2.7 Der allgemeine Kongruenzgenerator

Etwas komplizierter, aber nicht entmutigend, wird das Vorhersageverfahren für Kongruenzgeneratoren, bei denen auch der Modul unbekannt ist. Hier bringt die allgemeine Sprache der kommutativen Algebra nicht mehr viel, da sehr spezielle Eigenschaften der Ringe $\mathbb{Z}$ und $\mathbb{Z}/m\mathbb{Z}$ verwendet werden, insbesondere das „kanonische“ Repräsentantensystem $\{0, \ldots, m-1\} \subseteq \mathbb{Z}$ von $\mathbb{Z}/m\mathbb{Z}$.

Sei $X = \mathbb{Z}^r$, $\bar{X} = (\mathbb{Z}/m\mathbb{Z})^r$, $Z = \mathbb{Z}^k$, $\bar{Z} = (\mathbb{Z}/m\mathbb{Z})^k$. Gegeben seien die Abbildungen

$$\Phi^{(i)} : X^i \longrightarrow Z \text{ für } i \geq h,$$
$$\alpha : \bar{Z} \longrightarrow \bar{X} \text{ linear},$$

wobei α und m für die Kryptoanalyse als unbekannt behandelt werden. Mit Hilfe des kanonischen Repräsentantensystems wird $\bar{X}$ als Teilmenge $\{0, \ldots, m-1\}^r$ von X aufgefasst. Dann funktioniert die Erzeugung der Folge wie gehabt, und wir nennen das Verfahren einen **allgemeinen Kongruenzgenerator**, wenn die Berechnung aller $\Phi^{(i)}$ effizient möglich ist, d. h., mit einem Aufwand, der polynomial von r, k und $\log(m)$ abhängt. Insbesondere gibt es eine Schranke M für die Werte der $\Phi^{(i)}$ auf $\{0, \ldots, m-1\}^i$, die polynomial in r, k und $\log(m)$ ist.

Die Kryptoanalyse wird in zwei Phasen unterteilt. In der ersten Phase wird über dem Ring $\mathbb{Z}$ bzw. seinem Quotientenkörper $\mathbb{Q}$ gearbeitet und ein Vielfaches $\hat{m}$ des Moduls m bestimmt. In der zweiten Phase arbeitet man über dem Ring $\mathbb{Z}/\hat{m}\mathbb{Z}$. Bei der Voraussage von x_n sind jetzt drei Ereignisse möglich:

- $z_n \notin Z_{n-1}$; der ($\mathbb{Q}$- oder $\mathbb{Z}/m\mathbb{Z}$-) Modul Z_{n-1} muss zu Z_n erweitert werden, für x_n ist keine Vorhersage möglich.
- x_n wird korrekt vorhergesagt.
- x_n wird falsch vorhergesagt. Dann wird der Modul $\hat{m}$ korrigiert.

In der ersten Phase ist Z_{n-1} der $\mathbb{Q}$-Vektorraum, der von $z_h, \ldots, z_{n-1}$ aufgespannt wird, wobei man natürlich redundante z_i einfach weglässt.

1. Fall: $z_n \notin Z_{n-1}$. Dann wird $Z_n = Z_{n-1} + \mathbb{Q}z_n$ gesetzt und x_n nicht vorhergesagt. Dieser Fall kann höchstens k-mal auftreten.

2. Fall: $z_n = t_h z_h + \cdots + t_{n-1} z_{n-1}$. Dann wird $x_n = t_h x_h + \cdots + t_{n-1} x_{n-1}$ vorhergesagt (als Element von $\mathbb{Q}^r$). (Es treten höchstens k der z_i in der konstruierten Basis von Z_{n-1} auf, also auch höchstens k von 0 verschiedene Koeffizienten t_i.)

