a_2 - mq_2) \bmod m \\ &= (a_1 + a_2) \bmod m - [m(q_1 + q_2)] \bmod m = (r_1 + r_2) \bmod m \\ &= [r_1 + r_2]_m\end{aligned}$$

   b) sie ist außerdem sowohl assoziativ als auch kommutativ (A 'sch);

   $$[r_1]_m + ([r_2]_m + [r_3]_m) = ([r_1]_m + [r_2]_m) + [r_3]_m = [r_3]_m + [r_2]_m + [r_1]_m$$

[18] Die Operation führt immer wieder auf ein Element in R_m.

c) das neutrale Element $e_{(+)} = [0]_m = mq_0$ existiert;

$$[r]_m + [0]_m = (a - mq + mq_0) \bmod m = [a - m(q - q_0)] \bmod m = a \bmod m = [r]_m$$

d) jedes Element besitzt ein inverses Element $-[r]_m = (m - a) \bmod m$ mit

$$[r]_m + (-[r]_m) = a \bmod m + (m - a) \bmod m = m \bmod m = [0]_m.$$

2. Für die Multiplikation muß $(R_m, +, \cdot)$ eine Halbgruppe darstellen, also

a) ebenfalls abgeschlossen sein;

$$\begin{aligned}[r]_m = [r_1]_m \cdot [r_2]_m &= (r_1 - mq_1) \cdot (r_2 - mq_2) \bmod m \\ &= [r_1 r_2 + m(mq_1 q_2 - r_1 q_2 - r_2 q_1)] \bmod m = (r_1 \cdot r_2) \bmod m \\ &= [r_1 \cdot r_2]_m\end{aligned}$$

b) und das Assoziativgesetz für $[r_i]_m \in R_m$ erfüllen;

$$[r_1]_m \cdot ([r_2]_m \cdot [r_3]_m) = ([r_1]_m \cdot [r_2]_m) \cdot [r_3]_m$$

3. Außerdem muß für alle $[r]_m \in R_m$ das Distributivgesetz erfüllt sein.

$$\begin{aligned}[r_1]_m \cdot ([r_2]_m + [r_3]_m) &= (r_1 - mq_1) \cdot [r_2 + r_3 - m(q_2 + q_3)] \bmod m \\ &= (r_1 r_2 + r_1 r_3) \bmod m \\ &= [r_1]_m \cdot [r_2]_m + [r_1]_m \cdot [r_3]_m\end{aligned}$$

Da $(R_m, \cdot)$ ein neutrales Element $e_{(\cdot)}$ mit $[r]_m \cdot e_{(\cdot)} = [r]_m$ besitzt, handelt es sich sogar um einen *Ring mit Eins*.

$$[r]_m \cdot [1]_m = (r \cdot 1) \bmod m = r \bmod m = [r]_m$$

Aus Anwendungssicht ist sofort zu erkennen, daß insbesondere die Restklassen $R_m = \mathbb{Z}_m$ alle diese Bedingungen für $m \geq 2$ erfüllen und somit einen kommutativen Ring mit Einselement bilden.

3.3 Restklassenkörper

3.3.1 Existenz

Für den Übergang von einem Restklassenring $(R_m, +, \cdot)$ zu einem Restklassenkörper muß man zu jedem $r \in R_m \setminus \{0\}$ das Vorhandensein eines multiplikativ inversen Elements $[r]_m{}^{-1}$, mit $[r]_m \cdot [r]_m{}^{-1} = [1]_m$, fordern. Wir nehmen die Antwort vorweg und proklamieren, daß es ein solches Element in $(R_m, +, \cdot)$ nur dann gegeben kann, wenn m und r teilerfremd sind [PW72, Satz 6.4]. Bezieht man diese Aussage z. B. auf $\mathbb{Z}_m^* = \mathbb{Z}_m \setminus \{0\} = \{1, 2, 3, \ldots, m-1\}$, dann müssen (wenn keine weiteren Forderungen an m gestellt werden) alle Elemente r, für die $\gcd(r, m) = 1$

nicht erfüllt werden kann, ausgeschlossen werden. Sowohl im allgemeinen als auch speziellen Fall von $\mathbb{Z}_m^*$ ist diese Einschränkung grundsätzlich hinfällig, wenn es sich bei m um ein Primelement handelt (ein vollständiges Restklassensystem) – in Bezug auf $\mathbb{Z}_m^*$ also um eine Primzahl $p \in \mathbb{P}$, weshalb $\mathbb{F}_q := \mathbb{Z}_p$ (mit $q = p$). Im Fall des Ausschlusses von Elementen (ein reduziertes Restklassensystem) ist die Anzahl der teilerfremden Zahlen durch E 's Totient-Funktion $\phi(m)$ bestimmt und so die Ordnung der multiplikativen Gruppe $|\mathbb{Z}_m^*| = \phi(m)$, d. h. $\mathbb{F}_q \subseteq \mathbb{Z}_m$ (mit $q = \phi(m) + 1$).

Beweis Ist m kein primes Element, dann läßt es sich mindestens in zwei Faktoren $r, s \in R_m$ zerlegen (die nicht Vielfache von m sind, also $r, s \bmod m \neq 0$). Da nun $m \bmod m = 0$ ist, gilt für die Restklassenmultiplikation $[r]_m \cdot [s]_m = [m]_m = 0$. Soll aber $[r]_m$ eine inverse Restklasse $[r]_m^{-1}$ besitzen, dann kann man beide Seiten des Produktes mit $[r]_m{}^{-1}$ multiplizieren.

$$\underbrace{[r]_m{}^{-1}[r]_m}_{e_{(\cdot)}=1} \cdot [s]_m = [s]_m = [r]_m{}^{-1}[m]_m = 0$$

$[s]_m$ ist aber nach Voraussetzung nicht 0, demzufolge kann ein Inverses zu $[r]_m$ für den Fall dieser Zerlegung nicht existieren.

Im Gegenzug bleibt noch nachzuweisen, daß, wenn m ein Primelement ist, für jede Restklasse $[r]_m$ aus R_m ein inverses Element $[r]_m{}^{-1}$ auch wirklich existiert. Zu diesem Zweck betrachten wir alle $[r]_m \in R_m$, ausgenommen die Restklasse $[1]_m$, welche bei der Inversion auf sich selbst abgebildet wird. Da wegen der Modulo-Reduktion (wir verwenden jetzt wieder den Vertreter der Restklasse) immer $m > r$ gilt, kann nur r ein Teiler von m sein und nicht umgekehrt. Aber auch dies ist nicht möglich, wenn nach Voraussetzung m relativ prim zu r ist. Deshalb kann der größte gemeinsame Teiler von r und m nur das Einselement sein. Berücksichtigt man jetzt noch die aus dem euklidischen Algorithmus stammende Erkenntnis (Satz von B´ , vgl. Abschnitt 4.1), daß der größte gemeinsame Teiler $d = \gcd(r, s)$ zweier Elemente r, s in der Form $d = \alpha r + \beta s$ mit $\alpha, \beta \in \mathbb{Z}$ darstellbar ist, dann gilt:

$$\begin{aligned} \gcd(r, m) = 1 = \alpha r + \beta m \equiv \alpha r \bmod m \\ [1]_m = [\alpha]_m \cdot [r]_m \,. \end{aligned} \tag{3.3}$$

Das inverse Element von $[r]_m$ ist demzufolge

$$[r]_m{}^{-1} = [\alpha]_m,$$

wobei dessen Berechnung z. B. mit Hilfe des erweiterten euklidischen Algorithmus (siehe Abschnitt 4.1.4) möglich ist.[19]

[19] Diese Rechnung kann man auch in einem Ring ausführen, dann muß das Ergebnis jedoch nicht eindeutig sein (vgl. auch Bemerkungen auf Seite 43).

3.3.2 Multiplikative Gruppe

Da es sich bei den Restklassenkörpern um spezifische G -Körper handelt, kann man einige Formeln von Abschnitt 2.2 konkretisieren. Im folgenden sollen deshalb die Körperelemente und deren Ordnung im Zusammenhang mit der multiplikativen Gruppe $\mathbb{F}_q^*$ betrachtet werden.

Ordnung von Elementen Abgesehen von den allgemein geltenden Ordnungsrelationen (siehe Abschnitt 2.1) gibt es speziell für den Restklassenkörper $\mathbb{Z}_p$ noch eine erwähnenswerte Ausdrucksmöglichkeit für die von einem Körperelement generierte zyklische Untergruppe:

$$\langle r \rangle = \{ r^n \bmod (r^k - 1) \,\big|\, n \in \mathbb{N},\, r \in \mathbb{Z} \} \,. \tag{3.4}$$

Einsichtig wird die Schreibweise sofort, wenn man sie für jeden Exponent n expandiert.[20]

$$\begin{aligned}
r^0 \bmod (r^k - 1) &= r^0 \\
r^1 \bmod (r^k - 1) &= r^1 \\
r^2 \bmod (r^k - 1) &= r^2 \\
&\vdots \\
r^{k-1} \bmod (r^k - 1) &= r^{k-1} \\
r^k \bmod (r^k - 1) &= 1 = r^0 \\
&\vdots
\end{aligned}$$

Kleiner Satz von Fermat Eine für die praktische Anwendung von Restklassenkörpern sehr wichtige Folgerung aus Gleichung 2.10 ist der (für den Restklassenkörper $\mathbb{Z}_p$ geltende) kleine Satz von F [PD75, II]. Er resultiert sofort aus $r^{q-1} \equiv 1 \pmod{m}$, wenn man berücksichtigt, daß die Ordnung der multiplikativen Gruppe $\mathbb{Z}_p^* = \mathbb{Z}_p \setminus \{0\} = \{1, 2, 3, \ldots, p-1\}$ genau $q - 1 = p - 1$ ist.

$$r^{p-1} \equiv 1 \pmod{p} \tag{3.5}$$

Aus dieser Kongruenz kann man wegen $0 \equiv p \pmod{p}$ außerdem ableiten, daß p stets ein Teiler von $r^{p-1} - 1$ ist.[21]

$$r^{p-1} - 1 = np \equiv 0 \pmod{p} \tag{3.6}$$

[20] Letztlich eine Spezialisierung der Mengendarstellung 2.3 von Abschnitt 2.1.1 für den Fall $\mathbb{F}_q := \mathbb{Z}_p$.
[21] Oft auch in den Varianten $r^p \equiv r \pmod{p}$ oder $r^p - r \equiv 0 \pmod{p}$ verwendet.

Obwohl der kleine Satz von F ursprünglich auf $\mathbb{Z}_p$ bezogen war, wird sein Name oft auch in Verbindung mit Ausgangsformel 2.10 verwendet (siehe zum Beispiel [FHLM04]) — also ganz allgemein bezogen auf den endlichen Körper $\mathbb{F}_q$. Deshalb sollen an dieser Stelle noch weitere Berechnungsmöglichkeiten erwähnt sein, welche sich direkt aus $r^{q-1} = 1$ ergeben.

1. Multiplikation mit r^{-1} führt z. B. zur Möglichkeit ein Element zu invertieren.

$$r^{-1} = r^{q-2}$$

2. Multiplikation mit r^2 gibt uns eine weitere Ausdrucksmöglichkeit für das Quadrat eines Elements.

$$r^2 = r^{q+1}$$

3. Für einige Spezialfälle ermöglicht der kleine Satz von F sogar das Ziehen der Quadratwurzel aus einem Element.

   a) Beispielsweise kann man im Erweiterungskörper $\mathbb{F}_{2^n}$ (vgl. Abschnitt 3.4), in welchem $q = 2^n$ ja immer gerade ist, direkt folgende Formel angeben:

   $$\sqrt{r} = \sqrt{r^q} = r^{\frac{q}{2}} = r^{2^{n-1}}$$

   b) Aber auch im Körper $\mathbb{F}_p$ kann man, zumindest für den Fall $p \equiv 3 \bmod 4$, eine einfache Lösung präsentieren:[22]

   $$\sqrt{r} = r^{\frac{p+1}{4}} \pmod{p}\,.$$

   Mit einer kurzen Probe läßt sie sich schnell verifizieren:

   $$\left(r^{\frac{p+1}{4}}\right)^2 = r^{\frac{p+1}{2}} = \sqrt{r^{p+1}} = \sqrt{r^2 r^{p-1}} = \sqrt{r^2} = r \pmod{p}\,.$$

Satz von Wilson Betrachtet man die multiplikative Gruppe eines Restklassenkörpers $\mathbb{Z}_{p>2}$, so handelt es sich bei den Zahlen $r = 1, 2, \ldots, p-1$ um die $p-1$ Nullstellen α des Polynoms $\varphi(x) = x^{p-1} - 1 \pmod{p}$. Anwendung von Gleichung 2.15 auf $\mathbb{Z}_p^*$ führt folgerichtig zum Satz von W [Sho05, 2.8.1], [PD75, II], [Bun08, 2].[23]

$$(p-1)! \equiv -1 \equiv p-1 \pmod{p} \tag{3.7}$$

[22] Für eine generelle Lösungsmöglichkeit sei auf den T -S Algorithmus (siehe Abschnitt 4.4, Seite 51) verwiesen [Ton91, Sha73].

[23] Er ist sowohl notwendige als auch hinreichende Bedingung dafür, daß es sich bei p wirklich um eine Primzahl handelt.

Satz von E L. E hat für natürliche Zahlen eine sogenannte Totient-Funktion $\phi(m)$ definiert, welche die Anzahl der positiven Zahlen (größer als 0 und kleiner als m) teilerfremd zu m ausdrückt.[24] Mit Hilfe dieser Funktion hat er F 's kleinen Satz folgendermaßen verallgemeinert [Sho05, 2.6], [Bun08, 2], [PD75, II]:

> Sind zwei Zahlen $r, m \in \mathbb{N}$ relativ prim zueinander, d. h. sie haben keinen gemeinsamen Teiler (und so ist $\gcd(r,m) = 1$), dann gilt:

$$r^{\phi(m)} \equiv 1 \pmod{m} . \tag{3.8}$$

Der Beweis ist mit den Betrachtungen von Abschnitt 2.1 zur Ordnung der multiplikativen Gruppe in $\mathbb{Z}_m$ zu erbringen. Danach entspricht die Gruppenordnung $|\mathbb{Z}_m^*|$ der Anzahl invertierbarer Elemente (solche mit $\gcd(r,m) = 1$), d. h. mit E 's Totient-Funktion genau $|\mathbb{Z}_m^*| = \phi(m)$.[25] Berücksichtigt man jetzt noch die Gruppen nach Formel 2.10, dann bestätigt sich E 's Satz in Form von Gleichung 2.9.

$$r^{|\mathbb{Z}_m^*|} \equiv 1 \pmod{m}$$

Speziell im Restklassenkörper $\mathbb{Z}_p$ entspringt aus Kongruenz 3.8 mit $\phi(p) = |\mathbb{Z}_p^*| = p - 1$ sofort der kleine Satz von F .

Für einen speziellen Fall, nämlich das Produkt zweier Primzahlen $m = pq$, ist es sehr wünschenswert die Totient-Funktion zu kennen.[26] Sind p und q nach Voraussetzung Primzahlen, dann können nur die Zahlen (kleiner als m) gemeinsame Teiler mit m haben, die Vielfache von p *oder* q sind. Vielfache von p die kleiner als m sind, gibt es aber genau $q-1$, was für q äquivalent gilt (nämlich $p, 2p, 3p, \ldots, (q-1)p$ und $q, 2q, 3q, \ldots, (p-1)q$). Somit muß man von den $m-1$ Zahlen kleiner als m genau $p-1+q-1 = p+q-2$ subtrahieren, was zu

$$\begin{aligned} \phi(m) &= m - 1 - (p+q-2) \\ &= pq - (p+q) + 1 \\ &= (p-1)(q-1) \end{aligned}$$

führt.[27] In ähnlicher Art und Weise kann man auch die folgende Formel ableiten:

$$\phi(p^n) = \phi(p \cdot p^{n-1}) = (p-1)p^{n-1} .$$

[24] Die Zahl Eins wird immer als teilerfremd angesehen, d. h. $\phi(m) \geq 1$.

[25] Manchmal wird E 's Totient-Funktion auch als der multiplikativen Gruppe $\mathbb{Z}_m^*$ definiert.

[26] Dieser Fall hat besondere Bedeutung im Zusammenhang mit dem RSA-Algorithmus (siehe z. B. [Sch96, MvV92, Wel01] oder [RSA78]).

[27] Für den allgemeineren Fall zweier teilerfremder Zahlen gilt: $\phi(pq) = \phi(p)\phi(q)$. Man erkennt außerdem, daß in beiden Fällen $\phi(pq)$ dem kleinsten gemeinsamen Vielfachen entspricht: $\phi(pq) = \operatorname{lcm}(p,q)$.

3.3.3 Beispiele

Körper $\mathbb{Z}_2$ Der Körper $\mathbb{F}_2 := \mathbb{Z}_2$ ist von besonderer praktischer Bedeutung, denn er bildet häufig die Grundlage technischer Realisierungen. Die beiden Elemente von $\mathbb{Z}_2$ werden mit $\{[0]_2, [1]_2\}$ oder kürzer mit $0, 1$ bezeichnet. Wegen ihrer einfachen Implementierung sind die Operationen in $\mathbb{Z}_2$ besonders effizient.

1. Die Addition (modulo 2) ist mit

$$\begin{aligned} 0+0 &= 0 & 1+1 &= 2 \bmod 2 = 0 \\ 0+1 &= 1 & 1+0 &= 1 \end{aligned} \tag{3.9}$$

   geradezu primitiv und entspricht dem logischen *Exklusiv-Oder*.[28] Wie man sofort sieht, ist das neutrale Element die 0 und wegen Beziehung 3.9 das additiv Inverse die 1. Addition und Subtraktion sind in diesem Sinne gleichwertig, denn es gilt $-1 \equiv 1 \pmod 2$.

2. Ebenfalls eine einfache Operation ist die Multiplikation, denn sie entspricht dem logischen *Und*.

$$\begin{aligned} 0 \cdot 0 &= 0 & 1 \cdot 1 &= 1 \\ 0 \cdot 1 &= 0 & 1 \cdot 0 &= 0 \end{aligned}$$

3. Die Division erklärt sich mit Hilfe des inversen Elements in $\mathbb{Z}_2^* = \mathbb{Z}_2 \setminus \{0\} = \{1\}$, welches ja die Bedingung $1 \cdot 1^{-1} = e_{(\cdot)} = 1$ erfüllen muß. Einzig mögliche Schlußfolgerung ist die, daß es sich bei dem inversen Element 1^{-1} um 1 selbst handelt.

