On Sequences of Integers with No 4, or No 5 Numbers in Arithmetical Progression

By S. S. Wagstaff, Jr.

Let k be an integer greater than one. We denote by $A^{(k)}(x)$ the maximal number of integers that can be selected from among $1, 2, \cdots, x$ to form a set containing no k consecutive terms of an arithmetical progression (no k in a.p.).

Let $a^{(k)}(x) = (A^{(k)}(x))/x$.

The function $A^{(k)}(x)$ satisfies a triangle inequality:

$$A^{(k)}(x+y) \leqq A^{(k)}(x) + A^{(k)}(y)\,, \tag{1}$$

and, from this, it follows [7] that

$$\tau^{(k)} = \lim_{x\to\infty} a^{(k)}(x)$$

exists.

A well-known conjecture states that, for every k, we have

$$\tau^{(k)} = 0\,. \tag{2}$$

This would imply that every strictly increasing sequence of integers with positive upper density contains arbitrarily long arithmetical progressions. Although the conjecture (2) has been the subject of considerable interest (see references) over the past 30 years, our knowledge in relation to it remains very limited. That $\tau^{(2)} = 0$ is trivial. In 1952, K. F. Roth [4], [5] proved that $\tau^{(3)} = 0$. The cases of (2) for $k > 3$ remain unsettled.

Although the conjecture (2) has proved very resistant, it is not hard in principle to obtain upper bounds for $\tau^{(k)}$ (for any particular k).

For it is easily seen that (1) implies $\tau^{(k)} = \inf_x a^{(k)}(x)$, so that we need only compute $A^{(k)}(x)$ for particular x to obtain upper bounds for $\tau^{(k)}$. The first such computations, in the case $k = 3$, were carried out by Erdös and Turân [1] in 1936 (but see also [2] for a correction of these results). More precise results were obtained by L. Moser [3] in 1953. The purpose of the present paper is to obtain upper bounds for $\tau^{(4)}$ and $\tau^{(5)}$ similar to those for $\tau^{(3)}$ before $\tau^{(3)} = 0$ was known.

I wish to thank Professor Roth for suggesting the problem to me, and for his help and encouragement in preparing this paper for publication. Thanks also goes to E. Jensen and R. Howell of M.I.T., whose clever ideas helped my computer program run efficiently.

Using the PDP-1 computer at M.I.T., I computed values of $A^{(3)}(x)$ and $A^{(4)}(x)$ for $1 \leqq x \leqq 52$, and of $A^{(5)}(x)$ for $1 \leqq x \leqq 31$. See Table I. The values of $A^{(3)}(x)$ were computed for comparison with the other results.

The best results obtained were

$$\tau^{(4)} \leqq 1/2 \qquad \text{and} \qquad \tau^{(5)} \leqq 9/13\,.$$

Received February 3, 1967.

TABLE I

x	$A^{(3)}(x)$	$a^{(3)}(x)$	$A^{(4)}(x)$	$a^{(4)}(x)$	$A^{(5)}(x)$	$a^{(5)}(x)$
1	1	1.000	1	1.000	1	1.000
2	2	1.000	2	1.000	2	1.000
3	2	.667	3	1.000	3	1.000
4	3	.750	3	.750	4	1.000
5	4	.800	4	.800	4	.800
6	4	.667	5	.833	5	.833
7	4	.571	5	.714	6	.857
8	4	.500	6	.750	7	.875
9	5	.556	7	.778	8	.889
10	5	.500	8	.800	8	.800
11	6	.545	8	.727	9	.818
12	6	.500	8	.667	10	.833
13	7	.538	9	.692	11	.846
14	8	.571	9	.643	12	.857
15	8	.533	10	.667	12	.800
16	8	.500	10	.625	13	.813
17	8	.471	11	.647	14	.824
18	8	.444	11	.612	15	.833
19	8	.421	12	.632	16	.842
20	9	.450	12	.600	16	.800
21	9	.429	13	.619	16	.762
22	9	.409	13	.591	16	.727
23	9	.391	14	.609	16	.696
24	10	.417	14	.583	17	.708
25	10	.400	15	.600	18	.720
26	11	.423	15	.577	18	.692
27	11	.407	16	.593	19	.704
28	11	.393	17	.607	20	.714
29	11	.379	17	.586	21	.724
30	12	.400	18	.600	21	.700
31	12	.387	18	.581	22	.710
32	13	.406	18	.563		
33	13	.394	19	.576		
34	13	.382	20	.588		
35	13	.371	20	.571		
36	14	.389	20	.556		
37	14	.378	21	.568		
38	14	.368	21	.553		
39	14	.359	21	.538		
40	15	.375	22	.550		
41	16	.390	22	.537		
42	16	.381	22	.524		
43	16	.372	23	.535		
44	16	.364	23	.523		
45	16	.356	24	.533		
46	16	.348	24	.522		
47	16	.340	24	.511		
48	16	.333	25	.521		
49	16	.327	25	.510		
50	16	.320	26	.520		
51	17	.333	26	.510		
52	17	.327	26	.500		

