(3X + 1) / 2 PROBLEM AND ITS GENERALIZATION : A STOCHASTIC APPROACH

J.L. ROUET AND M.R. FEIX

Rapport Interne SUBATECH 00-03

Soumis pour publication à « EXPERIMENTAL MATHEMATICS «

$(3\,x+1)/2$ problem and its generalization : a stochastic approach

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

[1]Laboratoire de Mathématique, Applications et Physique Mathématique - UMR 6628,
Université d'Orléans, UFR des Sciences, F-45067 Orléans Cedex 2, France
jean-louis.rouet@labomath.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

Abstract The $(3\,x+1)/2$ problem is generalized into the n-furcation problem $(l_i\,x+m_i)/n$ where $i \in [0,1,\dots n-1]$ is the value modulo n of x. 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, first, to build a cipher and, second, to investigate a stochastic treatment of ensemble of large value seeds.

1 Introduction

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

$$\left\{ \begin{array}{lcll} x_{n+1} & = & x_n/2 & \text{if } x_n \text{ is even} \\ x_{n+1} & = & (3x_n+1)/2 & \text{if } x_n \text{ is odd} \end{array} \right. \tag{1}$$

Starting from x_0 (the seed) it has been checked [Leavens and Vermeulen 1992] that for all seeds up to 2^{40}, the sequence of iterates 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 [1], 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 [Conwey 1972].

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 purpose of this paper with a generalization of the bifurcation problem to an n-furcation. In this spirit another embedding can be found in the paper of Chamberland [Chamberland 1996] who extends to the real line the relation (1) only defined for integer numbers.

[1] While this paper is submitted, S. Fanelli [Fanelli 1999] claims that he is able to conclude. Nevertheless, our paper is not devoted to the proof of the conjecture but, rather, gives some applications of the generalized problem.

Another goal of our paper is to introduce a stochastic approach to this problem. First steps in this direction have been initiated by Terras [Terras 1976] and around 1990 this approach was independently studied by Lagarias [Lagarias and Weiss 1992] and ourselves [Feix et al. 1994, 1995]. This paper completes and precises the previous ones, and, moreover gives as possible application the building of a cipher.

Such a stochastic approach deserves obviously long comments. It may be pointed out that, since we have a perfectly deterministic process, a stochastic approach is unnecessary, irrelevant and may lead to errors. Statistical physicists have, since Boltzmann, faced this type of discussion (with sometimes very hot disputes). Of course in statistical physics we deal with the N-body problem which, as soon as $N > 2$, is totally non integrable while our present problem is integrable by pieces. But two kinds of results can be expected

- while the exact "trajectory" of a seed (i.e. the value of its n^{th} iterate) is totally lost in this stochastic approach, the behaviour of a large ensemble of seeds can be obtained. This is the usual statistical physic result.
- Deterministic sequences can look like random ones. This is the basic justification of all "pseudo-random numbers generators" used in computational physics. This is a very tricky problem and the statistical approach of sequence (1) will provide interesting remarks.

The paper is organized as follow : in section 2 we quickly remind the bijection theorem which provides partial justification for the stochastic approach. In 3 we give the generalization of the $(3x+1)/2$ problem introducing the n-furcation $(lx+m)/n$. Section 4 shows that the freedom in the choice of l, m and n allows the construction of cycles with period and sequence of iterates arbitrarily chosen. Section 5 consider the inverse problem (building the tree from the cycle 1-2 and its antecedent). In section 6, the structure of finite sequences of iterates is studied for different values of l, m and eventually n. It is shown that a cipher can be build as an application. We come back section 7 to a mixing mechanism (in fact a direct consequence of the bijection theorem given section 2) and introduce in section 8 the random walk game and its numerical simulation. Section 9 gives our conclusion.

2 The Bijection Theorem

The bijection theorem states that a bijection exists between the last k bits of a seed written in base 2 and the parities of the k first iterations given by (1). We simply show on an example the precise meaning of the theorem while proofs can be found in [Feix et al. 1994]. Table 1 gives the 16 numbers from 0 to $2^n - 1$ written in base 2 with $n = 4$ bits. The third column gives the parities of the first 4 iterates of the corresponding number with the following convention : an even iteration is noted by 0 and an odd iteration by 1, the number on the right in this column corresponding to the first iteration. For example, 1001 *i.e.* 9 gives the sequence $14, 7, 11$ and the parity sequences $IPII$ written in our convention 1101, i.e 13.

number		parity sequence	
base 10	base 2	convention	traduction
		even $\equiv 0$ odd$\equiv 1$	in base 10
0	0000	0000	0
1	0001	0101	5
2	0010	1010	10
3	0011	0011	3
4	0100	0100	4
5	0101	0001	1
6	0110	0110	6
7	0111	0111	7
8	1000	1000	8
9	1001	1101	13
10	1010	0010	2
11	1011	1011	11
12	1100	1100	12
13	1101	1001	9
14	1110	1110	14
15	1111	1111	15

Table 1: The 4 first sequences of parity corresponding to the 16 first positive integers.

Focusing our attention respectively on the last, the two last and three last bits of the numbers on one side and the first, the two first and three first iterations on the other side, we notice that the k last bits completely define the k first iterations (here for $k \leq 3$). Moreover, taking the 4 bits, we notice that there is a bijection between the 2^4 elements of the numbers running from 0 to 15 and the 2^4 possible sequences of 4 iterations.

The connexion with a stochastic treatment is based on this bijection. Selecting randomly the k bits of the seed produces a sequence of parities for the successive iterates which cannot be distinguished from the sequence obtained by tossing k times a coin. Notice that from the point of view of information theory the result is quite normal. The data of the k bits of the seed gives the same quantity of information as the data of the parities of the k first iterates. In fact the building of the seed from the sequence of parities is as simple as the direct process (see section 6 of this paper). The question is now to understand the influence of the finiteness of the seed. We will come back to this problem after the generalization of the next section.

3 Generalization : the $(l\,x + m)/n$ problem

Let us consider an iteration giving x_{p+1} as function of x_p accordingly the following rules where $x_p \in \mathbb{N}^+$, l_i, m_i, n are integers, positive for l_i and n, possibly negative for m_i, with $i = 0, \ldots, n-1$. We denote $x|_n$ the value of the integer x modulo n and

a_p l_i	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Table 2: Conditions for the bijection theorem for $n = 5$. The table gives all possible product $a_p l_i|_5$ when $a_p|_5$ and $l_i|_5$ take all possible values.

a_p l_i	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Table 3: Conditions for the bijection theorem for $n = 6$. The table gives all possible product $a_p l_i|_6$ when $a_p|_6$ and $l_i|_6$ take all possible values.

$$\text{if} \quad x_p\big|_n = i \quad \text{then} \quad x_{p+1} = (l_i\, x_p + m_i)/n \tag{2}$$

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

In addition we must add some constraints on m_i and l_i. First, x_{p+1} must belong to $\mathbb{N}^+$. Writing $x_p = a_p\, n + i$, with $i = 0, 1, 2, \dots n-1$, we have,

$$x_{p+1} = a_p\, l_i + \frac{i l_i + m_i}{n} \tag{3}$$

and a first condition is

$$(i l_i + m_i)\big|_n = 0 \tag{4}$$

Second, the bijection theorem imposes new constraints : after the first iteration given by (2), the n possibilities must exists for the next one. Writing the number x_p in base n and defining $\alpha_i = (i\, l_i + m_i)/n$, which is now an integer, because of the first condition, we have to check that when $a_p|_n$ goes from 0 to $n-1$ for each $l_i|_n$ the product $a_p\, l_i|_n$ goes from 0 to $n-1$. Notice that α_i plays a role in defining the bijection but not on its existence which only depends on $a_p\, l_i|_n$ with a_p and l_i going from 0 to $n-1$.