Körper $\mathbb{Z}_5$ Für den Körper $\mathbb{Z}_5$, also dem Fall einer multiplikativen Gruppe der $|\mathbb{Z}_5^*| = q - 1 = 4$, gilt für die der Elemente:

$$\begin{array}{llllll} r_1 = 1 & \langle r_1 \rangle = \{1\} & k = 1 & \phi(k) = 1 & r_1^k = 1 & r_1^{p-1} = 1 \pmod 5 \\ r_2 = 2 & \langle r_2 \rangle = \{1,2,4,3\} & k = 4 & \phi(k) = 2 & r_2^k = 16 \equiv 1 & r_2^{p-1} = 16 \equiv 1 \pmod 5 \\ r_3 = 3 & \langle r_3 \rangle = \{1,3,4,2\} & k = 4 & \phi(k) = 2 & r_3^k = 81 \equiv 1 & r_3^{p-1} = 81 \equiv 1 \pmod 5 \\ r_4 = 4 & \langle r_4 \rangle = \{1,4\} & k = 2 & \phi(k) = 1 & r_4^k = 16 \equiv 1 & r_4^{p-1} = 256 \equiv 1 \pmod 5 \,. \end{array}$$

3.4 Erweiterungskörper

3.4.1 Vorbetrachtungen

Ein Erweiterungskörper M/K ist ein Körper $(M, +, \cdot)$, der einen anderen Körper $(K, +, \cdot)$ als Teilkörper enthält [PW72, 6.5]. Der Grad der Körpererweiterung von M über K ist die Dimension

[28] Oftmals auch mit *XOR* oder dem Symbol $\oplus$ bezeichnet.

von M als (so genannter) K-Vektorraum und wird als $[M:K]$ bzw. $\dim_K M$ geschrieben. Jeder Vektor in M besteht entsprechend der Definition des Vektorraumes (vgl. Abschnitt 1.4) aus jeweils $[M:K]$ Tupeln in K. Bekannte Beispiele für Körpererweiterungen sind:

$[\mathbb{C}:\mathbb{R}] = 2$, die Erweiterung der Dimension um eine imaginäre Komponente;
$[\mathbb{R}:\mathbb{Q}] = \infty$, hier sind die rationalen Zahlen noch abzählbar.

3.4.2 Polynomringe

Ausgehend von den Vorbetrachtungen konstruieren wir jetzt einen endlichen Polynomring $(K[x],+,\cdot)$ auf dem Körper K.

$$K[x] := \left\{a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_2x^2 + a_1x + a_0\right\} = \{f(x)|n \in \mathbb{N},\ a_i \in K\} \tag{3.10}$$

$$f(x) = \sum_{\nu=0}^{n-1} a_\nu x^\nu \tag{3.11}$$

Darin seien die üblichen Polynomoperationen, wie Addition und Multiplikation gültig, weshalb man auch von einem Vektorraum der Polynome in der Unbestimmten x mit Koeffizienten aus dem Körper K spricht. Ist der Leitkoeffizient $a_{n-1} = 1$, dann wird das Polynom als normiert (monisch) bezeichnet, sonst ist $a_{n-1}x^{n-1}$ das so genannte Leitmonom.

Wird anschließend eine Restklassendivision dieser Polynome $f(x)$ durch ein Polynom $m(x)$ mit Grad n definiert, also $r(x) = f(x) \bmod m(x)$ mit $r(x) \in K[x]/m(x)$, dann bildet die Menge der darin enthaltenen Restklassen wieder einen Restklassenring [PW72, 6.], [MvV92, 2.5.4].

Der Polynom-Restklassenring auf dem Grundkörper $\mathbb{Z}_p$ Im Beispiel des Restklassenringes $\mathbb{Z}_p[x]/m(x)$ lassen sich die Eigenschaften eines Ringes (siehe Abschnitt 1.2) wiefolgt nachweisen:

1. Da bei einer Polynomaddition die einzelnen Koeffizienten unabhängig voneinander (und jeder für sich) addiert werden und außerdem nach Voraussetzung immer $\deg r(x) < \deg m(x) = n$ gilt, bildet $K[x]/m(x)$ eine additive A 'sche Gruppe.
   a) Das Assoziativgesetz gilt: $r(x)+[g(x)+h(x)] = [r(x)+g(x)]+h(x)$ mit $r(x), g(x), h(x) \in K[x]/m(x)$.
   b) Das neutrale Element $e_{(+)} = 0 \in K[x]/m(x)$ mit der Beziehung $r(x) + e_{(+)} = r(x)$ ist das Nullpolynom.
   c) Ein additiv inverses Element $-r(x) \in K[x]/m(x)$ mit $r(x) + [-r(x)] = 0$ ist vorhanden. Es ergibt sich aus den inversen Elementen der Koeffizienten $a_\nu \in K$ zu $-r(x) = \sum_{\nu=0}^{n-1} -a_\nu x^\nu = -\sum_{\nu=0}^{n-1} a_\nu x^\nu$.
2. $(K[x]/m(x),\cdot)$ ist eine multiplikative Halbgruppe, denn:

a) Aufgrund der Modulo-Reduktion ist $(K[x]/m(x), \cdot)$ abgeschlossen, d. h. wenn $r(x), g(x) \in K[x]/m(x)$ angenommen wird, dann gilt für die Multiplikation $r(x)\,g(x) \equiv h(x) \pmod{m(x)}$ gleichfalls $h(x) \in K[x]/m(x)$.

b) Auch das Assoziativgesetz $r(x)\,[g(x)\,h(x)] = [r(x)\,g(x)]\,h(x)$ ist in einem Restklassenring von Polynomen erfüllt.

3. Aus den bisherigen Erkenntnissen zu Restklassenringen ist im Zusammenhang mit Polynomoperationen zu schlußfolgern, daß das Distributivgesetz ebenfalls gilt: $r(x)\,[g(x) + h(x)] = r(x)g(x) + r(x)h(x)$.

3.4.3 Endlicher Erweiterungskörper

Der Übergang zu einem Körper wird möglich, wenn $m(x)$ ein Primelement in Bezug auf die Menge der Polynome $K[x]$ ist, es sich also um ein (so genanntes) irreduzibles Polynom handelt. Ein solches Polynom ist dadurch gekennzeichnet, daß es nicht weiter in Teilpolynome mit Koeffizienten aus K reduzierbar ist.[29]

Es sei nun $r(x)$ ein Restklassenpolynom mit Koeffizienten a aus dem endlichen Grundkörper $K := \mathbb{F}_p$ und $m(x)$ vom Grad n. Dann handelt es sich bei $\mathbb{F}_p[x]/m(x)$ um einen endlichen Körper mit $q = p^n$ Elementen [Bos96, 3.8]. Man spricht auch von einem Vektorraum V der Dimension n über $\mathbb{F}_p$, denn auf diese Weise wird (im Sinne von Abschnitt 1.4) jedem Vektor $\mathbf{v} = (v_1, v_2, v_3, \ldots, v_n)$ ein Polynom $r(x)$ vom Grad $n-1$ zugeordnet. Es handelt sich folglich nur um eine andere Darstellung der n Tupel des Vektors $\mathbf{v}$ in der Art $a_0 = v_1, a_1 = v_2, \ldots, a_{n-1} = v_n$. Die Potenzen $x^0, x^1, x^2, \ldots, x^{n-2}, x^{n-1}$ bilden die (Polynom-) Basis des Vektorraumes V über $\mathbb{F}_p$. Entsprechend ist die Dimension des Erweiterungskörpers $\mathbb{F}_q$ über dem Grundkörper $\mathbb{F}_p$ genau $[\mathbb{F}_q : \mathbb{Z}_p] = n$. Als Notation für einen solchen Körper wird deshalb auch $\mathbb{F}_{p^n}$ oder $\mathrm{GF}(p^n)$ verwendet.

Mit diesen Vorbemerkungen lassen sich alle Aussagen zu Restklassenkörpern, wie sie in Abschnitt 3.3 allgemein formuliert wurden, auf den Erweiterungskörper $\mathbb{F}_{q=p^n}$ anwenden:

1. Das Modul m (Primelement) wird nun als das irreduzible Polynom $m(x)$ interpretiert.
2. Die Restklassendivision ist definiert als $r(x) \equiv f(x) \bmod m(x)$, was grundsätzlich immer zu einem Grad kleiner als n für $r(x)$ führt.[30] Die Kennzeichnung der zu $r(x)$ gehörenden Restklasse erfolgt wie gewohnt mit $[r(x)]_{m(x)}$, wird meistens jedoch weggelassen. Das Element $r(x)$ steht also auch hier wieder als Restklassenvertreter aller Polynome $f(x)$, welche die Bedingung $f(x) = s(x)m(x) + r(x)$ mit $\deg r(x) < \deg m(x)$ erfüllen.
3. Das Nullelement ist das Nullpolynom $r(x) = 0 \pmod{m(x)}$ bzw. dessen Restklasse $[0]_{m(x)}$, das Einselement das Einheitspolynom $e_{(\cdot)} = x^0 = 1$.
4. Addition und Multiplikation (von Polynomen) in $\mathbb{F}_q$ sind wohldefiniert und abgeschlossen, die entsprechenden Gruppen $\mathbb{F}_q^*$ und $\mathbb{F}_q^+$ also existent.

[29] Ausführliche Betrachtungen zu irreduziblen und primitiven Polynomen sind z. B. in [MvV92, 4.5], Tabellen irreduzibler Polynome in [PW72, Anhang C] zu finden.

[30] Die Menge der Elemente r umfaßt alle Polynome mit einem Grad kleiner als n.

5. Die Anzahl der Elemente $r(x)$ im Restklassenkörper $\mathbb{F}_p[x]/m(x)$ ist aufgrund der Anzahl von möglichen Koeffizientenkombinationen p^n. Bei $\mathbb{F}_q$ handelt es sich folglich um einen G -Körper GF(p^n).[31] Wegen $|\mathbb{F}_q| = q = p^n$ hat dessen multiplikative Gruppe $\mathbb{F}_q^*$ die Ordnung $p^n - 1$.
6. Jedes Element $r(x) \in \mathbb{F}_q^*$ hat eine Ordnung bzw. Periode $k = |\langle r(x) \rangle|$, welche sich aus Gleichung 2.1 ableitet:

$$[r(x)]^k - 1 \equiv 0 \pmod{m(x)}\,.$$

7. Nach dem Satz von L teilt die Ordnung k des Elements die der multiplikativen Gruppe $\mathbb{F}_q^*$, d. h. $q - 1 = p^n - 1 = ik$ und demzufolge ist

$$[r(x)]^{q-1} - 1 \equiv 0 \pmod{m(x)}\,. \tag{3.12}$$

8. Jedes erzeugende Element $g(x)$ von $\mathbb{F}_q^*$ erfüllt die Bedingung der maximalen Zykluslänge (Index $i = 1$), hat also die Ordnung $|\langle g(x) \rangle| = p^n - 1$. Es ist damit geeignet, als Basiselement für die Erzeugung aller anderen Elemente $r(x) \in \mathbb{F}_q^*$ verwendet zu werden. Nimmt man das Nullelement $0 = g^q$ sowie das Einselement $1 = g^{q-1}$ hinzu, so gilt für die Menge der Elemente $\mathbb{F}_q = \{0, 1, g^1, g^2, g^3, \ldots, g^{p^n-3}, g^{p^n-2}\}$.
9. Aus Punkt 6 läßt sich (in Übereinstimmung mit Gleichung 3.6) schlußfolgern, daß für jedes Element $r(x) \in \mathbb{F}_q^*$ das Modul $m(x)$ ein Teiler von $[r(x)]^{q-1} - 1$ ist, also die Zerlegung $[r(x)]^{q-1} - 1 = h(x)\,m(x)$ hat. Setzt man insbesondere $r(x) = x$ als das kleinste Element (mit einem Grad größer als 0) aus $\mathbb{F}_q^*$, so erhält man die Beziehungen:

$$\begin{aligned} x^{q-1} - 1 &= h(x)\,m(x) \\ x^{q-1} - 1 &\equiv 0 \pmod{m(x)} \\ x^q - x &\equiv 0 \pmod{m(x)}, \end{aligned}$$

   welche auch die Kongruenz $h(x)\,m(x) \equiv 0 \pmod{x^{q-1} - 1}$ rechtfertigen.

3.4.4 Zerfällungskörper

Nach dem Hauptsatz der Zahlentheorie kann man für jede natürliche Zahl eine eindeutigen Primfaktorzerlegung der Form

$$m_1^{e_1}\, m_2^{e_2}\, m_3^{e_3} \cdots$$

[31] Ein Erweiterungskörper schafft dadurch die Möglichkeit, daß auch Potenzen von Primelementen noch als Körper zulässig sind.

finden. Gleiches trifft auch für Polynome in $\mathbb{F}_p[x]$ zu, nur daß es sich um irreduzible Polynome anstatt Primzahlen handelt.

$$f(x) = [m_1(x)]^{e_1}\ [m_2(x)]^{e_2}\ [m_3(x)]^{e_3} \cdots$$

Jedes der irreduziblen Polynome[32] $m_i(x)$ hat eine vom jeweiligen Grad n abhängige Anzahl von Nullstellen α, die entweder im Grundkörper $\mathbb{F}_p$ oder (sämtlich) in einem zugehörigen Erweiterungskörper $\mathbb{F}_{p^n}$ liegen. Nullstellen im Grundkörper, von denen $f(x)$ maximal $p = |\mathbb{F}_p|$ besitzen kann, lassen sich immer als einfache Faktoren der Art $m(x) = x - \alpha$ mit $\alpha \in \mathbb{F}_p$ darstellen (vgl. Beispiel-Faktorisierung von $x^3 - 1$ in Abschnitt 3.4.5). Liegen dagegen alle n Wurzeln von $m(x)$ in einem Erweiterungskörper, dann muß es sich um ein irreduzibles Polynom (höheren Grades) handeln.[33] Ein solcher Körper besteht aus p^n Elementen, welche die $q = p^n$ Nullstellen der zugeordneten Funktion $\psi(x) = x^q - x$ darstellen (vgl. Abschnitt 2.1.5). $\mathbb{F}_{p^n}$ nennt man deshalb auch den kleinsten Körper über den $\psi(x) \in \mathbb{F}_p[x]$ vollständig in Linearfaktoren zerfällt [Bos96, 4.5] bzw. kürzer: $\mathbb{F}_{p^n}$ sei der Zerfällungskörper von $x^q - x$.

$$\psi(x) = x^q - x = \prod_{\alpha \in \mathbb{F}_q} (x - \alpha), \qquad \psi(x) \in \mathbb{F}_p[x]$$

Ein weiteres Charakteristikum des Zerfällungskörpers $\mathbb{F}_{p^n}$ sind die konjugierten Nullstellen, d. h. bei Kenntnis einer Nullstelle $\alpha \neq 0$ des irreduziblen Polynoms $m(\alpha) = 0$ sind die restlichen $n - 1$ Nullstellen genau die Potenzen $\alpha^p, \alpha^{p^2}, \ldots, \alpha^{p^{n-2}}, \alpha^{p^{n-1}}$. Das irreduzible Polynom $m(x)$ zerfällt also bei Kenntnis nur *einer* Nullstelle α vollständig in seine n Linearfaktoren. Der Beweis dieses Satzes geht vom sogenannten „Anfänger-Traum“ (Freshmans Dream) aus:

$$(b + c)^p = b^p + c^p, \qquad b, c \in \mathbb{F}_{p^n} \tag{3.13}$$

und berücksichtigt dann, daß im Grundkörper $\mathbb{F}_p$ jedes Element a die Relation $a^p = a$ erfüllt (vgl. Abschnitt 3.3.2).

$$m(\alpha^p) = \sum_{i=0}^{n} a_i \alpha^{ip} = \sum_{i=0}^{n} a_i^p \alpha^{ip} = \sum_{i=0}^{n} \left(a_i \alpha^i\right)^p = \left(\sum_{i=0}^{n} a_i \alpha^i\right)^p = [m(\alpha)]^p = 0$$

Aus diesem Grund läßt sich für jedes der irreduziblen Polynome $m(x)$ die folgende Linearfaktordarstellung angeben:[34]

[32] Die Indizierung mit i wird im weiteren wieder weggelassen, da wir uns auf die Betrachtung einer Funktion $m(x) := m_i(x)$ beschränken. Man sollte sich aber immer bewußt sein, daß die Variablen n, q sowie die abgeleiteten Größen bzw. Funktionen $\psi(x)$ abhängig von $m(x)$ variieren.

[33] Und umgekehrt, denn $m(x)$ ist ja im Grundkörper nicht weiter faktorisierbar.

[34] Das irreduzible Polynom $m(x)$ wird auch Minimalpolynom von α über $\mathbb{F}_p$ genannt. Bei einer Minimalfunktion handelt es sich allgemein um ein normiertes Polynom $m(x) \in \mathbb{F}_p[x]$ *kleinsten* Grades, welches beim Einsetzen eines Elementes $\alpha \in \mathbb{F}_q^*$ die Gleichung $m(\alpha) = 0$ erfüllt [PW72, 6.5].

$$m(x) = \prod_{i=0}^{n-1} \left(x - \alpha^{p^i}\right) .$$

Die Nullstellen α^{p^i} kann man (wegen ihrer linearen Unabhängigkeit) verwenden, um statt einer Polynombasis eine so genannte Normalbasis des Vektorraumes (der Dimension n) über $\mathbb{F}_p$ zu definieren.

Ergänzungen zu Formel 3.13

1. Sie kommt zustande wenn man bei der formalen Anwendung des Binomischen Satzes berücksichtigt, daß im Binomialkoeffizient $\binom{p}{k} = \frac{p!}{k!(p-k)!}$ der Faktor $p!$ für $k \neq 0$ immer durch p teilbar ist und folglich $\binom{p}{k} \bmod p = 0$ gilt.

$$(b+c)^p = \sum_{k=0}^{p} \binom{p}{k} b^{p-k} c^k = b^p + c^p + \underbrace{\sum_{k=1}^{p-1} \binom{p}{k} b^{p-k} c^k}_{=0}$$

2. Wendet man sie mehrfach an, dann ergibt:

$$(b+c)^{p^i} = b^{p^i} + c^{p^i}, \qquad b, c \in \mathbb{F}_{p^n} . \tag{3.14}$$

3.4.5 Erweiterungskörper $\mathbb{Z}_2^n$

Für eine Erweiterung des Primkörpers $\mathbb{Z}_2$ auf n Dimensionen ist ein irreduzibles Polynom $m(x)$ vom Grad n notwendig. Der dadurch entstehende Körper $\mathbb{Z}_2^n$ soll am Beispiel des Polynoms $m(x) = x^2 + x + 1$ mit Koeffizienten aus $\mathbb{Z}_2$ (zu den Rechenoperationen vgl. Seite 24) jetzt kurz betrachtet werden. Nach Darstellung 3.10 gehören genau $q = p^n = 4$ Polynome zum Erweiterungskörper ($p = 2$, $n = 2$), nämlich:

$$r_0(x) = 0 \qquad r_1(x) = 1$$
$$r_2(x) = x \qquad r_3(x) = x + 1 .$$

Wie sich leicht feststellen läßt, ist $r_3(x)$ das erzeugende Element der multiplikativen Gruppe $\{r_1(x), r_2(x), r_3(x)\}$, denn die anderen Elemente ergeben sich als Potenzen $r_3^v(x) \bmod m(x)$.

$$(x+1)^0 = 1$$
$$(x+1)^1 = x + 1$$
$$(x+1)^2 = x^2 + x + x + 1 = x^2 + 1 \equiv x \pmod{x^2 + x + 1}$$

Außerdem sind alle Elemente (abgesehen von $r_0(x)$) wirklich Nullstellen des Polynoms $x^{q-1} - 1 = x^3 - 1 = (x-1)m(x)$. Auch gut erkennen läßt sich, daß $m(x)$ für jedes Element $r(x)$ ein Teiler von $[r(x)]^{q-1} - 1$ ist.

$$
\begin{aligned}
&[r_1(x)]^3 - 1 = 0 \\
&[r_2(x)]^3 - 1 = x^3 - 1 = (x-1)(x^2+x+1) \equiv 0 \pmod{x^2+x+1} \\
&[r_3(x)]^3 - 1 = (x+1)^3 - 1 = x^3 + x^2 + x = x(x^2+x+1) \equiv 0 \pmod{x^2+x+1}
\end{aligned}
$$

Äquivalent dazu ist das Produkt aller Elemente (konform zu Formel 2.15) genau das neutrale Element $e_{(\cdot)}$ der multiplikativen Gruppe.[35]

$$r_1(x) \cdot r_2(x) \cdot r_3(x) = 1 \cdot x \cdot (x+1) = x^2 + x \equiv 1 \pmod{x^2+x+1}$$

4 Algorithmen

4.1 GCD-Algorithmen

4.1.1 E 's Algorithmus

Der Algorithmus von E berechnet den größten gemeinsamen Teiler $d = \gcd(a,b)$ zweier natürlicher Zahlen $a, b \in \mathbb{N}$. Die grundlegende, iterativ angewendete Rechenoperation dabei ist modulare Division.[36] Algorithmus 1 beschreibt das klassische Verfahren [Knu98, Sho05, CP05, MvV92, Ber84]. Es beginnt unter der Voraussetzung $a > b$ mit einer Modulo-Division $c_2 = a \bmod b$. Im nächsten Schritt wird c_2 als Modul verwendet und $c_3 = b \bmod c_2$ berechnet, wonach $c_4 = c_2 \bmod c_3$ folgt usw. . Diese (wegen $c_{k+1} < c_k$) absteigende Sequenz endet wenn $c_{k+1} = 0$ wird[37] – das Ergebnis d befindet sich dann in c_k.