In contrast, the best result for $\tau^{(3)}$ was

$$\tau^{(3)} \leqq 8/25 .$$

Compare with Moser [3].

An analysis of the maximal deviations of $A^{(4)}(x)$ from $x/2$ and of $A^{(5)}(x)$ from $9x/13$ shows that

$$A^{(4)}(x) \leqq x/2 + 3 ,$$

with equality holding for $x = 10, 28, 30, 34$, and possibly other values of x congruent to these values (mod 52), and

$$A^{(5)}(x) \leqq 9x/13 + 37/13 ,$$

with equality holding only for $x = 19$, and possibly for 45, 71, 97, etc.

The following procedure was used in the computation. Suppose x and y are natural numbers with $y \leqq x$. We can decide whether $y \leqq A^{(k)}(x)$ or $y > A^{(k)}(x)$ by generating all possible selections of y distinct integers from $1, 2, \cdots, x$, and testing each selection for k numbers in a.p. If k numbers in a.p. are found in every selection, then $y > A^{(k)}(x)$. If, for some selection, there are no k in a.p., then $y \leqq A^{(k)}(x)$.

It suffices to consider selections

$$u_1, u_2, \cdots, u_y , \tag{3}$$

such that

$$1 \leqq u_1 < u_2 < \cdots < u_y \leqq x . \tag{4}$$

To test such a selection for k numbers in a.p. effectively, we check each of the numbers $u_k, u_{k+1}, \cdots, u_y$ for forming an a.p. with $k - 1$ of the numbers smaller than it. For each u_i, $k \leqq i \leqq y$, this check may be accomplished by considering the possible common differences

$$d = 1, 2, \cdots, [(u_i - 1)/(k - 1)] ,$$

and deciding whether the numbers

$$u_i - d, u_i - 2d, \cdots, u_i - (k - 1)d$$

are present among the u_j for $j < i$. If they are, then the numbers (3) contain the a.p. with highest term u_i, and common difference d.

If, for some integer m, it is known that $A^{(k)}(m) = n$, then $A^{(k)}(m + 1)$ can only be n or $n + 1$. Let $x = m + 1$, and $y = n + 1$. Then the above procedure tells us whether $n + 1 \leqq A^{(k)}(m + 1)$, or $n + 1 > A^{(k)}(m + 1)$, i.e., whether $n + 1 = A^{(k)}(m + 1)$, or $n = A^{(k)}(m + 1)$. In this manner $A^{(k)}(x)$ can be easily computed for consecutive natural numbers x.

Suppose x is chosen, $A^{(k)}(i)$ are known for $1 \leqq i < x$, and we want to compute $A^{(k)}(x)$. Then we set $y = 1 + A^{(k)}(x - 1)$, and apply the above procedure. The sequences (3) satisfying (4) must also satisfy

$$i \leqq u_i \leqq x - y + i, \quad \text{for } 1 \leqq i \leqq y . \tag{5}$$

For each natural number $n < y$, let $B^{(k)}(n)$ be the smallest number m such that

$n = A^{(k)}(m)$. Let $B^{(k)}(0) = 0$, and $B^{(k)}(y) = 1 + B^{(k)}(y - 1)$. Then we can improve the bounds on u_i in (5) to

$$B^{(k)}(i) \leqq u_i \leqq x - B^{(k)}(y - i)\,. \tag{6}$$

This narrowing of the bounds on u_i significantly reduces the number of sequences to be considered, and hence the computer running time. The computation of $A^{(4)}(x)$ for $x > 40$ would not have been feasible on the PDP-1 without these improved bounds.

