# Neues zum 3n+1-Problem

**Ingo Althöfer, 28. April 2020**
**ingo.althoefer@uni-jena.de**

Das 3n+1-Problem wurde 1937 von Lothar Collatz „in die Welt gebracht“. Die zugehörige Collatz-Vermutung ist nach wie vor unbewiesen. Je mehr Zeit vergeht, um so schwerer wird das Problem eingeschätzt. Prominente Mathematiker wie Paul Erdős und John Conway erklärten, es sei nicht lösbar mit den Techniken heutiger Mathematik (Lagarias, 2011). Eine andere Einordnung ist folgende: Das 3n+1-Problem ist das schwerste ungelöste mathematische Problem, was sich in nur zwei Minuten fast jedem Menschen erklären lässt.

Dieses Kapitel zerfällt in vier Teile. Im ersten Block, welcher die ersten drei Abschnitte umfasst, formulieren wir das Collatz-Problem, geben Beispiele an und erklären das Applet von Jürgen Dankert. Der zweite Block, mit den Abschnitten vier bis sechs, veranschaulicht theoretische Ergebnisse zum Collatz-Problem.

Die Abschnitte sieben bis neun bilden den dritten Block. Er beinhaltet deterministische Varianten des Collatz-Problems, eine weitere Variante mit Zufallskomponenten und einen Hinweis auf 3n+1-Daten in der Online-Enzyklopädie der Zahlenfolgen (OEIS).

Die beiden letzten Abschnitte bilden den vierten Block. Dort formulieren wir eine neue Vermutung zum 3n+1-Problem, das Verhalten der Collatz-Folgen bei benachbarten Startwerten betreffend. In den Schlussbemerkungen brechen wir eine Lanze für die experimentelle computerunterstützte Mathematik allgemein.

## Die 3n+1-Regeln und fünf Beispiele

Man muss nur die einfachen Rechenschritte „plus 1“, „mal 3“ und „geteilt durch 2“ können; und man muss merken können, ob eine Zahl gerade oder ungerade ist. Sogar Pippi Langstrumpf dürfte das schaffen: 3 mal 3 ist 9; widewidewitt; plus 1 ist zehne - und so weiter.

Aus einer natürlichen Anfangszahl wird durch wiederholte Anwendung der Rechenregeln eine endlose Folge von ganzen Zahlen erzeugt. Ist die aktuelle Zahl gerade, wird sie durch 2 geteilt. Ist sie dagegen ungerade, wird sie mit 3 multipliziert und 1 dazu gezählt. Mit der neuen Zahl verfährt man ebenso.

Angegeben ist in jedem Beispiel die Startzahl und, getrennt durch Pfeile, die sich ergebende Folge.

$$1 \to 4 \to 2 \to 1 \to 4 \to 2 \to 1 \to ... \tag{1}$$

Die Folge dreht sich also im Kreis; es kommen immer nur reihum 1, 4 und 2.

$$3 \to 10 \to 5 \to 16 \to 8 \to 4 \tag{2}$$

Ab hier brauchen wir nicht weiterzurechnen, weil wir aus Beispiel 1 wissen, dass nur noch gekreist werden wird (4, 2, 1, ...).

$$6 \to 3 \tag{3}$$

Also ein ganz einfacher Fall: Einmal durch 2 geteilt, und schon sind wir in einer Situation, die wir aus Beispiel 2 kennen. Als nächstes kommt aber ein mittelgroßer Kracher:

$$7 \to 22 \to 11 \to 34 \to 17 \to 52 \to 26 \to 13 \to 40 \to 20 \to 10 \tag{4}$$

Wie es von 10 aus weitergeht, kennen wir aus Beispiel 2. Und für die ganz Harten jetzt eine echte Geduldsübung mit Startwert 27. Aus Platzgründen sind nur die ungeraden Zahlen aus der Folge angegeben.

$$\begin{aligned} 27 &\to 41 \to 31 \to 47 \to 71 \to 107 \to 161 \to 121 \to 91 \to 137 \\ &\to 103 \to 155 \to 233 \to 175 \to 263 \to 395 \to 593 \to 445 \\ &\to 167 \to 251 \to 377 \to 283 \to 425 \to 319 \to 479 \to 719 \\ &\to 1079 \to 1619 \to 2429 \to 911 \to 1367 \to 2051 \to 3077 \\ &\to 577 \to 433 \to 325 \to 61 \to 23 \to 35 \to 53 \to ... \to 10 \to \text{usw} ... \end{aligned} \tag{5}$$

Es sind insgesamt 111 Schritte, ehe die Folge bei der 1 anlangt.

# Die Collatz-Vermutung

Schon Collatz vermutete, dass sich für jede mögliche Startzahl eine Folge ergibt, die früher oder später, aber jedenfalls nach endlich vielen Schritten, in dem Zyklus $4 \to 2 \to 1$ landet. Wieviele Beispiele er gerechnet hat, bevor er diese Vermutung äußerte, ist nicht überliefert.

Für jede Startzahl $2^k$, das heißt jede Zweier-Potenz, gilt die Vermutung natürlich. Es gibt also zumindest unendlich viele Zahlen, die bei $4 \to 2 \to 1$ enden.

Sollte es auch nur einen Startwert $n_0$ geben, für den die $4 \to 2 \to 1$ nicht erreicht wird, dann gäbe es automatisch unendlich viele mit dieser Eigenschaft: Nämlich mindestens alle Zahlen der Form $2^k \cdot n_0$.

