Journal of Statistical Physics, Vol. 107, Nos. 5/6, June 2002 (© 2002)

A Generalization of the Collatz Problem. Building Cycles and a Stochastic Approach

J. L. Rouet[1] **and M. R. Feix**[2]

Received March 20, 2001; accepted January 25, 2002

The $(3x+1)/2$ problem is generalized into the n-furcation problem $(l_i x+m_i)/n$ where $i \in [0, 1,..., n-1]$. It is shown that, under some constraints on l_i and m_i, the main bijection property between the k less significant digits of the seed, written in base n, and the sequence of generalized parities of the k first iterates is preserved. This property is used to investigate a stochastic treatment of ensemble of large value seeds. The bijection property predicts stochasticity for a number of iterations equals to the number of significant digits of the seed. In fact, the stochastic approach is valid for much larger numbers, a property which is more easily shown by using increasing sequences than decreasing ones. Finally, we extend the stochastic approach to cases where the bijection theorem does not hold, introducing the matrix giving the probability that a "j" number (where j is the last significant digit of this number written in base n) gives a "i" number iterate.

KEY WORDS: $3x+1$ problem; Collatz problem; mixing process; random walk; numerical simulation.

1. INTRODUCTION

The $(3x+1)/2$ problem[1] deals with the sequence of iterated positive integers defined in an iterative way by the relations

$$T(x) = \begin{cases} x/2 & \text{if } x \text{ is even} \\ (3x+1)/2 & \text{if } x \text{ is odd} \end{cases} \tag{1}$$

[1] Laboratoire de Mathématique, Applications et Physique Mathématique—UMR 6628, Université d'Orléans, UFR des Sciences, F-45067 Orléans Cedex 2, France; e-mail: jean-louis.rouet@univ-orleans.fr

[2] SUBATECH, Ecole des Mines de Nantes, La Chantrerie, 4 rue A. Kastler, B.P. 20722, F-44307 Nantes Cedex 3, France.

Starting from a positive x (the seed) it has been checked[2] that for all seeds up to 2^{40}, the sequence of iterates $(x, T(x), T^{(2)}(x),...)$ ends with the cycle 2, 1, 2, 1... and it has been conjectured that this remains true for all positive seeds. Unfortunately, neither the conjecture has been proved, nor a counter example given: this leads some people to believe that the problem may belong to the class of undecidable problems introduced in Mathematics by Godel and Turing.[3]

Experience shows that, for such hard problems, it may be interesting to embed them in a more general one and to see what happens (stability of the results, appearance of new properties...). This is one of the purposes of this paper with a generalization of the bifurcation problem to an n-furcation.

2. GENERALIZATIONS

Generalizations of the $(3x+1)/2$ problem have been already proposed (see refs. 1, 4–8). The generalization to an n-furcation is obtained in the following way. Let us consider the function U, acting on integers, where l_i, m_i, n are integers, positive for l_i and n, possibly negative for m_i, with $i = 0,..., n-1$, such that:

$$U: \mathbb{N} \to \mathbb{N}$$

$$U(x) = \frac{l_i x + m_i}{n} \quad \text{if} \quad x \equiv i \pmod{n} \quad \text{with} \quad (il_i + m_i) \equiv 0 \pmod{n} \tag{2}$$

As there are n possible issues, such a process will be called an n-furcation.

One of the property of the $(3x+1)/2$ problem is the bijection which exists between the k last bits of a seed x and its k first iterations. In order to recover this property in the generalized case, one must constraint the values of l_i and n accordingly the following theorem.

Theorem 1. A bijection exists between the k last n-digits (hereafter called nit) of a seed x (written in base n) and its k first iterations iff l_i and n are coprime for $i = 0,..., n-1$.

Proof. After the first step, the last significant nit (abbreviated LSN hereafter) of $U^{(1)}(x)$ determines the next iteration, or alternatively, the two last digits of x determine the two first iterations. Writing $x = kn^2 + jn + i$, with i and $j \in [0, n-1]$, we have

$$U^{(1)}(x) = kl_i n + l_i j + \alpha_i$$

with $\alpha_i \equiv il_i + m_i \pmod{n}$. When j goes from 0 to $n-1$, $l_i j \pmod{n}$ must take all values in $[0, n-1]$. This drives to $(l_i, n) = 1$ for all $i = 0, \ldots, n-1$.

