Beobachtungen an einer speziellen Baumstruktur der ungeraden Collatz-Folge

11.03.2010
von Mike Winkler

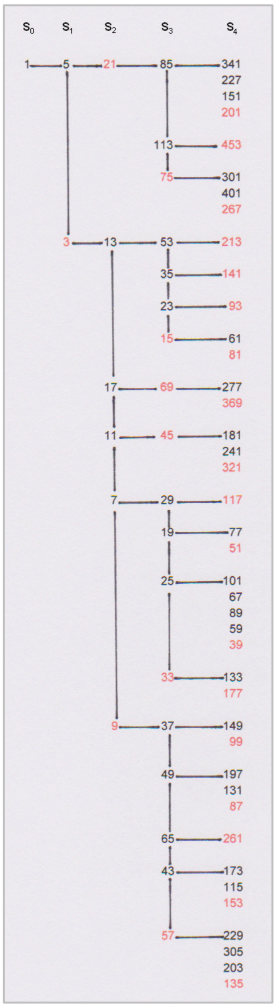
Das Konstruktionsprinzip der speziellen Baumstruktur

Für alle ungeraden Zahlen, für die die Collatz-Folge bei 1 endet, soll ein Graph konstruiert werden, der die Collatz-Iteration rückwärts verfolgt. Der so genannte **ungerade Collatz-Baum**. Für dessen Konstruktion wählen wir eine quadratische Gitterstruktur, so dass die Äste nur waagerecht und senkrecht angeordnet sind.

Die Menge der ungeraden Zahlen teilt sich modulo 3 in die Restklassen 1, 2 und 0. Wir definieren die Zahlen der Restklasse 0, also die Vielfachen von 3, als die **Blätter** des ungeraden Collatz-Baumes. Für eine bessere Übersicht werden diese stets rot geschrieben.

Im Gegensatz zu gewöhnlichen Collatz-Bäumen (siehe Anhang 3 und 4 bzw. Link 2), werden wir hier die **Blätter in den Baum integrieren**, so dass auch diese eine Fortleitung besitzen. Die eigentliche Iterations-Folge der ungeraden Zahlen wird dadurch nicht verfälscht. Möchte man diese rekonstruieren, müssen lediglich die Blätter auf dem Weg zur 1 übersprungen werden.

Wir beginnen den Baum oben links mit der 1 und konstruieren ihn dann nach rechts und unten weiter. Dabei besitzt jede ungerade Zahl eine Fortleitung nach rechts, gegeben durch $a_{n+1} = 4a_n+1$. Jedoch nur die ungeraden Zahlen der Restklasse 1 und 2 besitzen eine Fortleitung nach unten. Dadurch erhält der Baum eine **Dreiecks-Struktur**, weil die Äste nach unten bei jedem Blatt nur nach rechts verzweigen. Der einzige Nachteil dieser Struktur liegt in ihrer Darstellung. Wachstumsbedingt wird der Baum mit jeder weiteren Spalte um das dreifache nach unten gestreckt.



Das **Konstruktionsprinzip** des ungeraden Baumes lässt sich wie folgt formulieren. Die Pfeile geben die Konstruktionsrichtung für a_{n+1} nach rechts und unten an.

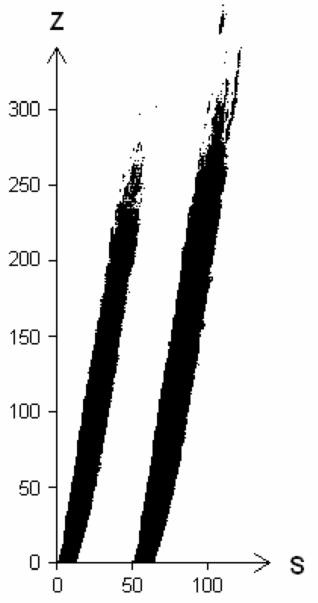
$$a_{n+1} = \begin{cases} 4a_n + 1 & \text{für } a_n \equiv 1 \bmod 2 \quad (\rightarrow) \\ \dfrac{4a_n - 1}{3} & \text{für } a_n \equiv 1 \bmod 3 \quad (\downarrow) \\ \dfrac{2a_n - 1}{3} & \text{für } a_n \equiv 2 \bmod 3 \quad (\downarrow) \end{cases}$$