## Computer-Ergebnisse und das Applet von Jürgen Dankert

Das 3n+1-Problem hat nicht nur den Reiz, dass es schwer und ungelöst ist, sondern auch den, dass man per Hand oder mit Computerhilfe für viele Startwerte durchrechnen kann, was passiert.

Etliche Mathematiker und Programmierer um Eric Roosendaal haben einige Jahre lang bis 2017 in einem Kraftakt mit Computerhilfe für jeden Startwert, der kleiner als $87 \cdot 2^{60}$ ist, gezeigt, dass am Ende immer der Zyklus $4 \to 2 \to 1$ steht (Roosendaal, 2020). Randbemerkung: $87 \cdot 2^{60} = 100.304.170.900.795.686.912 > 10^{20}$. Wer das Gefühl hat, dass er in seinem Leben nie mehr als 20-stellige Zahlen braucht, kann sich also bezüglich des Collatz-Problems entspannt zurücklehnen.

Dr. Jürgen Dankert, bis zu seinem Ruhestand Professor für Informatik an der HAW Hamburg, hat ein wunderbar kompaktes Applet geschrieben, mit dem sich das 3n+1-Problem für beliebig große Startzahlen berechnen lässt (Dankert, 2020). „Beliebig groß“ bedeutet praktisch, dass auch 1.000-stellige Zahlen für das Applet kein Problem sind.

Wie klappt das? Warum gibt es keinen Speicherüberlauf? Dankert speichert die aktuelle Zahl nicht als einen Wert, sondern als Folge einzelner Ziffern. Multiplikation, Division und Addition passieren dann so, wie es Grundschüler in der dritten und vierten Klasse lernen: Dezimalstelle für Dezimalstelle, mit Überträgen.

Wer ohne viel Zeitaufwand lange Zahlen ins Dankert-Applet eingeben will, kann einfach mal den Finger auf einer Ziffer im Kexboard gedrückt halten, zum Beispiel für 20 Sekunden. Dann ans Ende noch ein paar zufällige Ziffern und das Programm starten. Daaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaanke, Herr Daaaaaaaaaaaaaaaaaaaaankert!

Wir betrachten in Abb. 1 einen Block von 50 aufeinander folgenden Zahlen mit 21 Dezimalstellen. In jeder Zeile steht einer der Werte und mit etwas Abstand dahinter die Schrittzahl bis zum Erreichen der 1, ausgerechnet mit dem Dankert-Applet.

| | | | |
|---|---|---|---|
| 314.159.265.358.979.323.800 | 576 | 314.159.265.358.979.323.825 | 364 |
| 314.159.265.358.979.323.801 | 576 | 314.159.265.358.979.323.826 | 364 |
| 314.159.265.358.979.323.802 | 576 | 314.159.265.358.979.323.827 | 364 |
| 314.159.265.358.979.323.803 | 346 | 314.159.265.358.979.323 828 | 488 |
| 314.159.265.358.979.323.804 | 488 | 314.159.265.358.979.323.829 | 488 |
| 314.159.265.358.979.323.805 | 488 | 314.159.265.358.979.323.830 | 576 |
| 314.159.265.358.979.323.806 | 488 | 314.159.265.358.979.323.831 | 576 |
| 314.159.265.358.979.323.807 | 346 | 314.159.265.358.979.323.832 | 488 |
| 314.159.265.358.979.323.808 | 576 | 314.159.265.358.979.323.833 | 576 |
| 314.159.265.358.979.323.809 | 576 | 314.159.265.358.979.323.834 | 488 |
| 314.159.265.358.979.323.810 | 576 | 314.159.265.358.979.323.835 | 576 |
| 314.159.265.358.979.323.811 | 576 | 314.159.265.358.979.323.836 | 594 |
| 314.159.265.358.979.323.812 | 576 | 314.159.265.358.979.323.837 | 594 |
| 314.159.265.358.979.323.813 | 576 | 314.159.265.358.979.323.838 | 594 |
| 314.159.265.358.979.323.814 | 576 | 314.159.265.358.979.323.839 | 576 |
| 314.159.265.358.979.323.810 | 576 | 314.159.265.358.979.323.840 | 576 |
| 314.159.265.358.979.323.811 | 576 | 314.159.265.358.979.323.841 | 488 |
| 314.159.265.358.979.323.812 | 576 | 314.159.265.358.979.323.842 | 576 |
| 314.159.265.358.979.323.813 | 576 | 314.159.265.358.979.323.843 | 576 |
| 314.159.265.358.979.323.814 | 576 | 314.159.265.358.979.323.844 | 576 |
| 314.159.265.358.979.323.820 | 346 | 314.159.265.358.979.323.845 | 576 |
| 314.159.265.358.979.323.821 | 346 | 314.159.265.358.979.323.846 | 576 |
| 314.159.265.358.979.323.822 | 346 | 314.159.265.358.979.323.847 | 576 |
| 314.159.265.358.979.323.823 | 346 | 314.159.265.358.979.323.848 | 576 |
| 314.159.265.358.979.323.824 | 488 | 314.159.265.358.979.323.849 | 576 |

**Abb. 1** Schrittzahlen bis zum Erreichen der 1 für 50 aufeinander folgende Startwerte

Als Schrittzahlen treten in diesem Block nur die Werte 346, 364, 488, 576 und 594 auf. Solch eine nichttriviale Klumpung ist durchaus typisch. (Ist jemandem etwas an den vorderen Ziffern der 21-stelligen Zahlen aufgefallen? Wir waren faul und haben „einfach“ die 19 Startziffern von $\pi$ genommen.)