Tables 2 and 3 show the constraints for $n = 5$ and $n = 6$. For $n = 5$, we see that, except for the first line $a_p\, l_i$ brings the five digits $0, 1, 2, 3, 4$. Consequently, we must simply take all $l_i|_5 \neq 0$. But for $n = 6$ only the lines $l_i = 1$, $l_i = 5$ allow the continuation of the bijection and we must consequently have for all l_i, $l_i|_6 = 1$ or $l_i|_6 = 5$.

It is easily proven that

. for $n = 2$ we must have all l_i odd,

i	0	1	2	3	4
l_i	4	1	3	6	2
m_i	5	4	-1	2	-3

Table 4: An example of the coefficients l_i and m_i for a pentafurcation satisfying the bijection criterium : $il_i + m_i|_{n=5} = 0$.

. for $n = 3$ we must have all l_i such that $l_i|_3 \neq 0$,

. for $n = 4$ we must have all l_i such that $l_i|_4 \neq 2$ and $l_i|_4 \neq 0$,

. for $n = 5$ and 6 the constraints have been given above while for $n = 7$ the only constraint is $l_i|_7 \neq 0$.

Table 4 gives the example of a pentafurcation with the indicated values of l_i and m_i which satisfy (3) and the relation $l_i|_5 \neq 0$.

3.1 A consequence of the bijection theorem : divergence from or convergence to a small number (or a small cycle) of the iterated numbers

We suppose a very large seed where the k digits have been taken at random. If $k \to \infty$ the n issues will appear with the same frequency in the list of the iterations given by the seed. Supposing a large seed and consequently large subsequent iterates we see that, roughly speaking, the seed will be multiplied, after k iterations, by

$$l_0^{k/n} l_1^{k/n} l_2^{k/n} \dots l_{n-1}^{k/n} / n^k \tag{5}$$

The decrease or increase of the sequence at least for the k first steps will depend on the value of A given by

$$A = l_0 l_1 l_2 \dots l_{n-1} / n^n \tag{6}$$

If $A > 1$ the sequence diverges to infinity and if $A < 1$ it decreases. In both cases a trapping in a cycle is possible but the bijection theorem indicates that if the digit of the seed are taken randomly the falling in the trapping cycle has very small probability. This is no more true when we approach small numbers. It can be easily checked that the bijection $x/2$, $(3\,x + 1)/2$ decreases, $x/2$, $(5\,x + 1)/2$ increases and for the trifurcation $x/3$, $(4\,x + 2)/3$, $(7\,x + 4)/3$ for respectively $x|_3 = 0, 1, 2$, we have a slight increase since $4 \times 7/3^3 = 28/27$.

3.2 Guessing the nature of the sequence

Suppose you are given a sequence of parities (for a bifurcation) or a sequence of the nature $0, 1, \dots n - 1$ of the successive iterations (for a n-furcation), but you are told that, may be, the sequence has nothing to do with the n-furcation game and has been obtained by a truly random process. Then you are asked to guess how the sequence has been obtained.

If the rules of the (eventually used) game have been honestly communicated the best way is to use the bijection theorem to build the seed. If after a certain number of steps the digits obtained are systematically zero it is likely that we have obtained all the significant digits of a seed which is certainly finite. But the given sequence can be too short compared to the numbers of digits of the seed. In that case you cannot decide. The case where the number of digits of the seed is equal to the number of iterates corresponds to the possibility of building a cipher and will be studied section 6.

But things can be more difficult if you are not given the l_i and m_i of the game (n is specified) and even worst if you are given a wrong set of l_i, m_i. It would be interesting to study the sequence of generalized parities especially for diverging series and to compare their pseudo randomness to the sequences obtained with the actual random number generators used in computers.

4 An equation for building cycles

Although n-furcation can be also studied, we will treat only the bifurcation case $x/2$, $(l\,x+m)/2$ problem. Suppose we want a N-period cycle with a given sequence of parities. It is easily proved that x_0 the first term of the cycle, l, m and N are connected in the following equation

$$(2^N - l^J)\, x_0 = 2^N\, m\, K \tag{7}$$

In equation (7), J is the number of odd parities and K is given by applying the N iterations (either $x/2$ or $(l\,x+1)/2$) to the seed 0. Note that m has been taken out and figure explicitly in equation (7).

In the following example with $l = 3$, we look for m such that a cycle of period 5 having the sequence I, I, P, I, P for which I stands for odd and P for even. Consequently in equation (7) we have $J = 3$, $N = 5$.

The number K is computed in the following way. The I iterate of 0 is $1/2$, the second iterate is again I : applied to $1/2$ it gives $(3/2)(1/2) + 1/2 = 5/4$. The third is P : applied to $5/4$ it gives $5/8$. The fourth is I : Applied to $5/8$ it gives $15/16 + 1/2 = 23/16$. The fifth and last is P and gives $K = 23/32$. Equation (7) writes

$$5x_0 = 23m$$

Since 23 cannot be divided by 5 the solution with the smallest m is $m = 5$ and $x_0 = 23$ with the cycle $23, 37, 58, 29, 46, 23, 37, \ldots$.

Sometimes equation (7) can be simplified. It is the case in the example with $l = 3$ and the sequence $IIIIPIIIPPP$. Now $N = 11$, $J = 7$ and equation (7) writes

$$-139\,x_0 = 2363\,m$$

Since $2363/139 = 17$ we get $m = -1$ and $x_0 = 17$ a member of the well known 11 period cycle of the $(3x-1)/2$ problem.

A solution always exists for equation (7) : $m = 2^N - l^J$ and $x_0 = 2^N K$ (always an integer). If we could find N and J such that

$$2^N - 3^J = 1 \tag{8}$$

we would obtain a cycle for the $(3\,x + 1)/2$ problem ! Unfortunately the only solution is $N = 2$, $J = 1$ which leads to the trivial cycle $2, 1$. Of course, other solutions may exist if $2^N K$ can be divided by $2^N - 3^J$ (as in the above example leading to $m = -1$), but in that case we must guess the correct sequence. Consequently if equation (7) is certainly useful to obtain cycles of not too large period in the generalized $(l\,x + m)/2$ problem its usefulness in the study of the conjecture for $l = 3$, $m = 1$ is not obvious. Nevertheless equation (7) deserves further studies.

5 The inverse problem

Another approach consists in considering the "landing" on the cycle *i.e.* the antecedents of 1 : these numbers giving 1 after 1, 2, 3 . . . steps coming back to the usual $(3\,x+1)/2$ problem. Figure (1) gives the first 10 steps. The question is : how many "grand father" do we find at step k ?

Obviously each element y at the step $k - 1$ gives at step k its double $x = 2y$ but it can have another parent if $x = (2y-1)/3$ is an integer. Let γ be the last digit of the number y written in base 3 with $\gamma \in \{0, 1, 2\}$. Since $y = 3\beta + \gamma$ we obtain $x = 2\beta + (2\gamma - 1)/3$ and for x to be an integer we need $\gamma = 2$. Supposing an equiprobability between the 3 possible values of γ we see that in one third of the cases y will have 2 antecedents. This corresponds to the relation $F(k) = (4/3)F(k-1)$ where $F(k)$ is the number of antecedents at step k with, consequently an asymptotic formula $F(k) \sim (4/3)^k$. Figure (2) shows that for k up to 28 the curve $\log F(k)$ fits, indeed, a straight line the slope of which .2879 checks very nicely with the theoretical value of the stochastic model $\log 4/3 = .2877$. Moreover we have in the inverse problem the same kind of conservation of the equipartition of the last digits. We show for example that if we have equipartition of the last $p+1$ ternary digit in y at one step, we have equipartition of the last p ternary digits of the antecedent.