The proof is then established by recurrence. Let us consider the n numbers x_i, i going from 0 to $n-1$, which only differ by their significant digit k, such that

$$x_i = x_0 + in^k$$

After k iterations, these numbers give

$$U^{(k)}(x_i) = U^{(k)}(x_0) + il_0^{\beta_0} l_1^{\beta_1} \cdots l_{n-1}^{\beta_{n-1}}$$

with $\beta_0 + \beta_1 + \cdots + \beta_{n-1} = k$. The product $(il_0^{\beta_0} l_1^{\beta_1} \cdots l_{n-1}^{\beta_{n-1}}) \pmod{n}$ must take all values in $[0, n-1]$ when i goes from 0 to $n-1$, consequently

$$(il_0^{\beta_0} l_1^{\beta_1} \cdots l_{n-1}^{\beta_{n-1}}, n) = 1$$

which is already satisfy by the condition $(l_i, n) = 1$, $i = 0, \ldots, n-1$. ▮

In refs. 5 and 6, condition $(i\, l_i + m_i) \equiv 0 \pmod{n}$ is considered and the g.c.d of some l_i and n can be different of one. We will treat this case in Section 4.

3. BUILDING CYCLES

One of the possibility of solving the Collatz problem would be to find cycle, consequently failing the conjecture. The generalization brings new aspects on the question of the existence of cycles. Because the number of degrees of freedom has been increased, one question is now to find the l_i and m_i which allow the existence of a cycle (eventually an arbitrary one) of length N.

For a cycle of length N, and a n-furcation, the following relation must be satisfied by the l_i and m_i:

$$\left(n^N - \prod_{i=0}^{N-1} l_{p_i(x)}\right) x = \sum_{i=0}^{N-1} n^i m_{p_i(x)} \prod_{j=i+1}^{N-1} l_{p_i(x)} \tag{3}$$

where the vector $p_i(x)$, $i = 0, \ldots, N-1$, defines precisely the cycle by giving the successive iterations:

$$p_i(x) \equiv U^{(i)}(x) \pmod{n}$$

Notice that the right hand-side of Eq. (3) is the N-iterate of 0 (calculated using rational numbers), then multiplied by n^N.

Theorem 2. It is always possible to find a set of values l_i, m_i which allow to build an arbitrary cycle, that is for a given vector $(p_0(x),\ldots, p_{N-1}(x))$.

Proof. Choose arbitrary l_i (satisfying $(l_i, n) = 1$). Rewrite Eq. (3) factorizing the m_i in its right hand-side, calling α_i the coefficients of m_i, Eq. (3) reads

$$\alpha x = \sum_{i=0}^{n-1} \alpha_i m_i \tag{4}$$

where $\alpha = n^N - \prod_{i=0}^{N-1} l_{p_i(x)}$ is such that $(\alpha, n) = 1$ because $(l_i, n) = 1$. Let us now write $m_i = \alpha M_i$. The condition $il_i + m_i \equiv 0 \pmod n$ reads

$$il_i + \alpha M_i \equiv 0 \pmod n \tag{5}$$

In the most general case, $il_i \pmod n$ takes all values between 1 and $n-1$, and it must be consequently the same for αM_i. But this condition is fulfilled because $(\alpha, n) = 1$. Then it is always possible to find M_i satisfying Eq. (5). Now the seed is determined using Eq. (4)

$$x = \sum_{i=0}^{n-1} \alpha_i M_i \quad \blacksquare$$

Consider the following example where $n = 2$, $l_0 = 1$, $l_1 = 3$ and the cycle given by (1, 1, 0, 1, 0). A way to get the second member of Eq. (3) is to compute the iterates of 0 for the given cycle:

$$0 \xrightarrow{1} \frac{m_1}{2} \xrightarrow{1} \frac{3m_1}{4} + \frac{m_1}{2} \xrightarrow{0} \frac{3m_1}{8} + \frac{m_1}{4} + \frac{m_0}{2}$$

$$\xrightarrow{1} \frac{9m_1}{16} + \frac{3m_1}{8} + \frac{3m_0}{4} + \frac{m_1}{2} \xrightarrow{0} \frac{9m_1}{32} + \frac{3m_1}{16} + \frac{3m_0}{8} + \frac{m_1}{4} + \frac{m_0}{2}$$

and Eq. (3) reads

$$5x = 28m_0 + 23m_1 \tag{6}$$

Choosing $M_0 = 0$ and $M_1 = 1$, we get $m_0 = 0$, $m_1 = 5$ and the seed is 23.

4. THE RANDOM WALK GAME

Another goal of this paper is to complete and precise the stochastic approach of the Collatz problem. First steps in this direction have been initiated by Terras.[9] Matthews *et al.*[4, 5] deal with the ergodic properties of the generalized $(lx+m)/n$ problem and around 1990, Lagarias[10] and the present authors[7, 8] considered a similar approach. Our preceding work and the present one are based on the concepts and method of the statistical physics and here an operational and computational point of view is introduced giving precise operational meaning to some of the much more abstract concepts found in the work of Matthews.[4, 5]

Such a stochastic approach deserves some comments. It may be argued that, since we have a perfectly deterministic problem, (moreover one "integrable" by pieces) such an approach is unnecessary, irrelevant and, worst, may lead to errors. Since Boltzmann, numerous and sometimes very hot disputes have taken place. While philosophically interesting, they do not bring useful results. Here, as in the N-body problem of statistical physics, two kinds of results can be expected:

— While the "exact" trajectory of a seed (i.e., the exact value of the ith iterate) is totally lost in this stochastic approach, the behavior of a large number of seeds can be obtained (as in Brownian motion and diffusion problems).

