A Decomposition of the Collatz Tree

Syuji Miyazaki

Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan
E-mail address: miyazaki.syuji.8m@kyoto-u.ac.jp

(Received May 27, 2021; Accepted October 16, 2021)

The Collatz tree, or the directed graph of the $3x + 1$ problem, was demonstrated to be decomposed into three different subgraphs. An arbitrary positive integer is assigned uniquely to a specific position of the nodes of either of the three subgraphs. The manner of connecting a specific subgraph to its neighboring subgraphs is explicitly given. A comparison with complete chaos synchronization was made.
Key words: Collatz Conjecture, Collatz Tree, Directed Graph

1. Introduction

The Collatz problem, or the $3x + 1$ problem, is one of the unsolved mathematical problems (Andaloro, 2002; Chamberland, 2003). First, we defined the Collatz function $C(n)$ as $C(n) = 3n + 1$ $(n/2)$ for an odd (even) number n. The Collatz conjecture states that for each positive integer m, there is a positive integer k such that $C^{(k)}(m) = 1$, i.e., any positive integer will eventually iterate to 1, followed by the cycle $4 \to 2 \to 1 \to 4 \to \cdots$.

The mapping $C(n)$ is similar to that introduced as an analytical model (Hata-Miyazaki, 1997; Miyazaki-Hata, 1998) of intermittency caused by chaotic modulation (Fujisaka-Yamada, 1985; 1986), also known as on-off intermittency, which occurs when complete chaos synchronization (Pikovsky *et al.*, 2003) becomes slightly unstable. These intermittent time series have self-similar structure. At a slightly stable side of the vicinity of the critical point between complete chaos synchronization and on-off intermittency, a transient time from the initial values of the state variables leading to the attractor of the complete chaos synchronization depends self-similarly on the state point of the initial condition in the phase space (Inoue-Nishi, 1996). Whether the Collatz problem has such self-similarity related to complete chaos synchronization was the main focus of this paper.

This iteration is represented by a directed graph called the Collatz tree or Collatz graph, whose node denotes a positive integer. An even number $2N$ $(N = 1, 2, \cdots)$ is iterated to N and is shown as $2N \to N$. An odd number $2M - 1$ $(M = 1, 2, \cdots)$ is iterated to $3(2M - 1) + 1 = 6M - 2$ and is shown as $2M - 1 \to 6M - 2$. For $N = 6M - 2$, i.e., for a positive integer given by 4 mod 6, both $12M - 4$ and $2M - 1$ are iterated to $6M - 2$, and shown as a dichotomous branch $12M - 4 \to 6M - 2 \leftarrow 2M - 1$. The even number $6M - 2$ is further iterated to $3M - 1$ such that the Collatz tree always exhibits dichotomous branching at a positive integer given by 4 mod 6, and the input degree and output degree are equal to 2 and 1, respectively, at the node $6M - 2$. The other positive integers 0, 1, 2, 3, and 5 mod 6 are represented by the node whose input and output degrees are both equal to 1. In this paper, we discuss the structure between the neighboring dichotomous branching nodes.

We considered the distance between the neighboring dichotomous branching nodes. Starting from $6M - 2$ (positive integer 4 mod 6), we traced the backward iteration until the preceding positive integers were reached. There exist only three cases, which are distinguished by the remainder M when divided by 3, as $M = 3n - 2$, $3n - 1$, $3n$ $(n = 1, 2, \cdots)$.

Quadruplet: $M = 3n - 2$ (1 mod 3)
$\rightrightarrows 6(12n - 9) - 2$(the left nearest neighbor junction)
$\to 6(6n-4)-4 \to 6(3n-2)-2 \leftarrow 6n-5 \leftarrow 6(2n-1)-4$
$\leftarrow 6(4n - 3) - 2 \leftleftarrows$ (the right nearest neighbor junction)
Triplet: $M = 3n$ (0 mod 3)
$\rightrightarrows 6(12n - 1) - 2$(the left nearest neighbor junction)
$\to 6(6n) - 4 \to 6(3n) - 2 \leftarrow 6n - 1$
$\leftarrow 6(2n) - 2 \leftleftarrows$ (the right nearest neighbor junction)
Doublet: $M = 3n - 1$ (2 mod 3)
$\rightrightarrows 6(12n - 5) - 2$(the left nearest neighbor junction)
$\to 6(6n - 2) - 4 \to 6(3n - 1) - 2 \leftarrow$
$6n - 3 \leftarrow 6(2n - 1) \leftarrow 6(2n - 1) \times 2^1 \leftarrow$
$6(2n - 1) \times 2^2 \leftarrow \cdots$ (the Sharkovskii branch)