3. Fall: Genauso, aber $\hat{x}_n = t_h x_h + \cdots + t_{n-1} x_{n-1}$ stimmt nicht mit x_n überein. Sei dann $d \in \mathbb{N}$ der Hauptnenner von $t_h, \ldots, t_{n-1}$. Dann ist

$$d\hat{x}_n = \alpha(dt_h z_h + \cdots + dt_{n-1} z_{n-1}) = \alpha(dz_n) = dx_n$$

in $\bar{X}$, also mod m gerechnet. Damit ist gezeigt:

Hilfssatz 3 *(*BOYAR*) Der größte gemeinsame Teiler $\hat{m}$ der Komponenten von $d\hat{x}_n - dx_n$ im 3. Fall ist ein Vielfaches des Moduls m.*

Die erste Phase liefert also ein Vielfaches $\hat{m} \neq 0$ des Moduls m. Der Aufwand dafür beträgt:

- höchstens $k+1$ Versuche, ein lineares Gleichungssystem mit höchstens k Unbekannten über $\mathbb{Q}$ zu lösen,
- eine Bestimmung des größten gemeinsamen Teilers von r Zahlen.

Daneben wird eine unbestimmte Anzahl von Folgegliedern x_n korrekt vorhergesagt, was jeweils ebenfalls mit der Lösung eines solchen linearen Gleichungssystems bezahlt wird.

Wie groß kann $\hat{m}$ sein? Zur Abschätzung braucht man eine obere Schranke M für alle Komponenten aller $\Phi^{(i)}$ auf $\{0,\ldots,m-1\}^i \subseteq X^i$. Zur Herleitung wird die Ungleichung von HADAMARD verwendet: Für beliebige Vektoren $x_1,\ldots,x_k \in \mathbb{R}^k$ gilt

$$|\operatorname{Det}(x_1,\ldots,x_k)| \leq \|x_1\|_2 \cdots \|x_k\|_2$$

mit der euklidischen Norm $\| \bullet \|_2$.

Hilfssatz 4 *$\hat{m} \leq (k+1)\cdot m\cdot\sqrt{k^k}\cdot M^k$, insbesondere wächst $\log(\hat{m})$ höchstens polynomial mit k, $\log(m)$ und $\log(M)$.*

Beweis. Der Koeffizientenvektor t ist Lösung eines linearen Gleichungssystems aus höchstens k Gleichungen mit ebensovielen Unbekannten. Die Koeffizienten z_i dieses Gleichungssystems sind durch M beschränkt. Nach der Ungleichung von HADAMARD für die Determinante und der CRAMERschen Regel sind Zähler dt_i und Nenner d der Lösung durch

$$\prod_{i=1}^{k} \sqrt{\sum_{j=1}^{k} M^2} = \prod_{i=1}^{k} \sqrt{kM^2} = \sqrt{k^k} \cdot M^k$$

beschränkt. Die Komponenten von $d\hat{x}_n$ sind also durch

$$\|d\hat{x}_n\|_\infty = \|\sum dt_i x_i\|_\infty \leq \sqrt{k^k} \cdot M^k \cdot \sum \|x_i\|_\infty \leq km \cdot \sqrt{k^k} \cdot M^k$$

beschränkt, weil m eine Schranke für die Komponenten der x_i ist. Daraus folgt

$$\|d\hat{x}_n - dx_n\|_\infty \leq km \cdot \sqrt{k^k} \cdot M^k + \sqrt{k^k} \cdot M^k \cdot m = (k+1) \cdot m \cdot \sqrt{k^k} \cdot M^k,$$

wie behauptet. $\diamond$

Wie sieht das im Beispiel des gewöhnlichen linearen Kongruenzgenerators aus? Hier ist

$$z_1 = \begin{pmatrix} x_0 \\ 1 \end{pmatrix}, z_2 = \begin{pmatrix} x_1 \\ 1 \end{pmatrix}, z_3 = \begin{pmatrix} x_2 \\ 1 \end{pmatrix}, \ldots$$