Das Bild links zeigt die ersten fünf Spalten des ungeraden Collatz-Baumes, konstruiert nach dem vorgestellten Prinzip. Durch die Dreiecks-Struktur des Baumes ist die Anzahl an Zahlen in jeder Spalte begrenzt. Über sechs Spalten erstreckt sich der Baum in normaler Textgröße bereits über 2,5 Seiten. (siehe Anhang 1)

Die Spalten- und Zeilenzahl (Baum-Koordinaten)

Jeder ungeraden Zahl kann jetzt eine Spalten- und Zeilenzahl zugeordnet werden. Wir definieren die Anzahl der Schritte, die eine Zahl im Baum nach links bis zur 1 benötigt als ihre **Spaltenzahl S**. Die Schritte nach oben bis zum ersten waagerechten Ast als ihre **Zeilenzahl Z**. Die Spalten bezeichnen wir mit S_0, S_1, S_2, S_3, …, S_n.

Spalten- und Zeilenzahl sind jedoch nicht eindeutig einer ungeraden Zahl zuzuordnen. So besitzen 213, 227 und 453 jede die **Baum-Koordinaten** S=4 und Z=1. Eindeutig wird die Zuordnung im Baum erst durch die genaue Abfolge der Spalten- und Zeilenschritte, in der Art eines aufsteigenden Treppenmusters. Für 213 gilt [S=3, Z=1, S=1]. Für 227 gilt [Z=1, S=4] und für 453 gilt [S=1, Z=1, S=3]. Die Summe der S und Z ergibt dann wieder die Spalten- und Zeilenzahl. (vgl. Bild links)



Stellt man die ungeraden Zahlen (hier bis 10^8 und 10^9) in einem kartesischen Koordinatensystem durch ihre Baum-Koordinaten dar, ergibt sich folgender Graph, der nach rechts oben wächst, sich in der Breite aber kaum verändert. Der zweite Graph ist zum besseren Vergleich um 50 Spalten nach rechts verschoben. (Bild links)

Drei Gruppen von Blättern

Bedingt durch das Konstruktionsprinzip besitzen die einen Blätter eine Fortleitung nach oben, die anderen nicht. Dabei treten erstere in zwei verschiedenen Formen auf. Wir können somit **drei Gruppen von Blättern** unterscheiden. Die Zugehörigkeit eines Blattes zu einer Gruppe definieren wir dabei wie folgt. Dabei ist k eine natürliche Zahl.

Gruppe 1: Blätter **mit** Fortleitung nach oben der Form **12k+3** (3, 15, 27, 39, 51, 63, 75, 87,…)
Gruppe 2: Blätter **mit** Fortleitung nach oben der Form **24k+9** (9, 33, 57, 81, 105, 129, 153,…)
Gruppe 3: Blätter **ohne** Fortleitung nach oben der Form **24k+21** (21, 45, 69, 93, 117, 141, 165,…)

Die Blätter der Gruppe 1 und 2 werden wir auch als die **Eck-Blätter** des Baumes bezeichnen, die Blätter der Gruppe 3 als **Ast-Blätter**. Eine ungerade Zahl, die kein Blatt ist wird auch ein **Nicht-Blatt** genannt.

Die Spalten-Sprünge der Blätter

Da konstruktionsbedingt kein Blatt im Baum eine Fortleitung nach unten besitzt finden sich in den Spalten an diesen Stellen Lücken. Zum Beispiel in Spalte S_2 von 21 nach 13 oder in Spalte S_3 von 75 nach 53. Wir definieren als **Sprung** einer Spalte S_n die arithmetische Beziehung eines Blattes b zu der direkt unter ihr gelegenen Zahl a. Dabei ist a entweder ein Blatt der Gruppe 3 oder ein Nicht-Blatt.