— Deterministic sequences can look like random ones. This is the basic justification of all "pseudo random number" generators, a tricky problem upon which the statistical approach of the generalized sequences $(lx+m)/n$ may bring useful remarks.

In the generalization, introduced at the beginning of this paper, we impose $(l_i, n) = 1$ for all i. In that case, the bijection theorem holds. Consequently, the ensemble of the n^k integer seeds, from 0 to n^k-1, (formed with k nits), will give exactly, for the k first iterations, n^{k-1} iterates of each of the n types. Another aspect of this property is given by the matrix M (dimension $n \times n$), the elements of which are the probability that a number ending with the nit j will, after one iteration, gives a number for which the LSN is i.

An example will precise the concept. We consider a quadrifurcation ($n=4$) with $l_1 = 3$ and $m_1 = 1$. The iterate of a number ending with the nit 1, written consequently $1+4i+4^2j+\cdots$, with i and $j \in \{0, 1, 2, 3\}$, will be $1+3i+12j+\cdots$. Table I gives the two last nits of the iterate and it exhibits interesting properties. For seeds of type 1 (i.e., ending with 1), if the i (i.e., the second LSN) is equally partitioned between the n nits, then its iterate is equally partitioned between the n "type numbers." Moreover a further equirepartition of j (the third LSN of the seed) implies an equirepartition of the type of the second iterates.

Table I. The Two Last Nits of $1+3i+12j$

	i			
j	0	1	2	3
0	01	10	13	22
1	31	00	03	12
2	21	30	33	02
3	11	20	23	32

Since this properties are fulfilled for the k nits of the original set of seeds considered (integers from 0 to n^k-1), we can iterate k times and, at each iteration, consider that the number of type "i" is given by the product of a column vector, giving the number of type i ($i \in \{0, 1,..., n-1\}$) at the preceding iteration, by a matrix where all elements are $1/n$. This result is a consequence of the bijection theorem and bring two remarks.

First, this method of iterating the matrix M (all elements of which are equal to $1/n$) to obtain the number of "i" numbers at the successive iterations is no more valid after k iterations. Take, for example, the usual Collatz problem, $n=2$, $l_0=1$, $m_0=0$, $l_1=3$, $m_1=1$ and $k=3$. Then, the set of seeds is consequently $\{0, 1, 2,..., 7\}$. For these 8 seeds, it is easily checked that each of the 3 firts iterations exhibits 4 odds and 4 even iterates. This is no longer true if we push one iteration further. Then the last iteration exhibits 7 even and 1 odd iterate. To recover a balance of this fourth iteration we need to take $k=4$ and deal with a set of 16 seeds going from 0 to 15.

The second remark concerns mapping where some l_i are not coprime with n. In that case, the bijection theorem does not hold and the matrix has no more all its elements equal to $1/n$. We will come back to that point later on. But first we present the statistical physics nature conjecture.

Conjecture. Consider a great number of seeds randomly and uniformly distributed on an interval around a very large value. This interval is much smaller than the central value but much bigger than one (we take typically the interval $[10^9-10^7, 10^9+10^7]$). We conjecture that the behavior of this ensemble of seeds is equivalent to the behavior obtained by selecting randomly with equal probability among the n possible $U(x)$. Of course the successive x are no longer integers.

If we notice that, since $x \to \infty$, m_i can be neglected with respect to $l_i x$, a consequence of this conjecture is that the game for the evolution of the variable $u = \log_n(x)$ is a succession of independent random steps of respective length $\log_n(l_i) - 1$ each issue having the same probability $1/n$ to be selected. Consequently both the mean value and the mean square deviation taken on all the seeds vary as N, where N is the number of iterations.

Clearly this conjecture is a kind of generalization of the results given above concerning the n^k first integers. Two questions must be answered:

— Can we, at least for the first iterations, play this random walk game?

— For the n^k first integers, we notice that after k iterations the strict equality between the n possible issues stops. This suggests that after $K = \log_n(A)$, where A is the central value of the seeds, the random walk game does not apply any more.