## Historische Einordnung

Collatz selbst hat wohl nie ernsthaft einen Beweis seiner Vermutung gesucht. Er hat aber, speziell um 1950 herum, etlichen Mathematikern von dem Problem erzählt. Bei einigen, so auch bei seinem Hamburger Kollegen Hermann Hasse, fiel das Problem auf fruchtbaren Boden, und sie erzählten es in der halben Welt weiter: Hasse (1898-1979), Stanislaw Ulam (1909-1984), Shizuo Kakutani (1911-2004).

## Eine Umformulierung und Ergebnisse

Man muss doch nur zeigen, dass es für jede Startzahl $n_0 > 1$ eine Iteration $n_t$ gibt, so dass $n_t < n_0$ ist. Wiederholte Anwendung würde einen dann von $n_t$ zu einem noch kleineren $n_s$ führen usw., bis man endlich bei einem $n_T = 1$ wäre. In Bezug auf diese Formulierung gibt es theoretische Teilergebnisse, von denen das neueste von dem genialen Terence Tao stammt (geboren 1975; Fields-Medaille 2006; Löser etlicher schwerer Probleme).

**Abb. 2** Altmeister Paul Erdős und das Wunderkind Terence Tao im Jahr 1985


(i) Terras (1976) und Everett (1977): Fast alle Startzahlen $n_0$ liefern Folgen mit einem $n_t < n_0$ für mindestens ein $t > 0$.

(ii) Korec (1994): Fast alle $n_0$ liefern Folgen mit einem $n_t < {n_0}^{0,7925}$ (in Wirklichkeit sogar für Exponent $\log_4 3 \approx 0{,}79248$)

Bemerkung: das Vorkommen von 3 und 4 in der Formel für den Exponenten ist kein Zufall. Eine Erklärung würde den Rahmen des Kapitels sprengen, aber so viel: Die 3 kommt von dem „mal 3“, und die $4 = 2 \cdot 2$ daher, dass nach jedem „mal 3“-Schritt im Durchschnitt zwei Halbierungsschritte passieren.

Jetzt kommt ein Block mit einer ganzen Reihe von Ergebnissen, wobei jedes etwas stärker als das vorherige ist.

(iii-a) Tao (2019b): Fast alle $n_0$ liefern ein $n_t < \sqrt{n_0}$.

(iii-b) Tao (2019b): Fast alle $n_0$ liefern ein $n_t < \sqrt[3]{n_0}$.

(iii-c) Tao (2019b): Fast alle $n_0$ liefern ein $n_t < \log_{10}(n_0)$.

(iii-d) Tao (2019b): Fast alle $n_0$ liefern ein $n_t < \log_{10} \log_{10}(n_0)$.

Und nun das volle Ergebnis von **Terence Tao (2019b)**:

> Sei $g : \mathbb{N} \to \mathbb{R}$ eine Funktion mit $\lim\limits_{n\to\infty} g(n) = \infty$. Dann liefern fast alle Startwerte $n_0$ Collatz-Folgen, in denen ein $n_t$ vorkommt mit $n_t < g(n_0)$.

Das 49-Seiten-Preprint von Tao ist zum Teil technisch heftig. Es kommen beim Durchblättern aber immer ein paar vertraute Symbole vor, unter anderem der Faktor $\frac{3}{4}$. Tao nennt in seinem Artikel übrigens $\log\log\log\log(n)$ als langsam wachsende Beispielfunktion. Eine anschauliche Beschreibung von Taos Fortschritt findet sich in Davies und Hartnett (2019). Sehr interessant liest sich auch der Blog von Tao (2019a) zum Thema Collatz-Problem. Im Blog spekulierte Tao im Januar 2020, dass sein Ergebnis wahrscheinlich in der Richtung verschärft werden kann, dass der kleine $n_t$-Wert typischerweise zu einem Zeitpunkt t erreicht wird, der höchstens $\frac{2+\varepsilon}{\log \frac{4}{3}} \cdot \log(n_0)$ ist.

## Was bedeutet „fast alle“?

Dazu müssen wir etwas ausholen. Für eine Teilmenge der natürlichen Zahlen kann man ihre Dichte erklären. Wir beginnen mit zwei einfachen Beispielen.

Die geraden Zahlen sind $\{2, 4, 6, 8, 10, 12, ...\}$. Es ist also jede zweite natürliche Zahl gerade. Deshalb haben die geraden Zahlen Dichte $\frac{1}{2}$. Ebenso haben auch die ungeraden Zahlen Dichte $\frac{1}{2}$.

Was ist mit den Quadratzahlen $Q = \{1, 4, 9, 16, 25, 36, 49, 64, 81, 100,...\}$? Kleine Quadratzahlen gibt es eine Menge, größere längst nicht mehr so viele. Konkreter: Bis 100 gibt es 10 Quadratzahlen, das sind also 10%. Bis 10.000 gibt es genau 100 Quadratzahlen. $\frac{100}{10.000} = 0{,}01$, der Anteil beträgt also nur noch 1%. Bis 1.000.000 gibt es 1.000 Quadratzahlen, was einen Anteil von $0{,}001 = 1‰$ ausmacht. Aus diesen drei Häufigkeiten 0,1 und 0,01 und 0,001 erkennt man gefühlsmäßig, dass der Anteil der Quadratzahlen in 1, 2, ..., n gegen Null läuft, wenn n nach Unendlich strebt. Man sagt deshalb: Die Quadratzahlen haben in den natürlichen Zahlen Dichte 0.