Die Sprünge der Ast-Blätter sind arithmetisch leicht zu fassen.

Einem Blatt b der Gruppe 3 folgt immer eine kleinere Zahl a, gegeben durch $a = \frac{2b}{3} - 1$

Die Sprünge der Eck-Blätter hingegen scheinen sich einer einfachen Gesetzmäßigkeit zu entziehen.

Wir definieren als **Spaltenvorgänger** eines Eck-Blattes b in Spalte S_n diejenige Zahl a in Spalte S_{n-1} dessen Fortleitung nach rechts zu b führt. Die Anzahl an Zahlen zwischen a und b sei dabei die **Schrittzahl x**. Jedes Eckblatt besitzt genau einen eindeutigen, kleineren Spaltenvorgänger. Das Eck-Blatt 15 in Spalte S_3 besitzt zum Beispiel den Spaltenvorgänger 13 in Spalte S_2. Die Schrittzahl beträgt x=3, da zwischen 15 und 13 drei weitere Zahlen liegen.

Sprünge der Eck-Blätter der Gruppe 1

ist der Spaltenvorgänger ein Blatt und $x \leq 3$, dann folgt ein Sprung auf eine kleinere Zahl

ist der Spaltenvorgänger ein Nicht-Blatt und $x \leq 3$, dann folgt ein Sprung auf eine größere Zahl

Sprünge der Eck-Blätter der Gruppe 2

folgt ein Sprung auf eine kleinere Zahl mit $x \leq 3$, dann ist der Spaltenvorgänger ein Blatt

folgt ein Sprung auf eine größere Zahl mit $x \geq 3$, dann ist der Spaltenvorgänger ein Nicht-Blatt

Wie viele Zahlen besitzen die Spalten?

Wir wissen bereits, dass konstruktionsbedingt die Anzahl an Zahlen in jeder Spalte begrenzt ist. Doch wie viele Zahlen besitzt jede Spalte genau? Ihre genaue Anzahl ermöglicht uns einen Einblick in das Wachstumsverhalten des Baumes. Wir definieren A_n als die Anzahl aller ungeraden Zahlen in der Spalte S_n.

Spalte S_n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Anzahl A_n	1	2	6	18	41	130	399	1186	3591	10684	32132	96907	290511	869803

Vergleicht man die Anzahl A_n jeder Spalte miteinander, ergibt sich ein Vergrößerungsfaktor von Spalte zu Spalte von fast genau 3. (vgl. Tabelle 1, Anhang 7)

Es gilt die Beziehung: $\frac{A_n}{A_{n-1}} \approx 3$

Wie können vermuten, dass sich der Quotient A_n/A_{n-1} mit zunehmender Spaltenzahl immer genauer dem Wert 3 annähert oder immer ungefähr gleich 3 ist. Dieser Wert überrascht jedoch nicht. Er zeigt, dass sich jede Zahl, die kein Blatt ist, im Durchschnitt mit zwei Zahlen nach unten fortsetzt bis sie wieder ein Blatt erreicht. Wie die 85 mit 113 und 75, oder die 181 mit 241 und 321.

Die Realität des Baumes sieht natürlich anders aus. So folgt einer Zahl mal nur eine weitere, wie 61, 81. Einer anderen dagegen mal 17 Zahlen, wie 445, 593, 395, 263, 175, 233, 155, 103, 137, 91, 121, 161, 107, 71, 47, 31, 41, 27. (vgl. Anhang 2)

Wie viele Blätter besitzen die Spalten?

Konstruktionsbedingt besitzt jedes Blatt im Baum eine Fortleitung, meist über mehrere Zahlen, zu einer Zahl in der vorhergehenden Spalte. Die Anzahl der Blätter in einer Spalte ist somit immer gleich der Anzahl aller Zahlen in der Spalte davor. Wir definieren B_n als die Anzahl aller Blätter in der Spalte S_n. Das Ergebnis der Zählung überrascht daher nicht.

