Über die Länge nicht-trivialer Collatz-Zyklen

von Christian Hercher, Flensburg

Dieser Artikel ist das Ergebnis eines etwa einjährigen Projekts, welches sich aus einer Diskussion auf dem Forum Matheplanet ergeben hat, siehe [2]. Ich danke dabei allen Mitdiskutanten, die immer wieder neue Ideen ein- und damit das Vorhaben vorangebracht haben. Insbesondere möchte ich dabei Gerhard Pracht und den Autor „Amateur“ nennen.

Beim Collatz-Problem stellt sich die Frage, ob für jede Startzahl $k > 0$ die zugehörige Collatz-Folge C_k irgendwann in den trivialen Zyklus $1, 4, 2, \ldots$ übergeht. Diese Frage ist bisher unbeantwortet. In dieser Arbeit untersuchen wir die Mindestlänge eines hypothetischen nicht-trivialen Zyklus und zeigen folgende Ergebnisse:

1. Ein nicht-trivialer Collatz-Zyklus muss mindestens 17 Milliarden verschiedene Folgeglieder beinhalten. (Zuvor bestes bekanntes Ergebnis war ein Wert leicht unterhalb von 1,03 Milliarden Folgegliedern, siehe [1] in Verbindung mit [5].)
2. Für alle Startzahlen $k \leq 87 \cdot 2^{60}$, also insbesondere alle $k \leq 10^{20}$, geht die jeweilige Collatz-Folge in den trivialen Zyklus über. (Zuvor beste bekannte obere Schranke für diese Aussage war $5 \cdot 2^{60}$, siehe [5, 6].)

Dabei kamen neben hohem Rechnereinsatz und den algorithmischen und mathematischen Methoden von bisherigen Autoren auf diesem Gebiet wie Leavens und Vermeulen [3] sowie Oliveira e Silva [4, 5] auch neue Ideen zum Einsatz.

1 Collatz-Folgen und erste Abschätzungen zu nicht-trivialen Zyklen

Wir definieren den Collatz-Operator $C : \mathbb{N} \to \mathbb{N}$ via

$$n \mapsto \begin{cases} \frac{n}{2}\,, & \text{für gerade } n, \\ 3n+1\,, & \text{für ungerade } n, \end{cases}$$

der damit für jedes Folgeglied einer Collatz-Folge dessen Nachfolger berechnet. Unter dem trivialen Zyklus verstehen wir die Teilfolge $1, 4, 2$, da $C(1) = 4$, $C(4) = 2$ und schließlich wieder $C(2) = 1$ ist. Ist eine Collatz-Folge einmal bei diesem Zyklus angelangt, durchläuft sie diesen also periodisch und erzeugt dabei keine neuen Folgeglieder mehr.

Die *Collatz-Vermutung* besagt nun, dass für jede Startzahl $k > 0$ die zugehörige Collatz-Folge C_k irgendwann in den trivialen Zyklus übergeht. Diese

Vermutung ist bisher unbewiesen. Wäre sie falsch[1], müsste mindestens eine der beiden folgenden Aussagen eintreten:

- Für eine Startzahl k divergiert die zugehörige Collatz-Folge C_k unbeschränkt.
- Für eine Startzahl k geht die zugehörige Collatz-Folge in einen vom trivialen Zyklus verschiedenen Zyklus über. (Dieser heißt dann nicht-trivial.)

Wir widmen uns hier dem zweiten Punkt und fragen uns, welche Eigenschaften ein solcher nicht-trivialer Zyklus haben müsste, so er existiert.

Sei Ω die Menge aller Folgeglieder in einem solchen nicht-trivialen Zyklus sowie Ω_g und Ω_u die Mengen der geraden bzw. ungeraden Folgeglieder in Ω. Zur einfacheren Bezeichnung seien auch $p := |\Omega_\mathrm{g}|$ und $q := |\Omega_\mathrm{u}|$ als Anzahlen der geraden bzw. ungeraden Folgeglieder in diesem Zyklus eingeführt. Dann ist

$$\begin{aligned}\prod_{n\in\Omega} n &= \prod_{n\in\Omega} C(n) = \prod_{n\in\Omega_\mathrm{g}} C(n) \cdot \prod_{n\in\Omega_\mathrm{u}} C(n) = \prod_{n\in\Omega_\mathrm{g}} \frac{n}{2} \cdot \prod_{n\in\Omega_\mathrm{u}} (3n+1) \\ &= \prod_{n\in\Omega_\mathrm{g}} \left(n \cdot \frac{1}{2}\right) \cdot \prod_{n\in\Omega_\mathrm{u}} \left(n \cdot \left(3 + \frac{1}{n}\right)\right) = 2^{-p} \cdot \prod_{n\in\Omega} n \cdot \prod_{n\in\Omega_\mathrm{u}} \left(3 + \frac{1}{n}\right),\end{aligned}$$

also

$$2^p = \prod_{n\in\Omega_\mathrm{u}} \left(3 + \frac{1}{n}\right).$$

Dies kann man nun weiter abschätzen. Einerseits ist offensichtlich die rechte Seite $\prod_{n\in\Omega_\mathrm{u}} \left(3 + \frac{1}{n}\right)$ größer als $\prod_{n\in\Omega_\mathrm{u}} 3 = 3^q$. Umgekehrt ist sie aber aufgrund der Ungleichung zwischen arithmetischem und geometrischem Mittel