Ein spannendes Beispiel: Wir führen eine neue Menge von Zahlen ein und nennen sie „Multikulti-Zahlen“. Eine natürliche Zahl ist multikulti, wenn in ihr alle zehn Dezimalziffern vorkommen. Dabei seien führende Nullen nicht mitgezählt. Die kleinste Multikulti-Zahl ist 1.023.456.789. Sie ist also etwas größer als 1 Milliarde. Es folgen dann aufsteigend: 1.023.456.798, 1.023.456.879, 1.023.456.897, 1.023.456.978, ... , 1.098.765.432.

$1.098.765.432 - 1.023.456.788 \approx 75{,}3$ Millionen. In diesem Block enthalten sind $8! = 40.320$ Multikulti-Zahlen. $\frac{40.320}{75{,}3 \text{ Mio.}} \approx 0{,}0005$, also immer noch sehr wenige.

FRAGE: Welche Dichte haben die Multikulti-Zahlen asymptotisch? Die Antwort ist auf den ersten Blick überraschend: Fast alle langen Zahlen sind Multikulti-Zahlen. Das Argument für 1.001-stellige Zahlen mit führender 1 geht so: In dieser Menge gibt es weniger als $10 \cdot 9^{1.000}$ Zahlen, die nicht multikulti sind. Und $10 \cdot 9^{1.000}$ ist enorm viel kleiner als $10^{1.000}$.

Für Leser mit Würfelbegeisterung mag folgende Veranschaulichung helfen: 1.000 Ziffern seien zufällig und unabhängig voneinander mit einem fairen zehnseitigem Würfel geworfen. Mit „ziemlicher Sicherheit“ kommt jede der Ziffern 0 bis 9 dabei vor. (Bei 10.000 Ziffern wird es noch deutlicher.)

Das Beispiel der Multikulti-Zahlen soll vor allem eines zeigen: Eine Menge, die fast alle natürlichen Zahlen enthält, kann ganz „dünn“ loslegen, sogar für einen ziemlich langen Zeitraum, aber irgendwann „dominant“ werden.

Das Ergebnis von Tao sagt, dass, wenn die Startzahlen $n_0$ nur groß genug gewählt sind, fast alle $n_0$-Werte zumindest zwischenzeitlich zu Werten $n_t$ führen, die im Vergleich mit $n_0$ klitzeklein sind. Aber eine Warnung muss sein: Als ich neulich vor talentierten Schülern der Klassen 8 und 9 über das Thema vortrug, meldete sich ein schlauer Kopf: „Wir wissen ja, dass alles, was kleiner als $10^{20}$ ist, in die 1 läuft“. Man könnte doch jetzt die Funktion $g(\cdot)$ im Ergebnis von Tao so langsam wachsend wählen, dass man fast sicher unter die $10^{20}$ gelangt. Das Problem ist aber, dass das Ergebnis von Tao machtlos ist, wenn das „fast alle“ zu $g(\cdot)$ erst so spät einsetzt, dass man mit $g(n_0)$ gar nicht unter die $10^{20}$ gelangt.

Tao benutzt bei seiner Definition von „fast alle“ übrigens nicht diese natürliche Definition von Dichte, sondern eine etwas kompliziertere mit logarithmischen Gewichten. Mit der lassen sich leichter zwischenzeitliche Durststrecken (also Intervalle mit „falschen“ Werten) wegstecken. Die Details zu erklären, würde hier zu weit führen.

## Natürliche Varianten des Collatz-Problems

Beim 3n-1-Problem, also mit Minus statt Plus, wird bei geraden Zahlen auch durch 2 geteilt, und bei ungeraden Zahlen mit 3 multipliziert und dann 1 abgezogen. Bekannt sind bisher drei Zyklen, nämlich

$$1 \to 2 \to 1, \tag{6}$$

$$5 \to 14 \to 7 \to 20 \to 10 \to 5, \tag{7}$$

$$\begin{aligned} &17 \to 50 \to 25 \to 74 \to 37 \to 110 \to 55 \to 164 \to 82 \to 41 \\ &\quad \to 122 \to 61 \to 182 \to 91 \to 272 \to 136 \to 68 \to 34 \to 17. \end{aligned} \tag{8}$$

Die Vermutung ist: Jede Startzahl $n_0$ läuft nach endlichen vielen Schritten in einen dieser drei Zykel.

Beim 3n+5-Problem finden sich die Zykel

$$1 \to 8 \to 4 \to 2 \to 1, \tag{9}$$

$$5 \to 20 \to 10 \to 5, \tag{10}$$

$$19 \to 62 \to 31 \to 98 \to 49 \to 152 \to 76 \to 38 \to 19, \tag{11}$$

$$23 \to 74 \to 37 \to 116 \to 58 \to 29 \to 92 \to 46 \to 23, \tag{12}$$

$$171 \to \ldots \text{ein echter Kracherzyklus, mit 54 Elementen,} \tag{13}$$

$$187 \to \ldots \text{noch ein Kracherzyklus, mit 44 Elementen,} \tag{14}$$

$$347 \to \ldots\ . \tag{15}$$

Pfiffige Leser mögen sich selbst überlegen, warum das 3n+3-Problem nicht wirklich spannend ist, wenn man 3n+1 schon kennt. Das Dankert-Applet (Dankert, 2020) funktioniert übrigens für alle beliebigen ungeraden Zahlen m und das zugehörige 3n+m-Problem.