For example we consider the numbers $y = a\,3^3 + b$ where b runs from 0 to 26. We put in columns 1, 2, 3 of table 5 the three possible digits of b. On columns 4, 5, 6 we give the two last ternary digits of the antecedent $x = 2y$ for column 1, 2, 3 respectively. For example, the underlined b 211 corresponds to the number $27a + 22$, its double is $54a + 44$. We take this

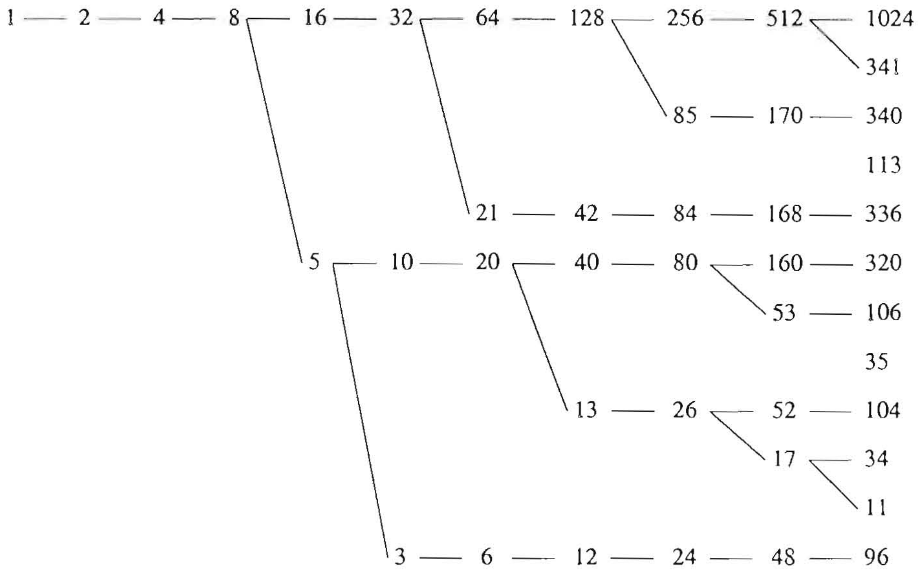

Figure 1: Tree of the "parent" numbers.

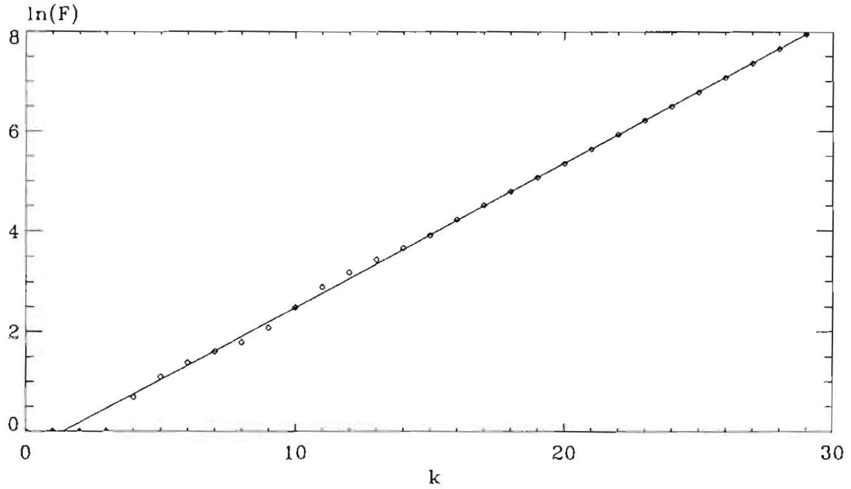

Figure 2: Logarithm of the numbers of parents at step k.

number modulo 9 to obtain the last two digits. The value is 8 written 22 in ternary units (underlined column 5 of table 5). In column 7 we give the nine 2 last ternary digits of the nine numbers which have an antecedent $x = (2y - 1)/3$. Since they must end by 2 they are the numbers listed column 3. For example, 122 corresponds to the number $a\,3^3 + 17$ which gives the antecedent $18a + 11$. This number has indeed 02 for its two last ternary digits (the number and its antecedent are inserted in squares).

These results and the results of the numerical investigation of the 28 antecedent steps shows the interest of a stochastic approach for the inverse problem. See [Feix et al. 1994] for further

1	2	3	4	5	6	7
000	001	002	00	02	11	01
010	011	012	20	22	01	10
020	021	022	10	12	21	12
100	101	102	00	02	11	21
110	111	112	20	22	01	00
120	121	$\boxed{122}$	10	12	21	$\boxed{02}$
200	201	202	00	02	11	11
210	$\underline{211}$	212	20	$\underline{22}$	01	20
220	221	222	10	12	21	22

Table 5: Column 1, 2, 3 : the three last "trits" of b, with $y = 27a + b$. Column 4, 5, 6 : the last 2 "trits" of the parent number $x = 2y$. Column 7 the 2 last "trits" of the parent number $x = (2y + 1)/3$.

details and section 7 of this paper for the direct problem.

6 An application to the $(l\,x + m)/2$ problem : building a cipher

Let us consider the $(l\,x + m)/2$ problem which, at each steps, exhibits 2 possibilities, characterized by the values of 4 numbers, l_0, m_0, l_1, m_1. If these numbers are chosen fulfilling the constraints given section 3, a bijection exists between the parities of the k first iterations and the k less significant bits (LSB) of the seed. In other words, the $(l\,x+m)/2$ problem gives a rule to obtain a permutation between two words of k bits taken among 2^k words. This suggest to use the $(l\,x + m)/2$ problem to build a cipher. Obviously, it is possible to consider a more general case based on the $(l\,x + m)/n$ problem but we will restrict the study to the case $n = 2$. The first reason is that it is important to study the structures between the ciphertext (k iteration bits) and the cleartext (the k seed bits) in the simplest case and the second one is that, from a practical point of view, computers working directly in binary, will provide an easy programming.

In order to determine the ciphertext we have to choose first the four values l_0, m_0, l_1, m_1. As already mentioned the l_0, m_0, l_1, m_1 have constraints for the bijection to exist and we have to take an even value for m_0 and odd values for l_0, l_1, m_1. In addition they are chosen in $\mathbb{N}^+$ in order to compute only positive integers. In fact this condition do not restrict the choice of the four values as it is shown thereafter. First, as the k first iteration depends only on the k last bits (a consequence of the bijection), it is not necessary to compute the iterates with more than the k LSB. Consequently, l_0, m_0, l_1 and m_1 can be large, giving very large iterates too. Without this property, the code would be of no practical value. Indeed, a large number of choices for the quadruplet l_0, m_0, l_1, m_1 must exist (to make the code difficult to crack). We will see that this numbers have, as the messages, around k significant bits. This implies that the k^{th} iterate will have around k^2 bits. Happily, all calculations will consider only k bit numbers (i.e. the k last bits of each iterate) In fact, if necessary, a further saving can take place since the first operation must be computed on k bits, the second one on $(k-1)$ bits and to be able to have the last iteration, the last bit of the k iterate is enough.