Spalte S_n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Anzahl A_n	1	2	6	18	41	130	399	1186	3591	10684	32132	96907	290511	869803
Anzahl B_n	0	1	2	6	18	41	130	399	1186	3591	10684	32132	96907	290511

Es gilt die Beziehung: $B_n = A_{n-1}$ und somit $\frac{A_n}{B_n} \approx 3$

Wie viele Zahlen der Spalte S_n sind größer als die Zahl a_n des Hauptastes?

Wir definieren als **Hauptast** des Baumes den obersten waagerechten Ast gegeben durch 1, 5, 21, 85, 341,... . Die Zahlen des Hauptastes sind gegeben durch die rekursive Weiterleitung $a_{n+1} = 4a_n+1$ mit $a_0=1$. Wir definieren G_n als die Anzahl aller ungerader Zahlen a in der Spalte S_n für die $a > a_n$ gilt.

Spalte S_n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Anzahl G_n	/	0	0	1	3	11	40	109	361	1144	3450	10385	30976	91980

Vergleicht man die Anzahl G_n jeder Spalte miteinander, scheint sich auch hier der Vergrößerungsfaktor von Spalte zu Spalte dem Wert 3 zu nähern. (vgl. Tabelle 2, Anhang 7)

Es gilt die Beziehung: $\frac{G_n}{G_{n-1}} \approx 3$

Wie viele Blätter der Spalte S_n sind größer als die Zahl a_n des Hauptastes?

Wir definieren H_n als die Anzahl aller Blätter b in der Spalte S_n für die $b > a_n$ gilt.

Spalte S_n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Anzahl H_n	/	0	0	0	2	6	10	47	120	378	1171	3448	10326	30598
Anzahl G_n	/	0	0	1	3	11	40	109	361	1144	3450	10385	30976	91980

Vergleicht man die Anzahl H_n jeder Spalte miteinander, scheint sich auch hier der Vergrößerungsfaktor von Spalte zu Spalte dem Wert 3 zu nähern. (vgl. Tabelle 3, Anhang 7)

Es gilt die Beziehung: $\frac{H_n}{H_{n-1}} \approx 3$

Die Zusammenhänge zwischen A_n, B_n, G_n und H_n stellen sich wie folgt dar. Es gelten die Beziehungen:

$$\frac{A_n}{B_n} \approx 3 \qquad \sqrt{\frac{A_n}{G_n}} \approx 3 \qquad \sqrt[3]{\frac{A_n}{H_n}} \approx 3 \qquad \frac{G_n}{H_n} \approx 3 \qquad \sqrt{\frac{B_n}{H_n}} \approx 3$$

Betrachten wir den natürlichen Logarithmus von A_n und G_n fällt eine enge Beziehung zur jeweiligen Spalte S_n auf. Dabei sind log (A_n) und log (G_n) auf eine Nachkommastelle gerundet. Die Differenz zwischen log (A_n) und log (G_n) beträgt dabei immer ungefähr 2,3.

S_n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
log (A_n)	0	0,7	1,8	2,9	3,7	4,9	6,0	7,1	8,2	9,3	10,4	11,5	12,6	13,7
log (G_n)	/	/	/	0	1,1	2,4	3,7	4,7	5,9	7,0	8,1	9,2	10,3	11,4

Es gilt die Beziehung: $\frac{\log A_n}{S_n} \approx 1$ und $\frac{\log G_n}{S_{n-2}} \approx 1$ bzw. $\log A_n - \log G_n \approx \log 10$