Beweis Jeder Iterationsschritt $c_{i+1} = c_{i-1} \bmod c_i$ kann in der Umkehrung (vgl. Restklassenbeziehung 3.2 in Abschnitt 3) als

$$c_{i-1} = q_{i+1} c_i + c_{i+1}, \qquad 0 \le c_{i+1} < c_i$$

geschrieben werden. So gesehen wird durch den Algorithmus der folgende Abstieg vorgenommen:

[35] Welches in der additiven Gruppe $(\mathbb{Z}_2, \cdot)$ gleich dem Inversen von 1 ist, d. h. zur Erinnerung $-1 \equiv +1$.

[36] Die Komplexität des klassischen Algorithmus wird in [MvV92, 2.4.2] mit $O(\log_2{}^2 n)$ angegeben, was in starkem Maße durch die fortlaufenden Divisionen bestimmt wird. Zur Komplexität verschiedener Varianten des GCD-Algorithmus siehe auch [Har06].

[37] Das diese stetig absteigende Sequenz mit dem Wert 0 endet, ist eine fundamentale Eigenschaft natürlicher Zahlen.

$$
\begin{aligned}
a = c_0 &= q_2 c_1 + c_2 && (c_2 < c_1)\\
b = c_1 &= q_3 c_2 + c_3 && (c_3 < c_2)\\
c_2 &= q_4 c_3 + c_4 && (c_4 < c_3)\\
&\vdots\\
c_{k-3} &= q_{k-1} c_{k-2} + c_{k-1} && (c_{k-1} < c_{k-2})\\
c_{k-2} &= q_k c_{k-1} + c_k && (c_k < c_{k-1})\\
c_{k-1} &= q_{k+1} c_k + 0 && (c_{k+1} = 0)
\end{aligned}
$$

bis c_{k+1} verschwindet.

Warum $d = c_k$ ein gemeinsamer Faktor von a und b ist, wird klar wenn man die Folge rückwärts betrachtet. In der letzten Zeile steht $c_{k-1} = q_{k+1} d$, also teilt d den Rest c_{k-1} (oder kürzer $d \mid c_{k-1}$). Wenn d aber als Faktor in c_{k-1} enthalten ist, dann kann man d auf der rechten Seite der vorletzten Gleichung ausklammern, weshalb es auch als Faktor in c_{k-2} vorkommen muß. Dies setzt sich bis in die erste Gleichung fort (sämtliche Divisionsreste $c_0, c_1, \ldots, c_k$ enthalten folglich d als Teiler), in der die Startbedingung $c_0 = a$ und $c_1 = b$ verankert ist. Deshalb ist der gemeinsame Teiler d sowohl in a als auch b enthalten.

Viel kürzer kann man damit argumentieren, daß die Modulo-Division $c_{i+1} = c_{i-1} \bmod c_i$ eine GCD erhaltende Operation ist. Denn mit der Zerlegung $c_i = \overline{c}_i d$ gilt ausgehend von Formel 3.2:

$$
\begin{aligned}
c_{i+1} &= c_{i-1} \bmod c_i\\
&= c_{i-1} - q_i c_i\\
&= \overline{c}_{i-1} d - q_i \overline{c}_i d\\
&= d(\overline{c}_{i-1} - q_i \overline{c}_i)\\
&= d(\overline{c}_{i-1} \bmod \overline{c}_i)
\end{aligned}
\tag{4.1}
$$

Algorithmus 1 Euklidischer Algorithmus $d = \gcd(a, b)$

Require: $a \geq b$

$c_0 \Leftarrow a$, $c_1 \Leftarrow b$
$k \Leftarrow 0$

repeat
  $k \Leftarrow k + 1$
  $c_{k+1} \Leftarrow c_{k-1} \bmod c_k$
until $c_{k+1} = 0$

$\gcd(a, b) \Leftarrow c_k$

d. h. der gemeinsame Teiler d in c_{i-1} und c_i ist auch in c_{i+1} wieder enthalten.[38]

$$\gcd(c_{i-1}, c_i) = \gcd(c_{i+1}, c_i) = \gcd(c_{i-1} \bmod c_i, c_i) \tag{4.2}$$

Damit d wirklich den größten gemeinsamen Teiler stellt, muß es überhaupt alle gemeinsamen Teiler enthalten. Mit dem Ziel dies nachzuweisen betrachten wir nochmals die Folge beginnend mit der vorletzten Gleichung, welche nach $c_k = d$ umgestellt wird. Ersetzt man darin c_{k-1} mit Hilfe der vorvorletzten Gleichung und fährt aufsteigend fort, so erhält man eine lineare Darstellung für $d = \gcd(a,b)$,

$$\begin{aligned} d &= c_{k-2} - q_k c_{k-1} \\ &= c_{k-2} - q_k(c_{k-3} - q_{k-1}c_{k-2}) = c_{k-2}(1 + q_k q_{k-1}) - c_{k-3}q_k \\ &= (c_{k-4} - q_{k-2}c_{k-3})(1 + q_k q_{k-1}) - c_{k-3}q_k = c_{k-4}(1 + q_k q_{k-1}) - c_{k-3}\left[q_{k-2}(1 + q_k q_{k-1}) + q_k\right] \\ &\vdots \\ &= \alpha c_0 + \beta c_1 \end{aligned}$$

welche auf ganzen Zahlen $\alpha, \beta \in \mathbb{Z}$ sowie $c_0 = a$ und $c_1 = b$ beruht.

$$d = \gcd(a,b) = \alpha a + \beta b \tag{4.3}$$

Mit Hilfe von Formel 4.3, welche auch Satz von B´ genannt wird,[39] kann jetzt relativ einfach bewiesen werden, daß d wirklich der größte gemeinsame Teiler ist. Nehmen wir dazu an, es gäbe einen weiteren gemeinsamen Teiler d'. Dann würde dieser (auf der rechten Seite von Gleichung 4.3) als Faktor von a und b auszuklammern sein und deshalb (wenn man die linke Seite betrachtet) als Teiler von d auftreten. Mit anderen Worten stecken alle weiteren Teiler (schon) in d, weshalb nur dieser der größte gemeinsame Teiler sein kann.

Hinweis Es existieren unendlich viele lineare Darstellungen für d als Linearkombination von a und b (vgl. auch die kurze Betrachtung zu diophantischen Gleichungen auf Seite 45). Jede Substitution der Art $\alpha := \alpha + n\overline{b}$ und $\beta := \beta - n\overline{a}$, mit

$$a = d\overline{a} \qquad b = d\overline{b} \qquad 1 = \gcd(\overline{a}, \overline{b}) = \alpha\overline{a} + \beta\overline{b} \tag{4.4}$$

[38] So gilt allgemein: $\gcd(a,b) = \gcd(a \bmod b, b)$.

[39] Die B´ -Identität nach Formel 4.3 kann man für $\alpha, \beta \in \mathbb{N}$ auch als $d = \pm\alpha a \mp \beta b$ schreiben.

erfüllt B´ 's Identität 4.3 ebenso.

$$\begin{aligned} d &= (\alpha + n\overline{b})a + (\beta - n\overline{a})b \\ &= \alpha a + \beta b + n(\overline{b}a - \overline{a}b) \\ &= \alpha a + \beta b + n(\overline{b}\overline{a}d - \overline{a}\overline{b}d) \\ &= \alpha a + \beta b \end{aligned}$$

Die kleinsten Werte für $|\alpha|$ und $|\beta|$ zeichnen sich folglich durch die Relationen $|\alpha| < \overline{b}$ und $|\beta| < \overline{a}$ aus. Um sie zu ermitteln kann man entweder $\overline{a}$ und $\overline{b}$ sukzessive von α und β subtrahieren/addieren oder man reduziert die B´ -Kofaktoren mit Hilfe von $q = \left\lfloor |\alpha|/\overline{b} \right\rfloor$ bzw. $q = \lfloor |\beta|/\overline{a} \rfloor$ und folgender Formeln:[40]

$$\alpha := \alpha \mp q\overline{b} \qquad\qquad \beta := \beta \pm q\overline{a}\,.$$

4.1.2 Erweiterter euklidischer Algorithmus

Der erweiterte euklidische Algorithmus erlaubt eine effiziente Berechnung der B´ -Kofaktoren α und β zusammen (und gleichzeitig) mit dem größten gemeinsamen Teiler (siehe auch [Knu98, Sho05, CP05, Ber84]). Dazu berücksichtigt er einige Erkenntnisse aus dem vorigen Abschnitt, insbesondere daß:

- der größte gemeinsame Teiler $d = \gcd(a,b)$ für $\alpha,\beta \in \mathbb{Z}$ als Linearkombination $\alpha a + \beta b$ darstellbar ist;
- d für $0 \le i \le k$ jeden Rest c_i teilt und damit die Darstellung $c_i = d\overline{c}_i = \alpha_i a + \beta_i b$ rechtfertigt;
- die Berechnung von c_{i+1} mit Hilfe von $q_{i+1} = \lfloor c_{i-1}/c_i \rfloor$ und $c_{i+1} = c_{i-1} - q_{i+1}c_i$ erfolgen kann.

Durch Induktion ergibt sich aus

$$c_i = c_{i-2} - q_i c_{i-1} = d\overline{c}_i = \alpha_i a + \beta_i b \tag{4.5}$$

wenn man $c_{i-1} = \alpha_{i-1}a + \beta_{i-1}b$ und $c_i = \alpha_i a + \beta_i b$ berücksichtigt:

$$\begin{aligned} c_{i+1} &= c_{i-1} - q_{i+1}c_i = \alpha_{i+1}a + \beta_{i+1}b \\ &= \alpha_{i-1}a + \beta_{i-1}b - q_{i+1}(\alpha_i a + \beta_i b) \\ &= (\alpha_{i-1} - q_{i+1}\alpha_i)a + (\beta_{i-1} - q_{i+1}\beta_i)b\,. \end{aligned}$$

Vergleich mit Ausgangsformel 4.5 erlaubt die Bestimmung von α_{i+1} und β_{i+1}.

[40] Sollte man nur an einem Kofaktor (beispielhaft α) interessiert sein, dann kann auf die Berechnung von q ganz verzichtet und statt dessen $\alpha := \alpha \bmod \overline{b}$ gerechnet werden.

$$\alpha_{i+1} = \alpha_{i-1} - q_{i+1}\alpha_i \qquad \beta_{i+1} = \beta_{i-1} - q_{i+1}\beta_i$$

Mit den Startwerten $\alpha_0 = 1$ und $\beta_0 = 0$ (gewährleistet $c_0 = a$) bzw. $\alpha_1 = 0$ und $\beta_1 = 1$ (ebenso für $c_1 = b$) kann man den erweiterten euklidischen Algorithmus 2 formulieren.

Algorithmus 2 Erweiterter euklidischer Algorithmus

Require: $a \geq b$

$(c_0, \alpha_0, \beta_0) \Leftarrow (a, 1, 0)$
$(c_1, \alpha_1, \beta_1) \Leftarrow (b, 0, 1)$
$k \Leftarrow 0$

repeat
  $k \Leftarrow k + 1$
  $q \Leftarrow \lfloor c_{k-1}/c_k \rfloor$ {Integer-Division}
  $(c_{k+1}, \alpha_{k+1}, \beta_{k+1}) \Leftarrow (c_{k-1}, \alpha_{k-1}, \beta_{k-1}) - q(c_k, \alpha_k, \beta_k)$ $\{c_{k+1} = c_{k-1} \bmod c_k\}$
until $c_{k+1} = 0$

$(d, \alpha, \beta) \Leftarrow (c_k, \alpha_k, \beta_k)$

Er endet ganz genauso wie Algorithmus 1 bei $c_{k+1} = 0$, ermittelt aber zusätzlich die B´ - Kofaktoren $\alpha = \alpha_k$ und $\beta = \beta_k$ sowie die teilerfremden Anteile $\overline{a} = a/d$ und $\overline{b} = b/d$. Letztere findet man wegen $c_{k+1} = \alpha_{k+1}a + \beta_{k+1}b = 0$ am Ende in β_{k+1} und α_{k+1}. Dies wird relativ schnell aus

$$\alpha_{k+1}a = -\beta_{k+1}b$$
$$|\alpha_{k+1}|\overline{a} = |\beta_{k+1}|\overline{b}$$

ersichtlich, wenn man auf beiden Seiten der letzten Gleichung das kleinste gemeinsame Vielfache anvisiert.

$$|\alpha_{k+1}| = \overline{b} \qquad |\beta_{k+1}| = \overline{a}$$

Hinweis Als Ergänzung zum Hinweis von Seite 32 kann man sogar feststellen, daß die „Wahl“ von q nicht unbedingt mit einer Modulo-Division verbunden sein muß. Etwas anders könnte man q einfach so wählen, daß

- die Folge der c_k weiterhin eine absteigende Sequenz ist;

- und in jedem Iterationsschritt $qc_k \leq c_{k-1}$ gilt (wodurch $c_{k+1} \geq 0$ gewährleistet ist).

Den steilsten Abstieg $\Delta = c_{k+1} - c_k$ erreicht man allerdings, wenn q im jeweiligen Iterationsschritt maximal ist. Dies ist aber unter der Voraussetzung $c_{k+1} \geq 0$ genau bei einer Modulo-Division $c_{k+1} = c_{k-1} \bmod c_k$ der Fall. Ansonsten ist jeder Wert $q = 0 \ldots \lfloor c_{i-1}/c_i \rfloor$ geeignet — einzig die Konvergenzgeschwindigkeit nimmt mit kleineren Werten von q ab.[41]

4.1.3 E 's Algorithmus für Polynome über $\mathbb{Z}_2$

Für Polynome mit Koeffizienten aus GF(2) kann man ganz ähnlich verfahren [HMV04, 2.3.6],[Sho05, 17.3], [CP05, 2.2.1], [Ber84, 2.1]. Mit Rückblick auf den Hinweis von Seite 34 muß man nur folgendes beachten:

1. Addition und Subtraktion sind gleichwertig (und die Addition wird zu einem *Exklusiv-Oder* $\oplus$).

2. An die Stelle eines wertbehafteten Vergleichs $c_{k-1} \geq c_k$ tritt ein Vergleich des Polynomgrades $\deg c_{k-1} \geq \deg c_k$.

3. Die Vorbedingung $\deg a \geq \deg b$ kann entfallen, da der Algorithmus selbst die Vertauschung der Argumente vornimmt.

4. Die Ermittlung von $q(x)$ läßt sich in folgende Einzelfälle zerlegen:

   - $\deg c_{k-1}(x) < \deg c_k(x)$ führt zu $q(x) = 0$, was nach Algorithmus 3 einer Vertauschung $(c_k, \alpha_k, \beta_k) \Longleftrightarrow (c_{k-1}, \alpha_{k-1}, \beta_{k-1})$ entspricht.
   - $\deg c_{k-1}(x) = \deg c_k(x)$ ergibt $q(x) = 1$, was mit $(c_{k+1}, \alpha_{k+1}, \beta_{k+1}) \Leftarrow (c_{k-1}, \alpha_{k-1}, \beta_{k-1}) + (c_k, \alpha_k, \beta_k)$ im nachfolgenden Fall aufgeht.
   - Für $\deg c_{k-1}(x) > \deg c_k(x)$ müßte man $q(x) = \lfloor c_{k-1}(x)/c_k(x) \rfloor$ berechnen. Mit der Option auch ein kleineres $q(x)$ zu verwenden (siehe Hinweis auf Seite 34), kann man sich eine vollständige Polynomdivision sparen. Stattdessen ist eine sinnvolle Vorgehensweise sich bei der Divison auf das höchstwertige Bit von $c_{k-1}(x)$ und $c_k(x)$ beschränken. In einer monischen Polynomrepräsentation (der höchstwertige Koeffizient ist 1) führt die Division so zu: $x^{\deg c_{k-1} - \deg c_k}$.

[41] Für den Fall $q < \lfloor c_{i-1}/c_i \rfloor$ bezeichnet man die Folge der c_i auch als eine Folge von Pseudo-Divisionsresten.

Algorithmus 3 Erweiterter euklidischer Algorithmus für Polynome in $\mathbb{Z}_2[x]$

$(c_0, \alpha_0, \beta_0) \Leftarrow (a, 1, 0)$
$(c_1, \alpha_1, \beta_1) \Leftarrow (b, 0, 1)$
$k \Leftarrow 0$

repeat
  $k \Leftarrow k + 1$
  if $\deg c_{k-1} \geq \deg c_k$ **then**
    $q \Leftarrow x^{\deg c_{k-1} - \deg c_k}$ $\{q(x) = 1$ im Fall: $\deg c_{k-1}(x) = \deg c_k(x)\}$
  else
    $q \Leftarrow 0$ $\{$führt letztlich zu: $(c_{k+1}, \alpha_{k+1}, \beta_{k+1}) \Leftarrow (c_{k-1}, \alpha_{k-1}, \beta_{k-1})\}$
  end if
  $(c_{k+1}, \alpha_{k+1}, \beta_{k+1}) \Leftarrow (c_{k-1}, \alpha_{k-1}, \beta_{k-1}) \oplus q(c_k, \alpha_k, \beta_k)$
until $c_{k+1} = 0$

$(d, \alpha, \beta) \Leftarrow (c_k, \alpha_k, \beta_k)$

4.1.4 Binärer GCD-Algorithmus

Der binäre GCD-Algorithmus kommt im Gegensatz zum klassischen euklidischen Algorithmus ohne Divisionen aus [CP05, Knu98, Ste67, Wel01]. Statt dessen wird (beginnend mit $a_0 = a$ und $b_0 = b$) durch Subtraktion und Halbierung eine stetige Reduktion der Argumente a und b vorgenommen, welche letztlich zum größten gemeinsamen Teiler $d = \gcd(a, b)$ führt.[42]

$$\begin{aligned} d &= \gcd(a_0, b_0) \\ &= \gcd(a_1, b_1) \\ &\vdots \\ &= \gcd(a_i, b_i) \\ &\vdots \\ &= \gcd(a_k, 0) = a_k \end{aligned}$$

Der Algorithmus endet spätestens dann, wenn im Schritt $i = k$ eines der beiden Argumente a_i oder b_i verschwindet (im obigen Fall beispielhaft für $b_k = 0$). Das Reduktionsschema zeigt Tabelle 4.1.