Now it is not necessary to look at all possible values of the quadruplet l_0, m_0, l_1, m_1 because for a similar reason only the k last bits of these numbers are relevant to perform the k first iterations. Let us compute more precisely the possible number of quadruplet (i.e. the complexity of the code). The calculation reads as follow. We consider the two sets of numbers l_0, m_0, l_1, m_1 and $\bar{l}_0, \bar{m}_0, \bar{l}_1, \bar{m}_1$ with $\bar{l}_i = l_i + 2^k\lambda_i$ and $\bar{m}_i = m_i + 2^k\mu_i$, with λ_i, $\mu_i \epsilon \mathrm{N}$, $i = 0, 1$. The first iterate x_1 of x_0 reads

$$x_1 = \frac{l_i x_0 + m_i}{2}$$

where $l_i = l_0$ or l_1 and $m_i = m_0$ or m_1 according to the parity of x_0. The first iterate $\bar{x}_1$ of x_0 using the other set of values $\bar{l}_i$ and $\bar{m}_i$ reads

$$\bar{x}_1 = \frac{\bar{l}_i x_0 + \bar{m}_i}{2} = x_1 + 2^{k-1}(\lambda_i x_0 + \mu_i).$$

As already mentioned, the $k-1$ next iterations only depends on the $k-1$ less significant bits of x_1 and $\bar{x}_1$ which are the same. Consequently, the iterates $\bar{x}_2$ and x_2 of the number x_1 will have the same $k-2$ last bits, and then the next iteration are the same for both. And so on for the next $k-2$ iterations.

At this step, we have a first limit on the number of possible quadruplets. For message of k bits l_0, m_0, l_1 and m_1 are also written on k bits : with m_0 even and l_0, l_1 and m_1 odd. But this set of different possible cases will be reduced by 4 taking into account the following considerations.

We first consider the iterates x_1 and $\bar{x}_1$ (points A and $\bar{A}$ on table 6(a)) of even seeds x_0 computed using the values l_0, m_0 and $\bar{l}_0, \bar{m}_0$ respectively, with $\bar{l}_0 = l_0 + 2^{k-1}$ and $\bar{m}_0 = m_0$. We have $\bar{x}_1 = x_1 + 2^{k-2}x_0$ and since $x_0 = 2\alpha$, x_1 and $\bar{x}_1$ taken modulo 2^{k-1} are identical.

We turn to odd iterations considering first l_1, m_1 and $\bar{l}_1 = l_1 + 2^{k-1}, \bar{m}_1 = m_1 + 2^{k-1}$ and their iterates x_1 and $\bar{x}_1 = x_1 + 2^{k-2}x_0 + 2^{k-2}$ (points B and $\bar{B}$ on table 6(b)). Since x_0 is now odd, $\bar{x}_1$ and x_1 have the same $k-1$ LSB. If now we consider x_1 and $\bar{x}_1$ (points C and $\bar{C}$ on table 6(b)) computed using the couples $l_1, \bar{m}_1$ and $\bar{l}_1, m_1$ respectively, we get $\bar{x}_1 = x_1 + 2^{k-2}x_0 - 2^{k-2}$ (the previously definition for $\bar{l}_1$ and $\bar{m}_1$ have been adopted). Again we check that x_1 and $\bar{x}_1$ have the same $k-1$ LSB.

The symmetries for l_0, m_0 on one hand, and l_1, m_1 on the other hand are given table 6. The conclusion is that we can let m_0 (respectively m_1) take all the even (respectively odd) values between 0 and 2^k but restrict l_0 and l_1 to the odd values between 0 and 2^{k-1}. Since we have now

	l_0	$\bar{l}_0$
m_0	A	$\bar{A}$

(a)

	l_1	$\bar{l}_1$
m_1	B	$\bar{C}$
$\bar{m}_1$	C	$\bar{B}$

(b)

Table 6: Points A and $\bar{A}$, B and $\bar{B}$, C and $\bar{C}$ represents quadruplets giving the same permutation.

2^{k-1} possibilities, for m_0 and m_1 and 2^{k-2} for l_0 and l_1, we obtain all together 2^{4k-6} possible quadruplets and consequently this is the number of all possible codes.

It is interesting to compare this value to the number of permutations possible with the k bits of a word. This kind of code has $k!$ possible permutations which correspond to the knowledge of $\log_2 k!$ bits of informations. If k is large enough this is equivalent to $k \log_2 k$ [2]. For the code based on the $(l\,x+m)/2$ rule we get $4k-6$ bits. This is less than the bit-permutation code but not so different if we consider the extremely slow increase of $\log_2 k$ with k. Now, the more general code is based on the permutations of words of the 2^k words which can be formed with k bits. It implies $2^k!$ permutations that is $\log_2 2^k! \sim 2^k \log_2 2^k = k2^k$ bits of information. This is much larger than our value, but in this case the table of all the $2^k!$ permutations is needed to decrypt a message although the $(l\,x+m)/2$ code needs the knowledge of the 4 numbers l_0, m_0, l_1 and m_1.

In the $(l\,x+m)/2$ cipher, the decoding process is as simple as the coding process. The coded message, written in base 2, is inspected from the LSB to the MSB (Most Significant Bit). The LSB of the coded message is directly the LSB of the original message. Supposing that the next bit is odd, the two iterations given by this seed are computed. If these two iterations are identical to the 2 corresponding last bits of the coded message the second bit is indeed odd, in the other case it is even. The process is repeated step by step for the k bits. As for the coding operation, it is possible to limit the computation on k bits where k is the number of bits of the message we want to decode. Moreover as for the decoding process, the encoding one is an operation implying k iterations.

The cryptanalysis of a ciphertext do not need the knowledge of the structure of the cipher. Nevertheless, these structures, if any, give insight on the complexity of the cipher and the amount of work necessary needed, in order to crack it. As already mentioned, because of limitation in the choice of l_i, m_i, there is 2^{4k-6} different permutations of words of k bits. First of all, an even (odd) cleartext gives an even (odd) ciphertext because the first iteration (LSB of the ciphertext) only depends on the parity of the cleartext. Then odd and even messages form two different groups. In the case $k=3$ bits the list of all possible ciphertext obtained for all cleartext and all different values of l_i and m_i shows that there is indeed $2^6 = 64$ different permutations. The inspection of this list shows that there is 8 different even permutations,

[2] This formula is valid as long as $\log k \gg 1$. We must consequently be careful when k is smaller than, say, 30.

l_0	m_0	l_1	m_1	0	2	4	6	1	3	5	7
1	0	1	1	0	6	4	2	7	5	3	1
1	0	1	5	0	6	4	2	3	1	7	5
1	0	3	3	0	6	4	2	3	5	7	1
1	0	3	7	0	6	4	2	7	1	3	5
3	0	1	3	0	6	4	2	5	7	1	3
3	0	1	7	0	6	4	2	1	3	5	7
3	0	3	1	0	6	4	2	5	3	1	7
3	0	3	5	0	6	4	2	1	7	5	3

Table 7: For 3 bits messages, the 8 odd coded messages related to the even coded message 0642 with the corresponding values of l_0, m_0, l_1 and m_1. In fact, in the case $k = 3$, all odd coded messages are connected to all even coded messages.

each related to all of the 8 different odd permutations. Table 7 gives one of these 8 even permutations with the related values of l_i, m_i and the 8 odd associated permutations. This structure in 8 odd x 8 even permutations is also observed in the case $k = 4$ and $k = 5$ bits (see figure 3). For $k = 3, 4$ and 5 the situation is resumed table 8.