The first three examples in the three cases are listed below, where the number in parentheses indicates the dichotomous branching node belonging to the nearest neighbor subgraph.

Quadruplet:
$\rightrightarrows (16) \to 8 \to [4 \leftarrow 1 \leftarrow 2 \leftarrow 4]$ (loop)
$\rightrightarrows (88) \to 44 \to 22 \leftarrow 7 \leftarrow 14 \leftarrow (28) \leftleftarrows$
$\rightrightarrows (160) \to 80 \to 40 \leftarrow 13 \leftarrow 26 \leftarrow (52) \leftleftarrows$
Triplet:
$\rightrightarrows (64) \to 32 \to 16 \leftarrow 5 \leftarrow (10) \leftleftarrows$
$\rightrightarrows (136) \to 68 \to 34 \leftarrow 11 \leftarrow (22) \leftleftarrows$
$\rightrightarrows (208) \to 104 \to 52 \leftarrow 17 \leftarrow (34) \leftleftarrows$
Doublet with the Sharkovskii branch:
$\rightrightarrows (40) \to 20 \to 10 \leftarrow 3 \leftarrow 6 \leftarrow 12 \leftarrow 24 \leftarrow \cdots$
$\rightrightarrows (112) \to 56 \to 28 \leftarrow 9 \leftarrow 18 \leftarrow 36 \leftarrow 72 \leftarrow \cdots$

doi:10.5047/forma.2021.004

$\rightrightarrows (184) \to 92 \to 46 \leftarrow 15 \leftarrow 30 \leftarrow 60 \leftarrow 120 \leftarrow \cdots$

The distance between the neighboring branching nodes is either 2 or 3 in the quadruplet case, always 2 in the triplet case, and either 2 or ∞ in the doublet case, such that the distances 2, 3, and ∞ appear at a ratio of 4 : 1 : 1. Positive integers 1, 5, and 3 mod 6 appear only in the quadruplet, triplet, and doublet case, respectively. A multiple of 6 (0 mod 6) can be uniquely factorized into the form $6(2n-1) \times 2^p$ $(p = 0, 1, \cdots)$ and appears only in the doublet case, which we termed the Sharkovskii branch because the sequence resembles Sharkovskii's sequence (Sharkovskii, 1964). Positive integers 2 mod 6 are further classified into $6(6n-4)-4$, $6(6n-2)-4$, $6(6n)-4$ and $6(2n-1)-4$. The even quotient $6(6n-4)-4$ and the odd quotient $6(2n-1)-4$ appear in the left and the right branch in the quadruple case, respectively. The single odd number $6n-5$ appears in the right branch of the quadruple case, and $6(6n)-4$ and $6(6n-2)-4$ appear in the left branch of the triplet and the doublet case, respectively.

In summary, every positive integer corresponds to a specific position in either the quadruplet, triplet, or doublet case such that the nodes of the Collatz tree cover the entirety of positive integers. Thus, the three cases are represented by three different V-shaped subgraphs and the entire Collatz tree can be constructed by the combination of the above-mentioned subgraphs, where the node $6M-2$ (4 mod 6) is a junction point. Note that the seemingly complicated Collatz tree, as shown in Fig. 1 of (Andaloro, 2002) or Fig. 2 of (Chamberland, 2003), can be decomposed into only pieces, which may be assigned to three different V-shaped subgraphs.