A first answer to these questions has been given in ref. 8 for the original $(3x+1)/2$ problem and a $n = 3$ case. Both suggest that the limit of validity exceeds the number of nits of the seeds (i.e., for iteration numbers larger than this value, the computed points were still on the straight lines characteristic of the random walk game). But these computer experiments were not totally conclusive since, as the number of iterations increases, a larger and larger set of numbers reached cycles or fixed points and then the independent random steps conjecture fails. To check the hypothesis of the validity for very large number of iterations it is consequently much more interesting to deal with ascending sequences hopping that iterates will escape cycles or fixed points. The conjecture gives the condition for which the walks move in the positive direction. All (l_i, n) being equal to 1, all types of iterations are equally probable. The condition writes consequently

$$\prod_{i=0}^{n-1} l_i > n^n \tag{7}$$

Figure 1 shows what happens in such a case with $n = 5$, $l_i = 7, 4, 6, 3, 7$ and $m_i = 0, 1, -2, 1, -3$ where i varies from 0 to 4. The seeds are uniformly and randomly distributed in the interval $10^9 - 10^7$, $10^9 + 10^7$. These values imply that after $\log_5 10^9 \sim 13$ iterations, the whole information concerning the seeds has been extracted. But this has no influence on the conjecture of independent random walk since we still have an excellent agreement after 200 iterations. Many other simulations have been performed always leading to an excellent agreement with the conjecture.

Now we turn to new cases where we relax the condition $(l_i, n) = 1$ for all i. As already mentioned an highly mathematical treatment has been given in refs. 4–6. Our approach follows the lines developed in the case

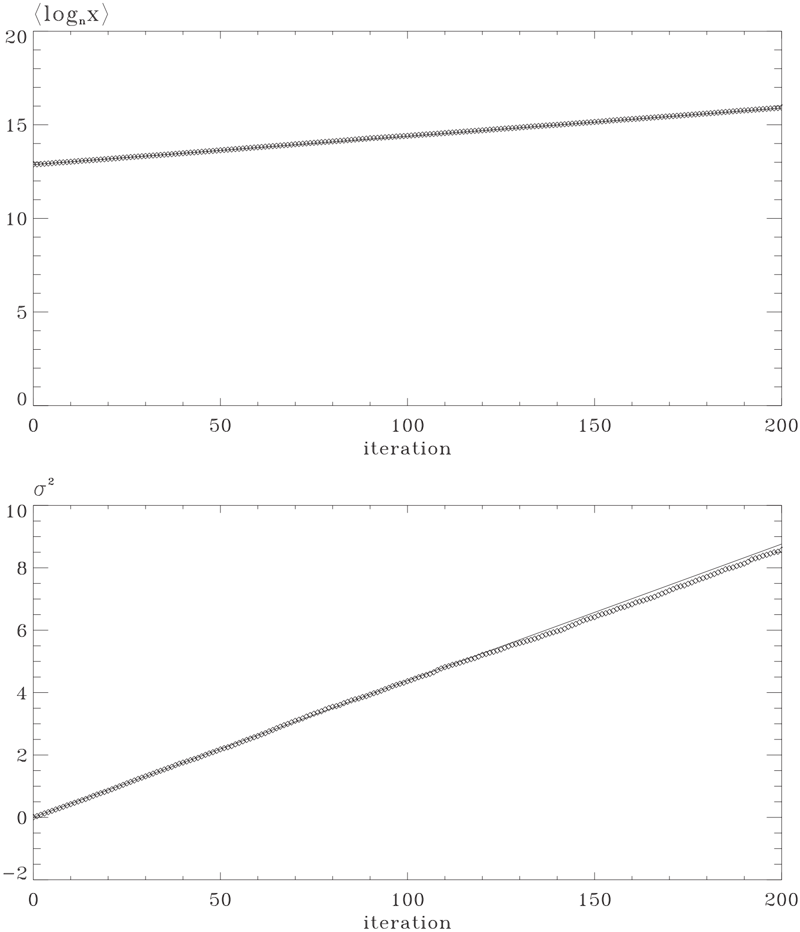

Fig. 1. $n = 5$, $l_i = 7, 4, 6, 3, 7$ and $m_i = 0, 1, -2, 1, -3$, with $i = 0, \ldots, 4$. Mean value (top) and standard deviation (bottom) of a population of 10^5 numbers, taken initialy at random in the interval $[10^9 - 10^7, 10^9 + 10^7]$, obtained after k iterations. The squares give the values obtained with an ensemble of numbers, the result of the random walk model is given by the continuous line.

$(l_i, n) = 1$ for all i. We first build the matrix M, the element of which $m_{i,j}$ (i line index, j column index) gives the probability that a number ending with the nit j gives, after one iteration a number ending with nit i.