In Jena haben wir uns andere Varianten des 3n+1-Problems ausgedacht. Beim 4n+x-Problem gibt es die folgenden zwei Typen von Schritten:

Wenn die Zahl durch 3 teilbar ist, so macht man das. Ist die Zahl nicht durch 3 teilbar, wird sie mit 4 multipliziert, und es wird dann entweder 1 oder 2 addiert, damit das Ergebnis durch 3 teilbar ist.

$$1 \to 4 \cdot 1 + 2 = 6 \to 2 \to 4 \cdot 2 + 1 = 9 \to 3 \to 1 \tag{16}$$

Also bildet $1 \to 6 \to 2 \to 9 \to 3 \to 1$ einen Zyklus.

$$\begin{aligned} 7 &\to 4 \cdot 7 + 2 = 30 \to 10 \to 4 \cdot 10 + 2 = 42 \to 14 \to 4 \cdot 14 + 1 = 57 \\ &\to 19 \to 4 \cdot 19 + 2 = 78 \to 26 \to 4 \cdot 26 + 1 = 105 \to 35 \\ &\to 4 \cdot 35 + 1 = 141 \to 47 \to 4 \cdot 47 + 1 = 189 \to 63 \to 21 \to 7 \end{aligned} \tag{17}$$

Also ist auch $7 \to 30 \to \ldots \to 21 \to 7$ ein Zyklus im 4n+x-Problem.

Vermutung (basierend auf vielen Rechenläufen des Studenten Nikita Yanyushkin): Es gibt beim 4n+x-Problem nur diese beiden Zyklen, und jeder Startwert $n_0$ läuft in einen der Zyklen.

Durch das +x nach einem „mal 4“-Schritt ist sichergestellt, dass nach jedem Multiplikations-Schritt mindestens ein Mal dividiert wird. Nimmt man an, dass nach einer Division durch 3 das Ergebnis wieder mit Wahrscheinlichkeit $\frac{1}{3}$ durch 3 teilbar ist, so wird sich bei großen Startzahlen insgesamt eine Abwärtstendenz ergeben: Auf jede Multiplikation mit 4 folgen im Durchschnitt eineinhalb Divisionen durch 3, und es ist $4 < 3^{\frac{3}{2}} \approx 5{,}196$.

Beim 5n+y-Problem gibt es die folgenden zwei Typen von Schritten: Wenn die Zahl durch 4 teilbar ist, so macht man das. Ist die Zahl nicht durch 4 teilbar, wird sie mit 5 multipliziert, und es wird dann entweder 1 oder 2 oder 3 addiert, damit das Ergebnis durch 4 teilbar ist.

$$1 \to 5 \cdot 1 + 3 = 8 \to 2 \to 5 \cdot 2 + 2 = 12 \to 3 \to 5 \cdot 3 + 1 = 16 \to 4 \to 1 \tag{18}$$

Also bildet $1 \to 8 \to 2 \to 12 \to 3 \to 16 \to 4 \to 1$ einen Zyklus.

$$5 \to 5 \cdot 5 + 3 = 28 \to 7 \to 5 \cdot 7 + 1 = 36 \to 9 \to 5 \cdot 9 + 3 = 48 \to 12 \to 3 \tag{19}$$

Damit sind wir mitten im Zyklus 18 gelandet.

$$\begin{aligned}23 &\to 116 \to 29 \to 148 \to 37 \to 188 \to 47 \to 236 \to 59 \to 296\\ &\to 74 \to 372 \to 93 \to 468 \to 117 \to 588 \to 147 \to 736\\ &\to 184 \to 46 \to 232 \to 58 \to 292 \to 73 \to 368 \to 92 \to 23\end{aligned} \tag{20}$$

Das ist also ein anderer Zyklus mit Länge 26 und kleinstem Element 23.

Hier ist die Vermutung, die auf vielen Rechenläufen von Nikita Yanyushkin basiert: Es gibt beim 5n+y-Problem nur die beiden Zyklen 18 und 20; und jeder Startwert $n_0$ läuft in einen der Zyklen.

Durch das +y nach einem „mal 5“-Schritt ist sichergestellt, dass nach jedem Multiplikations-Schritt mindestens ein Mal dividiert wird. Nimmt man an, dass nach einer Division durch 4 das Ergebnis wieder mit Wahrscheinlichkeit $\frac{1}{4}$ durch 4 teilbar ist, so wird sich bei großen Startzahlen insgesamt eine Abwärtstendenz ergeben: Auf jede Multiplikation mit 5 folgen im Durchschnitt $\frac{4}{3}$ Divisionen durch 4. Und es ist $5 < 4^{\frac{4}{3}} \approx 6{,}35$.

Die Idee mit den Varianten zu „mal 4“ und „mal 5“ funktioniert auch für allgemeines k > 1. Ist die aktuelle Zahl durch k teilbar, so teilt man durch k. Ist die Zahl aber nicht durch k teilbar, multipliziert man mit (k+1) und addiert eine Zahl $s \in \{1, 2, \ldots, k-1\}$, so dass das Ergebnis durch k teilbar ist.

Für jede Variante, in der mit k+1 multipliziert bzw. durch k dividiert wird, erwarte ich, dass alle großen Startzahlen nach und nach zu kleinen Zahlen (und damit in einen Zyklus) führen.