Next, we pieced a puzzle together. We omitted all integers except for $6M-2$ (4 mod 6) and reduced the representation of the quadruplet, triplet, and doublet cases with the right and left nearest neighbor junctions respectively as

$$\begin{aligned} &\rightrightarrows Q_l(n) \to Q(n) \leftarrow Q_r(n) \leftleftarrows, \\ &\rightrightarrows T_l(n) \to T(n) \leftarrow T_r(n) \leftleftarrows, \\ &\rightrightarrows D_l(n) \to D(n) \leftarrow \text{(the Sharkovskii branch)}, \end{aligned}$$

where

$$\begin{aligned} Q(n) &= 18n - 14 = 6(3n-2) - 2, \\ T(n) &= 18n - 2 = 6(3n) - 2, \\ D(n) &= 18n - 8 = 6(3n-1) - 2, \\ Q_l(n) &= 72n - 56 = 6(3(4n-3)) - 2 \\ &= 4Q(n) = T(4n-3), \\ Q_r(n) &= 24n - 20 = 6(4n-3) - 2 \end{aligned}$$

$$= \begin{cases} 18(4k-3) - 14 = Q(4k-3) \\ \qquad\qquad (\text{if} \quad n = 3k-2) \\ 18(4k-2) - 8 = D(4k-2) \\ \qquad\qquad (\text{if} \quad n = 3k-1) \\ 18(4k-1) - 2 = T(4k-1) \\ \qquad\qquad (\text{if} \quad n = 3k), \end{cases}$$

$$\begin{aligned} T_l(n) &= 72n - 8 = 6(3(4n) - 1) - 2 \\ &= 4T(n) = D(4n), \end{aligned}$$

$$T_r(n) = 12n - 2 = 6(2n) - 2$$

$$= \begin{cases} 18(2k-1) - 8 = D(2k-1) \\ \qquad\qquad (\text{if} \quad n = 3k-2) \\ 18(2k) - 14 = Q(2k) \quad (\text{if} \quad n = 3k-1) \\ 18(2k) - 2 = T(2k) \quad (\text{if} \quad n = 3k), \end{cases}$$

$$\begin{aligned} D_l(n) &= 72n - 32 = 6(3(4n-1) - 2) - 2 \\ &= 4D(n) = Q(4n-1), \end{aligned}$$

where $k = 1, 2, \cdots$. For example, $Q(n) = Q_r(m)$ is satisfied for $(n, m) = (5, 4)$ such that the n-th quadruplet is connected at the right nearest neighbor junction of the m-th quadruplet, which corresponds to the original Collatz sequence $Q(5) = 76 \to 38 \to 19 \to Q(4) = 58 \to \cdots$. The quotients of $Q(n)$, $T(n)$, and $D(n)$ when divided by 6 are $3n-2$, $3n$, and $3n-1$, respectively. On the other hand, the quotients of $Q_l(m)$, $T_l(m)$, and $D_l(m)$ when divided by 6 are $3(4m-3)$, $3(4m)-1$, and $3(4m-1)-2$, respectively. The remainders of these numbers when divided by 3 determine three possible combinations $Q(n) = D_l(m)$, $T(n) = Q_l(m)$, and $D(n) = T_l(m)$ for specific positive integers n and m. Furthermore, six more combinations, $Q(n) = Q_r(m)$, $Q(n) = T_r(m)$, $T(n) = Q_r(m)$, $T(n) = T_r(m)$, $D(n) = Q_r(m)$, and $D(n) = T_r(m)$, are possible for specific positive integers n and m. The relationship is summarized as follows:

$$\begin{aligned} Q(n) &= D_l(m) \text{ for } (n, m) = (3, 1), (7, 2), (11, 3), \cdots, \\ Q(n) &= Q_r(m) \text{ for } (n, m) = (1, 1), (5, 4), (9, 7), \cdots, \\ Q(n) &= T_r(m) \text{ for } (n, m) = (2, 2), (4, 5), (6, 8), \cdots, \\ T(n) &= Q_l(m) \text{ for } (n, m) = (1, 1), (5, 2), (9, 3), \cdots, \\ T(n) &= Q_r(m) \text{ for } (n, m) = (3, 3), (7, 6), (11, 9), \cdots, \\ T(n) &= T_r(m) \text{ for } (n, m) = (2, 3), (4, 6), (6, 9), \cdots, \\ D(n) &= T_l(m) \text{ for } (n, m) = (4, 1), (8, 2), (12, 3), \cdots, \\ D(n) &= Q_r(m) \text{ for } (n, m) = (1, 1), (3, 4), (5, 7), \cdots, \\ D(n) &= T_r(m) \text{ for } (n, m) = (2, 2), (6, 5), (10, 8), \cdots, \end{aligned}$$

where (n, m) in all cases are given by double arithmetic progressions. The first three relations, including $Q(n)$, imply that the quadruplet is connected to the quadruplet, the triplet, and the doublet at a ratio of 1:2:1. In a similar manner, both the triplet and doublet are connected to the quadruplet, the triplet, and the doublet at a ratio of 1:1:0. These connections from nodes $Q(n)$, $T(n)$, and $D(n)$ to the next junctions are expressed by the following transition matrix:

$$\begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & 0 & 0 \end{pmatrix}.$$

Its right eigenvalue $(a, b, c)^T$ corresponding to the unit eigenvalue satisfies $a : b : c = 0.4 : 0.5 : 0.1$, which implies that the quadruplets, triplets, and doublets appear at a ratio of $0.4 : 0.5 : 0.1$ in a long Collatz sequence starting from a fixed positive number. For example, the Collatz sequence starting from 27 is given in Table 1, where $QR(QL)$, $TR(TL)$, and $DR(DL)$ indicate that the number on the left belongs to the right(left) branch of the quadruplet, the triplet, and the doublet, respectively. Note that the odd number belongs to the right branches only and that this sequence contains many right branches. The number of the dichotomous branching node of the quadruplet, triplet, and doublet nodes are equal to 16, 28, and 5, respectively. The estimation from the transition matrix 0.4 : 0.5 : 0.1 was a good approximation of the ratio 16 : 28 : 5 = 0.326530612 : 0.571428571 : 0.102040816.

We clarified how many odd numbers appear from an initial positive integer to 1. In a long Collatz sequence, the ratio of visiting $Q_r(n) = 24n - 20$ and $Q_l = 72n - 56$ is 24 : 72 = 1/4 : 3/4. In a similar manner, the ratio of visiting $T_r(n) = 12n - 2$ and $T_l = 72n - 8$ is 12 : 72 = 1/7 : 6/7. Taking into account the visiting frequency of the quadruplet and triplet, we estimated the ratio of the right branch to the whole of the quadruplet and triplet as $4/10 \times 1/4 + 5/10 \times 1/7 = 6/35$. Next, we considered the relative ratio of the initial numbers on the Sharkovskii branch $6(2n-1) \times 2^p$ $(p = 0, 1, \cdots)$. The ratio of the odd number $3(2n - 1) = 6n - 3$ of the doublet to the initial numbers between 1 and N was given by $(N/6)/N = 1/6$ for $n \ll N$. The largest number on the Sharkovskii branch was estimated as $6(2n-1) \times 2^{p_{max}} = N$, leading to $p_{max} \propto \log N$ for $n \ll N$, which is much less than $N/6$. Ignoring the case of the initial number located at the Sharkovskii branch aside from the single odd number, the ratio of the initial number coinciding with the odd number $6n - 3$ was equal to 1/6, and 5/6 otherwise. In the latter case, the ratio of choosing the right branches including odd numbers was 6/35, as shown above. Thus, the ratio of odd numbers to the stopping time from the initial number to 1 was equal to $5/6 \times 6/35 + 1/6 = 13/42 \approx 0.31$. Starting from power-of-two numbers $n_p = 2^p$, we had $n_p(1/2)^T = 1$ for the stopping time T, i. e., $T = \log n_p / \log 2$. Starting from an arbitrary initial number, we had $n_p \lambda^T = 1$ for the roughly estimated stopping time T, where λ was estimated as $\lambda = \frac{42-13}{42}(-\log 2) + \frac{13}{42} \log 3 \approx -0.139$. The former and latter corresponded to the contribution from the even and the odd number, respectively. This averaged behavior was underestimated because we ignored the additional part