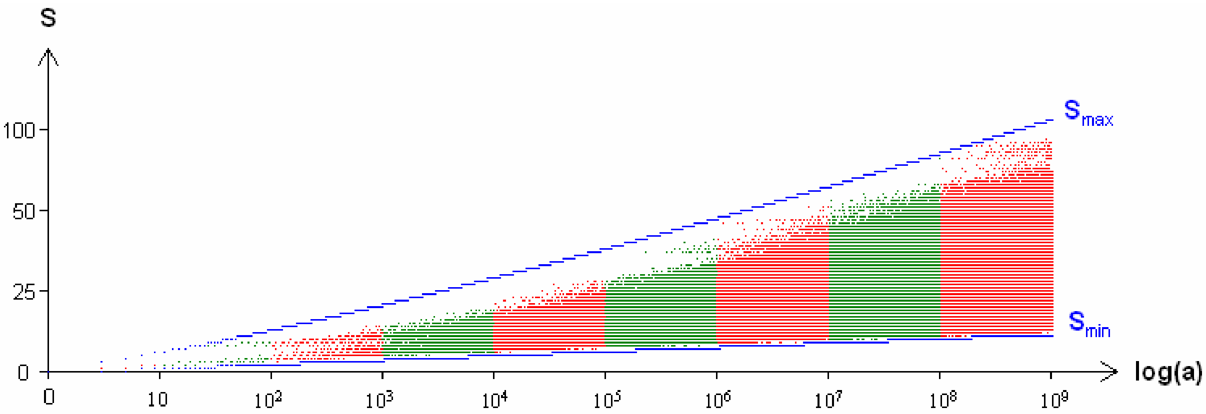
Zusammenfassung der Spaltenverteilungen

Zusammenfassend können wir für die Zahlenverteilung im Baum folgendes festhalten.

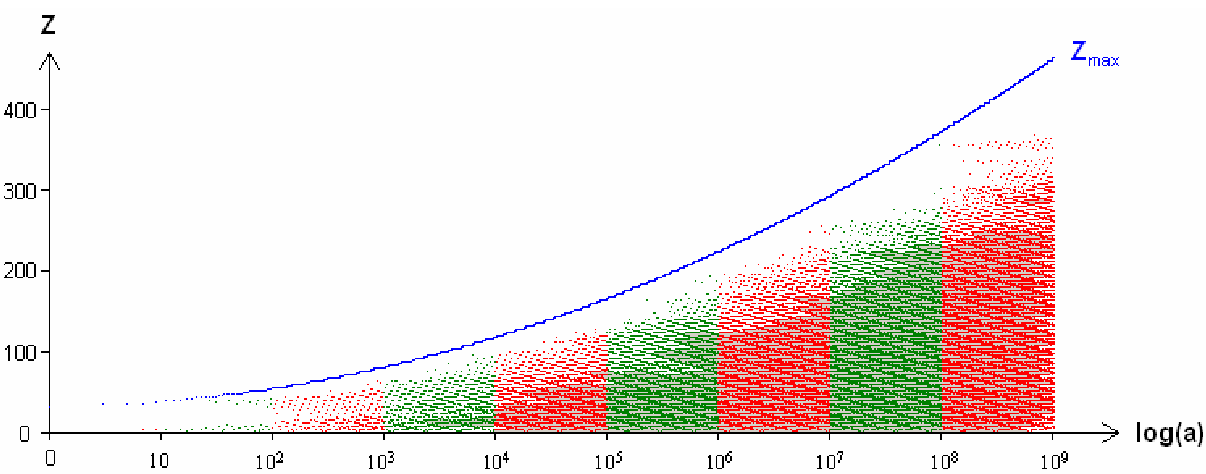
- die Anzahl an Zahlen nimmt von Spalte zu Spalte um ca. das dreifache zu
- die Anzahl an Blättern einer Spalte ist gleich der Anzahl aller Zahlen in der vorhergehenden Spalte
- ungefähr 1/3 der Zahlen einer Spalte sind Blätter
- ungefähr 1/9 der Zahlen einer Spalte sind größer als die zugehörige Zahl des Hauptastes
- ungefähr 1/9 der Blätter einer Spalte sind größer als die zugehörige Zahl des Hauptastes

Bis zu welcher Spalte und Zeile finden sich die ungeraden Zahlen?