The following example gives a way to build the matrix in the case where $n = 6$, $l_4 = 2$ and $m_4 = 4$. Then a number ending with the nit 4, written consequently $4 + 6i + 6^2 j + \cdots$ (i and $j \in \{0, 1, 2, 3, 4, 5\}$) has the iterate $2+2i+12j+\cdots$. The last significant nit of this iterate will be 2, 4, 0, 2, 4, 0 with i equal to 0, 1, 2, 3, 4, 5 respectively. Moreover, if these numbers (ending with 4) have their second LSN j equally distributed, we see that column number 4 will have $1/3$ on lines 0, 2, 4 and 0 elsewhere. If $l_4=3$ and $m_4=-6$, then column 4 will have $1/2$ on lines 1 and 4 and 0 elsewhere. If $l_4=6$ and $m_4=6$, then the element $m_{5,4}$ of matrix M equal 1 and, of course, all the other elements of column 4 are equal to 0.

Now let us consider again the n^k numbers (from 0 to n^k-1) which can be written with k nits. Can we obtain the number of the n "type i numbers" at the successive iterations by taking the powers of the matrix M as defined above? The answer is yes if the iteration preserves the property of equipartition for the second LSN. We checked that it is always true when $(l_i, n)=1$. Now we must check for each l_i. For the value $i=a$, let l_a and m_a be the associated values. x is written $a+ni+n^2j$. The iterate of x is $U^{(1)}(x)=(al_a+m_a)/n+l_ai+l_anj$. As a matter of fact, for this test on the preservation of the equirepartition property, the precise value of a does not play any role (it does of course in the building of the matrix M). We have just to build the table giving the two last nits of $li+lnj$ when both i and j vary from 0 to $n-1$.

Before giving a general rule, let us consider the following 4 examples for which $n=6$ and l take the values 2, 4, 3 and 6 respectively. As for Table I, the two last nits of the iterate of the seed written $a+6i+6^2j+\cdots$ are $2i+12j$ for $l=2$, $4i+24j$ for $l=4$, $3i+18j$ for $l=3$, $6i+36j$ for $l=6$ and are given Tables II–V respectively. Tables II and III show that in both cases the iterates ends with 0, 2 or 4. But in Table II, looking only at the upper half, which is identically reproduced by the lower one, we see that the six numbers ending with 0 (respectively 2 and 4) have as second LSN respectively 0, 1, 2, 3, 4, 5 assuming a conservation of the equirepartition property of the second LSN. This property is not found in Table III and

Table II. The Two Last Nits of $2i+12j$

	i					
j	0	1	2	3	4	5
0	00	02	04	10	12	14
1	20	22	24	30	32	34
2	40	42	44	50	52	54
3	00	02	04	10	12	14
4	20	22	24	30	32	34
5	40	42	44	50	52	54

Table III. The Two Last Nits of $4i+24j$

	i					
j	0	1	2	3	4	5
0	00	04	12	20	24	32
1	40	44	52	00	04	12
2	20	24	32	40	44	52
3	00	04	12	20	24	32
4	40	44	52	00	04	12
5	20	24	32	40	44	52

the 3 one third found on lines 0, 2 and 4 losses their meaning. Tables IV and V show that $l=3$ will imply two $1/2$ in the corresponding column of matrix M, while $l=6$ will produce only one element different from 0 in the corresponding column (this element being, of course, equal to 1). In both cases, the equipartition of the second LSN is fulfilled.

Theorem 3. Consider the ensemble of seeds with a given LSN, and all possible LSN+1 and LSN+2. Then the iterates which have the same LSN are such that their LSN+1 take all the values from 0 to $n-1$ iff $\bar{l}_i$ and n are coprime for $i=0,\dots, n-1$ with $l_i = k\bar{l}_i$ and $(l_i, n) = k_i$.

Proof. The iterate of $a+in+jn^2+\cdots$ is $(la+m)/n+il+jln+\cdots = a_0+a_1 n+\cdots$ where a_0 and a_1 are its two LSN. First, we look for the values i for which the LSN a_0 of the iterate is fixed. We compute

$$i\,l \equiv b_0 \pmod{n} \tag{8}$$

Table IV. The Two Last Nits of $3i+18j$

	i					
j	0	1	2	3	4	5
0	00	03	10	13	20	23
1	30	33	40	43	50	53
2	00	03	10	13	20	23
3	30	33	40	43	50	53
4	00	03	10	13	20	23
5	30	33	40	43	50	53

Table V. The Two Last Nits of $6i+36j$

	i					
j	0	1	2	3	4	5
0	00	10	20	30	40	50
1	00	10	20	30	40	50
2	00	10	20	30	40	50
3	00	10	20	30	40	50
4	00	10	20	30	40	50
5	00	10	20	30	40	50

with $b_0 \equiv a_0 - \frac{la+m}{n}$ (mod n). Suppose that $i = i_0$ is such that Eq. (8) is verified and take I with $I \in [0, n-1]$ such

$$(i_0 + I)\, l \equiv b_0 \pmod{n} \tag{9}$$

Then

$$Il \equiv 0 \pmod{n} \tag{10}$$

Taking into account that $(l, n) = k$ and introducing $l = \bar{l}k$ and $n = \bar{n}k$, we get