[1] Obwohl es einige, gute heuristische Argumente gibt, an deren Richtigkeit zu glauben: Fasst man etwa den $(3n+1)$-Schritt mit der darauf zwingend folgenden Halbierung zusammen, so kann man dies für genügend große Folgeglieder x als $< \frac{7}{4}x$ abschätzen. Mit dem heuristischen Ansatz, dass dadurch mit gleicher Wahrscheinlichkeit gerade bzw. ungerade Folgeglieder erhalten werden, kann man dann abschätzen, dass der Erwartungswert des Faktors von einem Folgeglied zum nächsten $\sqrt{\frac{7}{4} \cdot \frac{1}{2}} < 1$ ist. Damit fällt die Folge im Erwartungswert immer weiter (bis zum Erreichen der Schranke, unter der die gemachte Abschätzung für den $(3n+1)$-Schritt nicht mehr gilt) und man erwartet eine von der Startzahl k logarithmisch abhängende Anzahl an Folgegliedern, bevor die Folge mit gegen 1 gehender Wahrscheinlichkeit in den trivialen Zyklus eintritt, vgl. auch den Artikel *Einige Hilfsmittel für das* $(3n+1)$*-Problem* von Ingo Althöfer aus $\sqrt{\text{WURZEL}}$ 5/2018.

auch höchstens so groß wie

$$\prod_{n\in\Omega_\mathrm{u}}\left(3+\frac{1}{n}\right) = \left(\sqrt[q]{\prod_{n\in\Omega_\mathrm{u}}\left(3+\frac{1}{n}\right)}\right)^q \le \left(\frac{1}{q}\cdot\sum_{n\in\Omega_\mathrm{u}}\left(3+\frac{1}{n}\right)\right)^q = (3+\mu)^q\,,$$

mit $\mu := \frac{1}{q}\cdot\sum_{n\in\Omega_\mathrm{u}}\frac{1}{n}$. Zusammen erhalten wir also die Ungleichungskette

$$3^q < 2^p \le (3+\mu)^q$$

bzw. nach Logarithmieren zur Basis 2

$$q\cdot\log_2(3) < p \le q\cdot\log_2(3+\mu)$$

und nach Teilen durch q also

$$\log_2(3) < \frac{p}{q} \le \log_2(3+\mu)\,.$$

Demnach muss die rationale Zahl $\frac{p}{q}$ „in der Nähe“ der irrationalen Zahl $\log_2(3)$ liegen, und zwar umso näher, je kleiner das durchschnittliche Reziproke μ eines ungeraden Folgeglieds im nicht-trivialen Zyklus ist.

Aus der Zahlentheorie ist bekannt, dass man die besten rationalen Approximationen einer irrationalen Zahl α aus deren Kettenbruchentwicklung erhält. Anders formuliert: Weiß man, dass eine rationale Zahl $\frac{p}{q}$ näher an α liegt als ein endliches Anfangsstück $\frac{p_n}{q_n}$ der Kettenbruchentwicklung von α mit teilerfremden p_n und q_n, so muss $q > q_n$ (und damit dann auch $p > p_n$) gelten.

Wenn wir also für ein solches Anfangsstück $\frac{p_n}{q_n}$ der Kettenbruchentwicklung von $\log_2(3)$ nachweisen können, dass $\log_2(3+\mu) < \frac{p_n}{q_n}$ ist, so gilt

$$\log_2(3) < \frac{p}{q} \le \log_2(3+\mu) < \frac{p_n}{q_n}$$

und damit besteht dieser Zyklus aus mindestens $p+q > p_n+q_n$ Folgegliedern. Wir werden im Folgenden zeigen, dass für alle nicht-trivialen Collatz-Zyklen gilt, dass

$$\log_2(3+\mu) < \frac{p_{21}}{q_{21}} = \frac{10\,439\,860\,591}{6\,586\,818\,670}$$

mit dem 21. Kettenbruch $\frac{p_{21}}{q_{21}}$ von $\log_2(3)$ ist, woraus dann sofort folgt, dass dieser Zyklus aus mehr als $p_{21}+q_{21}$, also insbesondere mehr als 17 Milliarden Folgegliedern bestehen muss. Um dies zu zeigen, ist also eine genügend genaue Abschätzung von μ vonnöten. Genauer: Es muss $\mu < \frac{1}{2{,}17\cdot 10^{20}}$ gezeigt werden.

Weiß man, dass jede Collatz-Folge mit einer Startzahl $k \leq S$ in den trivialen Zyklus übergeht, so muss jedes Folgeglied eines nicht-trivialen Zyklus, also auch jedes Element von Ω_u, größer sein als S. Dementsprechend gilt dann die einfache Abschätzung

$$\mu = \frac{1}{q} \cdot \sum_{n\in\Omega_\mathrm{u}} \frac{1}{n} < \frac{1}{q} \cdot \sum_{n\in\Omega_\mathrm{u}} \frac{1}{S} = \frac{1}{S}\,.$$

Jedoch war bisher nur die beste Schranke $S_\mathrm{alt} = 5 \cdot 2^{60} \approx 5{,}76 \cdot 10^{18}$ bekannt, welche um einen Faktor 37 zu klein für den gewünschten Satz ist. Damit konnte nur die entsprechende Aussage für das 19. Anfangsstück der Kettenbruchentwicklung von $\log_2(3)$ gezeigt werden, was dann zu dem vorgenannten Ergebnis von mindestens einer Milliarde Folgegliedern führte.

