Untersuchung der Eigenschaften spezieller ungerader Zahlen auf Basis des Collatz-Gitters

Tobias Goltermann

7. März 2021

Gymnasium Carolinum Osnabrück

Kurzfassung

Das Collatz-Problem ist ein ungelöstes zahlentheoretisches Problem, mit dem sich seit den 1930er Jahren Mathematiker und Amateure aus aller Welt beschäftigen. Beim letztjährigen Jugend-Forscht Wettbewerb wurde das Collatz-Gitter vorgestellt, welches eine neue, übersichtliche Darstellung des Collatz-Problems ermöglicht. Damit konnten weiterführende Erkenntnisse gesammelt werden, die nun erneut aufgegriffen werden. Dabei werden Regelmäßigkeiten in dem Auftreten bestimmter ungerader Zahlen gesucht und überprüft. Schlussendlich können die Anzahl und die Dauer der Perioden beschrieben und mit Hilfe von Computerprogrammen berechnet werden. Auch die Zahldarstellung im Binärsystem wird untersucht, wobei die Ergebnisse der direkten Untersuchung des Collatz-Gitters verifiziert werden konnten.

Inhaltsverzeichnis

Kapitel 1

Was ist das Collatz-Problem?

Das Collatz-Problem ist ein zahlentheoretisches Problem der Mathematik. Es wurde in den 1930er Jahren von Lothar Collatz entwickelt, welcher sich mit zahlentheoretischen Funktionen beschäftigte. Mittlerweile ist es auch unter anderen Namen, wie zum Beispiel Katuni-, Syracuse-, oder 3x+1-Problem, bekannt.[1]
Im Mittelpunkt steht dabei die Collatz-Funktion $C(n), n \in \mathbb{N}$, die wie folgt definiert werden kann:

$$C(n) = \begin{cases} \frac{n}{2}, & n \equiv 0 \pmod 2 \\ 3n+1, & n \equiv 1 \pmod 2 \end{cases}$$

Da $C(n)$ in jedem Fall eine weitere natürliche Zahl erzeugt, lässt sich die Funktion mehrfach ausführen, also iterieren.[2]

$$C^0(n) = n, \quad C^1(n) = C(n), \ \ldots \ C^k(n) = C(C^{k-1}(n))$$

Eine Trajektorie der Zahl n ist dann die Folge $(n, C(n), C^2(n), C^3(n), \ldots)$.
Oftmals wird auch für den Fall $n \equiv 1 \pmod 2$ zusätzlich durch 2 dividiert, da $3n+1$ in jedem Fall eine gerade Zahl ergibt und so eine Iteration der Funktion eingespart werden kann. An dem Verhalten der Trajektorie ändert das jedoch nichts.
Die Trajektorien der ersten n sehen folgendermaßen aus:

$$\begin{aligned} n = 1 &: (1, 4, 2, 1, \ldots) \\ n = 2 &: (2, 1, 4, 2, 1, \ldots) \\ n = 3 &: (3, 10, 5, 16, 8, 4, 2, 1, \ldots) \\ n = 4 &: (4, 2, 1, \ldots) \\ n = 5 &: (5, 16, 8, 4, 2, 1, \ldots) \end{aligned}$$

Es ist zu erkennen, dass die Trajektorien in einen Zyklus, bestehend aus $(4, 2, 1, \ldots)$ gelangen. Daher kam Collatz zu der Vermutung, dass die Trajektorie jedes $n \in \mathbb{N}$ in eben dieser Periode endet. Bislang ist nicht bekannt, ob diese Vermutung wahr ist.[3]

[1] vgl. Lagarias 1985, S. 3.
[2] vgl. Wolke o.D., S. 1.
[3] vgl. Lagarias 1985, S. 3.

Kapitel 2

Bisherige Forschung

2.1 Stoppzeiten

Schon früh wurde das Wachstum der Zahlen durch die Collatz-Funktion untersucht. Dabei wurde vor allem darauf geachtet, wie stark die Startzahlen im Verlauf einer Trajektorie sinken. Meistens wird dafür eine modifizierte Form T der Collatz-Funktion verwendet, die gegenüber der ursprünglichen Funktion eine Iteration einspart.

$$T(n) = \begin{cases} \frac{n}{2}, & n \equiv 0 \pmod 2 \\ \frac{3n+1}{2}, & n \equiv 1 \pmod 2 \end{cases}, n \in \mathbb{N} \tag{2.1}$$

Die *Stoppzeit* $\sigma(n)$ ist dann das kleinste k, für das $T^k(n) < n$ gilt. Wenn es kein solches k gibt, gilt $\sigma(n) = \infty$. Wenn jede natürliche Zahl $n \geq 2$ eine endliche Stoppzeit hätte, müsste es immer ein $T^k(n) = 1$ geben, sodass die Collatz-Vermutung wahr wäre.[1]
Die Mathematiker Terras und Everett haben unabhängig voneinander bewiesen, dass der Grenzwert

$$F(k) = \lim_{x \to \infty} \frac{1}{x} \#\{n : n \leq x \text{ und } \sigma(n) \leq k\}$$

existiert und gegen 1 geht, wenn k gegen ∞ läuft. Das # zeigt an, dass die Anzahl der Elemente in der Menge gezählt wird. Das bedeutet, dass fast alle natürlichen Zahlen eine endliche Stoppzeit haben.[2]

2.2 Weitere Grenzen

Relativ neu ist das Ergebnis, welches Terence Tao erzielt hat. Er hat eine Begrenzung des Wachstums finden können, die für fast alle natürlichen Zahlen gültig ist. Für jede Funktion f, die gegen Unendlich wächst, gilt für fast alle n

$$T^k(n) < f(n).$$

Dabei ist wichtig zu betonen, dass es egal ist, wie langsam die Funktion f wächst.[3] Die Trajektorie von n wird also fast begrenzt. Die Collatz-Vermutung ist damit leider noch nicht bewiesen, aber „this is about as close as one can get to the Collatz conjecture without actually solving it.“(Tao 2020)

[1] Lagarias 1985, S. 4–5.
[2] Ebd., S. 6.
[3] Tao 2020, S. 54.