$$I\bar{l} \equiv 0 \pmod{\bar{n}} \tag{11}$$

But now, $(\bar{l}, \bar{n}) = 1$, then $I\bar{l}$ takes all values in $[0, \bar{n}-1]$ when I varies from 0 to $\bar{n}-1$. Consequently a_0 is fixed for $I = 0, \bar{n}, 2\bar{n},\ldots, (k-1)\,\bar{n}$ and take only $\bar{n}$ different values which are recover k times when I varies from 0 to $n-1$. For a fixed value of a_0, the carries are equal to $i'\bar{l}$ with $i' = 0,\ldots, k-1$ and the value a_1 reads

$$a_1 \equiv jl + i'\bar{l} \pmod{n} \tag{12}$$

$$\equiv j'\bar{l} \pmod{n} \tag{13}$$

with $j' = 0,\ldots, n-1$. a_1 will take all values $0,\ldots, n-1$ when j' varies from 0 to $n-1$ iff $(\bar{l}, n) = 1$ ∎

$(\bar{l}, n) = 1$ generalizes the preceeding bijection result $(l, n) = 1$. For the four examples given above, with $n = 6$, the theorem is not valid for $l = 4$, but holds for $l = 2$, 3 or 6, as expected.

As a consequence of the theorem, the number of each type number is computed using matrix M. Morover, the values in column j of the matrix, related to the "j" type of the seed, is determined by l_i and the k_i such $(l_i, n) = k_i$. The theorem indicates that the values $1/\bar{n}_i$ are recover $\bar{n}_i$ times in this column, spaced by $k_i - 1$ zeros. This is just an indication of the relative positions of these values. The presice position depends on m_i.

We, now, give two examples with $n = 6$ and $k = 5$. For the first one $l_i = 2, 3, 7, 6, 2, 5$ and $m_i = 0, 3, 4, 12, 10, 5$ with $i = 0,\ldots, 5$. The matrix takes the form:

$$M = \begin{bmatrix} 1/3 & 0 & 1/6 & 0 & 0 & 1/6 \\ 0 & 1/2 & 1/6 & 0 & 1/3 & 1/6 \\ 1/3 & 0 & 1/6 & 0 & 0 & 1/6 \\ 0 & 0 & 1/6 & 0 & 1/3 & 1/6 \\ 1/3 & 1/2 & 1/6 & 0 & 0 & 1/6 \\ 0 & 0 & 1/6 & 1 & 1/3 & 1/6 \end{bmatrix} \tag{14}$$

$k = 5$ gives $6^5 = 7776$ numbers with 1296 of each type which is the value of all the 6 components of the starting vector. The successive iterations of the matrix gives

$$\begin{aligned} M^0 v &= (1296, 1296, 1296, 1296, 1296, 1296) \\ M^1 v &= (864, 1512, 864, 864, 1512, 2160) \\ M^2 v &= (792, 1764, 792, 1008, 1548, 1872) \\ M^3 v &= (708, 1842, 708, 960, 1590, 1968) \\ M^4 v &= (682, 1897, 682, 976, 1603, 1936) \end{aligned} \tag{15}$$

It is easily checked that performing the iterations on the 7776 first integers gives, up to that point, strictly the same result. As predicted the agreement stops at the next iteration. Should we have forgotten it, a warning signal (at least in this case) would be given with components of $M^5 v$ no more integers.

For the second example we take again $n = 6$ with $l_i = 2, 3, 2, 4, 2, 5$ and $m_i = 0, 3, 2, 12, 10, 5$ with i running from 0 to 5. The method of raising M to successive powers must break down before $M^5 v$ since one of the component of l takes the value 4. Indeed

$$\begin{aligned} M^0 v &= (1296, 1296, 1296, 1296, 1296, 1296) \\ M^1 v &= (1080, 1728, 1080, 1080, 1728, 1080) \\ M^2 v &= (900, 1980, 900, 1116, 1764, 1116) \end{aligned} \tag{16}$$

Up to that point these values agree with those obtained by direct iteration of the 7776 first integers. Iterating once again, we obtain respectively

$$M^3 v = (858, 2064, 858, 1074, 1848, 1074) \tag{17}$$

while counting the number of "type i" in the set of the 7776 numbers obtained at the third iteration gives

$$M^3 v = (906, 1872, 906, 1026, 2040, 1026) \tag{18}$$

and, as expected, the results disagree.