Um Rechenaufwand zu sparen, wurden für den hier zu zeigenden Satz die folgenden beiden Teilaussagen bewiesen, die dann zusammen den gewünschten Sachverhalt liefern:

1. Alle Collatz-Folgen mit Startzahlen $S_\mathrm{neu} \leq 87\cdot 2^{60} \approx 1{,}003\cdot 10^{20}$ gelangen in den trivialen Zyklus.
2. Ist die erste Teilaussage wahr, so gilt für jeden nicht-trivialen Zyklus $\mu < \frac{1}{2{,}17\cdot 10^{20}}$.

Man bemerke, dass μ deutlich kleiner als $\frac{1}{S_\mathrm{neu}}$ ist bzw. umgekehrt S_neu nur knapp halb so groß, wie es sein müsste, um mit der naiven Abschätzung von μ die gewünschte Schranke zu unterschreiten. Nur durch einen neuen Ansatz sowie einige weitere algorithmische Verbesserungen konnte das Projekt mit den verfügbaren Ressourcen in akzeptabler Zeit bewältigt werden. Wie die beiden Teilaufgaben angegangen wurden, soll im Folgenden vorgestellt werden.

2 Algorithmische und Mathematische Überlegungen

Zuerst wollen wir uns der Frage widmen, wie man möglichst effizient für alle Startzahlen unterhalb einer gewissen Schranke nachweisen kann, dass die zugehörigen Collatz-Folgen in den trivialen Zyklus übergehen.

2.1 Effiziente Berechnung von Collatz-Folgen

Als erstes fällt vielleicht auf, dass nach einem ungeraden Folgeglied n direkt ein gerades $3n + 1$ folgt, was danach sofort halbiert wird. Also kann man den „$(3n+1)$-Schritt“ und den nachfolgenden Halbierungs-Schritt zu einem „$\frac{3n+1}{2}$-Schritt“ zusammenfassen. Auf diese Weise wird in jedem Schritt für gerade n die Halbierung $\frac{n}{2}$ und für ungerade n der $\frac{3n+1}{2}$-Schritt durchgeführt. Im Folgenden wollen wir nur noch über diese beiden Arten von Schritten sprechen. Dies ist möglich, da uns ja nicht die Anzahl der Schritte bis zum Eintritt in den trivialen Zyklus interessiert, sondern nur, dass die betrachtete Collatz-Folge in endlich vielen Schritten in diesen gelangt.

Die Restklasse modulo 2 gibt demnach an, welcher von beiden Schritten als nächstes folgt: $2m \mapsto m$ und $2m+1 \mapsto 3m+2$. Danach sind in beiden Fällen wieder beide Restklassen möglich, d. h., die Information wurde vollständig ausgeschöpft. Kennt man jedoch von einem Folgeglied die Restklasse modulo 2^2, so steht auch die Restklasse modulo 2 des Folgeglieds nach dem nächsten Schritt fest und damit auch die Art des übernächsten Schritts. Generell gilt, dass man aus der Kenntnis der letzten ℓ Bits eines Folgeglieds, d. h. dessen Restklasse modulo 2^ℓ, die Art der nächsten ℓ Schritte bestimmen kann. Oder anders formuliert: Jedes Folgeglied $a \cdot 2^\ell + b$ geht über in $a \cdot 3^{f(b)} + g(b)$, wobei die beiden Funktionen f und g nur von der Restklasse b der betrachteten Zahl modulo 2^ℓ abhängen.

Dies hat gleich mehrfache Auswirkungen. Zum einen kann man diesen Fakt nutzen und die Berechnung dadurch beschleunigen, dass man gleich mehrere Einzelschritte zu einem Mehrfachschritt zusammenfasst. Dazu berechne man zuvor für alle möglichen Restklassen $b \pmod{2^\ell}$ die entsprechenden Werte $f(b)$ und $g(b)$ und tabelliere diese. Während der eigentlichen Berechnung einer Collatz-Folge muss man dann nur die Restklasse des betrachteten Folgeglieds modulo 2^ℓ bestimmen (was sehr effizient geht, da dies genau die letzten ℓ Bits des Folgeglieds in seiner vom Computer nativ verwendeten Binärdarstellung sind) und kann dann in einer Rechnung daraus direkt den Wert des Folgeglieds dieser Collatz-Folge bestimmen, welcher ℓ Schritte später erscheint. Dies wird im verwendeten Programm mit $\ell = 10$ getan, wobei dieser Wert so gewählt ist, dass alle notwendigen Tabellen noch im L2-Cache des Prozessors zwischengespeichert werden können, also direkt der CPU ohne lange Ladezeiten zur Verfügung stehen.