Falls $x_1 = x_0$, sind wir im trivialen Fall der konstanten Folge. Andernfalls ist z_3 rationale Linearkombination $t_1 z_1 + t_2 z_2$: Die Lösung des Gleichungssystems

$$\begin{aligned} x_0 t_1 + x_1 t_2 &= x_2, \\ t_1 + t_2 &= 1 \end{aligned}$$

ist

$$t = \frac{1}{d} \cdot \begin{pmatrix} -x_2 + x_1 \\ x_2 - x_0 \end{pmatrix} \quad \text{mit } d = x_1 - x_0.$$

Vorhergesagt wird dann

$$\hat{x}_3 = t_1 x_1 + t_2 x_2 = \frac{-x_2 x_1 + x_1^2 + x_2^2 - x_2 x_0}{x_1 - x_0} = \frac{(x_2 - x_1)^2}{x_1 - x_0} + x_2.$$

Also ist $d(\hat{x}_3 - x_3) = (x_2 - x_1)^2 - (x_1 - x_0)(x_3 - x_2) = y_2^2 - y_1 y_3$ mit der Differenzenfolge (y_i). Falls $\hat{x}_3 = x_3$, müssen wir weiter machen. Sonst erhalten wir, wie aus Hilfssatz 1, $m | \hat{m} = |y_1 y_3 - y_2^2|$.

Im Standard-Beispiel $x_0 = 2134$, $x_1 = 2160$, $x_2 = 6905$, $x_3 = 3778$, also mit $y_1 = 26$, $y_2 = 4745$, $y_3 = -3127$, erhalten wir

$$\hat{m} = 4745^2 + 26 \cdot 3127 = 22596327.$$

In der zweiten Phase des Algorithmus wird das gleiche Verfahren, aber über dem Ring $\hat{R} = \mathbb{Z}/\hat{m}\mathbb{Z}$ durchgeführt. Da man die rationalen Ergebnisse aus der ersten Phase nicht einfach $\bmod \hat{m}$ reduzieren kann, startet man wieder neu bei z_h. Es gibt wieder drei Fälle für jeden Einzelschritt:

1. Fall: $z_n \notin \hat{Z}_{n-1} = \hat{R} z_h + \cdots + \hat{R} z_{n-1}$. Dann wird $\hat{Z}_n = \hat{Z}_{n-1} + \hat{R} z_n$ gesetzt (und dieser $\hat{R}$-Modul durch ein nicht-redundantes Erzeugendensystem $\{z_{j_1}, \ldots, z_{j_l}\}$ repräsentiert, wobei $z_{j_l} = z_n$). Hier wird x_n nicht vorhergesagt.

2. Fall: $z_n = t_h z_h + \cdots + t_{n-1} z_{n-1}$. Dann wird $x_n = t_h x_h + \cdots + t_{n-1} x_{n-1}$ vorhergesagt (als Element von $\hat{X} = (\mathbb{Z}/\hat{m}\mathbb{Z})^r$). Die Voraussage sei korrekt.

3. Fall: Genauso, aber die Voraussage $\hat{x}_n = t_h x_h + \cdots + t_{n-1} x_{n-1}$ stimmt in $\hat{X}$ nicht mit x_n überein. Dann wird $\hat{x}_n - x_n$ als Element von $\mathbb{Z}^r$ betrachtet:

Hilfssatz 5 *Der größte gemeinsame Teiler der Koeffizienten von $\hat{x}_n - x_n$ im 3. Fall ist ein Vielfaches von m, aber kein Vielfaches von $\hat{m}$.*

Beweis. Er ist ein Vielfaches von m, weil $\hat{x}_n \bmod m = x_n$ sein muss. Er ist kein Vielfaches von $\hat{m}$, weil sonst ja $\hat{x}_n = x_n$ in $\hat{X}$ wäre. $\diamond$

Im 3. Fall wird $\hat{m}$ durch den ggT dieses größten gemeinsamen Teilers mit $\hat{m}$ ersetzt und die ganze Kette der bisherigen z_j (soweit sie nicht schon redundant waren) $\bmod \hat{m}$ reduziert. Wegen der zweiten Aussage im Hilfssatz ist dabei $\hat{m}$ echt kleiner geworden.