Das heuristische Argument dafür ist: Nach jeder Multiplikation mit k+1 kommt zwingend eine Division durch k. Die Zahl, die sich daraus ergibt, ist mit Wahrscheinlichkeit $\frac{1}{k}$ (natürlich keine echte Wahrscheinlichkeit, sondern nur im Sinne der Heuristik) wieder durch k teilbar. Und falls das so ist, ist das neue Ergebnis wieder mit Wahrscheinlichkeit $\frac{1}{k}$ durch k teilbar, usw.

Insgesamt gibt es nach dem Multiplikationsschritt im Durchschnitt also $1 + \frac{1}{k} + \frac{1}{k^2} + \frac{1}{k^3} + \ldots$ viele Divisionsschritte. Die geometrische Reihe hat den Wert $\frac{k}{k-1}$. Also kommen nach jeder Multiplikation im Durchschnitt $\frac{k}{k-1}$ viele Divisionen. Das bedeutet: Auf $k-1$ viele Multiplikation (mit $k+1$) kommen im Durchschnitt k Divisionen durch k. Die entsprechenden Werte für k=1,...,10 sind in Tab. 1 veranschaulicht.

Für alle $k \geq 2$ ist $k^k$ echt größer als der Vergleichswert $(k+1)^{k-1}$. Dieses kann man auch leicht theoretisch beweisen, wobei man den Grenzübergang $\left(1 + \frac{1}{k}\right)^{k-1} \to e \approx 2{,}728$ benutzt.

| $k$ | $(k+1)^{k-1}$ | $k^k$ |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 3 | 4 |
| 3 | 16 | 27 |
| 4 | 125 | 256 |
| 5 | 1.296 | 3.125 |
| 6 | 16.807 | 46.656 |
| 7 | 262.144 | 823.543 |
| 8 | 4.782.969 | 16.777.216 |
| 9 | 100.000.000 | 387.420.489 |
| 10 | 2.357.947.691 | 10.000.000.000 |

**Tab. 1** Zur durchschnittlichen Größe der Multiplikationsschritte (mittlere Spalte) und der Divisionsschritte (rechts)

Für alle, die gerne per Hand und ohne elektronische Hilfe Beispieldaten sammeln wollen, eignen sich zwei Werte für k besonders: Bei $k = 10$ ergibt sich der Divisionsschritt durch Streichen der letzten 0, und das „mal 11" ist die einfache Addition von $n_t$ und $10\,n_t$, wobei man nach dem Addieren noch auf den nächsten Zehner aufrundet.

Beispiel: $1.137 \to 1.137 + 11.370 = 12.507$, aufgerundet zu 12.510. Streichen der 0 gibt dann den neuen Wert 1.251. Genommen über mehrere Iterationen ergibt sich die Ziehharmonika-Treppe

$$\begin{aligned} 1.137 &\to 12.507 \to 1.251 \to 13.761 \to 1.377 \to 15.147 \to 1.515 \\ &\to 16.665 \to 1.667 \to 18.337 \to 1.834 \to 20.174 \to 2.018 \\ &\to 22.198 \to 22.200 \to 2.220 \to 222 \to \ldots \end{aligned} \tag{21}$$

In der letzten gezeigten Runde durfte zwei Mal durch 10 geteilt werden. Die Zahlenfolge fiel sozusagen nicht nur eine Stufe, sondern zwei Stufen bergab. Ganz am Ende landet man im Zyklus $3 \to 33 \to 4 \to 44 \to 5 \to 55 \to 6 \to 66 \to 7 \to 77 \to 8 \to 88 \to 9 \to 99 \to 10 \to 1 \to 11 \to 2 \to 22 \to 3$.

Bei der Variante k = 9 ergibt sich das „mal 10" durch Anhängen einer Null. Danach bildet man die Quersumme und addiert zur Zahl so viel (einen Wert zwischen 1 und 8), dass sich als neue Quersumme ein Vielfaches von 9 ergibt. Die folgende Division durch 9 ist nicht ganz trivial ohne Taschenrechner, aber doch noch machbar, wenn die Zahl $n_t$ höchstens 4-stellig ist. Auch hierzu ein Beispiel. Der Startwert sei 158.

$$158 \to 1.584 \to 176 \to 1.764 \to 196 \to 1.962 \to 218 \to 2.187 \to 243 \to 27 \to 3 \tag{22}$$

Das letzte war also eine Glückssequenz (für den, der kleine Zahlen anstrebt) mit drei Divisionen in Folge. Die 3 gehört schon zum Zyklus, der in der unteren Ebene die Werte $3 \to 36 \to 4 \to 45 \to 5 \to 54 \to 6 \to 63 \to 7 \to 72 \to 8 \to 81 \to 9 \to 1 \to 18 \to 2 \to 27 \to 3$ annimmt. Die Doppeldivision im Stück $81 \to 9 \to 1$ sorgt dafür, dass der Zyklus sich schließt.

# Eine Variante des Collatz-Problems mit Zufallseinfluss

An einigen Stellen haben wir argumentiert, dass beim 3n+1-Problem wohl alle Startwerte in einen Zyklus mit kleinen Werten laufen dürften, weil im Durchschnitt auf jeden „mal 3“-Schritt zwei Schritte vom Typ „geteilt durch 2“ kommen.

Dieses „im Durchschnitt“ stimmt aber nicht im Sinne der Wahrscheinlichkeitsrechnung, weil das ganze System mit seinen Rechenregeln komplett deterministisch ist, ganz ohne Zufallsentscheidungen.