Zum anderen kann diese Beobachtung aber auch genutzt werden, um direkt ganze Restklassen von Startzahlen gemeinsam zu betrachten und große Anzahlen davon direkt auszuschließen: Dazu stelle man zuerst fest, dass es genügt, für jede Startzahl $k > 1$ nachzuweisen, dass die zugehörige Collatz-Folge C_k irgendwann ein Folgeglied $m < k$ enthält. Denn ab dann schließen sich dort die Glieder der Collatz-Folge C_m an, für welche nach dem Induktionsgedanken schon als gezeigt gelten kann, dass sie in den trivialen Zyklus mündet. Hat man also für alle Startzahlen $1 < k \leq S$ gezeigt, dass die jeweilige Collatz-Folge irgendwann einen Wert $< k$ annimmt, ist für all diese Startzahlen damit gezeigt, dass die Collatz-Folgen in den trivialen Zyklus übergehen.

Die Erkenntnis, dass man aus der Restklasse der Startzahl modulo 2^ℓ direkt den Wert nach ℓ Schritten berechnen kann, führt dazu, dass man z. B. auch direkt für alle Startzahlen einer bestimmten Restklasse feststellen kann, dass dieser neue Wert kleiner sein muss als die Startzahl, man diese also gar nicht mehr weiter betrachten muss. So geht etwa jede Startzahl der Form $a \cdot 2^1 + 0$ nach einem Schritt über in $a \cdot 3^0 + 0 < a \cdot 2^1 + 0$, muss also nach dem Induktionsgedanken nicht weiter betrachtet werden. Von den ungeraden Startzahlen fallen aber auch z. B. direkt die der Form $a \cdot 2^2 + 1$ heraus, da sie in zwei Schritten in $a \cdot 3^1 + 1$ übergehen, was für $a > 0$ echt kleiner als die Startzahl ist.

So kann man sukzessive Restklassen modulo immer größerer Zweierpotenzen anschauen: Gelangt eine Restklasse modulo 2^ℓ in den ersten ℓ Schritten zu einem Wert kleiner der Startzahl, kann man abbrechen. Wenn nicht, betrachtet man die zwei Restklassen modulo $2^{\ell+1}$, in die sich die betrachtete aufspaltet, und überprüft jeweils, ob diese im $(\ell+1)$-ten Schritt auf einen Wert kleiner der Startzahl führen. Diesen Prozess des Siebens von Restklassen führt man bis zur am Anfang vorgegebenen Maximaltiefe durch. Von den übrig bleibenden Restklassen sind dann alle Startzahlen $\leq S$ zu erzeugen und einzeln zu betrachten.

Dieser Sieb-Prozess verkleinert den Suchraum deutlich. So gibt etwa Oliveira e Silva in [4] an, dass von den etwas über 4 Milliarden Restklassen modulo 2^{32} nur ca. 41 Millionen *nicht* ausgesiebt werden. Dies entspricht einer Quote von ca. einem Prozent. Wir bemerken, dass durch dieses Vorgehen die Startzahlen in der Gesamtberechnung nicht in aufsteigender Reihenfolge bis S erzeugt werden. Dass eine Collatz-Folge C_k in den trivialen Zyklus übergeht, ist damit erst dann bewiesen, wenn die Betrachtung für alle Restklassen vollständig ist.

In der letzten Überlegung haben wir aus den Eigenschaften einer Zahl auf ihre Collatz-Nachfolger geschlossen. Man kann dies aber auch umgekehrt tun und die Collatz-Vorgänger betrachten. Auch hier gilt: Besitzt für eine Startzahl k ein Folgeglied n der zugehörige Collatz-Folge C_k einen „Collatz-Vorgänger“ $m < k$, d. h., auf m folgt in endlich vielen Schritten n, so kann auch hier die Berechnung abgebrochen werden, denn die Collatz-Folgen C_k und C_m sind ab dem Folgeglied n identisch. Also geht C_k genau dann in den trivialen Zyklus über, wenn dies C_m auch schon getan hat. Dies trifft etwa auf alle Startzahlen der Form $k = a \cdot 2^6 + 15$ zu. Zwar erhält man aus diesen in sechs Schritten die größeren Werte $n = a \cdot 3^4 + 20$. Da aber auch die Startzahlen der Form $m = a \cdot 2^5 + 7 < k$ in fünf Schritten zum gleichen Wert n gelangen, muss nur C_m weiter überprüft werden, während die weitere Berechnung von C_k nach dieser Feststellung abgebrochen werden kann.

Diese Untersuchung auf mögliche kleinere Collatz-Vorgänger wurde nach Wissen des Autors hier erstmals in dieser intensiven Form betrieben. Sie führt zu einer weiteren deutlichen Suchraum-Einschränkung, etwa bei einer Siebtiefe von 32 Bit auf nur noch ca. 24 Millionen verbliebene Restklassen, was einer Verbesserung um den Faktor 1,7 gegenüber der reinen „Vorwärtssuche“ entspricht. Bei höheren Siebtiefen verstärkt sich der Einfluss weiter.

Im für diese Suche geschriebenen und verwendeten Programm wird zuerst bis zur Tiefe von 32 Bit gesiebt. Diese Restklassen dienen dann zur Verteilung der Berechnung über verschiedene Prozesse auf einem PC sowie global über viele verschiedene Computer. Jede einzelne dieser Restklassen modulo 2^{32} wird weiter bis zur Tiefe von 58 Bit gesiebt, also in Restklassen modulo 2^{58} zerlegt. Von diesen (bezogen auf alle 2^{58}) wurden im Schnitt nur ca. ein Promille nicht ausgesiebt, d. h., es mussten dann alle Startzahlen $< S$ in dieser Restklasse gebildet und einzeln betrachtet werden. Für jede einzelne dieser Startzahlen werden nun die oben aufgeführten Mehrfachschritte durchgeführt, bis ein Folgeglied kleiner der Startzahl (oder eine maximale Schrittzahl) erreicht wurde. (Das Maximum an Schritten wurde aber nie erreicht, sodass alle Collatz-Folgen mit Startzahl $\leq S$ in den trivialen Zyklus münden.)