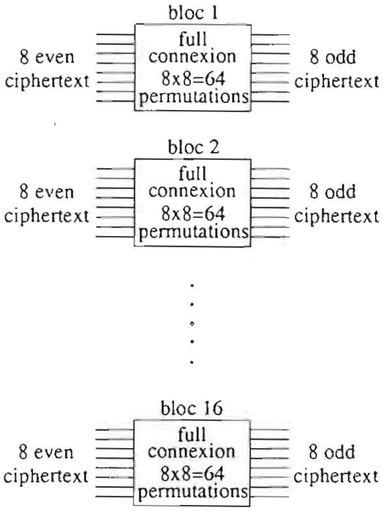

Figure 3: Case $k = 4$ bits. Correspondence between the 1024 odd and even permutations obtained with the 1024 possible different quadruplets l_0, m_0, l_1, m_1.

A point which must be precised is the strength of the cipher. A cracker, to recover the key of the code that is the l_0, l_1, m_0 and m_1, has to know some informations, that is some coded messages and the corresponding original ones. A couple cleartext-ciphertext of k bits gives $k-1$ bits of information (one bit is lost because the last bits of the couple is always the same). The 2^{4k-6} permutations corresponds to an uncertainty of $4k-6$ bits and the cracker needs, in the best case, 4 couples to recover the key (corresponding to the $4k-4$ bits of information). It can be proved that 4 pairs will be enough if and only if the two LSB of the 4 cleartexts are all

k	Number of permutations	Number of blocks with at each bloc
	2^{4k-6}	8 odd × 8 even permutations (see fig. 3)
3	64	1
4	1024	16
5	16384	256

Table 8:

differents that is correspond respectively to 00, 01, 10 and 11.

The cipher based on the $(l\,x + m)/2$ problem, here studied, is a first approach and can be complicated using the more general rules $(l\,x + m)/n$ or alternatively it is also possible to change the rules $(l\,x + m)/2$ while the message is coded, using two sets of number l_i, m_i.

7 Mixing properties

We return now to the classical $(3\,x + 1)/2$ problem. Consider a seed written in binary units. After a first iteration the last bit is lost. This means that the knowledge of the last bit of the seed do not give any information on the future iterations. But we have not lost the totality of the information attached to the two last bits of the seed since on these two bits of information only one has been lost.

Let us consider the four possibilities for the two last bits, 00, 01, 10, 11 and compute the following possibilities for the first iterate :

. a seed ending with 00 gives an iterate ending with 00 or 10,

. a seed ending with 01 gives an iterate ending with 00 or 10,

. a seed ending with 10 gives an iterate ending with 01 or 11,

. a seed ending with 11 gives an iterate ending with 01 or 11.

We build the probability that a number ending with 00, 01, 10 or 11 (the lines of matrix (9)) gives an iterate ending with 00, 01, 10 or 11 (the columns of matrix (9)).

$$M = \begin{array}{c|cccc} & 00 & 01 & 10 & 11 \\ \hline 00 & 1/2 & & 1/2 & \\ 01 & 1/2 & & 1/2 & \\ 10 & & 1/2 & & 1/2 \\ 11 & & 1/2 & & 1/2 \end{array} \tag{9}$$

In M an equipartition of the four possibilities for the two last bits of the seed has been supposed. To see what information on the two last bits of the seed has been lost after two iterations we just square M and find

$$M^2 = \begin{vmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{vmatrix} \tag{10}$$

Then after two iterations the two last bits of the seed have been totally lost. The result is generalized with a total lost of information for the k last bits of the seed after k iterations. In [Feix et al. 1994] the result has been generalized to trifurcation problem.

An interesting interpretation of the matrices M, M^2, M^3 ... is obtained using information theory. To see it clearly we need to consider the mixing mechanism for the three last bits. We build M as in the preceding (2 bits) case. For example on line 101 of matrix (11) we put 1/2 in columns 000 and 100 and zero elsewhere. Indeed a number ending with 101 can be written $8a + 5$ and gives as first iterate $12a + 8$. Now if a is even the iterate ends with 000 and with 100 if a is odd.

The matrix M takes the form

$$M = \frac{1}{2} \begin{array}{c|cccccccc} & 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \\ \hline 000 & 1 & & & & 1 & & & \\ 001 & & & 1 & & & & 1 & \\ 010 & & 1 & & & & 1 & & \\ 011 & & 1 & & & & 1 & & \\ 100 & & & 1 & & & & 1 & \\ 101 & 1 & & & & 1 & & & \\ 110 & & & & 1 & & & & 1 \\ 111 & & & & 1 & & & & 1 \end{array} \tag{11}$$

Computing M^2 and M^3 we obtain

$$M^2 = \frac{1}{4} \begin{vmatrix} 1 & & 1 & & 1 & & 1 & \\ & 1 & & 1 & & 1 & & 1 \\ 1 & & 1 & & 1 & & 1 & \\ 1 & & 1 & & 1 & & 1 & \\ & 1 & & 1 & & 1 & & 1 \\ 1 & & 1 & & 1 & & 1 & \\ & 1 & & 1 & & 1 & & 1 \\ & 1 & & 1 & & 1 & & 1 \end{vmatrix} \tag{12}$$

$$M^3 = \frac{1}{8} \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{vmatrix} \tag{13}$$

We notice that irrespectively of the last bits of the first iterate we have two possibilities for the three bits of the seed (two 1 in all columns of M) ; after two iterates (M^2) we have four possibilities (four 1 in all columns) and after three iterates (M^3) we get the 8 possibilities). These three situations correspond to an uncertainty of one bit (after one iteration) 2 bits (after two iterations) and of 3 bits (after three iterations) *i.e.* a complete lost of the three last bits. We have consequently generalized the trivial result of lost of the last bit after one iteration.

8 The random walk game

In the random walk game, the bifurcation and the n-furcations will be considered from a statistical point of view. In fact, the random walk game will take the point of view of the statistical physicist which, briefly speaking,is studying a system at two very different scales. At the first one, the system is observed at its microscopic level where all the precise details of the trajectories of all the particles are followed at each time. Moreover, it is also possible to compute these trajectories if the Hamiltonian of the system is known. The second level, which is really the level considered by the statistical physicist, is the macroscopic one for which the behavior of a large number of particles is studied. At this level, the details of the trajectories are forgotten and they are described by global averaged parameters. For example, in a hard collision between two particles, the trajectories after the collision crucially depends on the precise value of the impact parameter. The statistical approach introduces the cross section concept which describes the probability of deflexion into a given angle. At this level, it is only possible to obtain macroscopic quantities as the distribution function and derived quantities as temperature, pressure, density,...

The microscopic level will be given by the precise sequence of odd and even of the bifurcation process the rules of which are given by (1) and more generally by the sequence of the n-furcation given by (2).

The bijection theorem is the fundamental step of the random process. As in the statistical point of view, we are not interested in the precise sequence of a given seed but in the global behavior of the ensemble of sequences given by a large set of seeds. Starting from this large set of seeds for which the last k bits are taken at random, a consequence of the bijection theorem is that, for the k first iterations the probability of the n possibilities of the n-furcation process

	l_0	l_1	l_2	l_3	l_4	m_0	m_1	m_2	m_3	m_4	m	σ	k	k_{limit}
bifurcation	1	3	–	–	–	0	1	–	–	–	.21	.79	30	142
trifurcation	1	2	7	–	–	0	1	-2	–	–	.20	.73	19	64
pentafurcation	1	2	3	6	2	0	4	-1	2	-3	.55	.42	13	34

Table 9: Coefficients l_i and m_i of equation (2) adopted in the random walk simulation for the bifurcation, the trifurcation and the pentafurcation.

will be equal.