[42] Im Wesen unterscheidet er sich nicht vom euklidischen Algorithmus, der in ähnlicher Art und Weise die Argumente stetig reduziert und dabei den größten gemeinsamen Teiler d bewahrt. Die Halbierung eines der Argumente ist allerdings (durch Verschiebung um ein Bit) effizient zu realisieren.

Tabelle 4.1: Reduktionsschema des binären GCD-Algorithmus

a_i	b_i	$\gcd(a_{i+1}, b_{i+1}) =$	Begründung
gerade		$2\gcd\left(\frac{a_i}{2}, \frac{b_i}{2}\right)$	gemeinsamer Teiler ist 2
gerade	ungerade	$\gcd\left(\frac{a_i}{2}, b_i\right)$	2 ist kein gemeinsamer Teiler
ungerade	gerade	$\gcd\left(a_i, \frac{b_i}{2}\right)$	2 ist kein gemeinsamer Teiler
ungerade, $a_i \geq b_i$		$\gcd\left(\frac{a_i - b_i}{2}, b_i\right)$	$a_i - b_i$, wie auch $a_i + b_i$, enthalten den Teiler $\gcd(a_i, b_i) = d$, denn $a_i = \overline{a}_i d$, $b_i = \overline{b}_i d$ führt zu $\gcd(a_i - b_i, b_i) = \gcd[d(\overline{a}_i - \overline{b}_i), \overline{b}_i d]$. Außerdem sind sowohl Summe als auch Differenz gerade Zahlen ($a_i = 2\nu + 1$, $b_i = 2\mu + 1$ ergibt $a_i - b_i = 2(\nu - \mu)$), weshalb 2 als gemeinsamer Teiler wieder ausgeschlossen werden kann.
ungerade, $b_i \geq a_i$		$\gcd(a_i, \frac{b_i - a_i}{2})$	siehe Fall $a_i > b_i$
$a_i > 0$	0	a_i	a_i ist größter gemeinsamer Teiler mit 0
0	$b_i > 0$	b_i	b_i ist größter gemeinsamer Teiler mit 0

Anmerkungen

1. Bezüglich a_{k-1} gibt es genau eine Situation, die das Verschwinden von a_k hervorruft (genauso bezüglich b_{k-1} und b_k) . Betrachtet man dazu Tabelle 4.1, so wird a_{k-1} in folgenden Fällen reduziert:

   a) a_{k-1} und b_{k-1} sind gerade: Halbierung von $a_{k-1} = 1$ ergibt 0, bedeutet aber ungerades a_{k-1} (Widerspruch).

   b) a_{k-1} gerade, b_{k-1} ungerade: Halbierung von $a_{k-1} = 1$ ergibt 0, was ebenfalls ungerades a_{k-1} bedeutet (Widerspruch).

   c) a_{k-1} und b_{k-1} sind ungerade: Halbierung von $b_{k-1} - a_{k-1} = 1$ ergibt 0, aber unter der Voraussetzung $b_{k-1} = a_{k-1} + 1$ können niemals beide ungerade sein (Widerspruch).

   d) a_{k-1} und b_{k-1} sind ungerade: Halbierung von $b_{k-1} = a_{k-1}$ führt zu $a_k = 0$ und stellt damit den einzig möglichen Fall im Schritt $k - 1$ dar.

2. In jedem Iterationszyklus wird entweder a_i oder b_i um mindestens ein Bit reduziert.[43] Aus diesem Umstand läßt sich (im Fall $a_0 > b_0$) für die Anzahl der Iterationsschritte $\lceil \log_2 a_0 \rceil \leq k \leq 2\lceil \log_2 a_0 \rceil$ schlußfolgern (vgl. auch [Kal95, Theorem 2]).

[43] Wodurch zwar mehr Schritte als beim euklidischen Algorithmus nötig sind, sich die Ausführungszeit (durch Vermeidung von Langzahldivisionen) aber typischerweise verringert. Die Komplexität des Algorithmus bleibt unverändert quadratisch, kann aber durch eine systolische Implementierung nach [BK83, Jeb93, BB87] sogar bis auf $O(n)$ reduziert werden.

3. Die Halbierung des Arguments für den Fall, daß a_i und b_i ungerade sind, muß man nicht unbedingt im Iterationsschritt i ausführen. Es ist durchaus legitim, falls beispielsweise $a_i \geq b_i$ gilt, einfach nur $a_{i+1} = a_i - b_i$ zu berechnen. Die Differenz ergibt ja bekanntlich wieder eine gerade Zahl (dann ist a_{i+1} gerade, b_{i+1} ungerade) und es kommt im nächsten Iterationsschritt zu der gewünschten Halbierung.

4. Den Fall, daß sowohl a_i als auch b_i gerade sind, kann man grundsätzlich aus der Haupt-Iterationsschleife herausziehen. Denn wird einmal eines der beiden Argumente ungerade, dann kann dieser Fall niemals wieder eintreten (genau das ungerade Argument ist in allen anderen Reduktionsfällen unveränderlich). Solange beide Argumente gerade sind, kann man sie also kontinuierlich reduzieren, bis nach n Schritten entweder a_i oder b_i ungerade geworden ist. Danach kann auf $a_n = a/2^n$ und $b_n = b/2^n$ irgendein binärer (erweiterter) GCD-Algorithmus angewendet werden, der $\gcd(a_n, b_n) = \alpha_n a_n + \beta_n b_n$ ermittelt.

$$\begin{aligned} d &= 2^n \gcd\left(\frac{a}{2^n}, \frac{b}{2^n}\right) \\ &= 2^n(\alpha_n a_n + \beta_n b_n) = \alpha_n a + \beta_n b \end{aligned}$$

Der so ermittelte größte gemeinsame Teiler muß am Schluß nur noch mit 2^n multipliziert werden (also um n Bit verschoben), was auch Algorithmus 4 entsprechend wiedergibt.[44] Wir können deshalb in den Betrachtungen zum erweiterten binären GCD-Algorithmus (siehe nächste Abschnitte) immer voraussetzen, daß a oder b ungerade ist.

Algorithmus 4 Reduktion gerader Argumente

```
n ⇐ 0

while a gerade ∧ b gerade do
    a ⇐ a/2
    b ⇐ b/2
    n ⇐ n+1
end while

d ⇐ 2^n gcd(a,b)        {GCD-Algorithmus mit Voraussetzung: a oder b ungerade}
```

5. Für den Fall, daß entweder a oder b ungerade ist (oder beide), kann man ausgehend von $a = d\overline{a}$ und $b = d\overline{b}$ feststellen:[45]

   - Der größte gemeinsame Teiler d ist ungerade.
   - Sollte a ungerade sein, dann ist es dessen teilerfremde Anteil $\overline{a}$ ebenfalls.
   - Sollte b ungerade sein, dann ist es dessen teilerfremde Anteil $\overline{b}$ auch.

[44] Die B´ -Koeffizienten α und β müssen übrigens nicht korrigiert werden.

[45] Grundlage bildet die Tatsache, daß das Produkt zweier ganzer Zahlen nur dann ungerade ist, wenn beide Faktoren ungerade sind.

4.1.5 Erweiterter binärer GCD-Algorithmus

Algorithmus nach K [46] **[Kal95]** Betrachtet man das Reduktionsschema des binären GCD-Algorithmus, so kann man verschiedenste lineare Transformationen der Art

$$\begin{aligned} a_i &= u'_{i+1}a_{i+1} + v'_{i+1}b_{i+1} & a_{i+1} &= u''_i a_i + v''_i b_i \\ b_i &= s'_{i+1}a_{i+1} + t'_{i+1}b_{i+1} & b_{i+1} &= s''_i a_i + t''_i b_i, \end{aligned}$$

definieren, wobei sich die einzelnen Betrachtungsweisen durch unterschiedliche Werte in den Koeffizienten unterscheiden. Bei jedem Schritt sind so a_i und b_i als lineare Funktion von a_{i+1} und b_{i+1} darstellbar, die Ausgangsgrößen a_0 und b_0 deshalb als lineare Funktionen von a_i und b_i.

$$a_0 = u_i a_i + v_i b_i \qquad (u_0 = 1, v_0 = 0) \tag{4.6}$$
$$b_0 = s_i a_i + t_i b_i \qquad (s_0 = 0,\ t_0 = 1) \tag{4.7}$$

Durch Einsetzen von a_i und b_i in folgende Gleichung,

$$\begin{aligned} a_0 &= u_i a_i + v_i b_i = u_{i+1}a_{i+1} + v_{i+1}b_{i+1} \\ b_0 &= s_i a_i + t_i b_i \ = s_{i+1}a_{i+1} + t_{i+1}b_{i+1} \end{aligned}$$

gefolgt von einem Koeffizientenvergleich, kann man für die konkreten Reduktionsfälle die jeweilige Transformationen der Linearfaktoren ableiten (siehe Tabelle 4.2). Am Beispiel $a_{i+1} = (a_i - b_i)/2$, $b_{i+1} = b_i$ soll das Vorgehen exemplarisch verdeutlicht werden.

$$\begin{aligned} s_i a_i + t_i b_i &= s_{i+1}\frac{a_i - b_i}{2} + t_{i+1}b_i \\ &= \frac{s_{i+1}}{2}a_i + \left(t_{i+1} - \frac{s_{i+1}}{2}\right)b_i \\ &= \underbrace{\frac{s_{i+1}}{2}}_{s_i}\, a_i + \underbrace{(t_{i+1} - s_i)}_{t_i} b_i \end{aligned}$$

Aus den Transformationen nach Tabelle 4.2 kann man außerdem die Determinante

$$\mathbf{D_i} = u_i t_i - s_i v_i = \begin{vmatrix} u_i & v_i \\ s_i & t_i \end{vmatrix}, \text{ mit } \mathbf{D_0} = 1$$

Tabelle 4.2: Linearfaktoren beim binären GCD-Algorithmus nach K

a_i	b_i	$a_{i+1} =$	$b_{i+1} =$	$u_{i+1} =$	$v_{i+1} =$	$s_{i+1} =$	$t_{i+1} =$	$\mathbf{D_{i+1}} =$
gerade		$\frac{a_i}{2}$	$\frac{b_i}{2}$	$2u_i$	$2v_i$	$2s_i$	$2t_i$	$4\mathbf{D_i}$
gerade	ungerade	$\frac{a_i}{2}$	b_i	$2u_i$	v_i	$2s_i$	t_i	$2\mathbf{D_i}$
ungerade	gerade	a_i	$\frac{b_i}{2}$	u_i	$2v_i$	s_i	$2t_i$	$2\mathbf{D_i}$
ungerade, $a_i \geq b_i$		$\frac{a_i - b_i}{2}$	b_i	$2u_i$	$u_i + v_i$	$2s_i$	$s_i + t_i$	$2\mathbf{D_i}$
ungerade, $b_i \geq a_i$		a_i	$\frac{b_i - a_i}{2}$	$u_i + v_i$	$2v_i$	$s_i + t_i$	$2t_i$	$2\mathbf{D_i}$

bestimmen.

Um den Zusammenhang mit den B´ -Koeffizienten $\alpha, \beta \in \mathbb{Z}$ in $\gcd(a,b) = \alpha a_0 + \beta b_0$ herzustellen, stellen wir die Formeln 4.6 und 4.7 nach a_i und b_i um. Dazu werden beide Gleichungen mit den Linearfaktoren der jeweils anderen multipliziert und dann wechselweise voneinander subtrahiert.

$$\begin{aligned} s_i a_0 - u_i b_0 &= s_i(u_i a_i + v_i b_i) - u_i(s_i a_i + t_i b_i) \\ &= (s_i v_i - u_i t_i) b_i \\ &= -\mathbf{D_i}\, b_i \end{aligned} \qquad \begin{aligned} t_i a_0 - v_i b_0 &= t_i(u_i a_i + v_i b_i) - v_i(s_i a_i + t_i b_i) \\ &= (u_i t_i - s_i v_i) a_i \\ &= \mathbf{D_i}\, a_i \end{aligned}$$

Schließt man den Fall aus, daß a_0 *und* b_0 gerade sind ($\mathbf{D_i} = 2^i$, vgl. Anmerkungen auf Seite 37), dann kann für den letzten Iterationsschritt $i = k$, in Abhängigkeit davon ob a_k oder b_k zuerst verschwindet, folgendermaßen konkretisiert werden:

$$\begin{aligned} &\text{Fall: } b_k = 0 \\ &s_k a_0 - u_k b_0 = 0 \\ &t_k a_0 - v_k b_0 = 2^k a_k \end{aligned} \qquad \begin{aligned} &\text{Fall: } a_k = 0 \\ &s_k a_0 - u_k b_0 = -2^k b_k \\ &t_k a_0 - v_k b_0 = 0\,. \end{aligned}$$

An diesem Punkt stellen wir jedoch fest, daß es sich bei den Größen u_k, v_k, s_k und t_k nicht um die Kofaktoren von $\gcd(a,b)$, sondern um eine Linearfaktordarstellung von $2^k \gcd(a,b)$ handelt.[47]

[46] Auch *Right Shift Delayed Halving* (RSDH) oder *Almost Montgomery Inverse* Algorithmus genannt [Har06].

[47] Für Anwendungen, die im weiteren eine M -Reduktion vornehmen, kann dies sogar von Vorteil sein [Kal95].

Fall: $b_k = 0$	Fall: $a_k = 0$
$a_k = \gcd(a,b) = \alpha a_0 + \beta b_0$	$b_k = \gcd(a,b) = \alpha a_0 + \beta b_0$
$2^k \gcd(a,b) = t_k a_0 - v_k b_0$	$2^k \gcd(a,b) = -s_k a_0 + u_k b_0$

Um aus t_k, v_k und s_k, u_k die B´ -Koeffizienten α und β zu bestimmen, sind zusätzliche Korrekturschritte nötig, welche in [Kal95] auch Korrekturphase (oder Phase II) genannt werden. Ziel ist es dabei, die rechte Seite der letzten Gleichung durch 2^k zu dividieren (oder in k Schritten wiederholt durch 2). Da die linke Seite immer eine gerade Zahl ist, können folgende Schlußfolgerungen im Falle $b_k = 0$ gezogen werden (falls $a_k = 0$ war, äquivalent für $t_k \to s_k$ und $v_k \to u_k$):[48]

- Sind v und t gerade, dann ist eine ganzzahlige Halbierung $v := v^{(g)}/2$, $t := t^{(g)}/2$ möglich (d. h. führt wieder zu einer ganzen Zahl).
- In allen anderen Fällen (t bzw. v ungerade) kann man wegen

$$2^i \gcd(a,b) = ta - vb = (t+b)a - (v+a)b$$

die Reduktion $t := (t+b)/2$ bzw. $v := (v+a)/2$ vornehmen, ohne daß sich die Gleichung ändert. Man wandelt jedoch die ungerade Zahl t bzw. v in eine gerade Zahl, die dann (wie gewünscht) durch 2 teilbar ist. Die Ursache liegt einerseits darin begründet, daß die linke (und demzufolge auch rechte) Seite der Gleichung

$$2^i \gcd(a,b) = ta - vb$$

immer eine gerade Zahl repräsentiert und andererseits, entweder a oder/und b als ungerade vorausgesetzt wurden. Berücksichtigt man die folgenden Gesetzmäßigkeiten:

1. die Summe zweier ungerader oder zweier gerader Zahlen ist eine gerade Zahl;
2. die Summe einer ungeraden und einer geraden Zahl ist eine ungerade Zahl;
3. das Produkt zweier ungerader Zahlen ist eine ungerade Zahl;
4. alle anderen Kombinationen ergeben bei der Multiplikation eine gerade Zahl;

dann sind genau drei Fälle konstruierbar, in denen t oder v ungerade ist.

$$\begin{aligned}
\left[2^i \gcd(a,b)\right]^{(g)} &= \left[t^{(u)} a^{(u)}\right]^{(u)} - \left[v^{(u)} b^{(u)}\right]^{(u)} \\
\left[2^i \gcd(a,b)\right]^{(g)} &= \left[t^{(u)} a^{(g)}\right]^{(g)} - \left[v^{(g)} b^{(u)}\right]^{(g)} \\
\left[2^i \gcd(a,b)\right]^{(g)} &= \left[t^{(g)} a^{(u)}\right]^{(g)} - \left[v^{(u)} b^{(g)}\right]^{(g)}
\end{aligned}$$

In allen diesen Varianten führt aber die Addition $t+b$ bzw. $v+a$ zu einer geraden Zahl, womit sich das Korrekturverfahren bestätigt.