Trägt man die ungeraden Zahlen a in einem kartesischen Koordinatensystem logarithmisch auf die Abszissenachse und die zugehörige Spaltenzahl auf die Ordinatenachse auf, ergibt sich das Bild einer relativ gleichmäßig wachsenden Verteilung. Die Spaltenzahl bleibt dabei unter einer oberen Schranke S_{max}. Der untere Teil des Graphen lässt erkennen, dass die Spalten ab einer gewissen Größe abgeschlossen sind. Hier existiert eine untere Schranke S_{min}.



Trägt man in gleicher Weise die Zeilenzahl gegen die ungeraden Zahlen auf, ergibt sich ein ähnliches Bild. Auch hier bleibt die Zeilenzahl unter einer oberen Schranke Z_{max}.



Die Spalten- und Zeilenzahl unterliegt also einer groben Gesetzmäßigkeit. Die blauen Kurven geben die logarithmischen Approximationsfunktionen dafür an. Wir können damit den Aufenthaltsort einer Zahl im Baum einschränken und grob bestimmen. Es lässt sich folgende Aussage treffen.

Die ungerade Zahl a findet sich in der Baumstruktur in den Spalten S_{min} bis S_{max} mit der Zeilenzahl Z_1 bis Z_{max}.

Dabei gilt: $S_{min} = \text{floor}\left(\frac{\log a}{\sqrt{3}}\right)$ $\quad S_{max} = \text{ceil}\left(\log a \cdot \left(\sqrt[3]{\log a} + 1\right)\right)$ $\quad Z_{max} = \text{ceil}\left((\log a)^2 + 33\right)$

In ihrer Gesamtheit betrachtet sind die ungeraden Zahlen also sehr gleichmäßig in der speziellen Struktur des ungeraden Collatz-Baumes verteilt.

Die Beziehung zwischen den Blättern mit gleicher Spaltenzahl

Betrachtet man die Primfaktoren der Blätter einer Spalte fällt folgende Beziehung auf**.** Die Faktoren > 3 entsprechen den Zahlen in derselben oder einer anderen, meist höheren Spalte. Oft gilt sogar die Beziehung, dass sich die Faktoren der Blätter aus Spalte S_n in Spalte S_{2n} wieder finden, zum Teil in identischer Reihenfolge.

Dazu einige Beispiele. Auf die farbliche Kennzeichnung der Blätter wird in diesem Abschnitt verzichtet. Eine identische Reihenfolge ist durch Unterstreichung gekennzeichnet.

In **Spalte S_3** haben wir die Blätterfolge von oben nach unten 75, 15, 69, 45, 33, 57.
Dividiert durch 3 ergibt sich die Folge 25, 5, 23, 15, 11, 19.
Wir finden in Spalte S_3 die Zahlenfolgen 23, 15 und 19, 25.

In **Spalte S_4** haben wir die Blätterfolge 201, 453, 267, 213, 141, 93, 81, 369, 321, 117, 51, 39, 177, 99, 87, 261, 153, 135.

Dividiert durch 3 ergibt sich die Folge 67, 151, 89, 71, 47, 31, 27, 123(9·41), 107, 39, 17, 13, 59, 33(3·11), 29, 87, 51, 45.

Wir finden in Spalte S_2 die Zahlenfolge 13, 17, 11.
Wir finden in Spalte S_3 die Blätterfolge 45, 33.
Wir finden in Spalte S_4 die Zahlenfolge 67, 89, 59, 39, und die Zahlen 151, 51, 87.
Wir finden in Spalte S_8 die Zahlenfolge 107, 71, 47, 31, 41, 27.

In **Spalte S_5** haben wir die Blätterfolge 1365, 909, 537, 423, 1611, 633, 1605, 1425, 1137, 753, 441, 513, 303, 777, 2625, 483, 507, 1713, 555, 309, 273, 405, 105, 357, 237, 123, 315, 945, 597, 705, 789, 525, 465, 1857, 693, 363, 1089, 897, 1221, 813, 1281.