We suppose that all seeds are very large, consequently we can neglect m_i in front of $l_i x_p$ at each step and introducing $u_p = \log_n x_p$ we will have the same probability for the n choices, that is, using equation (7) (with $l_0 = 1$, $m_0 = 0$)

$$u_{p+1} = \log_n l_i + u_p - 1 \qquad \text{with a probability of} \quad 1/n \tag{14}$$

Consequently, the average decrease m of an ensemble of seeds from step p to step $p+1$ reads :

$$m =< u_{p+1} > - < u_p >= \frac{1}{n}\sum_{i=0}^{n} \log_n l_i - 1 \tag{15}$$

The same hypothesis of large numbers has also been used in equation (6) established to study the asymptotic behaviour of the sequence. And the standard deviation σ is given by

$$\sigma^2 =< (u_{p+1} - u_p)^2 - m^2 > \tag{16}$$

Starting from an ensemble of seeds, taken in a sharp distribution of numbers, the random process game will change this distribution into a Gaussian one, centered around $< x_0 > -nm$ with a variance $n\sigma^2$ after n iterations. Obviously, strictly speaking this model is valid on k iterations. Nevertheless, one can expect this game to be valid on a larger number of steps given at most by $k_{limit} = log_n \alpha / m$ iterations, for seeds taken around α and if numbers do not fall into cycles during these iterations.

The random walk process has been played on a distribution of 10^5 seeds taken at random in a range of 2×10^7 numbers centered on 10^9 for a bifurcation, a trifurcation and a pentafurcation. Table 9 summarizes the coefficients l_i and m_i of these n-furcations and the corresponding values of m, σ, k and k_{limit}.

Figures (4) give the evolution of the mean value and the standard deviation of this set of 10^5 numbers at each iteration in the $(3x+1)/2$ case. The continuous lines given by the random process game fits pretty well with the points on a much larger value than k. Nevertheless, the game is no longer valid for k_{limit} iterations for which a large set of numbers have already reach the cycle $2, 1$. Figure (5) gives the evolution of the distribution of the numbers. The change of the distribution into a Gaussian is clear.

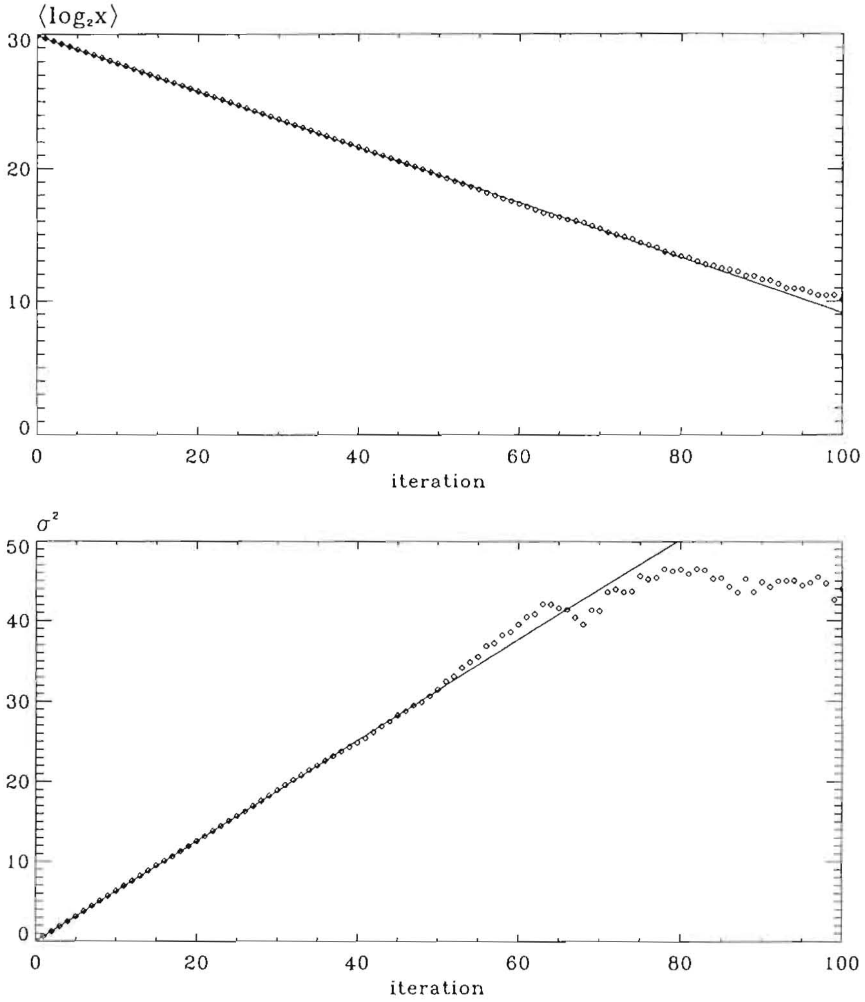

Figure 4: Bifurcation case : mean value of $\log_2$ of the population of numbers obtained after k iterations.

The random process is also valid for the trifurcation and pentafurcation (see figures (6), (7), (8) and (9)). But for these cases, the existence of cycles involving numbers much larger than 1, prevents the validity of the model up to k_{limit}. Consequently, the random process breaks more quickly.

9 Conclusion

What has been learned with the generalization to the $(l\,x+m)/n$ problem on one side and the stochastic approach on the other hand ? The generalization has shown that $(l\,x+m)/n$ can