Algorithmus 5 Algorithmus $d = \gcd(a,b) = \alpha a + \beta b$ nach K

Require: a ungerade $\vee$ b ungerade

{Phase I}
$(a_0, u_0, s_0) \Leftarrow (a, 1, 0)$
$(b_0, v_0, t_0) \Leftarrow (b, 0, 1)$
$f \Leftarrow 0$ {Flag, daß anzeigt, ob schlußendlich α oder β negativ ist}
$k \Leftarrow 0$

while $a_k > 0$ **do**
  if $b_k > a_k$ **then**
    $(b_k, v_k, t_k) \Longleftrightarrow (a_k, u_k, s_k)$ {gewährleistet $a_k \geq b_k$}
    $f \Leftarrow \overline{f}$ {invertiere Flag}
  end if
  if a_k ungerade $\wedge$ b_k ungerade **then**
    $(a_k, v_k, t_k) \Longleftarrow (a_k - b_k, v_k + u_k, t_k + s_k)$ {a_k ist jetzt gerade, b_k weiterhin ungerade}
  end if
  if a_k gerade **then**
    $(a_{k+1}, u_{k+1}, s_{k+1}) \Longleftarrow (a_k/2, 2u_k, 2s_k)$ {a_k gerade, b_k ungerade}
  else
    $(b_{k+1}, v_{k+1}, t_{k+1}) \Longleftarrow (b_k/2, 2v_k, 2t_k)$ {a_k ungerade, b_k gerade}
  end if
  $k \Leftarrow k + 1$
end while

$\gcd(a,b) \Leftarrow b_k$ {Teilergebnis $d = \gcd(a,b)$}
{Phase II}
$i \Leftarrow k$

while $i > 0$ **do**
  if s_k ungerade $\vee$ u_k ungerade **then**
    $s_k \Longleftarrow s_k + t_k$ {Addition von $t_k = \overline{b}$ (vgl. Bemerkung 3)}
    $u_k \Longleftarrow u_k + v_k$ {Addition von $v_k = \overline{a}$ (vgl. Bemerkung 3)}
  end if
  $s_k \Longleftarrow s_k/2$ {Korrektur α}
  $u_k \Longleftarrow u_k/2$ {Korrektur β}
  $i \Leftarrow i - 1$
end while

$(\alpha, \beta) \Longleftarrow (s_k, u_k)$ {Wenn $f = 0$, dann $\alpha < 0$, sonst β}

Als Ergebnis kann man Algorithmus 5 formulieren, wobei außerdem folgende Anmerkungen berücksichtigt wurden:

1. Aus den Gleichungen 4.6 und 4.7 läßt sich im letzten Iterationsschritt (von Phase I)

$$\begin{array}{ll} \text{Fall: } b_k = 0 & \text{Fall: } a_k = 0 \\ \quad a_0 = u_k a_k = u_k d & \quad a_0 = v_k b_k = v_k d \\ \quad b_0 = s_k a_k = s_k d & \quad b_0 = t_k b_k = t_k d \end{array}$$

   schlußfolgern, d. h. bei u_k und s_k handelt es sich im Fall $b_k = 0$ (gleichermaßen für v_k, t_k im Fall $a_k = 0$) um die teilerfremden Faktoren

$$\begin{array}{ll} \text{Fall: } b_k = 0 & \text{Fall: } a_k = 0 \\ \overline{a}_0 = u_k \le \gcd(a,b) \le a_0 & \overline{a}_0 = v_k \le \gcd(a,b) \le a_0 \\ \overline{b}_0 = s_k \le \gcd(a,b) \le b_0 & \overline{b}_0 = t_k \le \gcd(a,b) \le b_0\ . \end{array}$$

2. Wegen der stetigen Reduktion von a_i oder b_i müssen die Linearfaktoren u_i, s_i, v_i und t_i in Phase I schrittweise anwachsen, denn nur so können a_0 und b_0 nach Gleichung 4.6 und 4.7 konstant bleiben. In [Kal95, Theorem 1] wird bewiesen, daß keine dieser Größen während der Ausführung des Algorithmus den Maximalwert $2a_0 - 1$ überschreitet (für $a_0 > b_0$).[49]
3. Die in Phase II vorzunehmende Addition von a bzw. b kann (entsprechend Anmerkung 5 auf Seite 38) ersetzt werden durch eine Addition von $\overline{a}_0$ bzw. $\overline{b}_0$.[50]

$$\begin{aligned} 2^i \gcd(a,b) &= (t+\overline{b})a - (v+\overline{a})b \\ &= ta + \overline{b}a - vb + \overline{a}b \\ &= ta + \overline{b}\overline{a}d - vb - \overline{a}\overline{b}d \\ &= ta - vb \end{aligned}$$

Algorithmus nach P **[Knu98, Exercise 4.5.2.39], [MvV92, 14.4.3]** Dieser Algorithmus stellt a_{i+1} und b_{i+1} als lineare Funktion von a_i und b_i dar, letztlich wird also von (a_0, b_0) auf (a_i, b_i) geschlossen.

[48] Der Übersichtlichkeit halber wird auf die Indizierung jetzt verzichtet und statt dessen gerade Variablen/Terme mit (g) und ungerade mit (u) gekennzeichnet.

[49] Im Hinweis auf 32 wurde zwar für die kleinsten Werte der B´ -Koeffizienten $|\alpha| < \overline{b}$ und $|\beta| < \overline{a}$ geschlußfolgert, was jedoch nicht für die Zwischenwerte eines bestimmten Algorithmus gelten muß (hier anwendbar auf das Ende von Phase II). Theorem 1 in [Kal95] läßt sich aber auch nachvollziehen, wenn man mit Blick Algorithmus 5 die Anmerkung 2 (auf Seite 37) zur Anzahl der Iterationsschritte berücksichtigt.

[50] Und das nicht nur für den speziellen Fall $\gcd(a,b) = 1$, für welchen $a_0 = \overline{a}_0$, $b_0 = \overline{b}_0$ gilt.

$$a_i = u_i a_0 + v_i b_0 \qquad (u_0 = 1, v_0 = 0) \tag{4.8}$$
$$b_i = s_i a_0 + t_i b_0 \qquad (s_0 = 0,\ t_0 = 1) \tag{4.9}$$

Vorteilhaft wirkt sich aus, daß im letzten Reduktionsschritt (wenn a_k oder b_k verschwindet) u_k und v_k bzw. s_k und t_k direkt die gesuchten B´ -Kofaktoren α, β darstellen.

Fall: $b_k = 0$	Fall: $a_k = 0$
$a_k = \gcd(a,b) = \alpha a_0 + \beta b_0$	$b_k = \gcd(a,b) = \alpha a_0 + \beta b_0$
$= u_k a_0 + v_k b_0$	$= s_k a_0 + t_k b_0$

Durch Einsetzen von a_{i+1} und b_{i+1} (entsprechend Tabelle 4.1) in die Gleichungen

$$a_{i+1} = u_{i+1} a_0 + v_{i+1} b_0$$
$$b_{i+1} = s_{i+1} a_0 + t_{i+1} b_0$$

kann man für den jeweiligen Reduktionsfall die zugehörige Transformationen der Linearfaktoren ableiten. Wieder am Beispiel $a_{i+1} = (a_i - b_i)/2$, $b_{i+1} = b_i$ soll das Vorgehen veranschaulicht werden (a_i und b_i sind ungerade, $a_i \geq b_i$).

$$a_{i+1} = \frac{a_i - b_i}{2} = u_{i+1} a_0 + v_{i+1} b_0 \qquad b_{i+1} = b_i = s_{i+1} a_0 + t_{i+1} b_0$$
$$(u_i - s_i) a_0 + (v_i - t_i) b_0 = 2u_{i+1} a_0 + 2v_{i+1} b_0 \qquad s_i a_0 + t_i b_0 = s_{i+1} a_0 + t_{i+1} b_0$$

Vergleich von linker und rechter Seite läßt den Schluß zu:

$$u_{i+1} = \frac{u_i - s_i}{2} \qquad s_{i+1} = s_i$$
$$v_{i+1} = \frac{v_i - t_i}{2} \qquad t_{i+1} = t_i\ .$$

Die anderen Kombinationen können genau nach demselben Schema abgeleitet werden. Tabelle 4.3 faßt die Ergebnisse in übersichtlicher Form zusammen.[51]

[51] Der Fall, daß a_i und b_i gerade sind, kann ausgeschlossen werden, wenn dies auch für a_0 und b_0 vorausgesetzt wird (was man ohne Einschränkung kann, vgl. Anmerkungen zum binären GCD-Algorithmus auf Seite 37).

Tabelle 4.3: Linearfaktoren beim Algorithmus nach P

a_i	b_i	$a_{i+1} =$	$b_{i+1} =$	$u_{i+1} =$	$v_{i+1} =$	$s_{i+1} =$	$t_{i+1} =$
gerade	ungerade	$\frac{a_i}{2}$	b_i	$\frac{u_i}{2}$	$\frac{v_i}{2}$	s_i	t_i
ungerade	gerade	a_i	$\frac{b_i}{2}$	u_i	v_i	$\frac{s_i}{2}$	$\frac{t_i}{2}$
ungerade, $a_i \geq b_i$		$\frac{a_i - b_i}{2}$	b_i	$\frac{u_i - s_i}{2}$	$\frac{v_i - t_i}{2}$	s_i	t_i
ungerade, $b_i \geq a_i$		a_i	$\frac{b_i - a_i}{2}$	u_i	v_i	$\frac{s_i - u_i}{2}$	$\frac{t_i - v_i}{2}$

Dabei tritt allerdings wieder das bekannte Problem auf: Wie kann man die Ganzzahligkeit des jeweiligen Linearfaktoren bei der Halbierung wahren? Die Antwort haben wir schon beim vorangegangenen Algorithmus geliefert – indem die Ausgangsgleichungen 4.8 und 4.9 folgendermaßen erweitert:

$$a_i = (u_i \pm b_0)a_0 + (v_i \mp a_0)b_0$$
$$b_i = (s_i \pm b_0)a_0 + (t_i \mp a_0)b_0$$

und dadurch ungerade Zahlen u_i, v_i bzw. s_i, t_i in gerade umwandelt.

Beschränken wir uns auf die Kombinationen nach Tabelle 4.3, in denen u_i und v_i verändert werden (äquivalent für $u_i \to s_i$ und $v_i \to t_i$). Für den Fall, daß a_i gerade und b_i ungerade ist (zweite Zeile in 4.3), können wir auf die Argumentation von Seite 41 zurückgreifen (Phase II der Methode nach K). Sie erlaubt uns die Subtraktion/Addition von a_0 und b_0 für den Fall, daß u_i oder v_i ungerade ist. Die Situation, daß a_i und b_i ungerade sind (vorletzte Zeile in 4.3), kann man durch gedankliche Verzögerung der Halbierung in den nächsten Iterationsschritt erklären. Wählt man als modifizierten Einzelschritt $a_{i+1} = a_i - b_i$ (und entsprechend $u_{i+1} = u_i - s_i$, $v_{i+1} = v_i - t_i$), so reduziert sich die Fragestellung wieder auf eine gerade Zahl $a_{i+1} = u_{i+1}a_0 + v_{i+1}b_0$, also auf den vorangegangenen Fall.

Als Ergebnis der Ausführungen kann man Algorithmus 6 formulieren. Der Vorteil des Algorithmus (gegenüber K 's) liegt vor allem darin, daß keine Korrekturphase nötig ist. Nachteilig für eine praktische Umsetzung ist die notwendige Vorzeichen-Arithmetik.

4.2 Lineare diophantische Gleichungen

Gleichungen mit ausschließlich ganzzahligen Lösungen $x, y, \ldots \in \mathbb{Z}$ nennt man diophantisch, wobei sie im linearen Fall (für zwei Variablen) die folgende Form haben:

$$ax + by = c, \qquad \text{mit } a, b, c \in \mathbb{N}\,. \tag{4.10}$$

Algorithmus 6 Algorithmus $d = \gcd(a,b) = \alpha a + \beta b$ nach P

Require: b ungerade

$(a_0, u_0, v_0) \Leftarrow (a, 1, 0)$
$(b_0, s_0, t_0) \Leftarrow (b, 0, 1)$
$i \Leftarrow 0$

while $a_i > 0$ **do**
  if $b_i > a_i$ **then**
    $(b_i, s_i, t_i) \Longleftrightarrow (a_i, u_i, v_i)$ {gewährleistet $a_i \geq b_i$}
  end if
  if a_i ungerade $\wedge$ b_i ungerade **then**
    $(a_i, u_i, v_i) \Longleftarrow (a_i - b_i, u_i - s_i, v_i - t_i)$ {a_i ist jetzt gerade, b_i weiterhin ungerade}
  end if
  if a_i gerade **then**
    $a_{i+1} \Longleftarrow a_i/2$ {a_i gerade, b_i ungerade}
    if u_i gerade $\wedge$ v_i gerade **then**
      $u_{i+1} \Longleftarrow u_i/2$
      $v_{i+1} \Longleftarrow v_i/2$
    else
      $u_{i+1} \Longleftarrow (u_i + b)/2$
      $v_{i+1} \Longleftarrow (v_i - a)/2$
    end if
  else
    $b_{i+1} \Longleftarrow b_i/2$ {a_i ungerade, b_i gerade}
    if s_i gerade $\wedge$ t_i gerade **then**
      $s_{i+1} \Longleftarrow s_i/2$
      $t_{i+1} \Longleftarrow t_i/2$
    else
      $s_{i+1} \Longleftarrow (s_i + b)/2$
      $t_{i+1} \Longleftarrow (t_i - a)/2$
    end if
  end if
  $i \Leftarrow i + 1$
end while

$(d, \alpha, \beta) \Leftarrow (b_i, s_i, t_i)$ {$d = \gcd(a,b) = \alpha a + \beta b$}

Solche Gleichungen haben genau dann eine Lösung, wenn der größte gemeinsame Teiler $d = \gcd(a,b)$ den Wert c teilt. Im Zusammenhang mit E 's Algorithmus wurde dies praktisch schon nachgewiesen – auch daß es unendlich viele solcher Lösungen gibt, kam in Abschnitt 4.1.1 zur Sprache.

Zuerst bemerken wir, daß die Differenz zweier Lösungen (x_1,y_1) und (x_2,y_2) die homogene Gleichung $ax+by=0$ erfüllt.

$$(ax_1+by_1)-(ax_2+by_2) = a\underbrace{(x_1-x_2)}_{x}+b\underbrace{(y_1-y_2)}_{y} = 0$$

Mit $a = d\overline{a}$ und $b = d\overline{b}$ läßt sich sogar schreiben

$$\overline{a}x+\overline{b}y = 0$$

und es liegen die Lösungen $(x,y) = (n\overline{b},-n\overline{a})$, $n \in \mathbb{Z}$ auf der Hand (Einsetzen ergibt $\overline{a}x+\overline{b}y = \overline{a}n\overline{b}-\overline{b}n\overline{a} = 0$).[52] Findet man jetzt noch eine partikuläre Lösung (x_2,y_2), dann ergeben sich alle weiteren zu:

$$x_1 = x_2+x = x_2+n\overline{b} \qquad\qquad y_1 = y_2+y = y_2-n\overline{a}\,.$$

Partikuläre Lösungen (x_2,y_2) für $ax_2+by_2 = c$ haben wir aber schon mittels der erweiterten GCD-Algorithmen zur Verfügung, denn mit der B´ 's Identität (den Index 2 jetzt weggelassen)

$$\begin{aligned}\gcd(a,b) = d &= \alpha a+\beta b\\ d\overline{c} &= \alpha\overline{c}a+\beta\overline{c}b\end{aligned}$$

gilt:

$$ax+by = c = d\overline{c} = \alpha\overline{c}a+\beta\overline{c}b\,.$$

Eine partikuläre Lösung ($x_2 \stackrel{\wedge}{=} x$, $y_2 \stackrel{\wedge}{=} y$) kann deshalb mit Hilfe der B´ -Kofaktoren α,β gegeben werden.

$$x_2 = \alpha\overline{c} = \alpha\frac{c}{d} \qquad\qquad y_2 = \beta\overline{c} = \beta\frac{c}{d}$$

Die Vielfalt aller Lösungen stellt sich dadurch wiefolgt dar:

[52] Eine Untermenge der Lösungen stellt für $n = ld$, $l \in \mathbb{Z}$ natürlich $(x,y) = (lb,-la)$.

$$x_n = \alpha\frac{c}{d} + n\frac{b}{d} \qquad\qquad y_n = \beta\frac{c}{d} - n\frac{a}{d}\,. \tag{4.11}$$

Für den Spezialfall $d = \gcd(a,b) = 1$ entartet die Formel zu:

$$x_n = \alpha c + nb \qquad\qquad y_n = \beta c - na\,. \tag{4.12}$$

4.3 Chinesischer Restsatz

4.3.1 Hilfssatz für zwei Kongruenzen

Um sich dem Chinesischen Restsatz[53] zu nähern, betrachten wir zunächst einen etwas einfacheren Fall, der die grundsätzliche Fragestellung jedoch beinhaltet: Welche natürliche Zahl z erfüllt die folgenden beiden Kongruenzen:

$$z \equiv a \pmod{n} \qquad\qquad z \equiv b \pmod{m},$$

wenn vorausgesetzt wird, daß $n, m > 0$ relativ prim zueinander sind?

Um sie zu beantworten formulieren wir den Ausgangspunkt zuerst einmal entsprechend Restklassenbeziehung 3.2:

$$z = a + xn = b + ym, \qquad x, y \in \mathbb{N}\,. \tag{4.13}$$

Deren Darstellung als

$$c = b - a = xn - ym \tag{4.14}$$

zeigt mit Verweis auf die Form von 4.10, daß es sich um eine lineare diophantische Gleichung handelt. Wegen $\gcd(n,m) = 1$ könnte die zugehörige Lösungsformel 4.12 zwar sofort zur Anwendung kommen – naheliegend (da kurz) ist jedoch auch die Anwendung von B´ 's Identität. Denn multipliziert man $\gcd(n,m) = \alpha n + \beta m = 1$ mit c, dann kann aus

$$c = c(\alpha n + \beta m) = \alpha nc + \beta mc$$

[53] CRT – Chinese Remainder Theorem

durch Vergleich mit 4.14 sofort abgelesen werden, daß $x = \alpha c$ und $y = -\beta c$ gelten muß.[54] Mit der Erkenntnis aus Formel 4.12, daß es sich bei den Lösungen einer linearen diophantischen Gleichungen immer um eine ganze Lösungsmenge handelt, resultiert:[55]

$$x_k = \alpha c + km \qquad\qquad y_k = -\beta c - kn, \qquad k \in \mathbb{Z}\,.$$

Durch Einsetzen in Formel 4.13 erhält man

$$\begin{aligned} z_k &= a + x_k n \\ &= a + (\alpha c - km)n \\ &= a \underbrace{(1 - \alpha n)}_{\beta m} + \alpha bn - kmn \end{aligned} \qquad\qquad \begin{aligned} z_k &= b + y_k m \\ &= b - (\beta c + kn)m \\ &= b \underbrace{(1 - \beta m)}_{\alpha n} + \beta am - kmn \end{aligned} \tag{4.15}$$

und so eine geschlossene Darstellung für die Lösungsmenge.

$$z_k = \alpha bn + \beta am + kmn \tag{4.16}$$

Mit der von den Restklassen bekannten Kongruenz 3.3 für teilerfremde Zahlen:[56]

$$1 \equiv \beta m \pmod{n} \qquad\qquad 1 \equiv \alpha n \pmod{m}$$

können wir die Probe machen.

$$\begin{aligned} z_k \bmod n &= \underbrace{(\alpha b + km)n}_{0} + \beta am \bmod n \\ &= \beta am \bmod n = a \end{aligned} \qquad\qquad \begin{aligned} z_k \bmod m &= \alpha bn + \underbrace{(\beta a + kn)m}_{0} \bmod m \\ &= \alpha bn \bmod m = b \end{aligned}$$

Die Lösungsmenge z_k bildet demzufolge eine Restklasse $[z]_{mn}$.

$$z \equiv \alpha bn + \beta am \pmod{mn} \tag{4.17}$$

[54] H. L. G hat genau diesen Lösungsansatz für eine unbeschränkte Anzahl von Kongruenzen ausgebaut [Gar59].
[55] Die eigentliche Ursache für die Vielfalt von Lösungen ist im Hinweis zum euklidischen Algorithmus auf Seite 32 begründet.
[56] Hierbei wird ganz deutlich, daß es sich bei α und β um Inverse handelt: $\alpha \equiv n^{-1} \pmod{m}$, $\beta \equiv m^{-1} \pmod{n}$.