Dividiert durch 3 ergibt sich die Folge 455, 303, 179, 141, 537, 211, 535, 475, 379, 251, 147, 171, 101, 259, 875, 161, 169, 571, 185, 103, 91, 135, 35, 119, 79, 41, 105, 315, 199, 235, 263, 175, 155, 619, 231, 363, 121, 299, 407, 271, 427.

Die Zahlen finden sich in den Spalten S_3 bis S_{11} wieder, wobei die Maxima in den Spalten S_5, S_8 und S_{10} liegen. Im Folgenden gibt die erste Zahl die Spalte an, die Zahl in Klammern steht für die Anzahl gefundener Zahlen.

3(1) , 4(3) , **5(18)** , 6(2) , 7(1) , **8(8)** , 9(1), **10(6)** , 11(1)

Deutlich wird, dass sich die Faktoren der Blätter aus Spalte S_n auch in Spalte S_{2n} wieder finden.
(vgl. Anhang 1 und 2)

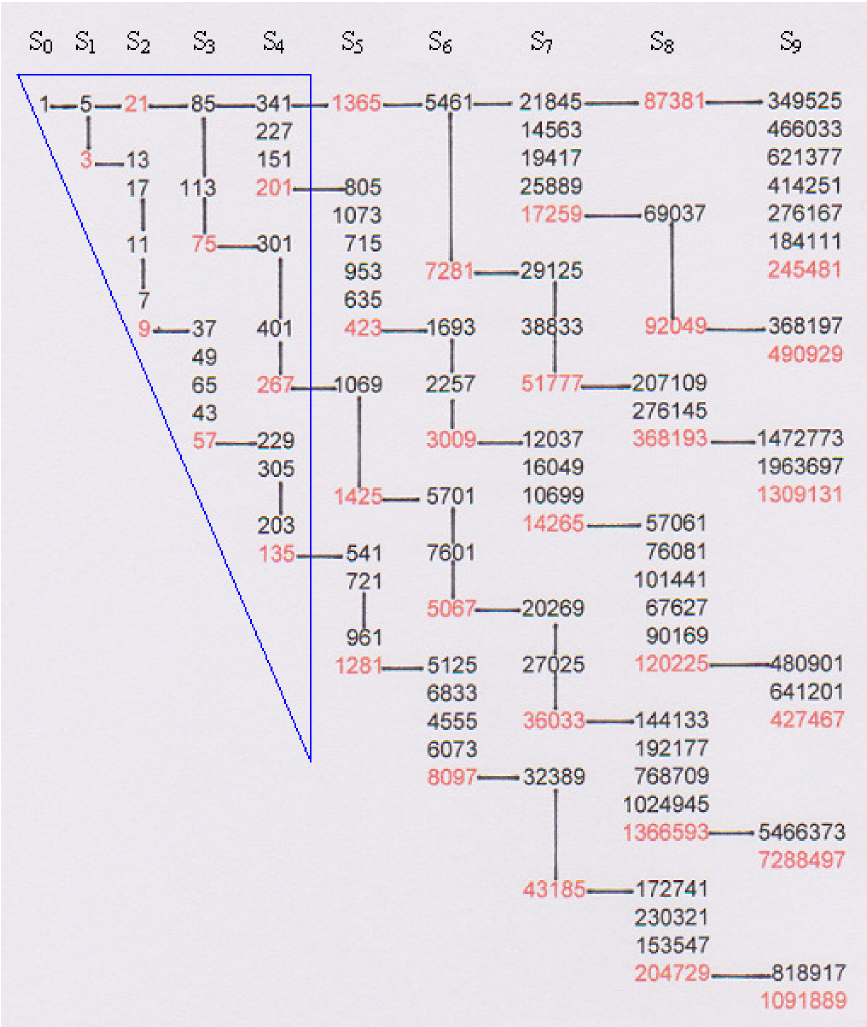
Die Blätter der ersten Äste mit gleicher Spaltenzahl (Das grobe Ordnungsprinzip)