Fügt man ein kleines bisschen an Zufall in das Modell ein, lässt sich tatsächlich beweisen, dass es für jeden Startwert $n_0$ mit Wahrscheinlichkeit 1 in der „Auffangmenge“ $\{1, 2, 4\}$ endet. Das Modell sieht konkret so aus:

Ist die aktuelle Zahl $n_t$ ungerade, so wird $n_{t+1} = 3 \cdot n_t + 1$. Ist $n_t$ dagegen gerade und größer als 2, wird eine faire Münze geworfen. Bei „Vorderseite“ setzt man $n_{t+1} = \frac{n_t}{2}$ und bei „Rückseite“ $n_{t+1} = \frac{n_t}{2} - 1$. Bei 2 wird keine Münze geworfen, sondern direkt zur 1 gegangen.

Beispiele: 16 ist gerade und geht deshalb mit Wahrscheinlichkeit $\frac{1}{2}$ über in 8 und mit $\frac{1}{2}$ über in 7. Von 4 aus geht es mit Wahrscheinlichkeit von je $\frac{1}{2}$ nach 2 und nach 1.

Ist die Folge erst einmal in der Menge $\{1, 2, 4\}$ angelangt, kommt sie nicht mehr hinaus.

Durch dieses „Zufallswackeln“ folgen auf jeden Multiplikationsschritt im Durchschnitt zwei Divisionsschritte, nämlich einer mit Wahrscheinlichkeit $\frac{1}{2}$, 2 mit Wahrscheinlichkeit $\frac{1}{4}$, 3 mit Wahrscheinlichkeit $\frac{1}{8}$, usw. Die unendliche Summe $\frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 3 + \frac{1}{16} \cdot 4 + ...$ hat den Wert 2.

## Die Online-Enzyklopädie der Zahlenfolgen (OEIS)

OEIS ist eine englische Abkürzung und steht im Original für „Online Encyclopedia of Integer Sequences“. Ursprünglich war es ein Nachschlagebuch aus Papier, ins Leben gerufen von Neil Sloane im Jahr 1964. Seit 1996 gibt es das Opus als Online-Datenbank. Das Tolle an dieser Datenbank ist, dass man nicht nur Zahlenfolgen oder Zahlenmengen in die Suchmaske eingeben kann, sondern auch beliebige Zeichenfolgen, zum Beispiel Worte oder Namen. Gibt man das Suchwort „collatz“ ein, werden (Stand 10. Februar 2020) 882 Treffer angezeigt, unter anderem folgende drei Folgen mit Bezug zum Collatz-Problem:

Folge A006370: 0, 4, 1, 10, 2, 16, 3, 22, 4, 28, 4, 34, 6, 40, 7, 46, 8, 52, 9, 58, 10, ... An Stelle i gibt diese Folge den Collatz-Nachfolger für i an, beginnend bei 0.

Folge A006577: 0, 1, 7, 2, 5, 8, 16, 3, 19, 6, 14, 9, 9, 17, 17, 4, 12, 20, 20, 7, 7, 15, 15, 10, 23, 10, 111, 18, ... Angegeben sind hier die Anzahl Schritte, bis man von $n_0$ zur 1 gelangt. Die Folge beginnt bei der Schrittzahl für 1. Bemerkenswert ist z. B. die große Zahl 111 für $n_0 = 27$.

Folge A005186: 1, 1, 1, 1, 1, 2, 2, 4, 4, 6, 6, 8, 10, 14, 18, 24, 29, 36, 44, 58, 71, 91, 113, 143, 179, 227, 287, 366, 460, 578, 732, 926, 1174, 1489, 1879, 2365, 2988, 3780, 4788, 6049, ... Angegeben ist an der Stelle i, wie viele Startzahlen $n_0$ es gibt, die nach genau i Schritten erstmals die 1 erreichen. Im Text zur Folge ist eine Strukturvermutung erklärt: Für hinreichend große i soll die Folge näherungsweise exponentiell wachsen, es gilt $a_{i+1} \approx 1{,}264 \cdot a_i$.

Noch etwas Statistik, wieder mit Stand vom 10. Februar 2020: Zu 3x+1 zeigt die OEIS-Suche 1481 Treffer, zu 3n+1 immerhin noch 1126 Treffer.

Zu etlichen Folgen aus der OEIS kann man mithilfe des Dankert-Applets einfach weitere Datenpunkte generieren. Durch die ungeheure Datenmenge und Vielfalt in der OEIS war meiner Meinung nach die Chance noch nie so groß wie heute, durch geschickte Vermutungen inneliegende 3n+1-Strukturen zu entdecken.

Vielleicht wird sogar, wenn es eines Tages einen Beweis gibt, dieser zu wesentlichen Teilen auf Beobachtungen in der OEIS basieren. Die gleichzeitige Existenz des Dankert-Applets und der OEIS gibt auch Nichtprogrammierern eine gute Chance, Strukturen (und vielleicht sogar einen Beweis) für das 3n+1-Problem zu finden.

## Collatzfolgen bei benachbarten Startzahlen

Hin und wieder (und ohne spezielles System) haben wir Blöcke von 100 aufeinander folgenden hundertstelligen Zahlen mit dem Dankert-Appplet angeschaut und dabei festgestellt, dass oft in solchen Blöcken nur genau eine Iterationszahl vorkommt. Es war also noch deutlich langweiliger als in der Abb. 1 mit den fünfzig 21-stelligen Zahlen und ihren fünf verschiedenen Iterationszahlen.

