SOME CURIOUS SEQUENCES CONSTRUCTED WITH THE GREEDY ALGORITHM

A. M. Odlyzko
Bell Laboratories
Murray Hill, New Jersey

R. P. Stanley
Massachusetts Institute of Technology
Cambridge, Massachusetts

January 1978

Given a fixed positive integer k, define a sequence $S(k)$ of positive integers $0 = a_0 < a_1 < \ldots$ by the conditions:

(i) $a_1 = k$

(ii) having chosen $a_0,\ a_1, \ldots, a_n$ $(n \geq 1)$, let a_{n+1} be the least integer such that $a_{n+1} > a_n$ and such that the sequence $a_1,\ a_1, \ldots, a_n, a_{n+1}$ contains no three terms (not necessarily consecutive) in an arithmetic progression.

For instance, for $k = 1, 2, 3, 4$ we obtain the sequence.

$$\begin{array}{ll} S(1): & 0,1,3,4,9,10,12,13,27,\ldots \\ S(2): & 0,2,3,5,9,11,12,14,17,\ldots \\ S(3): & 0,3,4,7,9,12,13,16,27,\ldots \\ S(4): & 0,4,5,7,11,12,16,23,26,\ldots \end{array}$$

One might hope that the sequence $S(k)$ (for a suitable value of k) has the least rate of growth among all sequences $0 = b_0 < b_1 < \ldots$ containing no three terms in an arithmetic progression. It remains an open problem to find this least possible rate of growth. A result of Roth [2] (see also [3]) implies that $\liminf_{n\to\infty} \frac{b_n}{n \log\log n} > 0$. In the other direction, it is known [1] that there exists a sequence with

$$b_n < n^{1+c(\log n)^{-1/2}} . \tag{1}$$

Unfortunately, it appears that the "greedy algorithm" which we have used to define a_n does not improve (1). This is certainly true for certain values of k (called *regular values*) for which the sequence a_n can be explicitly described. For other values of b (called *irregular*) it appears that a_n increases even more rapidly than in the regular case. We cannot prove this, although we can give a heuristic argument as to the rate of growth of a_n.

Despite the fact that the greedy algorithm does not seem useful for the original problem of minimizing the rate of growth of b_n, nevertheless the sequences $S(k)$ are of interest in their own right, largely because of the surprising dichotomy between regular and irregular values of k. There are also a host of related sequences, defined by rules similar to a_n, which appear to exhibit similar types of regular and irregular behavior. We will merely give one such example here, though it would be interesting to develop a general theory.

First we describe the sequences a_n when k is regular. The regular values of k are by definition integers of the form 2^m or $2 \cdot 3^m$.

Theorem 1 *Let $k = 3^m$, $m \geq 0$. Then a positive integer t is a member of the sequence if and only if the ternary expansion $t = \sum t_i 3^i$ of t satisfies the following properties:*

$$\text{(a)} \quad t_i = 0 \quad \text{or} \quad 1 \quad \text{if} \quad i \neq m$$

$$\text{(b)} \quad t_m = 0 \Rightarrow t_{m-1} = t_{m-2} = \cdots = t_0 = 0$$

$$\text{(c)} \quad t_m = 2 \Rightarrow \sum_{i=0}^{m-1} a_i > 0 \ .$$

Theorem 2 *Let $k = 2 \cdot 3^m$, $m \geq 0$. Then a positive integer t is a member of the sequence $S(k)$ if and only if the ternary expansion $t = \sum t_i 3^i$ of t satisfies the following properties:*

$$\text{(a)} \quad t_i = 0 \quad \text{or} \quad 1 \quad \text{if} \quad i \neq m, \quad m+1$$

$$\text{(b)} \quad t_m = 0 \quad \text{or} \quad 2$$

$$\text{(c)} \quad t_{m+1} = 2 \Rightarrow t_m = 0 \quad \text{and} \quad \sum_{i=0}^{m-1} t_i > 0 \ .$$

Theorems 1 and 2 an be proved by a routine though tedious case-by-case analysis.

Remark 1. The special case $k = 1$ is particularly simple to describe. The integer t is a member of the sequence $S(1)$ and only if the ternary expansion of t contains no 2's. It follows that a_n is obtained by writing n in binary and reading it in ternary. For instance, $1000 = 2^9 + 2^9 + 2^7 + 2^6 + 2^5 + 2^3$, so $a_{1000} = 3^9 + 3^8 + 3^7 + 3^6 + 3^5 + 3^3 = 29430$.