2.3 Bestehendes Jugend-Forscht Projekt

In dem Projekt „Erforschen der Struktur des Collatzproblems anhand einer neuen Darstellungsform“ habe ich das Collatz-Gitter (Abbildung 2.1) vorgestellt. Dieses Gitter ist eine unendlich große Tabelle, die in der ersten Spalte alle ungeraden Zahlen enthält, während die Felder nach rechts hin durch das Verdoppeln der Zahlen entstehen. Das Collatz-Problem kann dann mithilfe des Gitters dargestellt werden und es konnten Regelmäßigkeiten in der Art und Weise wie die Collatz-Funktion in dem Gitter arbeitet festgestellt werden.

1	2	4	8	16	32	64	128
3	6	12	24	48	96	192	384
5	10	20	40	80	160	320	640
7	14	28	56	112	224	448	896
9	18	36	72	144	288	576	1152
11	22	44	88	176	352	704	1408
13	26	52	104	208	416	832	1664
15	30	60	120	240	480	960	1920

Abbildung 2.1: Ausschnitt aus dem Collatz-Gitter

Das Collatz-Gitter wird gebildet, indem zu jeder ungeraden natürlichen Zahl i eine Menge M_i gebildet wird.

$$M_i = \{i \cdot 2^n : n \in \mathbb{N}_0\}$$

Dann werden alle Elemente einer bestimmten Menge M_i nach ihrer Größe geordnet von links nach rechts in eine Zeile geschrieben. Die verschiedenen Zeilen werden so untereinandergeschrieben, dass die erste Spalte nur die ungeraden Zahlen ihrer Größe nach geordnet von oben nach unten enthält. So entsteht das Collatz-Gitter. Es enthält dann alle natürlichen Zahlen, was auch bewiesen werden konnte.

Nun kann man in dem Collatz-Gitter den Verlauf einer Trajektorie verfolgen. Wenn eine natürliche Zahl n ungerade ist, dann bezeichnet man den Schritt zu $C(n)$ als *Sprung*. Wenn n jedoch gerade ist, liegt $C(n)$ genau ein Feld links von n. Diesen Schritt bezeichnet man dann als *Abstieg*. Die Anzahl der einzelnen Abstiege nach einem Sprung bis zur nächsten ungeraden Zahl bezeichnet man als *Abstiegslänge* $A(n)$. Da zuvor ein Sprung stattfindet, ist $A(n)$ nur für ungerade n definiert. Die nächste ungerade Zahl in der Trajektorie von n soll dann $I(n)$ heißen. Folglich gilt

$$I(n) = \frac{C(n)}{2^{A(n)}}$$

Die Anzahl an Reihen, die bei einem Sprung übersprungen werden, wird als *Sprunglänge* bezeichnet. Wenn man von der ungeraden natürlichen Zahl n ausgeht, ist die Sprunglänge

$$S(n) = \frac{I(n) - n}{2}. \tag{2.2}$$

Man beachte, dass die Sprunglänge negative Werte annimmt, wenn $I(n) < n$ gilt und der Sprung somit „nach oben“ stattfindet, also in eine Zeile, die näher an der 1 liegt als die Ausgangsreihe. Zu den erzielten Ergebnissen gehört die Erkenntnis, dass

$$I(n) = I(4n+1)$$

gilt. Alle Zahlen, für die $I(n) = 1$ gilt, können mithilfe der rekursiven Funktion

$$a_n = 4a_{n-1} + 1, \qquad a_0 = 1, n \in \mathbb{N} \tag{2.3}$$

erzeugt werden. Außerdem hat jede zweite ungerade Zahl die Abstiegslänge 1, jede vierte ungerade Zahl die Abstiegslänge 2, jede achte die Abstiegslänge 4, usw.. Dazu wurden Terme gefunden, die alle Zahlen mit einer minimalen Abstiegslänge beschreiben. Dabei ist $n_i \in \mathbb{N}_0$ eine beliebige natürliche Zahl.

$$\begin{aligned}
A(n) &\geq 1 : 2n_1 + 1 \\
A(n) &\geq 2 : 4n_2 + 1 \\
A(n) &\geq 3 : 8n_3 + 5 \\
A(n) &\geq 4 : 16n_4 + 5 \\
A(n) &\geq 5 : 32n_5 + 21
\end{aligned}$$

Diese Terme entstehen, indem in einem Term die Variable n_i abwechselnd durch $2n_{i+1}$ und $2n_{i+1} + 1$ ersetzt wird, um so eine beliebig große Abstiegslänge zu konstruieren. Bei einer ungeraden Zahl k ist die Zahl $l = k + 2^{A(k)+1}$ die nächste Zahl mit der gleichen Abstiegslänge. Die Differenz der Sprunglängen beider Zahlen ist dann

$$S(l) - S(k) = 3 - 2^{A(k)}.$$

Kapitel 3

Zielsetzung

Die Zielsetzung für diese Arbeit ist es, weitere Erkenntnisse zum Collatz-Problem zu sammeln und so der Lösung des Problems ein Stück näher zu kommen. Das soll auf Basis der letztjährigen Arbeit geschehen. Dort habe ich in einem Ausblick geschrieben, dass die Sprunglängen ein Ziel weiterer Forschung sein könnten, vor allem wenn man diese in Bezug auf unterschiedliche Abstiegslängen untersucht. Dahinter steckt der Gedanke, dass alle Zahlen, die eine Abstiegslänge von mindestens 2 haben, im Collatz-Gitter direkt nach oben führen. Alle anderen Zahlen, die immerhin noch die Hälfte aller ungeraden Zahlen ausmachen, führen nach unten und behindern damit eine einfache Lösung des Collatz-Problems. Ließe sich zeigen, dass auch diese übrigen Zahlen irgendwann nach oben führen, wäre das Collatz-Problem tatsächlich gelöst.

Diese Idee findet sich auch bei Lagarias wieder, der die Stoppzeiten beschrieben hat (siehe 2.1). Die Idee ist die gleiche, denn eine endliche Stoppzeit führt im Endeffekt zu einer Aufwärtsbewegung im Collatz-Gitter. Daher wird sich ein Teil dieser Arbeit damit beschäftigen, die Stoppzeit in Kombination mit den Erkenntnissen aus dem Collatz-Gitter zu untersuchen.