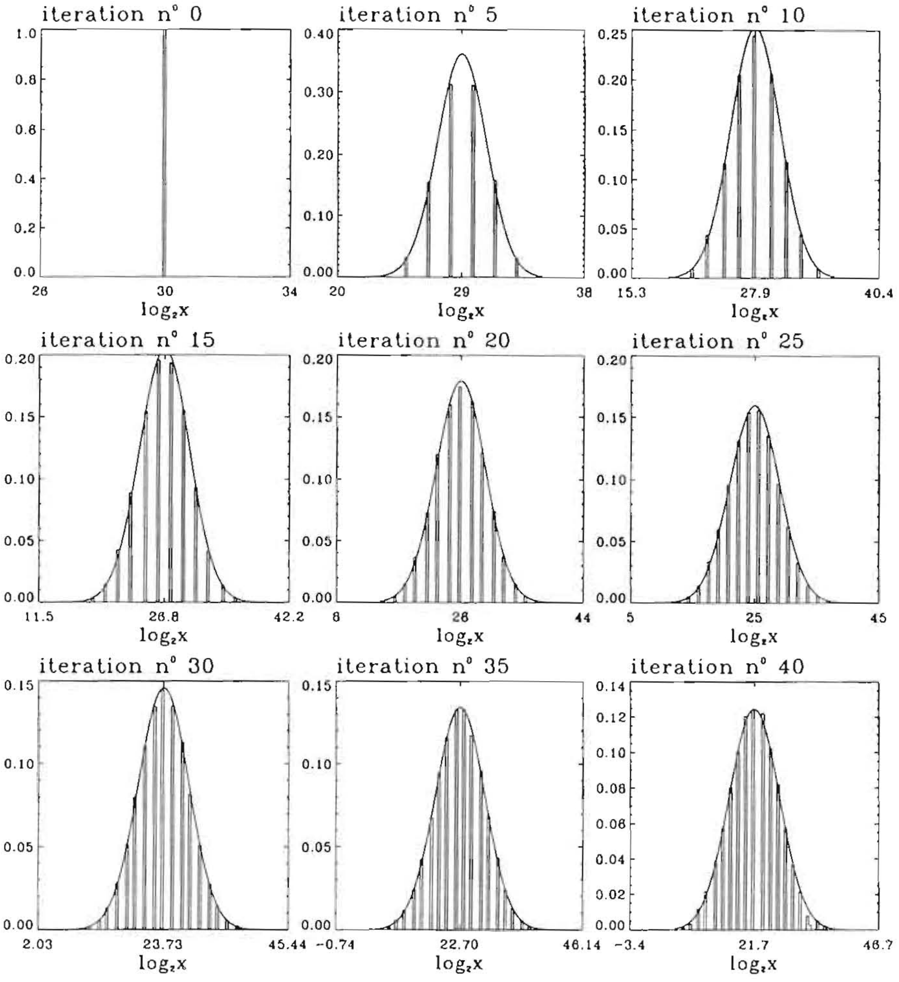

Figure 5: Bifurcation case : standard deviation of the population of numbers obtained after k iterations.

exhibit more than one cycle plus eventually a fixed point and that an arbitrary cycle can be built by a proper choice of m. Unfortunately no conclusion can be drawn for the $(3x+1)/2$ conjecture because of the difficulties to solve equation (7). But obviously numerical investigations are needed especially on these sequences where the successive iterates increase.

In converging sequences, after k iterations, the random model can be suspected and moreover iterates fall into cycles which generally involve not too large numbers. On the contrary, for diverging sequences, the k iterations brings the iterates to high values which may enforced the validity of the random model.

One of the result brought by the stochastic treatment is that the behavior of an ensemble

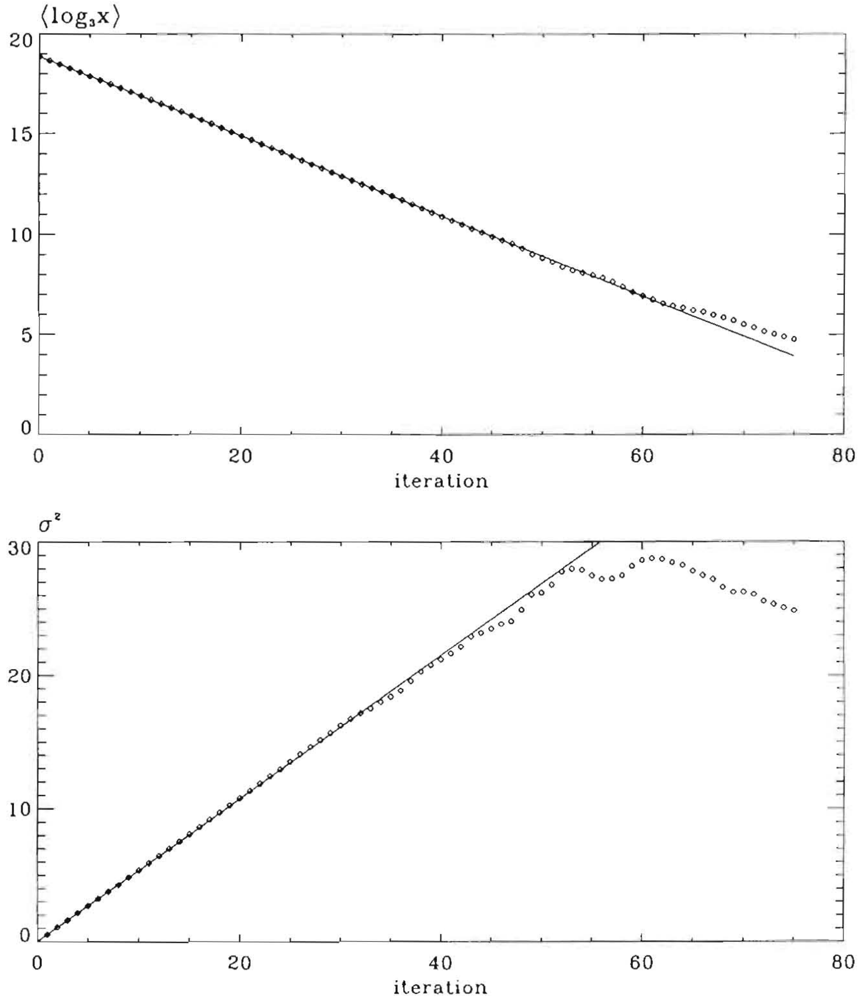

Figure 6: Trifurcation case : mean value of $\log_2$ of the population of numbers obtained after k iterations.

of seeds in the direct problem and of the antecedent in the inverse one is correctly described by probabilistic arguments. This is well in the spirit of statistical physics. On the other hand the randomness *a priori* limited to the significant digits of the seed may provide much longer "pseudo-random" sequences. Of course, this is a property claimed by all pseudo-random numbers generators. Do the sequences of parities provided by the n-furcaton is of better quality compared to the currently used generators ? This is certainly a difficult but practically important problem.

An application of the bijection is obtained turning to the study of limited sequences of k iterates. Starting from a k bit seed, we associate to the 2^k words, the 2^k sequences of iterates. All possible bijections between these numbers are obtained considering all possible values l_i,

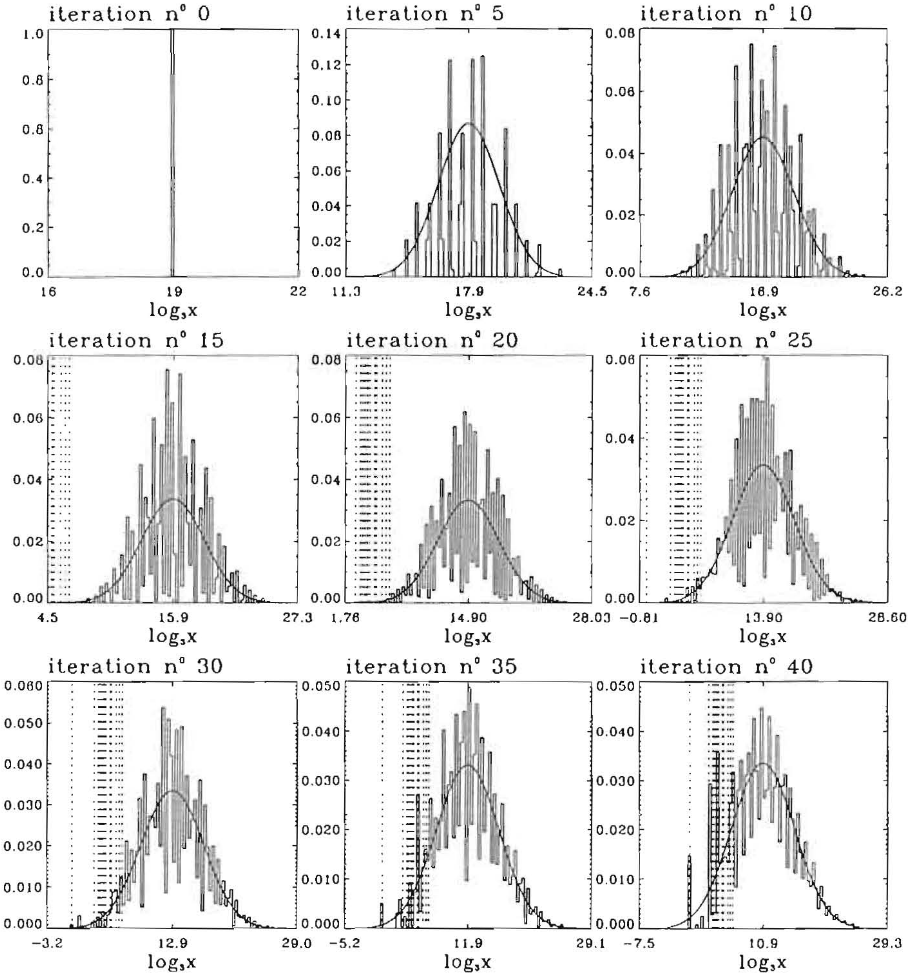

Figure 7: Trifurcation case : standard deviation of the population of numbers obtained after k iterations.

m_i. The building of a cipher is a possible application. Moreover, encryption and decryption are equally simple implying only the manipulation of k bits.