In the actual program, the upper bound in (6) was implemented, but the lower bound proved too difficult to program, so I used another technique. The numbers $1, 2, \cdots, y$ were deposited in y consecutive registers. This was the first sequence considered. Each sequence was tested as follows. First the kth number was checked for being the highest of k numbers in a.p., then the $(k + 1)$th number was so checked, then the $(k + 2)$th, etc., to the yth number. This check was performed essentially as above. The checking process stops as soon as an a.p. is found. If one is found, let the jth number be that which was being tested, i.e., the highest number in the a.p. found. If no a.p. is found, the sequence is typed, and we let $j = x$. Now let i be the largest number such that $i \leqq j$, and the ith number is not $x - B^{(k)}(y - i)$. If no such number i exists, then all sequences which could possibly have no k in a.p. have been examined, so the program halts. Otherwise, let m be the ith number. Then the modification made to the current sequence to produce the next one is to deposit the numbers

$$m + 1, m + 2, \cdots, m + y - i + 1$$

into the ith, $(i + 1)$th, $\cdots$, yth registers. If the program types any sequences, we know that $y \leqq A^{(k)}(x)$, and if none are printed, $y > A^{(k)}(x)$.

TABLE II

$k = 4, x = 8, y = 6$

1	2	3	4	5	6
1	2	3	5	6	7
1	2	3	5	6	8*
1	2	3	5	7	8
1	2	3	6	7	8*
1	2	4	5	6	7
1	2	4	5	6	8
1	2	4	5	7	8*
1	2	4	6	7	8
1	3	4	5	6	7
1	3	4	5	7	8
1	3	4	6	7	8*
2	3	4	5	6	7
2	3	4	6	7	8

HALT.

* Indicates sequences typed.

For an example of the above procedure, see Table II. The jth number in each sequence is underlined. The starred sequences are those typed. Note that only 14 sequences need be considered, while there are $C_{8,6} = 28$ sequences that satisfy (4) and (5).

Sequences were typed on-line, and, for $x < 20$, running time was negligible compared to typing time. The cycle time of the PDP-1 is 5 microseconds. For large x, the program ran for several hours. To save time, the program was stopped after one or two sequences were typed, for certain large x. Also, for certain small x, when it was clear that a large number of sequences would be typed, the program was stopped after 20 or 30 of them were typed.

Cornell University
Ithaca, New York 14850

1. P. Erdös & P. Turán, "On some sequences of integers," *Proc. London Math. Soc.* (2), v 11, 1936, pp. 261–264.

2. A. Makowski, "Remark on a paper of Erdös and Turán," *J. London Math. Soc.*, v. 34, 1959, p. 480. MR **21** #5596.

3. L. Moser, "On non-averaging sets of integers," *Canad. J. Math.* v. 5, 1953, pp. 245–252. MR **14**, 726.

→ K. F. Roth, "On certain sets of integers," *J. London Math. Soc.*, v. 28, 1953, pp. 104–109. MR **14**, 536.

5. K. F. Roth, "Sur quelques ensembles d'entiers," *C. R. Acad. Sci. Paris*, v. 234, 1952, pp. 388–390. MR **13**, 724.

→ R. Salem & D. C. Spencer, "On sets of integers which contain no three terms in arithmetic progression," *Proc. Nat. Acad. Sci. U.S.A.*, v. 28, 1942, pp. 561–563. MR **4**, 131.

7. R. Salem & D. C. Spencer, "On sets which do not contain a given number of terms in arithmetic progression," *Nieuw Arch. Wiskunde* (2), v. 23, 1950, pp. 133–143. MR **11**, 417.