4.3.2 Ein System von Kongruenzen

Nehmen wir jetzt ein ganzes System von Kongruenzen an:

$$\begin{aligned} x &\equiv a_1 \pmod{m_1} \\ x &\equiv a_2 \pmod{m_2} \\ &\vdots \\ x &\equiv a_n \pmod{m_n}, \end{aligned}$$

wobei $\gcd(m_i, m_k) = 1$ für $i \neq k$ gelten soll. Wie schon im Fall von zwei Kongruenzen stellt man die Frage, welche Lösung x kongruent zu allen a_i modulo m_i ist ($i = 1, \ldots, n$). Ohne einen exakten mathematischen Beweis anzutreten, scheint als Schlußfolgerung aus Abschnitt 4.3.1 einleuchtend, daß die Lösung(en) x eine Restklasse modulo $m = \prod_n m_i$ darstellen [Knu98, CP05].[57]

Bezeichnen wir mit $\overline{m}_i$ das Produkt aller Moduli ausgeschlossen m_i, also

$$\overline{m}_i = \frac{m}{m_i} = \frac{\prod_{j=1}^{n} m_j}{m_i} = \prod_{\substack{j=1 \\ j \neq i}}^{n} m_j,$$

dann gilt wegen der Teilerfremdheit der einzelnen Moduli $\gcd(\overline{m}_i, m_i) = 1$ bzw. mit dem Satz von B´ :[58]

$$\begin{aligned} \alpha_i \overline{m}_i + \beta_i m_i &= 1 \\ \alpha_i \overline{m}_i \bmod m_i &= 1\,. \end{aligned}$$

Im Gegensatz dazu verschwindet $\alpha_k \overline{m}_k \bmod m_i$ für $i \neq k$, denn m_i ist als Faktor in $\overline{m}_k$ enthalten. Die Orthogonalität beider Fälle kann mit Hilfe des K -Symbols δ_{ik} ausgedrückt werden:

$$\alpha_k \overline{m}_k \bmod m_i = \delta_{ik} = \begin{cases} 0 & (i \neq k) \\ 1 & (i = k) \end{cases}.$$

Die endgültige Lösungsidee besteht nun darin, für jede Kongruenz i den folgenden Ausdruck zu bilden:

[57] Da jedes $x - a_i$ ein Vielfaches des zugehörigen m_i sein muß, nach Voraussetzung aber alle m_i relativ prim zueinander sind, ist das kleinste gemeinsame Vielfache von $m_0, m_1, \ldots, m_n$ genau das Produkt $m = \prod_n m_i$.

[58] Der B´ -Koeffizient α_i kann wieder als Inverses aufgefaßt werden: $\alpha_i \equiv \overline{m}_i^{-1} \pmod{m_i}$.

$$\sum_{k=1}^{n} \alpha_k a_k \overline{m}_k \bmod m_i = \sum_{k=1}^{n} a_k \delta_{ik} \bmod m_i = a_i \bmod m_i \equiv x \pmod{m_i}$$

Wie zu sehen, ist die Summe auf der linken Seite kongruent zu a_i modulo m_i, was

$$x \equiv \sum_{k=1}^{n} \alpha_k a_k \overline{m}_k \pmod{m} \tag{4.18}$$

als finales Ergebnis rechtfertigt.

Für den einfachen Fall $n = 2$ von Abschnitt 4.3.1 kann man mit $\overline{m}_1 = m_2$, $\overline{m}_2 = m_1$ sowie

$$\begin{aligned} 1 &= \alpha_1 \overline{m}_1 + \beta_1 m_1 \\ &= \alpha_1 m_2 + \beta_1 m_1 \end{aligned} \qquad \begin{aligned} 1 &= \alpha_2 \overline{m}_2 + \beta_2 m_2 \\ &= \alpha_2 m_1 + \beta_2 m_2 \end{aligned}$$

$$\alpha_1 = \beta_2 \qquad \alpha_2 = \beta_1$$

relativ einfach Kongruenz 4.17 verifizieren.[59]

$$x \equiv \alpha_1 a_1 \overline{m}_1 + \alpha_2 a_2 \overline{m}_2 = \alpha_1 a_1 m_2 + \alpha_2 a_2 m_1 = \alpha_1 a_1 m_2 + \beta_1 a_2 m_1 \pmod{m_1 m_2}$$

4.4 Quadratwurzeln in $\mathbb{F}_p$

4.4.1 Vorbetrachtungen

Das Bestimmen der Quadratwurzel kann man im Körper $\mathbb{F}_p$ recht effizient mit dem T - S Algorithmus erledigen. Bevor man aber dazu übergehen kann den Algorithmus zu erläutern, muß man sich unbedingt über einige Fakten in Bezug auf das Quadrieren von Körperelementen klar werden. Ohne wesentlich an Allgemeinheit zu verlieren, betrachten wird dazu die multiplikative Gruppe $\mathbb{Z}_p^* = (\mathbb{Z}_p \setminus \{0\})$ und formulieren zuerst die folgenden Grundaussagen in Bezug auf:[60]

$$y = x^2 \pmod{p}$$

1. Aus der Beziehung $y = x^2 \pmod{p}$ ersieht man sofort, daß immer zwei Elemente x, x' genau zum selben y führen – ein „Positives“ und ein „Negatives“: $x' \equiv -x \pmod{p}$.

[59] Ein echtes Rechenbeispiel für $n = 3$ kann man z. B. in [CP05, 2.1.3] finden.
[60] Siehe z. B. auch [Ros99, 9.1] oder [NZM91, 3.1].

2. Da $\mathbb{Z}_p^*$ genau $p-1$ Elemente enthält (besitzt die Ordnung $|\mathbb{Z}_p^*| = |\mathbb{Z}_p| - 1 = p-1$), können $(p-1)/2$ Elemente $y \in \mathbb{Z}_p^*$ keine Quadratwurzel besitzen. Denn umgekehrt betrachtet sind ja die verfügbaren $p-1$ Elemente x wegen der Paarbildung beim Quadrieren (welche zu $(p-1)/2$ Elementen y führt) schon „aufgebraucht“.

3. Für die $(p-1)/2$ Elemente y, welche der Beziehung $y = x^2 \pmod p$ genügen, gilt wegen des kleinen Satzes von F (vgl. Abschnitt 3.3.2):

$$y^{\frac{p-1}{2}} = (x^2)^{\frac{p-1}{2}} = x^{p-1} \equiv 1 \pmod p . \tag{4.19}$$

4. Für die restlichen $(p-1)/2$ Elemente y, welche keine zugeordnete Quadratwurzel x besitzen, gilt hingegen:[61]

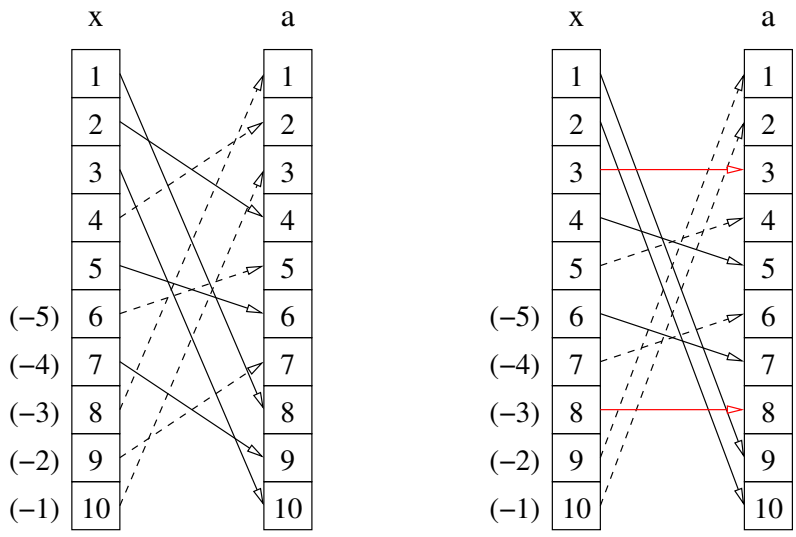
$$y^{\frac{p-1}{2}} \equiv -1 \pmod p . \tag{4.20}$$

Für einen kurzen Beweis nehmen wir uns ein bestimmtes (konstantes) Element y heraus und betrachten die Zerlegung $y = ax \pmod p$, welche es in $\mathbb{Z}_p^*$ immer geben muß: $a = x^{-1}y \pmod p$. Wenn man nun für x alle möglichen Werte $1 \ldots p-1$ setzt, dann muß auch a entsprechend variieren. Wegen:

- der Abgeschlossenheit der Multiplikation;
- der Eindeutigkeit des Inversen;
- und der Kommutativität von Elementen

in $\mathbb{Z}_p^*$ wird a einerseits alle möglichen Werte $1 \ldots p-1$ annehmen und andererseits wird jedes Produkt ax insofern doppelt vorkommen, daß a und x nur die Rolle getauscht haben. In Abbildung 4.1 sind mögliche Produkte für $y = 8$ (hat keine Wurzel) und $y = 9$ (Wurzeln sind ± 3) beispielhaft in $\mathbb{Z}_{11}^*$ dargestellt (wobei gestrichelte Linien Produkte kennzeichnen, bei denen a und x nur die Rolle getauscht haben).

[61] Äquivalent ist die folgende Schreibweise: $y^{(p-1)/2} \equiv p-1 \pmod p$.



(a) $y = 8$ (b) $y = (\pm 3)^2 = 9$

Abbildung 4.1: Beispiele für Produkte in $\mathbb{Z}_{11}^*$

Multiplizieren wir alle Produkte ax miteinander und schließen dabei die Doppelungen aus (d. h. berücksichtigen nur die Produkte in Abbildung 4.1a, welche mit durchgezogenen Linien dargestellt sind), dann hat man alle Elemente $x \in \mathbb{Z}_p^*$ miteinander multipliziert.

$$\prod_{i=1}^{\frac{p-1}{2}} a_i x_i = \prod_{i=1}^{\frac{p-1}{2}} a_i \cdot \prod_{i=1}^{\frac{p-1}{2}} x_i = \prod_{i=1}^{p-1} x_i = (p-1)!$$

Bei den $(p-1)/2$ Produkten $a_i x_i$ handelt es sich aber genau um die Potenz

$$y^{\frac{p-1}{2}} = \prod_{i=1}^{\frac{p-1}{2}} y_i = \prod_{i=1}^{\frac{p-1}{2}} a_i x_i$$

und so kann man mit dem Satz von W (siehe Formel 3.7) den Beweis abschließen:

$$y^{\frac{p-1}{2}} = (p-1)! \equiv -1 \ .$$

Noch bevor man also versucht die Quadratwurzel $x = \sqrt{y}$ zu berechnen, kann man mit Hilfe des gerade eingeführten (E -) Kriteriums prüfen, ob diese überhaupt existiert. Die Nomenklatur für das E -Kriterium stammt allerdings von A.-M. L und wird als L -Symbol bezeichnet:

$$\left(\frac{y}{p}\right) = y^{\frac{p-1}{2}} = \begin{cases} 1 & \text{wenn } y \text{ eine Quadratwurzel in } \mathbb{Z}_p \text{ besitzt;} \\ -1 & \text{wenn } y \text{ keine Quadratwurzel in } \mathbb{Z}_p \text{ besitzt.} \end{cases} \tag{4.21}$$

4.4.2 Der Spezialfall $p \bmod 4 = 3$

Im speziellen Fall $p \equiv 3 \pmod 4$ kann man eine sehr einfache Lösung für die Quadratwurzel $x = \sqrt{y}$ erhalten, sofern sie denn existiert [Ros99, Theorem 9.3], [ANS05, D.1.4]:

$$x = y^{s+1} \pmod p, \tag{4.22}$$

mit $s = (p-3)/4$.

Ausgangspunkt für einen kurzen Beweis könnte die folgende Darstellung für das Modul p sein:[62]

$$p = 4s+3 = 2(2s+1)+1$$
$$\frac{p-1}{2} = 2s+1\,.$$

Wird sie im Zusammenhang mit dem E -Kriterium von Gleichung 4.19 angewendet, dann bestätigt sich Formel 4.22 wiefolgt:

$$x^2 = y^{2(s+1)} = y \cdot y^{2s+1} = y \cdot y^{\frac{p-1}{2}} \equiv y \pmod p\,.$$

4.4.3 Der T -S Algorithmus

Idee Der T -S Algorithmus nach [Ton91, Sha73] reduziert, ausgehend von einem Anfangswert $x_0 \in \mathbb{Z}_{p>2}$, den Exponenten 2^{t_i} im Ausdruck

$$\left(\frac{x_i^2}{y}\right)^{2^{t_i}} \equiv 1 \pmod p \tag{4.23}$$

immer weiter, bis schließlich $t_n = 0$ wird und deshalb $x_n^2 \equiv y \pmod p$ die gesuchte Lösung darstellt (siehe z. B. auch [NZM91, 2.9], [CP05, 2.3.2]). Mit 2^{t_i} soll es sich aber nicht um irgendeine Zweierpotenz handeln, sondern um die Ordnung des Elements $z_i \equiv x_i^2 y^{-1} \pmod p$.

Es gilt also einen Algorithmus zu beschreiben, bei dem sowohl für den Startwert z_0 als auch für alle Werte z_i in

[62] In einer binären Darstellung für p gilt so: Bit 1 = 1 (auch Bit 0 = 1 gilt, da $p > 2$ als Primzahl ungerade ist).

$$z_i^{2^{t_i}} \equiv 1 \pmod{p} \tag{4.24}$$

die Ordnung $|z_i| = 2^{t_i}$ einerseits (und überhaupt) eine Zweierpotenz ist und andererseits (auch noch) abnimmt.[63]

Startwerte Als Anfangswerte setzt man:

$$\begin{aligned} x_0 &= y^{\frac{s+1}{2}} \bmod p & t_0' &= r-1 \\ x_0^2 &\equiv y^{s+1} \bmod p & 2^{t_0'} &= 2^{r-1} \\ (z_0 &\equiv y^s \bmod p)\,. \end{aligned}$$

Dabei entspricht der Wert $2^{t_0'}$ nicht zwingendermaßen der Ordnung $|z_0| = 2^{t_0}$, er befriedigt aber Kongruenz 4.24 grundsätzlich:

1. Da es sich bei p um eine Primzahl mit $p > 2$ handeln soll (p ist zwingendermaßen ungerade), kann man

   $$p - 1 = 2^r s \tag{4.25}$$

   setzen (mit $r \geq 1$, maximal) und erhält so:

   $$\frac{p-1}{2} = 2^{r-1} s\,.$$

2. Soll die Quadratwurzel $x = \sqrt{y}$ wirklich existieren (wovon wir ausgehen) dann gilt mit dem E -Kriterium 4.21:

   $$\left(\frac{y}{p}\right) = y^{\frac{p-1}{2}} \equiv 1 \pmod{p}\,.$$

3. Faßt man beide Punkte zusammen, dann bestätigt sich:

   $$z_0^{2^{t_0'}} = \left(\frac{x_0^2}{y}\right)^{2^{r-1}} = \left(\frac{y^{s+1}}{y}\right)^{2^{r-1}} \equiv y^{2^{r-1}s} \equiv y^{\frac{p-1}{2}} \equiv 1 \pmod{p}\,. \tag{4.26}$$

Interpretieren wir Kongruenz 4.26 jetzt mit dem Wissen aus Abschnitt 2.1.2, dann muß es sich bei $2^{t_0'}$ entweder um die Ordnung $|z_0|$ selbst oder aber um ein Vielfaches davon handeln: $2^{t_0'} = K|z_0|$. Deshalb können K und $|z_0|$ nur Zweierpotenzen sein (was $K = 1$ nicht ausschließt), d. h. für die Ordnung kann man ohne weitere Bedenken $|z_0| = 2^{t_0}$ ansetzen (mit $t_0 \leq t_0'$). Um nun aus t_0' den exakten Exponenten t_0 bzw. die Ordnung $|z_0|$ zu ermitteln, kann man z_0 einfach so oft quadrieren bis (nach F 's kleinem Satz) die Bedingung $z_0^{|z_0|} \equiv 1 \pmod{p}$ erfüllt ist.

[63] Im weiteren soll mit t_i' ein Exponent bezeichnet sein, bei dem $2^{t_i'}$ entweder die Ordnung des Elements z_i selbst oder aber ein Vielfaches dessen darstellt.

Iteration Bevor wir zum iterativen Teil des Algorithmus übergehen können, benötigen wir noch die Mithilfe irgendeines (zufällig gewählten) Elements $\alpha \in \mathbb{Z}_p$, welches keine Wurzel in $\mathbb{Z}_p^*$ haben darf. Für α muß das E -Kriterium 4.21 dann den Wert

$$\left(\frac{\alpha}{p}\right) \equiv -1 \pmod{p}$$

annehmen. Aus diesem Element α erzeugen wir ein neues Element $\beta = \alpha^s \pmod{p}$ mit den Eigenschaften:

- potenziert zu 2^r ergibt sich mit F 's kleinem Satz:

$$\beta^{2^r} = \alpha^{2^r s} = \alpha^{p-1} \equiv 1 \pmod{p};$$

- aus dem E -Kriterium für α läßt sich schlußfolgern:

$$\beta^{2^{r-1}} = \alpha^{2^{r-1} s} = \alpha^{\frac{p-1}{2}} \equiv -1 \pmod{p}\,. \tag{4.27}$$

Nun wird die Näherung schrittweise nach folgender Vorschrift verbessert:[64]

$$x_{i+1} = x_i \beta^{2^{r-t_i-1}} \pmod{p}, \tag{4.28}$$

wobei sich der Exponent t_i jeweils um 1 verringert ($t'_{i+1} = t_i - 1$, mit $2^{t'_{i+1}}$ einem mglw. Vielfachen der Ordnung $|z_{i+1}|$).

Um die Verringerung des Exponenten nachzuweisen bilden wir einfach $z_{i+1}^{2^{t_i-1}}$ mit Hilfe von Iterationsformel 4.28:

$$z_{i+1}^{2^{t'_{i+1}}} = \left(x_{i+1}^2 y^{-1}\right)^{2^{t'_{i+1}}} = \left(x_i^2 \beta^{2 \cdot 2^{r-t_i-1}} y^{-1}\right)^{2^{t_i-1}} = \left(x_i^2 y^{-1}\right)^{2^{t_i-1}} \left(\beta^{2^{r-t_i}}\right)^{2^{t_i-1}} = z_i^{2^{t_i-1}} \beta^{2^{r-1}} \pmod{p}\,.$$

Beide Faktoren auf der rechten Seite sind aber vom Wert $-1 \pmod{p}$, was letztlich

$$z_{i+1}^{2^{t'_{i+1}}} \equiv 1 \pmod{p}$$

bedeutet. Zur Begründung folgende Argumentation:

- $\beta^{2^{r-1}} \equiv -1 \pmod{p}$ gilt wegen des E -Kriteriums für α (siehe Kongruenz 4.27).

[64] Wobei maximal $\log_2(p)$ Iterationen nötig sind (die Komplexität des Algorithmus wird z. B. in [Knu98] mit $O(\log_2{}^3 p)$ angegeben).

- Der Beweis für $z_i^{2^{t_i-1}} \equiv -1 \pmod{p}$ stützt sich auf die Darstellung:

$$z_i^{2^{t_i}} = \left(z_i^{2^{t_i-1}}\right)^2,$$

  welche ja bei jeder Iteration als $+1 \pmod{p}$ vorausgesetzt wird. Aus diesem Grund kann der Wert innerhalb des Klammerausdrucks nur $+1$ oder -1 sein [NZM91, Hilfssatz 2.10]. Wenn 2^{t_i} aber die Ordnung von z_i derstellt, dann muß t_i der kleinste Wert sein, welcher zu $z_i^{2^{t_i}} \equiv +1 \pmod{p}$ führt. Deshalb kann 2^{t_i-1} nicht (ebenfalls) die Ordnung von z_i sein und es bleibt nur die Schlußfolgerung: $z_i^{2^{t_i-1}} \equiv -1 \pmod{p}$.