Dies führt uns zu der Vermutung: Bei großen Startzahlen $n_0$ und $n_0+1$ ergibt sich mit sehr großer Wahrscheinlichkeit die genau gleiche Schrittzahl bis zum Erreichen der 1. Das ist zumindest etwas erstaunlich, weil ja eine der beiden Zahlen gerade und die andere ungerade ist, so dass sich durch die allererste Iteration schon eine Diskrepanz ergibt: „mal 3“ gegen „dividiert durch 2“.

Sei $G(n)$ = Anzahl der $n_0$ in $\{1, 2, ..., n\}$, so dass die Startwerte $n_0$ und $n_0 + 1$ die gleiche Schrittzahl bis zum erstmaligen Erreichen der 1 haben. Dann kann man vermuten, dass $\lim\limits_{n\to\infty} \frac{G(n)}{n} = 1$. Oder als Dichteformulierung: Die asymptotische Dichte der Startwerte $n_0$, bei denen $n_0 + 1$ die gleiche Iterationszahl hat, ist 1. Experimentell kann man, auch mit Monte-Carlo-Experimenten, ermitteln, wie schnell die Konvergenz der Dichte gegen 1 ist.

## Toller Repräsentant und Botschafter für die Experimentelle Mathematik

Das Collatz-Problem ist nur eines von vielen aus einer großen Problemklasse. Ausgehend von einer natürlichen Zahl macht man abwechselnd Multiplikations- und Divisions-Schritte. Dabei ist der Multiplikator größer als der Teiler. Als Ausgleich kommen die Multiplikationsschritte nur einzeln vor und Divisions-Schritte des öfteren in „Pulks“. Das System ist so austariert, dass „im Durchschnitt“ stärker dividiert als multipliziert wird.

Das 5n+1-Problem gehört nicht zu der Klasse: Nach jeder Multiplikation mit 5 passieren im Durchschnitt „nur“ zwei Divisionen durch 2. Und weil $\frac{5}{4}$ größer als 1 ist, laufen viele Startwerte mit ihren Zahlenfolgen in Richtung Unendlich.

Das 3n+1-Problem ist so etwas wie der beste Repräsentant: Es hat, wenn unser aller Intuition stimmt, genau einen Zyklus, und dieser ist so einfach, dass man ihn von jedem Drittklässler – oder auch von Pippi Langstrumpf – durchrechnen lassen kann.

Auch Menschen, die keine formale mathematische Ausbildung haben, aber gerne mit Zahlen spielen, können sich mit Gewinn mit dem Collatz-Problem beschäftigen. Insbesondere lassen sich mit etwas Vorbildung im Bereich Computer-Programmierung einfache Programme schreiben, mit denen man für diesen oder jenen Aspekt oder auch diese oder jene Variante des Problems Daten erzeugen und dann Vermutungen aufstellen kann. Mit der OEIS gibt es auch einen Ort im Internet, wo interessante gesammelte Daten (zum Collatz-Problem, aber auch zu vielen anderen Problemen) angehäuft sind und zum Beispiel auch von theoretischen Mathematikern bei ihren Versuchen für Beweise durchgeschaut werden können.

# Danksagung

Dr. Raphael Thiele, Dr. Roland Voigt, Dr. Andreas Rüdinger, Frau Stella Schmoll und Dr. Torsten Sillke haben eine frühere Version des Artikels kritisch gelesen und konstruktive Anmerkungen gemacht. Raphael Thiele half wesentlich beim Tippen und der TeX-Formatierung. Der Masterstudent Nikita Yanyushkin führte mehrere Proberechnungen zu Varianten des Collatz-Problems durch. Ohne das tolle Applet von Prof. Dr. Jürgen Dankert wäre ich nie so tief experimentell in das 3n+1-Problem eingedrungen.

# Literaturverzeichnis


Althöfer I.: Lothar Collatz zwischen 1933 und 1950 - eine Teilbiographie. 3-Hirn-Verlag, 2. Auflage, Lage (2019)

Dankert J.: Applet für das Collatz-Problem. Online verfügbar unter http://www.juergendankert.de/spezmath/html/collatzm.html.

Davies J. und Hartnett K.: Mathematician Proves Huge Result on 'Dangerous' Problem. Online verfügbar unter https://www.quantamagazine.org/mathematician-terence-tao-and-the-collatz-conjecture-20191211/

Everett C.J.: Iteration of the number-theoretic function f(2n)=n, f(2n+1)=3n+2. Advances in Mathematics 25, 42-45 (1977)

Korec I.: A density estimate for the 3x + 1 problem. Mathematica Slovaca 44, 85-89 (1994)

Lagarias J.C.: The Ultimate Challenge: The 3x+1 Problem. American Mathematical Society, Nachdruck (2011)

Roosendaal E.: Webseite zum 3n+1-Problem. Online verfügbar unter http://ericr.nl/wondrous/

Tao T.: Almost all Collatz orbits attain almost bounded values (2019a). Online verfügbar unter https://terrytao.wordpress.com/2019/09/10/almost-all-collatz-orbits-attain-almost-bounded-values/

Tao T.: Almost all orbits of the Collatz map attain almost bounded values, arXiv (2019b). https://arxiv.org/abs/1909.03562

Terras R.: A stopping time problem on the positive integers. Acta Arithmetica 30, 241-252 (1976)


Alle Weblinks wurden am 21. Januar 2020 überprüft.