Ein weiterer Teil dieser Arbeit setzt sich mit der Binärdarstellung der Collatz-Schritte auseinander, da diese einen neuen Blick auf das Problem ermöglichen können. In wie weit die einzelnen Ansätze ausgeführt werden können, lässt sich am Anfang noch nicht sagen, denn das hängt davon ab, welche Ideen man selber zur Untersuchung hat.

Kapitel 4

Neue Forschung

Bei der weiteren Betrachtung des Collatz-Problems sind also nur die ungeraden Zahlen von Interesse, da jede gerade Zahl auf eine solche reduziert werden kann, indem man genügend oft durch 2 teilt. In der Untersuchung wird somit zunächst nur die erste Spalte des Collatz-Gitters betrachtet.

4.1 Modifizierte Stoppzeiten

4.1.1 Suche nach Mustern

Für jede ungerade natürliche Zahl $n > 1$ mit $A(n) \geq 2$ gilt $I(n) < n$:

$$I(n) = \frac{C(n)}{2^{A(n)}} \leq \frac{3n+1}{2^2} = \frac{3}{4}n + \frac{1}{4} < n \implies \sigma(n) \text{ ist endlich}$$

Wenn man nun für alle ungeraden natürlichen Zahlen n mit $A(n) = 1$ entscheiden könnte, ob die Stoppzeit endlich ist oder nicht, wäre das Collatz-Problem gelöst. Zu diesem Zweck wird die *modifizierte* Stoppzeit $\sigma_i(n)$ verwendet. Sie beschreibt für eine ungerade natürliche Zahl n die kleinste Anzahl von Sprüngen, bis eine ungerade Zahl der Trajektorie von n kleiner ist als n selbst. Beispielsweise ist $\sigma_i(3) = 2$, da $I(3) = 5$ und $I(5) = 1 < 3$.

Nun kann man jeder ungeraden Zahl eine modifizierte Stoppzeit zuordnen. (Abbildung 4.1)

n	$\sigma_i(n)$	n	$\sigma_i(n)$	n	$\sigma_i(n)$	n	$\sigma_i(n)$	n	$\sigma_i(n)$	n	$\sigma_i(n)$	n	$\sigma_i(n)$
1	∞	29	1	57	1	85	1	113	1	141	1	169	1
3	2	31	35	59	4	87	3	115	2	143	4	171	3
5	1	33	1	61	1	89	1	117	1	145	1	173	1
7	4	35	2	63	34	91	28	119	3	147	2	175	5
9	1	37	1	65	1	93	1	121	1	149	1	177	1
11	3	39	5	67	2	95	5	123	5	151	3	179	2
13	1	41	1	69	1	97	1	125	1	153	1	181	1
15	4	43	3	71	32	99	2	127	9	155	25	183	3
17	1	45	1	73	1	101	1	129	1	157	1	185	1
19	2	47	34	75	3	103	26	131	2	159	13	187	4
21	1	49	1	77	1	105	1	133	1	161	1	189	1
23	3	51	2	79	5	107	3	135	4	163	2	191	8
25	1	53	1	81	1	109	1	137	1	165	1	193	1
27	37	55	3	83	2	111	19	139	3	167	18	195	2

Abbildung 4.1: Die ersten 96 modifizierten Stoppzeiten

Dabei fallen einige Dinge auf: Für jede zweite Zahl gilt $\sigma_i(n) = 1$. Das sind die eben betrachteten Zahlen mit $A(n) \geq 2$. Allerdings fällt auch auf, dass die meisten modifizierten Stoppzeiten nur kleine Zahlwerte wie 2, 3 oder 4 annehmen, während es in unregelmäßigen Abständen Ausreißer gibt, wie zum Beispiel $\sigma_i(27) = 37,\ \sigma_i(31) = 35,\ \sigma_i(47) = 34, \ldots$. Die kleineren Stoppzeiten sind dagegen regelmäßiger vorhanden. Die Stoppzeit 2 kommt beispielsweise in jedem 8. Eintrag vor. Es soll überprüft werden, ob das generell gültig ist.
Zur Überprüfung wird eine mögliche Konstruktionsvorschrift vorgeschlagen, mit denen die bekannten Werte 3, 19, 35, …erzeugt werden können. In der Tabelle sind nur ungerade Zahlen gegeben, die die Form $n = 2n_1 + 1, n_1 \in \mathbb{N}_0$ haben. Da jede 8. ungerade Zahl die Stoppzeit 2 hat, gilt $n_1 = 8n_2 + k, n_2 \in \mathbb{N}_0$, wobei das k die Platzierung des ersten n mit $\sigma_i(n) = 2$ angibt. Da $2 \cdot 0 + 1 = 1$ muss mit dem Zählen bei 0 begonnen werden, weshalb $k = 1$ gilt.

Der entstehende Term ist also $n = 2(8n_2 + 1) + 1 = 16n_2 + 3$. Nun kann man den Verlauf der Trajektorie manuell berechnen:

$$\begin{aligned} &16n_2 + 3 \text{ ist ungerade} \\ C(16n_2+3) &= 48n_2 + 10 \text{ ist durch 2 teilbar} \\ C(24n_2+5) &= 72n_2 + 16 \text{ ist durch 8 teilbar} \\ &9n_2 + 2 < 16n_2 + 3 \\ &\implies \sigma_i(16n_2+3) = 2 \end{aligned} \tag{4.1}$$