Our test on the two LSN of $li + lnj$ gives a sufficient condition but not a necessary one. To gain some insight, we take the following examples. $n = 3$, $k = 3$, $l_i = 9, 2, 1$ and $m_i = 6, 1, 4$ for i running from 0 to 2. The value $l = 9$ turns the iterate of $3i + 9j$ into $2 + 9i + 27j$. A one is put in the column 0, line 2 of M and the equirepartition of the second LSN is not fulfilled. Indeed, starting from the initial vector (9, 9, 9), we obtain

$$\begin{aligned} M^0 v &= (9, 9, 9) \\ M^1 v &= (6, 6, 15) \\ M^2 v &= (7, 7, 13) \end{aligned} \tag{19}$$

while counting the number of type i in the set of the 27 numbers obtained at the three first iterations gives respectively

$$\begin{aligned} &(9, 9, 9) \\ &(6, 6, 15) \\ &(4, 4, 19) \end{aligned} \tag{20}$$

The matrix M being

$$M = \begin{bmatrix} 0 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 \\ 1 & 1/3 & 1/3 \end{bmatrix} \tag{21}$$

The usual discrepancy due to the finite $k = 3$ value should take place only at the next iteration. This is due to the $l_0 = 9$ value. But if now we take $m_0 = 0$ having all the other values unchanged, we obtain both, for the direct iterations and the matrix iteration (9, 9, 9), (15, 6, 6),(19, 4, 4). The explanation here is simple. In the case $m_0 = 6$, the one in column zero is on line 2 and the "type 0" numbers are transformed on "type 2" numbers which do not possess the equipartition property of the second LSN. Consequently these numbers will not provide, at the next step, equal numbers of "type 0," "1," and "2" numbers. In the case $m_0 = 0$, the one of column 0 has moved to line 0. "0 type" numbers give automatically "0 type" numbers irrespectively of the conservation of the equipartition property. The matrix method is valid up to iteration k.

5. ASYMPTOTIC PROPERTIES WHEN THE NUMBER OF ITERATIONS GOES TO INFINITY

Now, we want to generalize the conjecture given in the case of all $(l_i, n) = 1$ and to distinguish between ascending and descending sequences. When the M matrix can be used, we just need to study the properties of M^p when $p \to \infty$. These properties are given by the largest (in modulus) eigenvalue and the associated eigenvector. Here, the result is particularly simple. M being a stochastic matrix (in each column the sum of the element is equal to one) the largest eigenvalue is equal to one and the eigenvector associated to this value has all its components positive. We call α_0, α_1,..., α_{n-1} these components normalized in such a way that $\sum_{i=0}^{n-1} \alpha_i = 1$. Then, when the matrix method can be applied, the average value of the increase of $u = \log_n(x)$ is given by

$$\langle \Delta u \rangle = \frac{1}{n} \sum_{i=0}^{n-1} \alpha_i \log_n \left(\frac{l_i}{n} \right) \tag{22}$$

and the ascending or descending nature of the sequence depends on the value of $\prod_{i=0}^{n-1} (l_i/n)^{\alpha_i}$ compared to one (ascending if greater, descending if smaller).

For all $(l_i, n) = 1$, all the α_i are equal to $1/n$ and we recover the previous formula. We now show a connection with the work of Matthews *et al.* They introduced a family of matrix $Q(m)$ where m is an integer. When $m = n$, we checked that for all examples they give, the matrix $Q(n)$ is identical with our matrix M. Moreover, they use the eigenvector associated to the eigenvalue one to determine the ascending or descending nature of the sequence. They do not mention the meaning which can be given to M in the determination of the number of "type i" numbers for the successive iterations. Neither they show the precise meaning of this matrix M in the treatment of the k first iterations of the n^k first integers. A deeper understanding of the connection between the two approaches is certainly needed and very likely will unveil new aspects of the problem.

As before, we conjecture that we can transfer results obtained for a set of n^k seeds experiencing k iterations to an ensemble of seeds centered around a very large value and uniformly and randomly distributed around this large value. And, at least if the sequence is ascending, there is no limit in the number of iterations. Obviously, successive iterations are not uncorrelated and we cannot apply the central limit theorem and multiply by N (number of iterations) the mean square deviation at one step for the variable $u = \log_n(x)$. Consequently, we will only check the variation with the number of iterations of the average value of $\log_n(x)$. After a transient, this variation must go as a straight line of slope

$$\frac{1}{n}\sum_{i=0}^{n-1} \alpha_i \log_n\left(\frac{l_i}{n}\right) \tag{23}$$

To check the conjecture, we will proceed in a slightly different way (with respect to the case where the bijection theorem holds). We simply compute at each iteration how many "0 type," "1 type,"..., "$n-1$ type" numbers have been obtained and compare with the results given by the successive powers of matrix M. The initial 10^4 seeds are selected, as before,