The stochastic approach does not say much about the conjecture except that it confirms that cycles, if they exist, will have very small basin of attraction. Being certainly extremely rare events, a statistical approach is certainly not very appropriate for the study of the conjecture.

On the other hand if we want to study the possibility of cycles for very large seed 10^{100} for example, we may have to give up the systematic trial of all the numbers and sample intelligently this huge set of cardinality 10^{100}. Monte-Carlo methods and random walk approximations may be useful.

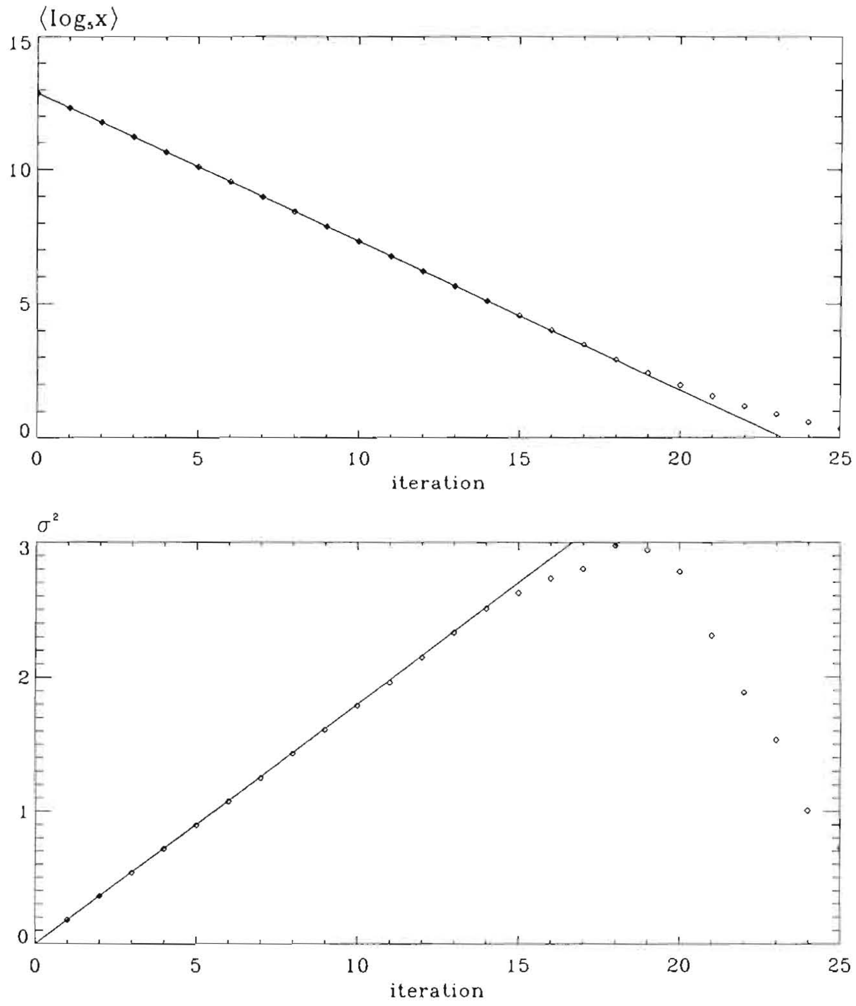

Figure 8: Pentafurcation case : mean value of $\log_2$ of the population of numbers obtained after k iterations.

References

[1] G.T. Leavens and M. Vermeulen, *$3x + 1$ search programs*, Computers and Mathematics with Application **24** (11), 79–99, 1992

[2] S. Fanelli *The solution of the 3x+1 problem*, Preprint, Roma2 University, Tor Vergata, Centro Vito Volterra, 382 , 1999

[3] J.H. Conwey, *Unpredictable iterations*, Proceedings of the number theory conference, University of Colorado, Boulder, 1972

[4] M. Chamberland, *A continuous extension of the 3x+1 Problem to the Real Line*. Dynamics of Continuous, Discrete and Impulsive Systems **2**, 495–509, 1996

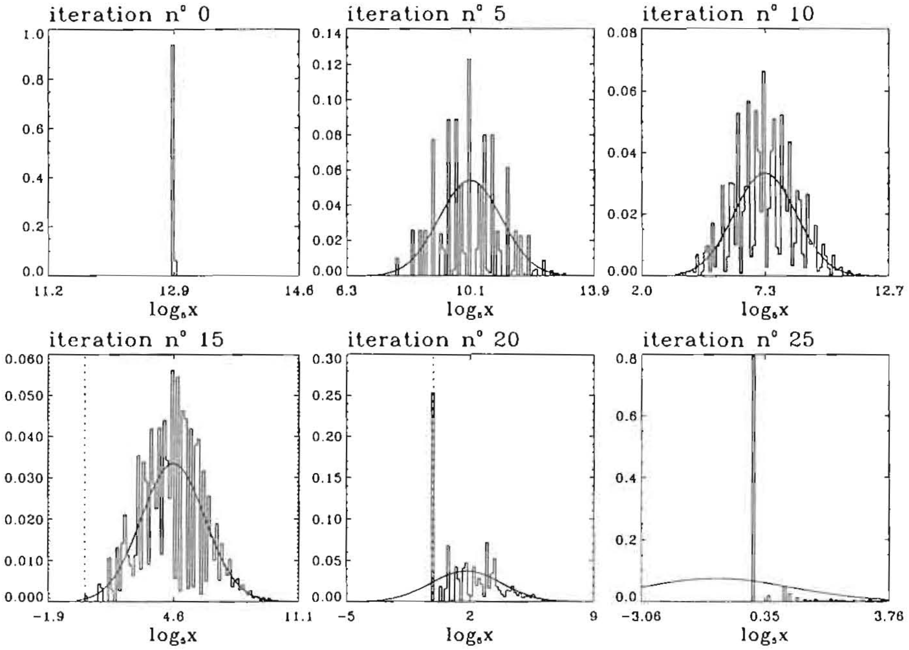

Figure 9: Pentafurcation case : standard deviation of the population of numbers obtained after k iterations.

[5] R. Terras, *A stopping time problem on the positive integers*, Acta Arithmetica **30**, 241–252, 1976

[6] J.C. Lagarias and A. Weiss, *The* $3x+1$ *problem : Two stochastic models*, Annals of Applied Probability **2**, 229–261, 1992

[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, D. Merlin Eds., World Scientific, 289–300, 1991

[8] M.R. Feix, A. Muriel and J.-L. Rouet, *Statistical properties of an iterated mapping*, Journal of Statistical Physic **76**, 725–741, 1994