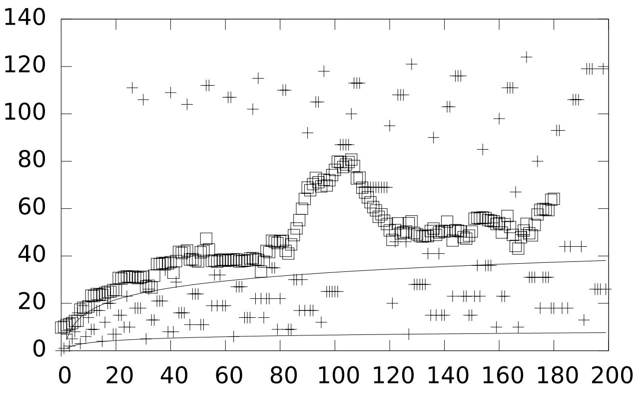

Fig. 1. The stopping time is plotted against the initial number with symbol + and that of the moving average over the succeeding 20 plots with symbol □. The shortest stopping time for power-of-two numbers (dashed line) and the rough averaged stopping time (solid line) are shown.

1 of $3x + 1$ of the iteration of an odd number x. The stopping time (symbol +), that of the moving average over the preceding 20 points (symbol □), the shortest stopping time for power-of-two numbers (dashed line), and the rough averaged stopping time (solid line) are shown in Fig. 1.

The quadruplets, triplets, and doublets seem randomly arranged on the Collatz tree. However, for a specific dichotomous branching node $54M - 38$ (positive integer 16 mod 54), there always exists a periodic backward sequence in order of the quadruplet, triplet, and doublet tracking the left-most branches Q_l, T_l, and D_l. The periodic sequence located at the nearest neighbor of the final loop $4 \to 2 \to 1 \to 4$ is given by $\cdots \to 262144 = D(14564) \to 65536 = T(3641) \to 16384 = Q(911) \to 4096 = D(228) \to$

Table 1. Collatz sequence from 27 to 4.

27	DR	445	QR	1154	QR
82	D(5)	1336	Q(75)	577	QR
41	TR	668	DL	1732	Q(97)
124	T(7)	334	D(19)	866	QR
62	QR	167	TR	433	QR
31	QR	502	T(28)	1300	Q(73)
94	Q(6)	251	TR	650	QR
47	TR	754	T(42)	325	QR
142	T(8)	377	TR	976	Q(55)
71	TR	1132	T(63)	488	DL
214	T(12)	566	QR	244	D(14)
107	TR	283	QR	122	QR
322	T(18)	850	Q(48)	61	QR
161	TR	425	TR	184	Q(11)
484	T(27)	1276	T(71)	92	DL
242	QR	638	QR	46	D(3)
121	QR	319	QR	23	TR
364	Q(21)	958	Q(54)	70	T(4)
182	QR	479	TR	35	TR
91	QR	1438	T(80)	106	T(6)
274	Q(16)	719	TR	53	TR
137	TR	2158	T(120)	160	T(9)
412	T(23)	1079	TR	80	QL
206	QR	3238	T(180)	40	Q(3)
103	QR	1619	TR	20	DL
310	Q(18)	4858	T(270)	10	D(1)
155	TR	2429	TR	5	TR
466	T(26)	7288	T(405)	16	T(1)
233	TR	3644	QL	8	QL
700	T(39)	1822	Q(102)	4	Q(1)
350	QR	911	TR		
175	QR	2734	T(152)		
526	Q(30)	1367	TR		
263	TR	4102	T(228)		
790	T(44)	2051	TR		
395	TR	6154	T(342)		
1186	T(66)	3077	TR		
593	TR	9232	T(513)		
1780	T(99)	4616	QL		
890	QR	2308	Q(129)		