Dieses Vorgehen kann auch für andere Stoppzeiten angewandt werden, wie zum Beispiel im Fall $\sigma_i(n) = 3$. Dieser Fall tritt ein für $n = 11;\ 23;\ 43;\ 55;\ 75;\ \ldots$ Also immer abwechselnd nach 6 bzw. 10 Zahlen. Das kann man aufteilen, sodass für jede 16. ungerade Zahl n $\sigma_i(n) = 3$ gilt, beginnend mit $n = 11$ und nochmal mit $n = 23$. Es gibt also zwei Terme, die es zu überprüfen gilt: $n = 2(16n_2 + 5) + 1$ und $n = 2(16n_2 + 11) + 1$:

$$\begin{aligned} 2(16n_2+5)+1 &= 32n_2 + 11 \text{ ist ungerade} \\ C(32n_2+11) &= 96n_2 + 34 \text{ ist durch 2 teilbar} \\ C(48n_2+17) &= 144n_2 + 52 \text{ ist durch 4 teilbar} \\ C(36n_2+13) &= 108n_2 + 40 \text{ ist durch 4 teilbar} \\ &27n_2 + 10 < 32n_2 + 11 \\ &\implies \sigma_i(32n_2+11) = 3 \end{aligned} \tag{4.2}$$

$$\begin{aligned} 2(16n_2+11)+1 &= 32n_2 + 23 \text{ ist ungerade} \\ C(32n_2+23) &= 96n_2 + 70 \text{ ist durch 2 teilbar} \\ C(48n_2+35) &= 144n_2 + 106 \text{ ist durch 2 teilbar} \\ C(72n_2+53) &= 216n_2 + 160 \text{ ist durch 8 teilbar} \\ &27n_2 + 20 < 32n_2 + 23 \\ &\implies \sigma_i(32n_2+23) = 3 \end{aligned} \tag{4.3}$$

Mit diesen Berechnungen lassen sich die Vermutungen also bestätigen. Das Problem bei dieser Art der Untersuchung ist, dass ich noch keine Regelmäßigkeit in den Perioden erkennen konnte, die Vorhersagen über weitere Perioden zulässt. Im Kontrast dazu konnten in der letztjährigen Arbeit die Muster für die Abstiegslängen relativ einfach hergeleitet werden und ließen so weitere Aussagen zu, ohne jemals die Werte errechnet zu haben.

4.1.2 Suche nach der Anzahl der Perioden

Um ein Muster in den Periodenlängen zu erkennen, kann man probieren, alle möglichen Wege zu finden, mit denen eine Zahl eine bestimmte modifizierte Stoppzeit erlangen kann. Die Anzahl der verschiedenen Wege beschreibt dann auch die Anzahl der unterschiedlichen Startwerte, mit denen die Perioden beginnen können.

Dabei entsteht folgende Überlegung: Wenn eine ungerade Zahl n besonders groß ist, dann ist $C(n) \approx 3n$, da der Wert 1 nur einen Bruchteil des Wertes von n ausmacht. Das ist besonders nützlich, da alle Schritte der Collatz-Funktion dann aus einer Multiplikation von 3 oder $\frac{1}{2}$ bestehen. Es müssen also nur Kombinationen von Sprüngen und Abstiegen gesucht werden, sodass

$$C^k(n) \approx \frac{3^{\sigma_i(n)}}{2^{\sigma_i(n)+l}} n < n \implies \frac{3^{\sigma_i(n)}}{2^{\sigma_i(n)+l}} < 1, l \in \mathbb{N} \tag{4.4}$$

Gesucht wird dann das kleinste l. Da die Addition am Ende des $3n+1$-Schritts nun ausgelassen wird, aber durch weitere Multiplikation vergrößert werden kann, muss man immer daran denken, dass es sich hier um eine Annäherung handelt. Ob tatsächlich $C^k(n) < n$ gilt, wenn der Bruch kleiner als 1 ist, ist nicht für alle n bekannt.[1]

Um die möglichen Abfolgen von Sprüngen und Abstiegen möglichst einfach zu untersuchen, kann man sie als eine Zahlenfolge f notieren. Jede Zahl steht dann für eine Abstiegslänge, vor jeder Zahl findet ein Sprung statt. Beispielsweise beschreibt die Folge $f = (1,2,5)$ drei Sprünge, die von einem, zwei und fünf Abstiegen gefolgt werden. Eine Abschätzung des Ergebnisses ist dann

$$\frac{3^3}{2^{1+2+5}} = \frac{27}{256}$$

Diese Schreibweise findet sich auch in der Arbeit von Lagarias wieder.[2] Dort sind in der Zahlenfolge allerdings nur 1 und 0 erlaubt, um die Teilbarkeit mit 2 für die Zwischenergebnisse der Trajektorie zu beschreiben. Invertiert man diesen Vektor, addiert alle aufeinander folgenden Einsen und lässt dann die Nullen weg, ergibt sich die kompakte Schreibweise für die Abstiege.

$$(1,0,1,0,0,1,0,0,0,0,0) \to (0,1,0,1,1,0,1,1,1,1,1) \to (1,2,5)$$

Nun kann man die Anzahl der Möglichkeiten für eine gegebene modifizierte Stoppzeit errechnen. Der Fall $\sigma_i(n) = 1$ ist einfach:

$$\frac{3^1}{2^{1+l}} < 1 \implies l = 1$$

Die zugehörige Zahlenfolge ist dann $f^1_{max} = (2)$, da $1+1=2$. Wichtig ist dabei, dass die 2 eine untere Grenze darstellt. Jede größere Abstiegslänge erzeugt die gleiche modifizierte Stoppzeit.

Für $\sigma_i(n) = 2$ gilt:

$$\frac{3^2}{2^{2+l}} < 1 \implies l = 2$$

Dafür gibt es auch genau eine mögliche Folge $f^2_{max} = (1,3)$, da eine höhere Zahl an erste Stelle bedeuten würde, dass die Folge (2) enthalten ist und somit $\sigma_i(n) = 1$ gelten müsste.

Für $\sigma_i(n) = 3$ gilt:

$$\frac{3^3}{2^{3+l}} < 1 \implies l = 2$$

Wieder darf die vorherige Folge nicht in der neuen Folge enthalten sein. Die Folge beginnt also mit $f^3_{max} = (1,2,\dots)$. Allerdings ist es auch möglich, dass nach dem zweiten Sprung nur ein Abstieg erfolgt und nicht zwei. Die 2 ist also als obere Grenze zu verstehen und die dritte Zahl der Folge muss diese Möglichkeit erlauben. Sie bleibt eine untere Grenze für den Fall, dass alle vorderen Abstiege nur aus einem Schritt bestehen. Daraus folgt, dass die Folge aus

[1] Lagarias 1985, S. 11.

[2] Ebd., S. 7.

$f^3_{max} = (1,2,3)$ besteht. Die verschiedenen Möglichkeiten, eine modifizierte Stoppzeit von 3 zu erreichen, lassen sich daraus ablesen.