Da die Folgeglieder sehr schnell den Zahlenbereich bis 2^{64}, welcher mit moderner 64-Bit-Prozessor-Architektur sehr effizient zu bearbeiten ist, verlassen, müsste eigentlich mit höherer Präzision, etwa 128-Bit-Arithmetik, gerechnet werden. Dies ist durch den verwendeten Compiler gcc auch durchaus möglich, da dieser über den Standard hinaus auch 128-Bit-Ganzzahl-Datentypen zur Verfügung stellt. Jedoch sind diese nur Soft- und nicht Hardware-seitig implementiert, haben also deutliche Performance-Nachteile gegenüber 64-Bit-Operationen.

Um diesen Flaschenhals zu umgehen, sollte die Anzahl an notwendigen 128-Bit-Operationen möglichst stark minimiert werden. Dies gelingt dadurch, dass für die Durchführung der nächsten 60 Einzelschritte (also 6 Mehrfachschritte, da je 10 Einzel- zu einem Mehrfachschritt zusammengefasst wurden) nur die Kenntnis der letzten 60 Bit eines Folgeglieds von Nöten ist. All diese Operationen können also modulo 2^{64} durchgeführt werden, was sehr effizient möglich ist. Das garantiert, dass die Art der nächsten 60 Schritte (ob Halbierung oder „$\frac{3n+1}{2}$“) korrekt berechnet wird, nicht aber, wie die davor liegenden höherwertigen Bits korrekt lauten. Insbesondere erhält man also auf diese Weise keine Aussage über die Größe der jeweils aktuell betrachteten Folgeglieder. Dem jedoch kann man beikommen, indem man parallel diese Folgeglieder auch näherungsweise durch Gleitkomma-Werte darstellt: Aus der Ganzzahlberechnung mit den letzten Bits wird die Art der nächsten Schritte klar, womit dann die Gleitkomma-Berechnung näherungsweise durchgeführt und die Größe des jeweiligen Folgeglieds bestimmt werden kann. Ist dieser Näherungswert kleiner als ein bestimmter Schwellenwert unterhalb der Startzahl (der so gewählt wird, dass die betrachtete Zahl trotz möglicherweise auftretender Rundungsfehler während der Berechnung trotzdem beweisbar sicher kleiner ist als die Startzahl), kann die Berechnung mit dem obigen Argument an dieser Stelle abgebrochen werden. Ist dies aber auch nach den 60 Einzelschritten noch nicht der Fall, muss die Berechnung mittels exakter Ganzzahl-Arithmetik wiederholt werden, sodass wieder die volle Information über das jeweilige Folgeglied für die nächsten Schritte vorliegt. Da dies aber nur in ca. 2 % aller Fälle nötig wurde, ergibt sich trotz des ggf. doppelten Aufwands in diesen Fällen insgesamt eine deutliche Zeitersparnis. Auch dieser Ansatz scheint erstmals im vorliegenden Projekt intensiver verfolgt worden zu sein.

Auch weitere Optimierungen auf eher niedrigem Abstraktionslevel (wie die Feststellung, dass Multiplikation oder Division mit/durch Zweierpotenzen effizienter durch einen Links- bzw. Rechts-Shift der Bits durchgeführt werden, oder man ggf. Schleifen mit fester Länge abrollt, um mögliche Verzweigungen, ob der Laufindex schon seine obere Schranke erreicht hat, zu vermeiden und so Instruktionen zu sparen) fanden Eingang in das C-Programm, welches Schritt für Schritt weiter optimiert wurde. Auf diese Einzelheiten soll hier aber nicht weiter eingegangen werden.

Schlussendlich wurde das Programm am 14. August 2017 dem Distributed-Computing-Projekt *yoyo@home* des Rechenkraft e. V. hinzugefügt [8], sodass sich über das Internet insgesamt über 1000 Freiwillige an den Berechnungen beteiligen konnten. In einer Zeit von 40 Tagen wurden schließlich etwa $2{,}5 \cdot 10^{19}$ Rechenoperationen durchgeführt[2], was etwa 33 Jahren Rechenzeit auf einem Prozessor-Kern eines gängigen Büro-PCs entspricht. Dabei wurde jede Aufgabe von zwei unterschiedlichen Rechnern ausgeführt und zum Vergleich validiert, um zufällige Berechnungsfehler ausschließen zu können. (Zum Vergleich: Oliveira e Silva benötigte ca. 40 CPU-Jahre zur Berechnung und weitere 40 zur Validierung seiner Ergebnisse aus dem Jahr 2008, welche einen ca. 17 mal kleineren Suchraum betrachtete. Dabei

[2] Das heißt, im Schnitt wurde für jede Startzahl bis 10^{20} nur eine Viertel-Rechenoperation benötigt um nachzuweisen, dass die jeweilige Collatz-Folge in den trivialen Zyklus mündet.

ist festzustellen, dass sich die Rechenleistung der einzelnen CPU-Kerne in den letzten 10 Jahren kaum noch gesteigert hat, sondern die höhere Rechenleistung seitdem insbesondere durch eine höhere Anzahl an Kernen je CPU erreicht werden konnte. Der weitaus größte Teil der Verringerung der benötigten Rechenzeit geht auf die algorithmischen Verbesserungen sowie den standardmäßigen Einsatz einer 64-Bit-Architektur, die sich damals erst etabliert hatte, zurück.)