Wegen Hilfssatz 4 kann der dritte Fall insgesamt nicht zu oft auftreten; die Anzahl der Vorkommnisse ist polynomial in k, $\log(m)$ und $\log(M)$. Ist das richtige m erreicht, kann dieser Fall gar nicht mehr vorkommen. Der erste Fall kann in der zweiten Phase insgesamt wegen Satz 2 höchstens ${}^2\log(\#(\mathbb{Z}/\hat{m}\mathbb{Z})^k) = k \cdot {}^2\log(\hat{m})$ Mal vorkommen, und diese Schranke ist polynomial in k, $\log(m)$ und $\log(M)$.

Anmerkung. Die Gemeinsamkeit von erster und zweiter Phase besteht darin, dass beide Male über dem vollen Quotientenring gerechnet wird: Der volle Quotientenring von $\mathbb{Z}$ ist der Quotientenkörper $\mathbb{Q}$. In einem Restklassenring $\mathbb{Z}/m\mathbb{Z}$ dagegen sind die Nicht-Nullteiler genau die zu m teilerfremden Elemente, also die Einheiten. Daher ist $\mathbb{Z}/m\mathbb{Z}$ sein eigener voller Quotientenring.

Im Standard-Beispiel haben wir nach der ersten Phase $\hat{m} = 22596327$ und müssen nun das lineare Gleichungssystem (1) $\bmod \hat{m}$ lösen. Es wird $-26t_1 = 4745$, also $26t_1 = 22595853$ (alles in $\mathbb{Z}/\hat{m}\mathbb{Z}$). Der größte gemeinsame Teiler von $\hat{m}$ und 26 ist 13, und 22595853 ist kein Vielfaches davon. Also ist der erste Fall eingetreten, wir müssen den $\hat{R}$-Modul Z_3 mit dem Erzeugendensystem $z_1 = \binom{2134}{1}, z_2 = \binom{2160}{1}, z_3 = \binom{6905}{1}$ bilden und mit $z_4 = \binom{3778}{1}$ weiterarbeiten. Da $z_2 = 15515735z_1 + 7080593z_3$, wird Z_3 schon von z_1 und z_3 erzeugt. Zu lösen ist also (in $\mathbb{Z}/\hat{m}\mathbb{Z}$)

$$\begin{aligned} 2134t_1 + 6905t_3 &= 3778, \\ t_1 + t_3 &= 1. \end{aligned}$$

Elimination von t_3 ergibt

$$4771t_1 = 3127.$$

Der Koeffizient 4771 von t_1 ist durch 13 teilbar, aber 3127 nicht. Also gibt es keine Lösung. Wir sind im 1. Fall und bilden Z_4. Da $z_1 = 15492972z_3 + 7103356z_4$, kann man z_1 weglassen. Nun versuchen wir, $z_5 = \binom{8295}{1}$ als Linearkombination darzustellen:

$$\begin{aligned} 6905t_3 + 3778t_4 &= 8295, \\ t_3 + t_4 &= 1. \end{aligned}$$

Elimination von t_4 ergibt

$$3127t_3 = 4517.$$

Das Inverse $\bmod \hat{m}$ von 3127 ist 11316229, also $t_3 = 2514719$, $t_4 = 20081609$. Daraus wird x_5 vorhergesagt:

$$\hat{x}_5 = t_3x_3 + t_4x_4 = 6975053.$$

Da $x_5 = 5543$, sind wir im 3. Fall und haben $\hat{m}$ zu korrigieren:

$$\mathrm{ggT}(\hat{x}_5 - x_5, \hat{m}) = \mathrm{ggT}(6969510510, 22596327) = 8397.$$

Von jetzt an wird nur noch der dritte Fall auftreten, d. h., der Rest der Folge wird korrekt vorhergesagt.