Eine Möglichkeit ist $f = (1,1,3)$. Außerdem möglich ist $f = (1,2,2)$. Diese beiden einzigen Möglichkeiten passen zu den beiden Termen $n = 32n_2 + 11$ und $n = 32n_2 + 23$, für die $\sigma_i(n) = 3$ gilt.

Für $\sigma_i(n) = 4$ gilt $f^4_{max} = (1,2,2,4)$. Die einfachste ist wieder $f = (1,1,1,4)$. Die zweite Abstiegslänge kann aber auch 2 sein, mit $f = (1,2,\dots)$. Dann müssen die weiteren Grenzen aber um 1 reduziert werden, da man bei der Berechnung bislang davon ausging, dass alle bisherigen Abstiegslängen 1 sind. Die neuen Grenzen sind also $f^4_{max} = (\dots,1,3)$, sodass für diesen Fall nur noch eine Lösung übrig bleibt, nämlich $f = (1,2,1,3)$. Auch die dritte Abstiegslänge kann 2 betragen, sodass eine weitere Möglichkeit entsteht. Es gibt also drei mögliche Folgen $f = (1,1,1,4)$, $f = (1,2,1,3)$ und $f = (1,1,2,3)$. In der Tat gibt es auch drei Terme für n: $n = 2(64n_2 + 3) + 1$, $n = 2(64n_2 + 7) + 1$ und $n = 2(64n_2 + 29) + 1$. Für alle Terme gilt $\sigma_i(n) = 4$.

4.1.3 Verallgemeinerung der Folgen

Allgemein lassen sich die Folgen für eine modifizierte Stoppzeit immer aus der vorigen Folge entwickeln. Dann ist die neue Folge

$$f^n_{max} = (f^{n-1}_{max_1}, \dots, f^{n-1}_{max_{n-1}} - 1, l+1), \tag{4.5}$$

wobei $l \in \mathbb{N}$ die kleinste Lösung ist für

$$\frac{3^{\sigma_i(n)}}{2^{\sigma_i(n)+l}} < 1. \tag{4.6}$$

Die Anzahl der Möglichkeiten, die modifizierte Stoppzeit zu erhalten, lässt sich rekursiv berechnen. Das Verfahren funktioniert wie für f^4_{max} beschrieben und lässt sich als Algorithmus beschreiben (Listing 4.2). Hierfür wird die Programmiersprache Python verwendet, zusammen mit der Bibliothek „Numpy“, die Vektor-Mathematik bereitstellt. Man kann auch die nächsten Maximalfolgen berechnen, wie in Formel 4.5 beschrieben. (Listing 4.1)

Mit dem Algorithmus kann man die Anzahl der verschiedenen Folgen für eine Maximalfolge berechnen (Abbildung 4.2). Da die berechneten Anzahlen schnell sehr groß werden, dauert die Berechnung zunehmend sehr lange.

$\sigma_i(n)$	Anzahl	$\sigma_i(n)$	Anzahl	$\sigma_i(n)$	Anzahl	$\sigma_i(n)$	Anzahl
1	1	6	12	11	961	16	108950
2	1	7	30	12	2652	17	312455
3	2	8	85	13	8045	18	663535
4	3	9	173	14	17637	19	1900470
5	7	10	476	15	51033	20	5936673

Abbildung 4.2: Anzahl der Möglichkeiten für die modifizierten Stoppzeiten

Im Idealfall steht jede errechnete Zahl auch für die Anzahl der Terme, die verschiedene $n \in \mathbb{N}$ mit einer bestimmten modifizierten Stoppzeit $\sigma_i(n)$ beschreiben können. Da es sich um eine Annäherung handelt, kann es natürlich sein, dass die Werte nicht ganz der Realität entsprechen. Jedoch kann man davon ausgehen, dass der Trend des starken Wachstums richtig ist. Das bedeutet, dass die Abstände, in denen die Zahlen einer bestimmten modifizierten Stoppzeit auftreten, in ihrer Anzahl immer größer werden. Beispielsweise gibt es für $\sigma_i(n) = 4$ genau drei

```python
  def berechneL(stoppzeit):
  l = 0
  # solange hochzaehlen, bis der passende Exponent erreicht ist
  while 2**(stoppzeit+l) < 3**stoppzeit:
    l += 1
  return l

def berechneFolge(stoppzeit):
  if stoppzeit == 0:
     # keine Folge mehr moeglich
    return []

  # fruehere Folge ausrechnen
  folge = berechneFolge(stoppzeit-1)

  if stoppzeit > 1:
    # letztes Element der Folge verringern
    folge[-1] -= 1

  # l+1 an die Folge anhaengen
  l = berechneL(stoppzeit)
  folge.append(l + 1)

  return folge
```

Listing 4.1: Bestimmung der f_{max}

```python
import numpy as np

def anzahlFolgen(folge):

  if len(folge) == 1:
    # es gibt nur noch eine weitere Moeglichkeit
    return 1

  # erstelle einen Vektor, mit dem man rechnen kann
  f = np.array(folge)

  # die bislang berechnete Anzahl
  anzahl = 0
  for i in range(1, f[0]+1): # 1 <= i <= f[0]

    # Elemente der Restfolge um 1 verringern
    neueMaxFolge = f[1:] - (i-1)

    # berechne Folgenanzahl fuer neue Maximalfolge
    anzahl += anzahlFolgen(neueMaxFolge)

  return anzahl
```

Listing 4.2: Bestimmung der Folgenanzahl

Abstände, die sich immer wiederholen und ansonsten keine offensichtliche Beziehung zueinander haben. Für $\sigma_i(n) = 6$ gibt es schon 12 verschiedene Abstände, die in der ersten Tabelle 4.1 nicht alle enthalten sind. Dadurch kann man die Regelmäßigkeit nicht erkennen, weshalb die Stoppzeiten so zufällig verteilt wirken. Es scheint aber Perioden zu geben, jedoch wirken die Anfangszahlen, bei denen die periodischen Abstände beginnen, tatsächlich zufällig.

4.1.4 Die Periodenlänge

Da nun die Anzahl der Perioden abgeschätzt werden kann, ist die Länge der Perioden ein interessanter Blickpunkt. Schließlich kann man für $\sigma_i(n) = 2$ die Periode noch von Hand herausfinden (siehe Formel 4.1), doch für größere modifizierte Stoppzeiten ist das nicht mehr ohne weiteres möglich, da die Periodendauern sehr groß werden. Wie kann man dennoch die Periodendauer berechnen?