Wir definieren als **erste Äste** des Baumes die Weiterleitungen der Zahlen des Hauptastes nach unten, die bei jedem Blatt jeweils nur eine Spalte nach rechts verzweigen. Der erste Ast der Spalte S_1 wird also durch die Zahlen 5, 3, 13, 17, 11, 7, 9, 37, 49, 65, 43, 57, 229,… gebildet. Der erste Ast der Spalte S_3 durch 85, 113, 75, 301, 401, 267, 1069,…, usw. Von den Blätter des Hauptastes, 21, 1365, 87381,…, verzweigen keine ersten Äste. Die ersten Äste sind also genau die Äste im Baum deren Eck-Blätter als Spaltenvorgänger wieder Eck-Blätter besitzen.

Betrachtet man die Zahlenverteilung im Baum anhand dieser ersten Äste, lässt sich eine weitere grobe Ordnungsstruktur beobachten. Die Blätter der ersten Äste mit gleicher Spaltenzahl bilden dabei Wachstumsgrenzen im Baum, die mit der Zahl des Hauptastes derselben Spalte in enger Beziehung stehen. Die Dreiecks-Struktur des Baumes schafft die Grundlage für dieses Ordnungsprinzip. Die nachfolgende Grafik und das Beispiel verdeutlichen dies. Das **Ordnungsprinzip** lautet *zunächst.*

Alle ungeraden Zahlen von 1 bis a_n finden sich in den Weiterleitungen der ersten Äste bis zu ihren Blättern mit der Spaltenzahl S_n.

Der ungerade Collatz-Baum reduziert auf die ersten Äste des Hauptastes



Beispiel für a_4=341

Alle ungeraden Zahlen von 1 bis a_4=341 finden sich in den Weiterleitungen der ersten Äste der Spalte S_1 bis zur 135, der Spalte S_3 bis zur 267 und der Spalte S_4 bis zur 201. (vgl. Anhang 2)

Das Dreieck (blau) verdeutlicht die Struktur dieses Prinzips. Wir können auch sagen, dass sich in allen Weiterleitungen der ersten Äste nach 135, 267 und 201 keine Zahlen mehr finden, die kleiner als 341 sind.

Gültigkeitsbereich des Ordnungsprinzips und erste Ausreißer

Das Ordnungsprinzip besitzt uneingeschränkte Gültigkeit bis zur Spalte S_8 mit a_8=87381. Nach dieser Grenze muss entweder eine sehr geringe Fehlerquote mit einbezogen oder das Dreiecks-Schema um eine oder mehr Spalten erweitert werden. Leicht *modifiziert* bleibt es aber **das grobe Ordnungsprinzip** im ungeraden Collatz-Baum in der hier vorgestellten Anordnung.

Alle ungeraden Zahlen von 1 bis a_n finden sich in den Weiterleitungen der ersten Äste bis zu ihren Blättern mit der Spaltenzahl S_{n+k} mit $k \geq 0$.

Das Verhalten von k bleibt für größere Zahlenbereiche zu untersuchen.

Es folgt eine **Liste der ersten Ausreißer**. Diese finden sich ab der Spalte S_9 mit a_9=349525. Die Ausreißer sind jeweils kleiner als die zugehörige Zahl des Hauptastes a_n und finden sich jeweils erst in den Weiterleitungen der nächsten Spalte.

Spalte S_9	(349525)	Ausreißer: 300223, 263271, 316283, 210855
Spalte S_{10}	(1398101)	Ausreißer: 843421, 1124561, 749707, 999609
Spalte S_{11}	(5592405)	Ausreißer: 3998437, 5331249
Spalte S_{12}	(22369621)	Ausreißer: 21324997
Spalte S_{13}	(89478485)	Ausreißer: 20665775, 13777183, 18369577, 21324997
Spalte S_{14}	(357913941)	Ausreißer: 336607519, 326201919, 299206683, 244651439, 217467945, 163100959