Remark 2. Let $\alpha = \frac{\log 3}{\log 2}$. For any regular value of k, it follows from Theorems 1 and 2 that

$$\frac{1}{2} = \liminf_{n \to \infty} \frac{a_n}{n^\alpha} < \limsup_{n \to \infty} \frac{a_n}{n^\alpha} = 1 \ . \tag{2}$$

For all values of k except 3^m and $2 \cdot 3^m$, empirical evidence suggests that the sequences $S(k)$ behave very erratically and have no simple description. However, they seem to have roughly the same rate of growth. The following table of selected values of a_n illustrates this phenomenon.

	regular				irregular			
$n \setminus k$	1	2	3	6	4	5	7	8
31	121	122	124	127	203	179	179	220
32	243	243	243	243	220	190	194	227
63	364	365	367	370	580	540	541	600
64	729	729	729	729	598	541	549	601
127	1093	1094	1096	1099	1771	1935	1597	1716
128	2187	2187	2187	2187	1792	2003	1638	1746
255	3280	3281	3283	3286	6086	6039	7219	5163
256	6561	6561	6561	6561	6107	6089	7295	5183

We have no idea of how to prove that the sequences $S(k)$ for k irregular indeed don't have a simple description and do have the same rate of growth.

Conceivably we have overlooked some irregular value that deserves to be classified as regular. It is also conceivable that for some irregular value the sequence $S(k)$ behaves erratically up to 10^{100} but then "straightens out". Assume, however, that $S(k)$ does behave "randomly" for irregular k.

We now want to define a sequence $p_0, p_1, \ldots$ of real numbers, where we think of p_n as the "probability" that n appears in $S(k)$ for irregular k. Now n will appear in $S(k)$ if and only if $n-i$ and $n-2i$ do not appear for $1 \leq i \leq n/2$. If these events are independent then we obtain

$$p_n = \prod_{i=1}^{[n/2]} (1 - p_{n-i} p_{n-2i}), \quad n \geq 2 \ . \tag{3}$$

For initial conditions set $p_0 = p_1 = .5$, which corresponds to a "random" start. A simple heuristic argument (which should not be difficult to make rigorous) suggests that

$$p_1 + \cdots + p_N \sim C\sqrt{n \log n}$$

for some constant c. Hence for a sequence $S(k)$ for k irregular we would expect $n \sim C\sqrt{a_n \log a_n}$, or

$$a_n \sim \frac{c' n^2}{\log n} \ . \tag{4}$$

This estimate agrees quite well with the numerical evidence. Note that (4) yields a faster rate of growth than (2).

The sequences $S(k)$ contain no solution to $x + y = 2z$ for distinct x, y, z. We can define sequences analogous to $S(k)$ for other equations. We merely describe one such sequence here, in order to suggest that there may be a general theory waiting to be developed. Let $0, 1, 2, 3, 4, 12, 13, 14, 15, 16, 48, \ldots$ be the sequence $C_0 < C_1 < C_2 < \ldots$ defined by the rules:

(i) $C_0 = 0$

(ii) having chosen $C_0, \ldots, C_n$, let C_{n+1} be the least integer such that $C_{n+1} > C_n$ and such that the sequence C_0, $C_1, \ldots, C_{n+1}$ contains no four *distinct* terms satisfying $x + y + z = 3w$.

It then can be shown that a positive integer t is in the above sequence if and only if the base 4 expansion $t = \sum t_i 4^i$ satisfies:

$$\text{(a)} \quad i \geq 1 \Rightarrow t_i \neq 2$$

$$\text{(b)} \quad t_i = 1 \Rightarrow t_{i-1} = t_{i-2} = \cdots = t_0 = 0 \ .$$

Another potentially interesting question has to do with the robustness of the regularity of the $S(k)$. That is, if $0 = b_0 < \cdots < b_r$ contain no three term progression, let $T(b_0, \ldots, b_r)$ denote the sequence $\{a_n\}$ constructed as follows:

(i) $a_n = b_n$ for $0 \leq n \leq r$

(ii) having chosen $a_0, \ldots, a_n$ $(n \geq r)$, let a_{n+1} be the least integer such that $a_{n+1} > a_n$ and such that $a_0, \ldots a_{n+1}$ contains no three term progression.

If $b_0, \ldots b_r$ are in some sense close to an initial segment of a regular $S(k)$, we might expect that $T(b_0, \ldots, b_r)$ will again be regular. We have tested some approximations to the regular sequence $S(1) = \{0, 1, 3, 4, 9, 10, 12, 13, 27, \ldots\}$. The following seemed to be regular (in the sense of membership in them being simply describable in terms of ternary expansions):

$$T(0,1,4),\ T(0,1,3,9),\ T(0,1,3,4,10),\ T(0,1,3,4,11),\ T(0,1,3,4,27),\ T(0,3,4,9,10)\ .$$

On the other hand, the following seemed to be irregular:

$$T(0,1,3,7),\ T(0,1,3,4,14),\ T(0,3,4,9,11),\ T(0,1,3,4,9,29),\ T(0,1,3,4,9,11)\ .$$

References

[1] L. Moser, On non-averaging sets of integers, *Canad. J. Math.* **5** (1953), 245–252.

[2] K. F. Roth, On Certain sets of integers. II, *J. London Math. Soc.* **29** (1954), 20–26.

[3] E. Szemerdi, On sets of integers containing no k elements in arithmetic progression, *Proc. 1974 Internat. Congress Math.*, Vancouver, Vol. 2, 503–505.