Generell hat der Term, der eine Periode beschreiben kann, die Form

$$n = a \cdot b + c, \quad a, b, c \in \mathbb{N}$$

wobei b die Rolle der Laufvariablen hat, während a und c für einen Term fest sind. Nun soll zu einem beliebigen c, dessen modifizierte Stoppzeit bekannt ist, ein passendes a konstruiert werden. Zu diesem Zweck soll der Term eine weitere Bedingung erfüllen, nämlich

$$C^k(n) \equiv C^k(c) \pmod{2}, \quad 0 \leq k \leq \sigma_i(c) - 1. \tag{4.7}$$

Wird die Collatz-Funktion nun mit dieser Bedingung ausgeführt, gibt es zwei Möglichkeiten:

$$C(n) = 3 \cdot (a \cdot b + c) + 1 = 3 \cdot a \cdot b + C(c)$$

und

$$C(n) = \frac{a \cdot b + c}{2} = \frac{a}{2} \cdot b + C(c).$$

In beiden Fällen entsteht ein neuer Term der gleichen Form

$$C(n) = a' \cdot b + C(c).$$

Aufgrund der Bedingung 4.7, kann der Collatz-Schritt genauso wiederholt werden. Es folgt

$$C^k(n) = a^{(k)} \cdot b + C^k(c).$$

Weiterhin folgt aus der Bedingung

$$a^{(k)} \cdot b + C^k(c) \equiv C^k(c) \pmod{2} \implies a^{(k)} \cdot b \equiv 0 \pmod{2}, \quad 0 \leq k \leq \sigma_i(c) - 1$$

Also soll $a^{(k)}$ durch 2 teilbar sein, sodass auch $a^{(k)} \cdot b$ auf jeden Fall gerade ist. Wenn also insgesamt 3 $\frac{n}{2}$-Schritte in den ersten k Schritten der Trajektorie auftreten, muss a durch $2^3 = 8$ teilbar sein. Der kleinstmögliche Wert ist daher $a = 8$.

Die Anzahl der $\frac{n}{2}$-Schritte bei einer gegebenen modifizierten Stoppzeit ist $\sigma_i(c) + l$, wobei $l \in \mathbb{N}$ die kleinste Lösung für die Gleichung 4.6 ist. Die Schrittanzahl entspricht also auch der Summe aller Elemente einer konkreten Folge f für c. Für $\sigma_i(n) = 2$ gilt $l = 2$. Also ist a durch $2^{2+2} = 16$ teilbar, womit die Periodenlänge auch bekannt ist. Das kleinste n, für das $\sigma_i(n) = 2$ gilt, ist $n = 3$. Der konstruierte Term ist also $n = 16b + 3$, was mit dem Term 4.1 übereinstimmt. Man kann also auch die Periodendauern für die modifizierten Stoppzeiten berechnen (Abbildung 4.3). Dabei fällt auch auf, dass die Periodenlänge stärker wächst als die Anzahl der möglichen Perioden, welche in Abbildung 4.2 dargestellt ist. Dies erscheint auch soweit logisch, schließlich müssen zwischen zwei Zahlen, die die gleiche Stoppzeit haben und durch den gleichen Term gebildet werden, auch alle anderen Terme für die gleiche Stoppzeit einen Wert liefern. Beispielsweise haben die Zahlen 7 und 135 die gleiche Stoppzeit von 4 und sie werden durch den Term $n = 128b + 7$ beschrieben. Die beiden anderen Terme für die gleiche Stoppzeit, $n = 128b + 15$ und $n = 128b + 59$, liefern die Werte 15 und 59, welche zwischen 7 und 135 liegen.

$\sigma_i(n)$	Periode	$\sigma_i(n)$	Periode	$\sigma_i(n)$	Periode	$\sigma_i(n)$	Periode
1	4	6	1024	11	262144	16	67108864
2	16	7	4096	12	1048576	17	134217728
3	32	8	8192	13	2097152	18	536870912
4	128	9	32768	14	8388608	19	2147483648
5	256	10	65536	15	16777216	20	4294967296

Abbildung 4.3: Periodenlängen zu gegebenen Stoppzeiten

4.2 Binärzahlen

Man kann jede Zahl auch als Binärzahl darstellen. Ein offensichtlicher Vorteil ist die einfache Repräsentation der Halbierung von Zahlen, da diese einer einfachen Verschiebung der Ziffern nach rechts entspricht, wobei die Endnull gelöscht wird. Indem der $3n+1$-Schritt auch im Binärsystem dargestellt wird, können möglicherweise weitere Auffälligkeiten gefunden werden. Im Allgemeinen kann eine positive ganze Zahl p im Binärsystem dargestellt werden mit

$$p = \sum_{i=0}^{\infty} z_i \cdot 2^i, z_i \in \{0, 1\}.$$

Beispielsweise ist

$$13 = 1101_2 = 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0.$$

Die Division mit 2 entspricht dann nur einer Verschiebung der Zahl nach rechts, ähnlich wie die Division mit 10 im Dezimalsystem.

$$\frac{p}{2} = \frac{\sum_{i=0}^{\infty} z_i \cdot 2^i}{2} = \sum_{i=0}^{\infty} z_i \cdot \frac{2^i}{2} = \sum_{i=0}^{\infty} z_i \cdot 2^{i-1}$$

Die Multiplikation mit 3 kann als Addition mit dem Doppelten aufgefasst werden, da $3p = 2p + p$ gilt. $13 \cdot 3$ sieht also so aus:

$$\begin{array}{r} 1\,1\,0\,1_2 \\ +1\,1\,0\,1\,0_2 \\ \hline 1\,0\,1\,1\,1_2 \end{array}$$

Nun kann man überlegen, wie eine ungerade natürliche Zahl n aussehen muss, für die $I(n) = m$ gilt. Die Abstiegslänge $A(n) = a$ ist dabei eine beliebige natürliche Zahl. Dann ist

$$C(n) = m \cdot 2^a \iff 3n = m \cdot 2^a - 1.$$

Nun kann man untersuchen, wie die Binärdarstellung von $3n$ aussieht. Man beachte, dass m in jedem Fall ungerade ist.