Aus dem letzten Punkt ergibt sich die Notwendigkeit, daß man am Anfang des Iterationszyklus jedesmal die exakte Ordnung $|z_i|$ bzw. den zugehörigen Exponenten $t_i \leq t_i'$ bestimmt (vgl. auch die Erläuterungen zu den Startwerten).

Algorithmus Bevor wir in Algorithmus 7 alle Erkenntnisse zusammenfassen, noch zwei Hinweise zur Umsetzung:

1. Die Berechnung des inversen Elements y^{-1} kann man einmalig zu Beginn des Algorithmus durchführen.
2. Setzt man für die Reduktion der Ordnung in jedem Schritt $\Delta_i = t_{i-1} - t_i$ und führt in Formel 4.28 die Abkürzungen

$$\gamma_i = \beta^{2^{r-t_i-1}} \qquad\qquad x_{i+1} = x_i\gamma_i \pmod{p}$$

   ein, dann kann man auch γ_i rekursiv berechnen:

$$\gamma_i = \beta^{2^{r-t_{i-1}-1+\Delta_i}} = \left(\beta^{2^{r-t_{i-1}-1}}\right)^{2^{\Delta_i}} = \gamma_{i-1}^{2^{\Delta_i}}.$$

   Um den korrekten Startwert $\gamma_0 = \beta^{2^{r-t_0-1}}$ zu erzielen, vergleichen wir jetzt noch die Werte für $i = 0$:

$$\beta^{2^{r-1-t_0}} = \gamma_{-1}^{2^{\Delta_0}} = \gamma_{-1}^{2^{t_{-1}-t_0}}$$

   und erhalten:

$$t_{-1} = r - 1 \qquad\qquad \gamma_{-1} = \beta\,.$$

Algorithmus 7 Quadratwurzel eines Elements $y \in \mathbb{Z}_p$

Require: $p > 2$
Ensure: $x = \sqrt{y}$

$r, s \Leftarrow$ aus $p - 1 = 2^r s$, mit r maximal
repeat
  $\alpha \Leftarrow$ zufälliges Element aus $\mathbb{Z}_p$
until $\alpha^{(p-1)/2} \bmod p = p - 1$ {E -Kriterium: $\alpha^{(p-1)/2} \equiv -1 \pmod p$}

$\gamma \Leftarrow \alpha^s \bmod p$ $\{\gamma_{-1} = \beta\}$
$x \Leftarrow y^{(s+1)/2} \bmod p$ $\{x_0\}$
$y_{\text{inv}} = y^{-1} \bmod p$
$t' \Leftarrow r - 1$
$t \Leftarrow$ aus $2^t = \text{ord}(x^2 y_{\text{inv}} \bmod p)$ {Ordnung von z_0}

while $t > 0$ **do**
  $\Delta \Leftarrow t' - t$
  $\gamma \Leftarrow \gamma^{2^\Delta} \bmod p$ $\{\gamma_i = \beta^{2^{r-t_i-1}}\}$
  $x \Leftarrow x\gamma \bmod p$
  $t' \Leftarrow t$
  t aus $2^t = \text{ord}(x^2 y_{\text{inv}} \bmod p)$ {Ordnung von $z_i \equiv x_i^2 y^{-1} \pmod p$}
end while

{Probe (siehe auch [DM11, Appendix C]),
anstatt zu Beginn das E -Kriterium $(y|p) = 1$ zu verifizieren}
$y' \Leftarrow x^2 \bmod p$
if $y' = y$ **then**
  return y
else
  return keine Lösung
end if

4.5 Quadratische Gleichungen in $\mathbb{F}_{2^n}$

4.5.1 Problemstellung

Quadratische Gleichungen (mit $a \neq 0$) der Form

$$ay^2 + by + c = 0, \qquad a, b, c \in \mathbb{F}_{2^n}$$

kann man bezüglich y in $\mathbb{F}_{2^n}$ nicht so einfach lösen wie in $\mathbb{R}$. Wie noch zu sehen sein wird, hat diese Gleichung in $\mathbb{F}_{2^n}$:

- möglicherweise keine Lösung;

- genau eine Lösung für den Fall $b = 0$;
- oder sogar zwei Lösungen.

Zuersteinmal kann man sie unter der Voraussetzung $a, b \neq 0$ mit Hilfe der Substitution $z = ay/b$ und der naheliegenden Abkürzung $\beta = ac/b^2$ in eine allgemeine Form überführen:

$$\begin{aligned} \frac{b^2}{a}z^2 + \frac{b^2}{a}z + c &= 0 \\ z^2 + z + \frac{a}{b^2}c &= 0 \\ z^2 + z + \beta &= 0\,. \end{aligned}$$

Davon ausgehend können wir uns den mathematischen Grundlagen einer Lösung zuwenden [Ber84, Sho05].

4.5.2 Trace

Zuallererst soll eine Hilfsformel eingeführt werden, welche in der einen oder anderen Form immer wieder benötigt wird.[65] Dazu sei die folgende Summe λ_i für ein Element $\alpha \in \mathbb{F}_{p^n}$ definiert und mit $i \in \mathbb{N}$ indiziert (vorerst soll $i \neq 0$ gelten):

$$\lambda_i = \sum_{k=0}^{i-1} \alpha^{p^k}$$

Dann berechnen wir unter Zuhilfenahme des „Anfänger-Traums“ nach Formel 3.13 bzw. 3.14 die p-te Potenz:

$$\lambda_i^p = \left[\sum_{k=0}^{i-1} \alpha^{p^k}\right]^p = \sum_{k=0}^{i-1} \left(\alpha^{p^k}\right)^p = \sum_{k=0}^{i-1} \alpha^{p^{k+1}} = \sum_{k=1}^{i} \alpha^{p^k} = \alpha^{p^i} - \alpha + \sum_{k=0}^{i-1} \alpha^{p^k}$$

und stellen fest:

$$\lambda_i^p - \lambda_i = \alpha^{p^i} - \alpha, \text{ mit } \lambda_i = \sum_{k=0}^{i-1} \alpha^{p^k}\,. \tag{4.29}$$

[65] Interessant für die Lösung des Problems der quadratischen Gleichung in $\mathbb{F}_{2^n}$ ist eigentlich der Fall $p = 2$ bzw. $\alpha \in \mathbb{F}_{2^n}$, trotzdem wird der Trace hier zuerst allgemein eingeführt.

Definition Der Trace ist die Summe der Konjugierten eines Elements im Körper $\mathbb{F}_{p^n}$.

$$\mathrm{tr}(\alpha) = \sum_{i=0}^{n-1} \alpha^{p^i} = \alpha + \alpha^p + \alpha^{p^2} + \alpha^{p^3} + \cdots + \alpha^{p^{n-1}}$$

Mit Formel 4.29 ist der Trace auch darstellbar als:

$$\mathrm{tr}(\alpha) = \lambda_n \,. \tag{4.30}$$

Eigenschaften Mit Hilfe der Polynomdarstellung für ein Element $\alpha \in \mathbb{F}_{p^n}$

$$\alpha = \sum_{j=0}^{n-1} \alpha_j x^j \qquad (\alpha_j \in \mathbb{F}_p)$$

kann man einige interessante Eigenschaften und Beziehungen für den Trace ableiten:

1. Aus den Formeln 4.29 und 4.30 ergeben sich mit dem kleinen Satz von F $(\alpha^{p^n} - \alpha = 0)$ die Beziehungen:

$$\begin{aligned} \lambda_n^p &= \lambda_n, \text{ mit } \lambda_n = \mathrm{tr}(\alpha) \\ \mathrm{tr}^p(\alpha) &= \mathrm{tr}(\alpha), \end{aligned} \tag{4.31}$$

   d. h. der Trace ist eine lineare Abbildung $\mathbb{F}_{p^n} \to \mathbb{F}_p$. Der Grund liegt ganz einfach darin, daß Gleichung 4.31 nur im Körper $\mathbb{F}_p$ Gültigkeit hat (nicht in $\mathbb{F}_{p^n}$) – und deshalb: $\mathrm{tr}(\alpha) \in \mathbb{F}_p$.

2. Der Trace eines Elements wird nicht verändert, wenn man das Element vorher (mehrfach) zur Potenz p potenziert:

$$\begin{aligned} \mathrm{tr}(\alpha^p) &= \mathrm{tr}(\alpha) \\ \mathrm{tr}(\alpha^{p^i}) &= \mathrm{tr}(\alpha) \end{aligned} \tag{4.32}$$

   Warum dies so ist, erkennt man durch Anwendung der im Punkt 1 hergeleiteten Formel 4.31 sowie (wieder) des „Anfänger-Traums“ nach Formel 3.14:

$$\mathrm{tr}(\alpha) = \mathrm{tr}^p(\alpha) = \left[\sum_{i=0}^{n-1} \alpha^{p^i}\right]^p = \sum_{i=0}^{n-1} \left(\alpha^{p^i}\right)^p = \sum_{i=0}^{n-1} (\alpha^p)^{p^i} = \mathrm{tr}(\alpha^p) \,.$$

3. Der Trace ist linear:

a) $\mathrm{tr}(\alpha+\beta) = \mathrm{tr}(\alpha) + \mathrm{tr}(\beta)$, mit $\alpha,\beta \in \mathbb{F}_{p^n}$
In ausführlicher Schreibweise des Trace und (wieder) mit Formel 3.14 kann man auch dies schnell beweisen.

$$\mathrm{tr}(\alpha+\beta) = \sum_{i=0}^{n-1}(\alpha+\beta)^{p^i} = \sum_{i=0}^{n-1}(\alpha^{p^i}+\beta^{p^i}) = \sum_{i=0}^{n-1}\alpha^{p^i} + \sum_{i=0}^{n-1}\beta^{p^i} = \mathrm{tr}(\alpha)+\mathrm{tr}(\beta)$$

b) $\mathrm{tr}(c\alpha) = c\,\mathrm{tr}(\alpha)$, mit $c \in \mathbb{F}_p$, $\alpha \in \mathbb{F}_{p^n}$
Ausklammern von $c = c^p = c^{p^i}$ aus der Summendarstellung des Trace führt zu:

$$\mathrm{tr}(c\alpha) = \sum_{i=0}^{n-1}(c\alpha)^{p^i} = \sum_{i=0}^{n-1}\left(c^{p^i}\alpha^{p^i}\right) = \sum_{i=0}^{n-1}\left(c\alpha^{p^i}\right) = c\,\mathrm{tr}(\alpha)\,.$$

Im Körper $\mathbb{F}_{2^n}$ Speziell im Binärkörper $\mathbb{F}_{2^n}$ entsteht aus Formel 4.29:

$$\lambda_i^2 + \lambda_i = \alpha^{2^i} + \alpha \tag{4.33}$$

und mit der Rekursion $\lambda_i = \lambda_{i-1} + \alpha^{2^{i-1}}$ die (insbesondere für einen Berechnungsalgorithmus) interessante Beziehung:[66]

$$\lambda_i = \lambda_{i-1} + \alpha^{2^{i-1}} = \lambda_{i-1}^2 + \alpha\,. \tag{4.34}$$

Insbesondere aus den Punkten 1 und 2 auf Seite 60 ergeben sich weitere spezialisierte Eigenschaften:

1. Quadrieren des Arguments α in $\mathrm{tr}(\alpha)$ ändert nach Formel 4.32 nichts am Wert des Trace.

$$\mathrm{tr}(\alpha) = \mathrm{tr}(\alpha^2) = \mathrm{tr}(\alpha^{2^k}), \qquad k \in \mathbb{N} \tag{4.35}$$

2. Das Potenzieren eines Trace ändert ebenfalls nichts an seinem Wert.[67]

$$\mathrm{tr}^i(\alpha) = \mathrm{tr}(\alpha), \qquad i \in \mathbb{N} \tag{4.36}$$

Grundlage des Beweises bilden die folgenden beiden Reduktionsfälle:

$$\mathrm{tr}^i(\alpha) = \begin{cases} \mathrm{tr}^{i/2}(\alpha) & (i \text{ gerade}) \\ \mathrm{tr}^{(i+1)/2}(\alpha) & (i \text{ ungerade})\,. \end{cases}$$

Sie erlauben es, den Exponenten iterativ bis auf 1 zu reduzieren, was letztlich zu $\mathrm{tr}^i(\alpha) = \mathrm{tr}(\alpha)$ führt.

[66] Ausgehend von $\lambda_1 = \alpha$ kann man damit auch einen „Pseudowert“ $\lambda_0 = \lambda_0^2 = \lambda_1 + \alpha = 0$ festlegen.

[67] Diese Aussage ergibt sich schon logisch aus $\mathrm{tr}(\alpha) \in \mathbb{F}_2$ bzw. $\mathrm{tr}(\alpha) = \{0,1\}$. Im Gegensatz dazu gilt für $i \neq 2^k$ $(k \geq 1)$: $\mathrm{tr}(\alpha^i) \neq \mathrm{tr}(\alpha)$, was in $\alpha \in \mathbb{F}_{2^n}$ begründet ist.

In einer kurzen Fallbetrachtung für i gerade:

$$\mathrm{tr}^i(\alpha) = [\mathrm{tr}^{i/2}(\alpha)]^2 = \mathrm{tr}^{i/2}(\alpha)$$

bzw. i ungerade (und so $i-1$ gerade):

$$\mathrm{tr}^i(\alpha) = \mathrm{tr}(\alpha)\,\mathrm{tr}^{i-1}(\alpha) = \mathrm{tr}(\alpha)[\mathrm{tr}^{(i-1)/2}(\alpha)]^2 = \mathrm{tr}(\alpha)\,\mathrm{tr}^{(i-1)/2}(\alpha) = \mathrm{tr}^{(i+1)/2}(\alpha)$$

kann man die Begründung finden.

3. Aus der Linearität des Trace und aus Formel 4.32 ergibt sich direkt:

$$\mathrm{tr}(\alpha^2+\alpha+\beta) = \underbrace{\mathrm{tr}(\alpha^2)+\mathrm{tr}(\alpha)}_{0}+\mathrm{tr}(\beta) = \mathrm{tr}(\beta)\,.$$

4. Eine bemerkenswerte Formel, insbesondere im Zusammenhang mit der Lösung quadratischer Gleichungen in $\mathbb{F}_{2^n}$, ist die folgende:

$$\alpha\,\mathrm{tr}(\beta)+\beta\,\mathrm{tr}(\alpha) = \gamma^2+\gamma,\ \text{mit}\ \gamma = \sum_{i=1}^{n-1}\beta^{2^i}\lambda_i\ \text{und}\ \lambda_i = \sum_{k=0}^{i-1}\alpha^{2^k}\,. \tag{4.37}$$

Ihr Beweis startet, indem man auf die äußere Summe den „Anfänger-Traum“ anwendet und danach mit $i := i-1$ umindiziert. Mittels Formel 4.34 wird im Anschluß λ_{i-1}^2 substituiert.

$$\gamma^2 = \sum_{i=1}^{n-1}\beta^{2^{i+1}}\lambda_i^2 = \sum_{i=2}^{n}\beta^{2^i}\lambda_{i-1}^2 = \sum_{i=2}^{n}\beta^{2^i}(\lambda_i+\alpha)$$

Läßt man die Summe bei $i=1$ starten, dann ist zwar eine Korrektur um $\beta^2(\lambda_1+\alpha)$ vonnöten, letztlich ändert sich wegen $\lambda_1=\alpha$ aber überhaupt nichts ($\lambda_1+\alpha=0$). Jetzt noch den Summanden für $i=n$ herausziehen, ausmultiplizieren und $\beta^{2^n}=\beta$ sowie $\lambda_n=\mathrm{tr}(\alpha)$ entsprechend Formel 4.30 berücksichtigen:

$$\begin{aligned}\gamma^2 &= \sum_{i=1}^{n}\beta^{2^i}(\lambda_i+\alpha)\\ &= \beta^{2^n}(\alpha+\lambda_n)+\sum_{i=1}^{n-1}\beta^{2^i}(\alpha+\lambda_i)\\ &= \beta\alpha+\beta\,\mathrm{tr}(\alpha)+\alpha\sum_{i=1}^{n-1}\beta^{2^i}+\sum_{i=1}^{n-1}\beta^{2^i}\lambda_i\,.\end{aligned}$$

Im letzten Schritt wird noch der Summand für $i = 0$ hinzugenommen: $\sum_{i=1}^{n-1} \beta^{2^i} = \beta + \sum_{i=0}^{n-1} \beta^{2^i} = \beta + \mathrm{tr}(\beta)$ und man erkennt, daß der ganz rechte Term genau γ darstellt.

$$\begin{aligned}\gamma^2 &= \beta\alpha + \beta\,\mathrm{tr}(\alpha) + \alpha \sum_{i=1}^{n-1} \beta^{2^i} + \sum_{i=1}^{n-1} \beta^{2^i} \lambda_i \\ &= \beta\alpha + \beta\,\mathrm{tr}(\alpha) + \alpha\beta + \alpha\,\mathrm{tr}(\beta) + \sum_{i=1}^{n-1} \beta^{2^i} \lambda_i \\ &= \beta\,\mathrm{tr}(\alpha) + \alpha\,\mathrm{tr}(\beta) + \gamma\end{aligned}$$

Algorithmus Die Idee für einen Algorithmus steckt in der Darstellung 4.30 für den Trace bzw. hinter Rekursionsformel 4.34.

Algorithmus 8 Trace eines Elements $\alpha \in \mathbb{F}_{2^n}$

Ensure: $\lambda = \mathrm{tr}(\alpha)$

$\lambda \Leftarrow \alpha$

for $i = 1$ to $n - 1$ **do**
  $\lambda \Leftarrow \lambda^2 + \alpha$ $\{\lambda_i = \lambda_{i-1}^2 + \alpha\}$
end for

4.5.3 Halb-Trace

Der Halb-Trace ist eine wichtige Funktion für den Fall, daß n für den Binärkörper $\mathbb{F}_{2^n}$ ungerade ist.

Definition

$$\mathrm{htr}(\alpha) = \sum_{i=0}^{(n-1)/2} \alpha^{2^{2i}} = \sum_{i=0}^{(n-1)/2} \alpha^{4^i} = \sum_{i=0}^{(n-1)/2} \left(\alpha^{2^i}\right)^{2^i}$$

Eigenschaften

1. Der Halb-Trace ist genauso linear wie der Trace.

$$\mathrm{htr}(\alpha + \beta) = \mathrm{htr}(\alpha) + \mathrm{htr}(\beta)$$

2. $\mathrm{htr}(\alpha^2) = \mathrm{htr}^2(\alpha)$, es gilt jedoch hier: $\mathrm{htr}(\alpha^2) \neq \mathrm{htr}(\alpha)$, wehalb $\mathrm{htr}(\alpha)$ kein Element aus $\mathbb{F}_2$ (im Gegensatz zum Trace), sondern aus $\mathbb{F}_{2^n}$ ist.