Ein **Vorhersageverfahren** für einen allgemeinen Kongruenzgenerator ist ein Algorithmus, der als Eingabe die Startwerte $x_0, \ldots, x_{h-1}$ erhält, dann Schätzungen für $x_h, x_{h+1}, \ldots$ auswirft und diese anschließend mit dem jeweiligen wahren Wert vergleicht; bei einer Fehlvorhersage werden unter Verwendung des wahren Werts die Parameter des Verfahrens adjustiert. Das Vorhersageverfahren ist **effizient**, wenn

(a) der Aufwand für die Vorhersage jedes x_n polynomial in r, k und $\log(m)$ ist,

(b) die Zahl der Fehlvorhersagen durch ein Polynom in r, k und $\log(m)$ beschränkt ist, ebenso der Aufwand für die Parameteradjustierung im Fall einer Fehlvorhersage.

Der Algorithmus von Boyar/Krawczyk, den wir in diesem Abschnitt behandelt haben, erfüllt (b). Er erfüllt auch (a), da das Lösen linearer Gleichungssysteme über Restklassenringen $\mathbb{Z}/m\mathbb{Z}$ effizient möglich ist, wie schon früher gezeigt. Damit ist bewiesen:

Hauptsatz 1 *Für einen beliebigen effizienten Kongruenzgenerator ist der Algorithmus von* Boyar/Krawczyk *ein effizientes Vorhersageverfahren.*

Die Anwendung auf nichtlineare Generatoren wird an einem weiteren einfachen Zahlenbeispiel gezeigt. Von einem quadratischen Generator der Form

$$x_n = ax_{n-1}^2 + bx_{n-1} + c \bmod m$$

sei die Zahlenfolge

$$63, 96, 17, 32, 37, 72$$

erzeugt worden. Wir verwenden also $X = \mathbb{Z}$, $Z = \mathbb{Z}^3$, $h = 1$. In der ersten Phase spannen

$$z_1 = \begin{pmatrix} 3969 \\ 63 \\ 1 \end{pmatrix} z_2 = \begin{pmatrix} 9216 \\ 96 \\ 1 \end{pmatrix} z_3 = \begin{pmatrix} 289 \\ 17 \\ 1 \end{pmatrix}$$

schon ganz $\mathbb{Q}^3$ auf, denn ihre Determinante ist 119922. Die Auflösung von

$$z_4 = \begin{pmatrix} 1024 \\ 32 \\ 1 \end{pmatrix} = t_1 z_1 + t_2 z_2 + t_3 z_3$$

ergibt $t_1 = \frac{160}{253}, t_2 = -\frac{155}{869}, t_3 = \frac{992}{1817}$ mit Hauptnenner $d = 11{\cdot}23{\cdot}79 = 19987$. Die Voraussage ist $\hat{x}_4 = \frac{1502019}{19987} \neq x_4$. Der erste geschätzte Modul ist also

$$\hat{m} = d\hat{x}_4 - dx_4 = 762500.$$

Das gleiche lineare Gleichungssystem wird jetzt über $\mathbb{Z}/\hat{m}\mathbb{Z}$ aufgelöst und ergibt

$$t_1 = 161520, t_2 = 436805, t_3 = 164176,$$

$$\hat{x}_4 = 735237, \hat{x}_4 - x_4 = 735200.$$

Also wird $\hat{m}$ korrigiert zu

$$\mathrm{ggT}(762500, 735200) = 100.$$

Da schon x_0 größer als die Hälfte davon ist, ist mit Sicherheit $m = 100$, und es wird keine falschen Vorhersagen mehr geben. Da die Determinante von z_1, z_2, z_3 mod 100 nicht invertierbar ist, könnte es aber noch Lücken in den Voraussagen geben!

Für die Vorhersage von x_5 ergibt sich (mod100)

$$t_1 = 10, t_2 = 40, t_3 = 51, x_5 = 10 \cdot 96 + 40 \cdot 17 + 51 \cdot 32 = 72.$$

Analog für x_6:

$$t_1 = 15, t_2 = 85, t_3 = 51, x_6 = 15 \cdot 96 + 85 \cdot 17 + 51 \cdot 32 = 17.$$

Die Folge hat also mit Sicherheit die Periode 4 (nach Vorperiode 2) und ist daher komplett vorhersagbar.