Neben der Erkenntnis, dass alle Collatz-Folgen mit Startzahlen $\leq 87 \cdot 2^{60}$ in den trivialen Zyklus übergehen, was den bisherigen Maximalwert um den Faktor 17,4 übertrifft, konnten durch den erfolgreichen Abschluss des Projekts insgesamt vier neue Pfad-Rekorde ermittelt werden, d. h. Startzahlen k, deren maximales Element in ihrer Collatz-Folge C_k größer ist als das aller Collatz-Folgen mit Startzahlen $< k$, siehe dazu [6]. Darüber hinaus konnte *en passent* auch eine weitere Schranke verbessert werden: Fasst man wie in der bisherigen Betrachtung einen „$(3n+1)$-Schritt“ mit dem darauf zwingend folgenden Halbierungs-Schritt zusammen, so kann man, beginnt man einen nicht-trivialen Zyklus o. B. d. A. beim kleinsten Folgeglied, diesen in aufeinanderfolgende Abschnitte von wechselseitig streng monoton steigenden bzw. fallenden Teilfolgen unterteilen. (Das müssen dann natürlich gleich viele jeder Sorte sein.) Mit der hier neu gezeigten oberen Schranke $S \approx 1{,}003 \cdot 10^{20}$, für die alle Collatz-Folgen mit Startzahl $\leq S$ in den trivialen Zyklus münden, konnte die Anzahl der mindestens in einem nicht-trivialen Zyklus vorkommenden streng monoton steigenden (und analog fallenden) Abschnitte von 75 auf 77 erhöht werden, siehe dazu [7]. (Die Autoren von [7] zeigen 2005 eine Schranke von mindestens 68 solcher Abschnitte, halten aber eine Verbesserung der Abschätzung auf 77 für quasi unmöglich, da sie diese erst für das Jahr 2419 prognostizieren. Zum Glück hat es seitdem einige deutliche Weiterentwicklungen sowohl in der Hardware wie auch den Algorithmen in der Software gegeben, sodass deutlich bessere Ergebnisse mittlerweile möglich sind.)

2.2 Abschätzungs-Verbesserung zu μ

Wir erinnern uns, dass $\mu = \frac{1}{q} \cdot \sum_{n \in \Omega_\mathrm{u}} \frac{1}{n}$ die Größe des durchschnittlichen Reziproken $\frac{1}{n}$ der ungeraden Folgeglieder n eines nicht-trivialen Zyklus ist, dass wir $\mu < \frac{1}{2{,}17 \cdot 10^{20}} =: s$ zeigen wollen und dafür voraussetzen dürfen, dass alle Collatz-Folgen mit Startzahl $\leq S = 87 \cdot 2^{60} \approx 1{,}003 \cdot 10^{20}$ in den trivialen Zyklus münden, dies also auch der Fall ist für alle Collatz-Folgen, die mindestens ein Folgeglied $\leq S$ besitzen.

Die Idee ist nun, dass wir den nicht-trivialen Zyklus so in Abschnitte in der Folge unterteilen, dass wir für jeden dieser Abschnitte zeigen können, dass der Durchschnitt der Reziproken der darin enthaltenen ungeraden Folgeglieder kleiner als s ist. Denn gilt dies für alle Abschnitte des Zyklus, so natürlich auch für den gesamten Zyklus.

Wir können o. B. d. A. annehmen, dass jeder dieser Abschnitte mit einem ungeraden Folgeglied anfängt. (Für Abschnitte ohne ungerades Folgeglied gilt die Aussage natürlich sofort auch.) Die einzelnen Abschnitte können dabei eine variable Länge kleiner oder gleich einem vorgegebenen Maximalwert besitzen.

Ist das erste Folgeglied n eines solchen Abschnitts z. B. kongruent $5 \pmod 8$, also $n = 8k + 5$, so lautet das nächste Folgeglied $\frac{3n+1}{2} = 12k + 8$, was dann nach zwei weiteren Halbierungen auf $3k + 2$ führt. Da also auch $3k + 2$ Teil des Zyklus ist und damit $3k + 2 > S > 1{,}003 \cdot 10^{20}$ gilt, ist demnach $n = 8k + 5 > \frac{8}{3} \cdot 10^{20} > \frac{1}{s}$. Bricht man also den Abschnitt schon nach n ab, so gilt für diesen, dass das durchschnittliche Reziproke eines ungeraden Folgeglieds in diesem Abschnitt genau $\frac{1}{n}$ und damit kleiner als s ist, was wir hier zeigen wollten. Der nächste Abschnitt startet dann mit dem nächsten ungeraden Folgeglied.