   Einsetzen der Polynomdarstellung von α führt in ähnlicher Art und Weise wie beim Trace zu:

$$\mathrm{htr}(\alpha^2) = \sum_{i=0}^{(n-1)/2} \left(\alpha^2\right)^{2^{2i}} = \sum_{i=0}^{(n-1)/2} \left(\alpha^{2^{2i}}\right)^2 = \left[\sum_{i=0}^{(n-1)/2} \alpha^{2^{2i}}\right]^2 = \mathrm{htr}^2(\alpha)$$

3. $\mathrm{htr}(\alpha^2) + \mathrm{htr}(\alpha) = \alpha + \mathrm{tr}(\alpha)$

   Etwas ausführlicher

$$\begin{aligned}\mathrm{htr}(\alpha^2) + \mathrm{htr}(\alpha) &= \sum_{i=0}^{(n-1)/2} \left(\alpha^2\right)^{2^{2i}} + \sum_{i=0}^{(n-1)/2} \alpha^{2^{2i}} \\ &= \sum_{i=0}^{(n-1)/2} \alpha^{2^{2i+1}} + \sum_{i=0}^{(n-1)/2} \alpha^{2^{2i}}\end{aligned}$$

   Umindizierung der beiden Summen mit $l := 2i + 1$ bzw. $l := 2i$ führt unter zusätzlicher Berücksichtigung von $\alpha^{2^n} = \alpha$ zum Ergebnis:

$$\begin{aligned}\mathrm{htr}(\alpha^2) + \mathrm{htr}(\alpha) &= \sum_{l=1,3,5,\ldots}^{n} \alpha^{2^l} + \sum_{l=0,2,4,\ldots}^{n-1} \alpha^{2^l} \\ &= \alpha^{2^n} + \sum_{l=1,3,5,\ldots}^{n-2} \alpha^{2^l} + \sum_{l=0,2,4,\ldots}^{n-1} \alpha^{2^l} \\ &= \alpha + \sum_{l=0}^{n-1} \alpha^{2^l} \\ &= \alpha + \mathrm{tr}(\alpha)\end{aligned}$$

4. Aus den beiden vorangegangenen Punkten ergibt sich:

$$\alpha + \mathrm{tr}(\alpha) = \mathrm{htr}^2(\alpha) + \mathrm{htr}(\alpha) = \gamma^2 + \gamma \tag{4.38}$$

Algorithmus Der Algorithmus arbeitet ähnlich wie beim Trace (vgl. Algorithmus 8), außer daß der zusätzlichen Potenzierung Rechnung getragen wird.

Algorithmus 9 Halb-Trace eines Elements $\alpha \in \mathbb{F}_{2^n}$

Ensure: $h = \mathrm{htr}(\alpha)$

$h \Leftarrow 0$

for $i = 0$ to $(n-1)/2$ **do**
  $h \Leftarrow h^2$
  $h \Leftarrow h^2 + \alpha$
end for

4.5.4 Lösung

Ist $\beta = 0$ dann sind die Lösungen 0 und 1.

Im allgemeinen gilt:

$$\begin{aligned} z^2 + z + \beta &= 0 \\ \mathrm{tr}(z^2 + z + \beta) &= \mathrm{tr}(0) \\ \mathrm{tr}(z^2) + \mathrm{tr}(z) + \mathrm{tr}(\beta) &= 0 \\ \mathrm{tr}(\beta) &= 0 \end{aligned}$$

d. h. eine Voraussetzung für die Existenz der zwei Lösungen ist $\mathrm{tr}(\beta) = 0$.

Für den Spezialfall, daß n ungerade ist, kann man mit der Grundvoraussetzung $\mathrm{tr}(\beta) = 0$ aus Formel 4.38 direkt eine Lösung der quadratischen Gleichung ablesen:[68]

$$\mathrm{htr}^2(\beta) + \mathrm{htr}(\beta) = \beta \tag{4.39}$$

Mit $\mathrm{htr}^2(\beta) + \mathrm{htr}(\beta) = [\mathrm{htr}(\beta) + 1]\,\mathrm{htr}(\beta)$ sind die zwei Lösungen demzufolge $z_1 = \mathrm{htr}(\beta)$ und $z_2 = \mathrm{htr}(\beta) + 1$.

Ist n jedoch gerade (bzw. im allgemeinen Fall), dann besteht eine Berechnungsmöglichkeit durch Anwendung von Formel 4.37. Setzt man nämlich als Grundvoraussetzung $\mathrm{tr}(\beta) = 0$, dann gilt:

$$\beta\,\mathrm{tr}(\alpha) = z^2 + z, \text{ mit } z = \sum_{i=1}^{n-1} \beta^{2^i} \lambda_i \text{ und } \lambda_i = \sum_{k=0}^{i-1} \alpha^{2^k}$$

d. h. findet man ein Element α mit $\mathrm{tr}(\alpha) = 1$, wodurch

$$\beta = z^2 + z$$

dann kann man die Lösung z aus diesem Element α berechnen. Glücklicherweise besitzt im Mittel die Hälfte aller Elemente $\alpha \in \mathbb{F}_{2^n}$ einen Trace von 1, so daß man α durchaus zufällig wählen kann (mit 50% Wahrscheinlichkeit für $\mathrm{tr}(\alpha) = 1$).

Algorithmus Für den allgemeinen Fall (bzw. wenn n gerade ist) kann man auf Algorithmus 10 zurückgreifen, um die quadratische Gleichung in $\mathbb{F}_{2^n}$ zu lösen [ANS05, IEE00].

Beide Summen können vereint werden, wenn man λ_i iterativ in der äußeren Summe mitberechnet (nach derselben Methodik wie in Algorithmus 8), was dann im letzten Durchlauf $\lambda_n = \mathrm{tr}(\alpha)$ ergibt.

[68] Entweder man testet $\mathrm{tr}(\beta) = 0$ im Vorfeld oder man muß am Schluß nocheinmal $z^2 + z = \beta$ testen.

Und es gilt ja nach Formel 4.37, unter der Voraussetzung $\mathrm{tr}(\beta) = 0$ und $\mathrm{tr}(\alpha = 1)$, immer:

$$\gamma = \beta + \gamma^2, \text{ mit } \gamma = \sum_{i=1}^{n-1} \beta^{2^i} \lambda_i \text{ und } \lambda_i = \sum_{k=0}^{i-1} \alpha^{2^k}$$

Algorithmus 10 Lösung der quadratischen Gleichung in $\mathbb{F}_{2^n}$

Ensure: $z = \sum_{i=1}^{n-1} \beta^{2^i} \lambda_i$, mit $\lambda_i = \sum_{k=0}^{i-1} \alpha^{2^k}$

repeat
  $\alpha \Leftarrow$zufälliges Element aus $\mathbb{F}_{2^n}$
  $z \Leftarrow 0$
  $\lambda \Leftarrow \alpha$ $\{\lambda_1 = \alpha\}$
  for $i = 1$ to $n-1$ **do**
    $z \Leftarrow z^2 + \lambda^2 \beta$ $\{z = z^2 + \beta$ für $i = n-1$ (bzw. $\lambda_n = \mathrm{tr}(\alpha) = 1)\}$
    $\lambda \Leftarrow \lambda^2 + \alpha$ $\{\lambda_i = \lambda_{i-1}^2 + \alpha\}$
  end for
until $\lambda = 1$ $\{\mathrm{tr}(\alpha) = 1\}$

Optimierung Schreibt man dazu unter Zuhilfenahme der Polynomdarstellung für das Element α und noch unter Berücksichtigung von Formel 3.13:

$$\mathrm{tr}(\alpha) = \sum_{i=0}^{n-1} \left(\sum_{j=0}^{n-1} \alpha_j x^j \right)^{2^i} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \left(\alpha_j x^j \right)^{2^i} = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \alpha_j^{2^i} \left(x^{2^i} \right)^j .$$

Umsortieren der Summen und mit $\alpha_j^l = \alpha_j$ in $\mathbb{F}_2$ und dann wieder mit Formel 3.13

$$\mathrm{tr}(\alpha) = \sum_{j=0}^{n-1} \alpha_j \sum_{i=0}^{n-1} \left(x^{2^i} \right)^j = \sum_{j=0}^{n-1} \alpha_j \sum_{i=0}^{n-1} \left(x^j \right)^{2^i} = \sum_{j=0}^{n-1} \alpha_j \, \mathrm{tr}(x^j)$$

Den Wert für $\mathrm{tr}(x^j)$ kann man vorausberechnen.

4.6 M -Potenzierung

Bei der Methode nach M handelt es sich eigentlich um eine Multiplikationsmethode, bei der die Modulo-Reduktion des (Zwischen-) Ergebnisses ohne echte Langzahldivision erfolgen kann [Mon85], [CP05, 9.2.1], [HMV04, 2.2.4]. Da allerdings zuerst eine Transformation beider Faktoren in den M -"Raum" vorgenommen werden muß (welche zwei Modulo-Division erfordert) und außerdem einmal der erweiterte GCD-Algorithmus bemüht werden muß, kommen die Geschwindigkeitsvorteile nur bei der Potenzierung wirklich zum tragen.

Um einen leicht verständlichen Zugang zu finden, konzentrieren wir uns auf eine einzelne Multiplikation $c = ab \pmod m$. Dazu soll eine Zahl r vorausgesetzt werden, die keinen gemeinsamen Teiler mit dem Modul m hat und für die $r > m$ gewährleistet ist. Die M -Multiplikation kann man nun, unter der Voraussetzung $\gcd(r,m) = 1$ (bzw. mit dem Satz von B´ $\alpha r - \beta m = 1$), in folgenden Einzelschritten darstellen:

1. Berechnung der B´ -Koeffizienten α, β mit Hilfe des erweiterten euklidischen Algorithmus;
2. Eingangstransformation der Faktoren a und b zu $\hat{a} = ar \bmod m$ und $\hat{b} = br \bmod m$ (eine normale Modulo-Division);
3. (Wiederholte) Multiplikation im M -Bereich:
   a) Berechnung des Produkts $x = \hat{a} \cdot \hat{b}$ (man beachte, daß hierbei keine Modulo-Reduktion vorgenommen wird);
   b) M -Reduktion von $x = (ar \bmod m) \cdot (br \bmod m)$ zu $\hat{c} = abr \bmod m = cr \bmod m$;
      Die nötige Korrektur[69] $\hat{c} = \mathrm{M}(x,r,m) = xr^{-1} \bmod m$ führt zu einer M -Darstellung für c (als Voraussetzung für den nächsten Teilschritt).[70]
   c) Eine weitere (optionale) Multiplikationen im M -"Raum", die $\hat{c}$ als einen der Faktoren verwendet;
4. Rücktransformation des Ergebnisses $\hat{c}$ zu $c = \hat{c}r^{-1} \bmod m$, ebenfalls eine M -Reduktion: $c = \mathrm{M}(\hat{c},r,m)$.

M -Reduktion Zur Herleitung von M 's effizienter Berechnungsmethode für $\mathrm{M}(x,r,m) = xr^{-1} \bmod m$ wählen wir als Ausgangspunkt:

$$\begin{aligned} m(\beta x \bmod r) &= m(\beta x - ur), \quad \text{mit } u = \lfloor \beta x / r \rfloor \\ r(\alpha x \bmod m) &= r(\alpha x - vm), \quad \text{mit } v = \lfloor \alpha x / m \rfloor . \end{aligned}$$

Subtraktion beider Gleichungen ergibt

$$\begin{aligned} m(\beta x \bmod r) - r(\alpha x \bmod m) &= m(\beta x - ur) - r(\alpha x - vm) \\ &= rm(v-u) - x(\alpha r - \beta m), \end{aligned}$$

was mit $\gcd(r,m) = \alpha r - \beta m = 1$ zu

[69] Der Ausdruck $xr^{-1} \bmod m = \hat{a}\hat{b}r^{-1} \bmod m$ wird auch als M -Produkt von $\hat{a}$ und $\hat{b}$ bezeichnet.

[70] Im Gegensatz zur Multiplikation, für welche $\hat{a} \cdot \hat{b} \neq \hat{c} \pmod m$ gilt, verhält sich die Addition regulär: $\hat{a} + \hat{b} = ar + br = (a+b)r$.

$$m(\beta x \bmod r) + x = (v-u)rm + r(\alpha x \bmod m)$$

führt.[71]

Mit dem Wissen, daß es sich bei α um das multiplikativ inverse Element von r im Restklassensystem modulo m handelt ($\alpha = r^{-1} \pmod m$, $\beta = -m^{-1} \pmod r$), stellen wir noch um:

$$\alpha x \bmod m = \frac{m(\beta x \bmod r) - (v-u)rm + x}{r}$$

und gewinnen letztlich M 's Reduktionsformel.

$$\mathrm{M}(x,r,m) = xr^{-1} \bmod m = \frac{m(\beta x \bmod r) + x}{r} - (v-u)m \tag{4.40}$$

Praktisch benötigt man u und v nicht, sondern geht meist nach folgendem Algorithmus vor:

Algorithmus 11 M -Reduktion

Ensure: $y = xr^{-1} \bmod m$

$t \Leftarrow \beta x \bmod r$
$y \Leftarrow \dfrac{mt + x}{r}$
if $y \geq m$ **then**
  $y \Leftarrow y - m$
end if

Bei geschickter Wahl von r zum Beispiel als $r = 2^s$ sind für alle (Modulo-) Divisionen in Formel 4.40 nur logische oder Schiebeoperationen nötig.[72] Um $\gcd(2^s, m) = 1$ zu garantieren ist die einfachste Bedingung die, m als ungerade vorauszusetzen.[73]

Literatur

[ANS05] *Public Key Cryptography For The Financial Services Industry: The Elliptic Curve Digital Signature Algorithm (ECDSA).* American National Standard for Financial Services X9.62, ANSI, September 2005.

[BB87] B , A. und R. P. B : *A systolic algorithm for extended GCD computation.* Computers & Mathematics with Applications, 14(4):233–238, 1987.

[71] Wegen des Vorkommens von r in allen Summanden der rechten Seite muß die linke Seite durch r teilbar sein, d. h. $m(\beta x \bmod r) + x$ ist ein Vielfaches von r.

[72] Effiziente numerische Algorithmen zur M -Reduktion findet man z. B. in [Koç94b, KAK96, DK91].

[73] Oftmals kann eine solche Einschränkung hingenommen werden, falls nicht siehe z. B. [Koç94a].

[Ber84] B , E R.: *Algebraic Coding Theory, Revised 1984 Edition*. Aegean Park Press, Laguna Hills, CA, 1984.

[BK83] B , R. P. und H. T. K : *Systolic VLSI arrays for linear time GCD computation*. In: A , F. und E. J. A (Herausgeber): *Proceedings of International Conference on Very Large Scale Integration (VLSI 83)*, Seiten 145–154. International Federation of Information Processing, Elsevier Science Publishers B.V., 1983.

[Bos96] B , S : *Algebra*. Springer, Berlin, 1996.

[Bun08] B , P : *Einführung in die Zahlentheorie*. Springer, Berlin Heidelberg, 6. Auflage, 2008.

[CP05] C , R und C P : *Prime Numbers. A Computational Perspective*. Springer, 2. Auflage, 2005.

[DK91] D ´, S R. und B S. K , J .: *A Cryptographic Library for the Motorola DSP56000*. In: *Advances in Cryptology — EUROCRYPT '90*, Band 473 der Reihe *Lecture Notes in Computer Sciences*, Seiten 230–244, New York, 1991. Springer.

[DM11] D. M G , K. I , M. S : *Fundamental Elliptic Curve Cryptography Algorithms*. RFC 6090, IETF, Februar 2011.

[FHLM04] F , K., D. H , J. L und A. M : *Field inversion and point halving revisited*. IEEE Trans. Comput., 53(8):1047–1059, August 2004.

[Gar59] G , H L.: *The Residue Number System*. IRE Transactions on Electronic Computers, 8(2):140–147, Juni 1959.

[Har06] H , L : *Modular Inverse Algorithms Without Multiplications for Cryptographic Applications*. EURASIP Journal on Embedded Systems, 2006.

[HMV04] H , D , A M und S V : *Guide to Elliptic Curve Cryptography*. Springer, New York, 2004.

[IEE00] *Standard Specifications For Public-Key Cryptography*. Std 1363-2000, IEEE, 2000.

[Jeb93] J , T : *Systolic Normalization of Rational Numbers*. In: D , L. und B. W (Herausgeber): *ASAP'93 — International Conference on Application Specific Array Processors*, Seiten 502–513, Venice, Italy, Oktober 1993. IEEE Computer Society Press. Also: Technical Report 93-45, RISC-Linz, Johannes Kepler University, Linz, Austria, August 1993.

[KAK96] K ¸, Ç K , T A und B S. K , J .: *Analyzing and Comparing Montgomery Multiplication Algorithms*. IEEE Micro, 16(3):26–33, Juni 1996.

[Kal95] K , J ., B S.: *The Montgomery Inverse and Its Applications*. IEEE Transactions on Computers, 44(8):1064–1065, August 1995.

[Knu98] K , D E.: *Seminumerical Algorithms*, Band 2 der Reihe *The Art of Computer Programming*. Addison-Wesley, Reading, Massachusetts, 3. Auflage, 1998.

[Koç94a] K ¸, Ç. K.: *Montgomery Reduction with Even Modulus*. IEE Proceedings on Computers and Digital Techniques, 141(5):314–316, September 1994.

[Koç94b] K ¸, Ç K : *High-Speed RSA Implementation*. Technischer Bericht TR-201, RSA Laboratories, 1994. Version 2.0.

[Kör88] K¨ , T.W.: *Fourier Analysis*. Cambridge University Press, 1988.

[Mon85] M , P.: *Modular multiplication without trial division*. Mathematics of Computation, 44(170):519–521, 1985.

[MvV92] M , A J., P. C. O und S A. V : *Handbook of applied cryptography*. CRC Press Series on Discrete Mathematics and Its Application. CRC Press, 2. Auflage, 1992.

[NZM91] N , I., H.S. Z und H.L. M : *An Introduction to the Theory of Numbers*. John Wiley & Sons, 5. Auflage, 1991.

[PD75] P , H und H G. D : *The Theory of Algebraic Numbers*. The Mathmatical Association of America, 2. Auflage, 1975.

[PW72] P , W. W und E. J. W , J .: *Error-Correcting Codes*. The MIT Press, Cambridge Massachusetts and London, England, 2. Auflage, 1972.

[Ros99] R , K H.: *Elementary Number Theory and Its Applications*. Addison-Wesley, 4. Auflage, 1999.

[RSA78] R , R. L., A. S und L. M. A : *A Method for Obtaining Digital Signatures and Public-Key Cryptosystems*. Communications of the ACM, 21(2):120–126, Februar 1978.

[Sch96] S , B : *Applied Cryptography*. John Wiley & Sons, 2. Auflage, 1996.

[Sha73] S , D : *Five Number-theoretic Algorithms*. In: R S. D. T , H C W (Herausgeber): *Proceedings of the Second Manitoba Conference on Numerical Mathematics*, Band VII der Reihe *Congressus Numerantium*, Seiten 51–70, Winnipeg, 1973. Utilitas Mathematica Publishing, Inc.

[Sho05] S , V : *A Computational Introduction to Number Theory and Algebra*. Cambridge University Press, 2005.

[Ste67] S , J.: *Computational problems associated with Racah algebra*. Journal of Computational Physics, 1(3):397–405, 1967.

[Ton91] T , A : *Bemerkung über die Auflösung quadratischer Congruenzen*. Nachrichten von der Königlichen Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen, Seiten 344–346, 1891.

[Wel01] W , M : *Kryptographie in C und C++*. Springer, Berlin Heidelberg New York, 2. Auflage, 2001.

Index