$$3n = (m-1+1) \cdot 2^a - 1 = (m-1) \cdot 2^a + 2^a - 1 = (m-1) \cdot 2^a + \sum_{i=0}^{a-1} 2^i$$

Also endet die Binärdarstellung von $3n$ auf a Einsen.

$$3n = \underbrace{1\ldots0}_{m-1}\underbrace{1\ldots1_2}_{\text{a-mal}} \tag{4.8}$$

Man beachte, dass bei der Addition von 1 durch den Übertrag am Ende der Binärdarstellung a Nullen entstehen, weil die Abstiegslänge a ist. Nun soll untersucht werden, wie eine Zahl $n \in \mathbb{N}$ in Binär aussehen muss, damit die Binärdarstellung von $3n$ auf einer Reihe von Einsen endet. Dabei wird zunächst nur der Fall betrachtet, in dem $m-1=0$ gilt. Aus der Multiplikation entsteht also eine Zahl, deren Ziffern ausschließlich Einsen sind.

$$3n = \frac{\begin{array}{r} z_i\ z_{i-1}\ \ldots\ z_2\ z_1\ z_0 \\ +z_i\ z_{i-1}\ \ldots\ z_2\ z_1\ z_0\ 0_2 \end{array}}{1\ 1\ \ldots\ldots\ 1\ 1\ 1\ 1_2}$$

Damit nun die letzte Ziffer des Ergebnisses, die Einerstelle, eine 1 ist, muss $z_0 = 1$ sein. Da z_0 auch in dem zweiten Summanden an der Zweierstelle enthalten ist, muss weiterhin $z_1 = 0$ gelten. Auch z_1 ist im zweiten Summanden vorhanden, als Viererstelle. Daraus folgt $z_2 = 1$. Dieses Muster setzt sich fort, sodass n am Ende aus einer Folge alternierender Nullen und Einsen besteht.

$$\frac{\begin{array}{r} z_i\ z_{i-1}\ \ldots\ z_2\ z_1\ 1_2 \\ +z_i\ z_{i-1}\ \ldots\ z_2\ z_1\ 1\ 0_2 \end{array}}{1\ 1\ \ldots\ldots\ 1\ 1\ 1\ 1_2} \rightarrow \frac{\begin{array}{r} z_i\ z_{i-1}\ \ldots\ z_2\ 0\ 1_2 \\ +z_i\ z_{i-1}\ \ldots\ z_2\ 0\ 1\ 0_2 \end{array}}{1\ 1\ \ldots\ldots\ 1\ 1\ 1\ 1_2} \rightarrow \cdots \rightarrow \frac{\begin{array}{r} 1\ 0\ \ldots\ 1\ 0\ 1_2 \\ +1\ 0\ \ldots\ 1\ 0\ 1\ 0_2 \end{array}}{1\ 1\ \ldots\ 1\ 1\ 1\ 1_2}$$

Alle z_k mit geraden k sind also 1, alle z_k mit ungeradem k sind 0. Da z_i und z_0 gleich 1 sein müssen, muss es eine ungerade Anzahl an Ziffern geben, sodass das Ergebnis immer eine gerade Anzahl von Einsen enthält. Allerdings wurde hier nur ein Sonderfall betrachtet, von dem aus weiter fortgeführt werden kann. Beachtet man 4.8, so folgt

$$3n = \frac{\begin{array}{r} z_i\quad \ldots\ z_{a+1}\ z_a\quad z_{a-1}\ \ldots\ 1_2 \\ +z_i\ z_{i-1}\ \ldots\ z_a\quad z_{a-1}\ z_{a-2}\ \ldots\ 0_2 \end{array}}{\underbrace{1\ \ldots\ldots\ldots\ldots\ 0}_{m-1}\ \underbrace{1\ \ldots\ 1_2}_{\text{a-mal}}}$$

Nun gibt es zwei Fälle: a kann gerade oder ungerade sein. Wenn a gerade ist, dann folgt daraus, dass $z_{a-1} = 0$, da $a-1$ ungerade ist. Also muss auch $z_a = 0$ sein. Sollte a jedoch ungerade sein, so folgt $z_{a-1} = 1$ und $z_a = 1$.

a gerade

$$\frac{\begin{array}{r} z_i\quad \ldots\ z_{a+1}\ 0\quad 0\quad \ldots\ 1_2 \\ +z_i\ z_{i-1}\ \ldots\ 0\quad 0\quad z_{a-2}\ \ldots\ 0_2 \end{array}}{\underbrace{1\ \ldots\ldots\ldots\ 0}_{m-1}\ \underbrace{1\ \ldots\ 1_2}_{\text{a-mal}}}$$

a gerade

$$\frac{\begin{array}{r} z_i\quad \ldots\ z_{a+1}\ 1\quad 1\quad \ldots\ 1_2 \\ +z_i\ z_{i-1}\ \ldots\ 1\quad 1\quad z_{a-2}\ \ldots\ 0_2 \end{array}}{\underbrace{1\ \ldots\ldots\ldots\ 0}_{m-1}\ \underbrace{1\ \ldots\ 1_2}_{\text{a-mal}}}$$

Also können auch ungerade a erreicht werden und jede Zahl n, für die $3n = m \cdot 2^a$ gilt, muss in der Binärdarstellung auf einer Reihe alternierender Nullen und Einsen enden, wobei an der letzten Stelle eine 1 steht. Die Ziffern z_a und z_{a-1} sind genau dann eine 0, wenn a gerade ist. z_{a-1} ist dann die erste Ziffer des Musters. Die so entstehenden Zahlen sehen so aus:

$$\begin{aligned} A(n) = 1 :&\quad 1\ldots1\ 1_2 \\ A(n) = 2 :&\quad 1\ldots0\ 0\ 1_2 \\ A(n) = 3 :&\quad 1\ldots1\ 1\ 0\ 1_2 \\ A(n) = 4 :&\quad 1\ldots0\ 0\ 1\ 0\ 1_2 \\ A(n) = 5 :&\quad 1\ldots1\ 1\ 0\ 1\ 0\ 1_2 \end{aligned}$$