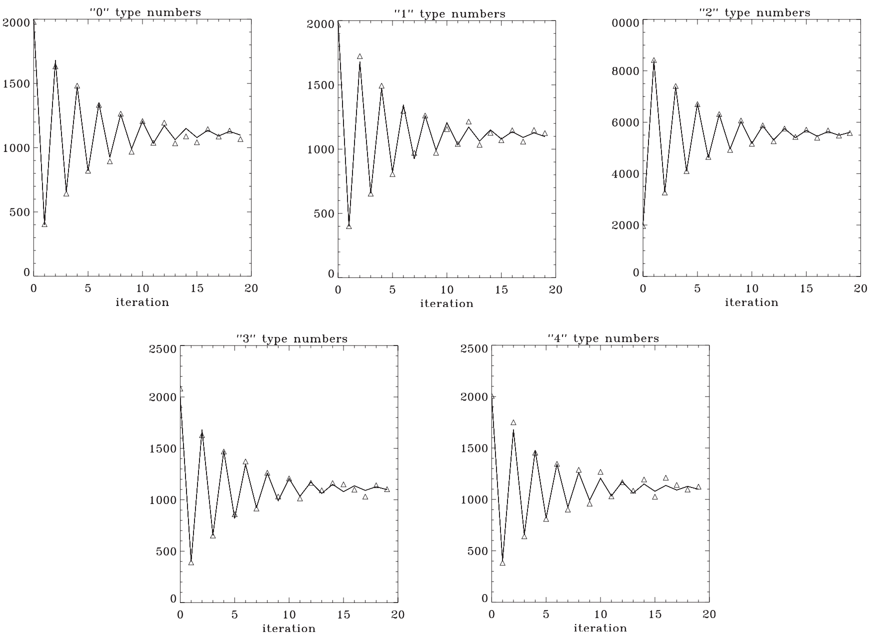

Fig. 2. Number of each type number for $n = 5$, $l_i = 5, 10, 4, 10, 5$ and $m_i = 10, 0, 2, 5, 15$, with $i = 0,\ldots, 4$. The values computed at each iteration, using a set of 10,000 seeds taken at random around 10^{10}, are connected with a solid line, while the triangles give the values computed using the matrix M.

in the interval $[10^{10}-10^6, 10^{10}+10^6]$ and are randomly and uniformly distributed in this interval. We must see both the transient and the asymptotic state given by the eigenvectors associated to the eigenvalue 1. Figure 2 shows, for the 5 cases, the results given by performing the exact calculations (triangles) compared to the values obtained using the matrix M (solid line). The values of l_i and m_i $(i=0,\ldots,n-1)$ were chosen in order to bring important oscillations with the following structure of the matrix:

$$M = \begin{bmatrix} 0 & 0 & 1/5 & 0 & 0 \\ 0 & 0 & 1/5 & 0 & 0 \\ 1 & 1 & 1/5 & 1 & 1 \\ 0 & 0 & 1/5 & 0 & 0 \\ 0 & 0 & 1/5 & 0 & 0 \end{bmatrix} \tag{24}$$

The agreement is excellent and show clearly the validity of the matrix M method for the macroscopic treatment of a large set of seeds. Many other simulations have always confirmed this results.

ACKNOWLEDGMENTS

The authors thank anonymous referee for bringing to their attention the work of K. R. Matthews and collaborators.

REFERENCES

1. J. C. Lagarias, The $3x+1$ problem and its generalizations, *American Mathematical Monthly* **92**:3–23 (1985).
2. G. T. Leavens and M. Vermeulen, $3x+1$ search programs, *Computers and Mathematics with Application* **24**(11):79–99 (1992).
3. J. H. Conwey, *Unpredictable Iterations*, Proceedings of the number theory conference (University of Colorado, Boulder, 1972).
4. K. R. Matthews and A. M. Watts, A generalization of Hasse's generalization of the Syracuse algorithm, *Acta Arithmetica* **XLIII**:167–175 (1984).
5. K. R. Matthews and A. M. Watts, A Markov approach to the generalized Syracuse algorithm, *Acta Arithmetica* **XLV**:29–42 (1985).
6. R. N. Buthsworth and K. R. Matthews, On some Markov matrices arising from the generalized Collatz mapping, *Acta Arithmetica* **LV**:43–57 (1990).
7. M. R. Feix, A. Muriel, D. Merlini, and R. Tratini, The $(3x+1)/2$ problem: A statistical approach, in *Stochastic Process, Physics and Geometry, Locarno, Switzerland, 24–26 June 1991*, S. Albervio, U. Cattaneo, and D. Merlin, eds. (World Scientific, 1991), pp. 289–300.
8. M. R. Feix, A. Muriel, and J.-L. Rouet, Statistical properties of an iterated mapping, *J. Statist. Phys.* **76**:725–741 (1994).
9. R. Terras, A stopping time problem on the positive integers, *Acta Arithmetica* **30**:241–252 (1976).
10. J. C. Lagarias and A. Weiss, The $3x+1$ problem: Two stochastic models, *Annals of Applied Probability* **2**:229–261 (1992).