$1024 = T(57) \to 256 = Q(15) \to 64 = D(4) \to 16 = T(1) \to$, which is termed the quarter sequence because the number is multiplied by 1/4 along the Collatz sequence. In a similar manner, there exists a Sharkovskii branch $2^p(6\times1-3)$ $(p = 0, 1, \cdots)$ connected to $10 = D(1)$ followed by the triplet $16 = T(1)$. The variety of branches are extending between the quarter and Sharkovskii branches of the Collatz tree. A countable number of V-shaped subgraphs consisting of the right Sharkovskii and left quarter branches are embedded in the Collatz tree. The Sharkovskii branch located at the second nearest neighbor of the final loop is given by $2^p(6 \times 13 - 3)$ $(p = 0, 1, \cdots)$ connected to $226 = D(13)$, followed by the triplet $340 = T(19)$. The corresponding quarter branch is given by $\cdots \to 87040 = D(4836) \to 21760 = T(1209) \to 5440 = Q(303) \to 1360 = D(76) \to 340 = T(19)$. The V-shaped structure located at the third and fourth nearest neighbors of the final loop leads to $70 = T(4)$ and $7252 = T(403)$, respectively. The first, second, and third V-shaped combinations of the quarter and Sharkovskii branches (SB) are as follows:

```
262144 87040 1120   SB
|      |        |   /
65536 21760   280 46
|      |        | /
16384 5440 SB  (70)
|      |   /    |
4096 1360 226   106
|      | /    /
1024 (340) 160
|   /    /
256    40  SB
|      |  /
64    10
|   /
(16)
```

The figures in parentheses indicate the beanching nodes between the quarter and Sharkovskii branches.

To sum up our main results, the Collatz tree can be decomposed into three pieces, which were referred to as the quadruplet, triplet, and doublet with the Sharkovskii branch in this paper. An arbitrary positive integer is assigned uniquely to a specific position of the nodes of either of the three subgraphs. The manner of connecting a specific subgraph to its neighboring subgraphs was explicitly discussed in this paper. As for comparison with chaos synchronization, the value $\lambda \approx -0.139$, which corresponds to the transverse Lyapunov exponent in this research field, suggests that this system is located at a deeply stable region far from the critical point $\lambda = 0$. Thus, we concluded that self-similar structure in the context of chaos synchronization is hardly observable in the Collatz problem.

References

Andaloro, P. J. (2002) THE $3x + 1$ PROBLEM AND DIRECTED GRAPHS, *Fibonacci Quarterly* **40**, 43–54.

Chamberland, M. (2003) An Update on the $3x + 1$ Problem, [Una actualizacio del problema 3x+1 (Catalan, translated by Toni Guillamon i Grabolosa)] *Butlleti de la Societat Catalana de Matematiques* **18**, 19–45.

Fujisaka, H. and Yamada, T. (1985) A New Intermittency in Coupled Dynamical Systems, *Prog. Theor. Phys.* **74**, 918–921.

Fujisaka, H. and Yamada, T. (1986) Stability Theory of Synchronized Motion in Coupled-Oscillator Systems. IV: Instability of Synchronized Chaos and New Intermittency, *Prog. Theor. Phys.* **75**, 1087–1104.

Hata, H. and Miyazaki. S. (1997) Exactly Solvable Maps of On-off Intermittency, *Phys. Rev.* E **55**, 5311–5314.

Inoue, M. and Nishi, Y. (1996) Highly Complicated Basins of Periodic Attractors in Coupled Chaotic Maps, *Prog. Theor. Phys.* **95**, 685–690.

Miyazaki, S. and Hata, H. (1998) Universal scaling law of the power spectrum in the on-off intermittency, *Phys. Rev.* E **58**, 7172–7175.

Pikovsky, A., Rosenblum, M. and Kurths, J. (2003) Synchronization: a universal concept in nonlinear sciences, Cambridge University Press.

Sharkovskii, A. N. (1964) Co-existence of cycles of a continuous mapping of the line into itself, *Ukrainian Math. J.* **16**, 61–71.