Dieses Ergebnis stimmt mit der früheren Jugend-Forscht-Arbeit überein. Die Zahlen für eine bestimmte Abstiegslänge konnten da als Terme formuliert werden mit $k \in \mathbb{N}$:

$$
\begin{array}{ll}
A(n) \geq 1 : 2k+1 & A(n) = 1 : 4k+3 \\
A(n) \geq 2 : 4k+1 & A(n) = 2 : 8k+1 \\
A(n) \geq 3 : 8k+5 & A(n) = 3 : 16k+13 \\
A(n) \geq 4 : 16k+5 & A(n) = 4 : 32k+5 \\
A(n) \geq 5 : 32k+21 & A(n) = 5 : 64k+53
\end{array}
$$

Die immer größer werdenden Koeffizienten vor dem k entsprechen den immer länger wachsenden Binärdarstellungen. Außerdem kann man leicht erkennen, dass alle n mit $n = 32k+21$ auch die Form $n = 16k+5$ haben, da die Binärdarstellung der Zahl $5 = 101_2$ auch in der Binärdarstellung der Zahl $21 = 10101_2$ enthalten ist. Diese Konstanten erzeugen das Muster. Außerdem ist von früher bekannt, dass die Zahlen 1, 5, 21, ... durch die rekursive Funktion 2.3 erzeugt werden können. Durch die Multiplikation mit 4 wird eine Zahl in binär um zwei Stellen nach links verschoben, die addierte 1 sorgt dafür, dass die letzte Stelle eine 1 wird. So entstehen dann die Muster aus Nullen und Einsen.

$$
\begin{aligned}
&a_0 = 1_2 \\
&a_1 = 100_2 + 1_2 = 101_2 \\
&a_2 = 10100_2 + 1_2 = 10101_2 \\
&a_3 = 1010100_2 + 1_2 = 1010101_2 \\
&\ldots
\end{aligned}
$$

Kapitel 5

Ergebnisdiskussion und Ausblick

Die Ergebnisse lassen sich kurz zusammenfassen: Zu jeder modifizierten Stoppzeit lässt sich eine Abschätzung treffen, wie viele mögliche Terme die Zahlen beschreiben, für die diese Stoppzeit gelten. Für diese Abschätzung wurde ein Algorithmus gefunden und die Anzahl der Terme wächst stark. Außerdem können mithilfe dieser Abschätzungen die Periodenlängen gefunden werden, also die Abstände, die zwischen zwei Zahlen des gleichen Terms stehen. Auch die Periodenlängen wachsen sehr schnell. Die Grundlage dieser Abschätzungen bildet der Bruch

$$\frac{3^{\sigma_i(n)}}{2^{\sigma_i(n)+l}},$$

welcher in abgewandelter Form zusammen mit weiteren Vektoren in Lagarias' Arbeit wiederzufinden ist.[1] Auch in weiteren Arbeiten finden sich Verweise auf diesen Bruch, meistens in logarithmierter Form.[2,3] Dieser Bruch erklärt also die Periodenlängen und -anzahlen. Gleichzeitig wirft er aber auch weitere Fragen auf, da die Startzahlen der Perioden immer noch zufällig verteilt wirken. Außerdem können die Folgen bislang nur mit rekursiven Algorithmen berechnet werden, jedoch wäre eine einfachere mathematische Beschreibung wünschenswert, da dies die Arbeit erleichtern würde und womöglich schneller weitere Regelmäßigkeiten gefunden werden könnten.

Ein weitere Aspekt, der untersucht wurde, ist die Übertragung des Collatz-Problems ins Binärsystem. Dabei wurde untersucht, wie eine Zahl für eine bestimmte Abstiegslänge aussehen muss. Das gefundene Muster ist leider nicht neu, da schon in der vorigen Jugend-Forscht-Arbeit Terme für die Abstiegslängen gefunden wurden. Dies zeigt, dass auch über unterschiedliche Wege die gleichen Ergebnisse möglich sind. Außerdem ist das ein Beispiel für Forschung, die nicht direkt neue Erkenntnisse bringt. Allerdings kann man nicht immer von Anfang an sehen, welche Ergebnisse über einen Ansatz entstehen können.

Dennoch kann man sagen, dass beide Ansätze gute Ergebnisse gebracht haben, die insgesamt einen Fortschritt für das Collatz-Problem bedeuten. Außerdem lässt sich auf Basis dieser Ergebnisse weitere Forschung betreiben. Eine Untersuchung der Folgen und Perioden erscheint sinnvoll, mit der Intention, die scheinbare Zufälligkeit der Startzahlen zu durchschauen. Auch im Binärsystem ist weitere Forschung möglich. Man könnte untersuchen, wie sich die Binärzahlen während einer Trajektorie verändern. Vielleicht ergeben sich auch Ziffernreihen, die durch mehrere Nullen voneinander getrennt sind und so zunächst unabhängig voneinander untersucht werden könnten. Die Arbeit am Collatz-Problem ist noch lange nicht abgeschlossen und wird bestimmt auch in Zukunft noch viele Menschen beschäftigen. Die hier erarbeiteten Ergebnisse können wiederum für weitere Forschung genutzt werden oder als Inspiration dienen. Die schlussendliche Lösung des Problems ist aber noch immer nicht in Sicht.

[1] Lagarias 1985, S. 7.

[2] Wolke o.D., S. 3.

[3] Lagarias 1985, S. 8.

Literatur

Lagarias, Jeffrey C. (1985). „The 3x + 1 problem and its generalizations“. In: *American Mathematical Monthly*. URL: https://www.maa.org/sites/default/files/pdf/upload_library/22/Ford/Lagarias3-23.pdf.

Tao, Terence (2020). *The Notorious Collatz conjecture*. URL: https://terrytao.files.wordpress.com/2020/02/collatz.pdf (besucht am 27. 09. 2020).

Wolke, Dieter (o.D.). „Das Collatz–Problem“. In: (). URL: http://home.mathematik.uni-freiburg.de/didaktik/material_download/collatzproblem.pdf.

Unterstützer

Ich bedanke mich bei Herrn Crystalla und Herrn Dr. Striethorst vom Gymnasium Carolinum Osnabrück für die Unterstützung im Rahmen der Jugend-Forscht-AG.