Auf diese Weise wollen wir für die ersten Folgeglieder der Abschnitte eine vollständige Fallunterscheidung nach ihren Restklassen modulo Zweierpotenzen durchführen, wobei jede ungerade natürliche Zahl in eine dieser Restklassen fällt. In jedem Fall wird dabei folgen, dass, liegt das erste Folgeglied eines Abschnitts in der gerade betrachteten Restklasse, so folgt, dass das durchschnittliche Reziproke der ungeraden Folgeglieder in diesem Abschnitt kleiner als s ist. Den Fall, dass das erste Folgeglied eines Abschnitt kongruent 5 modulo 8 ist, haben wir im vorgenannten Beispiel betrachtet.

Generell gehen wir hier ähnlich vor wie beim Sieben der Collatz-Folgen nach den Restklassen ihrer Startzahlen modulo aufsteigenden Zweierpotenzen im vorherigen Teil dieses Artikels: Wir betrachten für das erste Folgeglied eines Abschnitts eine Restklasse $r \pmod{2^\ell}$. Dann kennen wir auch die ℓ nachfolgenden Folgeglieder. Können wir aufgrund der Größenabschätzungen nun schließen, dass es ein $1 \le t \le \ell+1$ gibt, sodass das durchschnittliche Reziproke der ungeraden Folgeglieder unter den ersten t hier betrachteten kleiner ist als s, so brechen wir den Abschnitt nach dem t-ten Folgeglied ab und haben damit für diese Restklasse des ersten Folgeglieds dieses Abschnitts das Gewünschte gezeigt. (Der Abschnitt besteht dann aus t Folgegliedern.) Andernfalls erhöhe man ℓ um Eins und betrachte die beiden Teilfälle, dass das erste Folgeglied kongruent $r \pmod{2^{\ell+1}}$ bzw. kongruent $r + 2^\ell \pmod{2^{\ell+1}}$ ist. (Dies geschieht solange, bis entweder alle Fälle abgedeckt sind, oder aber eine maximale Zweierpotenz als Modul erreicht wurde. Im konkreten Fall wurden die Werte so gewählt, dass die Berechnung durch Abdeckung aller Teilfälle abgeschlossen werden konnte.)

Neben der Vorwärts-Betrachtung, welcher die Größe der nächsten Folgeglieder einbezieht, wurde auch wieder eine Rückwärts-Betrachtung durchgeführt: Kann man für ein erstes Folgeglied n bzw. einen seiner Nachfolger einen anderen Collatz-Vorgänger $m < n$ finden, dann führt also auch die Collatz-Folge mit Startzahl m in den betrachteten nicht-trivialen Zyklus. Also muss auch schon $m > S$ gelten. Das dazu im vorherigen Teil des Artikels beschriebene Beispiel reicht aus, um damit die Abschnitte, die mit Folgegliedern der Form $n = a \cdot 64 + 15$ beginnen, erfolgreich zu betrachten: Diese führen in sechs Schritten auf die Folgeglieder $a \cdot 84 + 20$, die auch in fünf Schritten aus $m = a \cdot 32 + 7 < \frac{n}{2}$ folgen. Da damit auch $m > S$ gelten muss, ist insbesondere $n > 2S > 2 \cdot 10^{20}$ und das nächste ungerade Folgeglied $\frac{3n+1}{2} > 3 \cdot 10^{20}$. Brechen wir den Abschnitt nach diesen beiden Folgegliedern ab, ist also das durchschnittliche Reziproke eines ungeraden Folgeglieds in diesem Abschnitt kleiner als $\frac{1}{2} \cdot \left(\frac{1}{2 \cdot 10^{20}} + \frac{1}{3 \cdot 10^{20}}\right) = \frac{1}{2{,}4 \cdot 10^{20}} < s$, wie gewünscht.

Diese beiden Strategien führen dazu, dass bei der Betrachtung der noch nicht abgeschlossenen Restklassen modulo 2^ℓ, in welchen sich das erste Folgeglied eines Abschnitts befinden kann, von dem noch nicht die gewünschte Eigenschaft gezeigt werden konnte, für einige der Nachweis nun gelingt, während die übrigen jeweils zwei Unterfälle bei der Betrachtung modulo $2^{\ell+1}$ erzeugen. Tatsächlich steigt die Anzahl der offenen Restklassen auf diese Weise exponentiell mit ℓ an, wenn auch mit einer Basis deutlich kleiner Zwei.

Jedoch wurde hierbei bisher das Zusammenspiel aus absoluter Größe und Restklasse des betrachten ersten Folgeglieds eines Abschnitts nicht weiter beachtet: In allen Fällen ging es um Vielfache dieses ersten Folgeglieds, von denen das kleinste dann nach unten durch S abgeschätzt wurde und so die gewünschte Aussage folgte. Tatsächlich wissen wir aber mehr:

Gilt für ein erstes Folgeglied n eines Abschnitts, dass es kongruent $r \pmod{2^\ell}$ mit $0 < r \leq 2^\ell$ ist, so wissen wir nicht nur, dass es (als Teil des nicht-trivialen Zyklus) größer als S ist. Wir wissen z. B. auch, dass $n \geq r$ ist. Ist die Zweierpotenz, modulo der wir gerade die Betrachtung durchführen, schon groß, so kann auch die Restklasse r hier große Werte annehmen. Und ist $r > \frac{1}{s}$, so sind wir mit dieser Restklasse fertig, da $\frac{1}{n} \leq \frac{1}{r} < s$ gilt, der betrachtete Abschnitt also nach einem Folgeglied abgebrochen werden kann und dann das Gewünschte erfüllt.

Und sollte für dieses $n \equiv r \pmod{2^\ell}$ zwar nicht $r > \frac{1}{s}$ gelten, sondern $r \leq S = 87 \cdot 2^{60} \approx 1{,}003 \cdot 10^{20}$, so kann nicht $n = r$ gelten, da wir ja wissen, dass es dann in den trivialen Zyklus führen würde. Also muss für diese Restklassen ein erstes Folgeglied eines Abschnitts mindestens den Wert $r+2^\ell$ besitzen. Und dieser Wert ist für genügend große ℓ wiederum groß genug, um mittels der gerade zuvor durchgeführten Überlegung das Gewünschte für diesen Abschnitt zu zeigen.

Auch lassen sich diese Zusammenhänge nicht nur für das erste Folgeglied, sondern natürlich auch die weiteren, und deren Vorgänger anwenden. Zusammen führt dies dazu, dass die Zahl der noch offenen Restklassen bei diesen Betrachtungen bis $\ell = 60$ exponentiell ansteigt, dann bis $\ell = 70$ kaum noch wächst und schließlich wieder zu sinken beginnt. (Insbesondere kann ab $\ell = 68$ die Zahl der noch offenen Restklassen gar nicht mehr steigen, da dann für eine noch offene Restklasse $r \pmod{2^\ell}$ der zweite Unterfall $r + 2^\ell \pmod{2^{\ell+1}}$ wegen $2^\ell > \frac{1}{s}$ automatisch erfüllt wird, also höchstens ein Unterfall mod $2^{\ell+1}$ übrig bleibt.)

Auf einem Server wurde in einer Zeit von einer Woche nun nachgeprüft, dass tatsächlich alle Restklassen sukzessive abgeschlossen werden können, also für jedes mögliche erste Folgeglied eines Abschnitts folgt, dass das durchschnittliche Reziproke dieses Abschnitts kleiner ist als der gewünschte Maximalwert s. Dabei wurde den Folgen eine Maximallänge von 250 Folgegliedern zugebilligt (d. h. eine Untersuchung der Restklasse des ersten Folgeglieds eines Abschnitts bis maximal modulo der Zweierpotenz 2^{250}), wobei dies nie erreicht wurde. Alle Abschnitte brechen nach weniger als 250 Folgegliedern ab.

Da dies für alle Abschnitte gezeigt werden konnte, gilt auch für den gesamten Zyklus (bzw. auch mehrere Durchläufe davon hintereinander), dass das durchschnittliche Reziproke der ungeraden Folgeglieder darin kleiner ist als s, unter der Annahme, dass jede Collatz-Folge mit Startzahl $\leq S = 87 \cdot 2^{60} \approx 1{,}003 \cdot 10^{20}$ in den trivialen Zyklus übergeht.[3]

Da wir aber diese Voraussetzung im vorherigen Teil dieses Artikels schon gezeigt haben, folgt mit den dort angegebenen weiteren Überlegungen der

Satz. *Ein nicht-trivialer Collatz-Zyklus besteht aus mehr als 17 026 679 261, also mehr als 17 Milliarden Folgegliedern.*

Literatur

[1] Shalom Eliahou: *The $3x+1$ problem: new lower bounds on nontrivial cycle lengths*. Discrete Mathematics 118(1–3), S. 45–56, 1993.

[2] Christian Hercher (cyrix), Gerhard Pracht (gonz), (Amateur) et al.: *MP-Forum: Die Collatz-Folge programmieren (Matroids Matheplanet)*. http://matheplanet.com/matheplanet/nuke/html/viewtopic.php?topic=223140&start=0

[3] Mit einer höheren Rechendauer wäre sicherlich auch noch eine weitere Verkleinerung der Schranke S möglich gewesen, um den Satz zu zeigen. Es erschien aber auch aus Gründen der Außendarstellung sinnvoll, bei einem Wert von knapp über 10^{20} zu bleiben, sodass hierfür keine weiteren Anstrengungen unternommen wurden.

[3] Gary T. Leavens, Mike Vermeulen: $3x + 1$ *search programs*. Computers & Mathematics with Applications 24(11), S. 79–99, 1992.

[4] Tomás Oliveira e Silva: *Maximum excursion and stopping time record-holders for the* $3x+1$ *problem: Computational results*. Mathematics of Computation 68(225), S. 371–384, 1999.

[5] Tomás Oliveira e Silva: *Empirical verification of the* $3x+1$ *and related conjectures*. In: *The Ultimate Challenge: The* $3x + 1$ *Problem*, hg. von Jeffrey C. Lagarias. S. 189–207, American Mathematical Society, 2010.

[6] Eric Roosendaal: $3x + 1$ *Path Records*. http://ericr.nl/wondrous/pathrecs.html

[7] John Simons, Benne de Weger: *Theoretical and computational bounds for* m*-cycles of the* $3n + 1$*-problem*. Acta Arithmetica 117(1), S. 51–70, 2005.

[8] Yoyo@home: Nontrivial Collatz Cycle wu status.
http://www.rechenkraft.net/yoyo/y_status_col.php