MATHEMATICS OF COMPUTATION, VOLUME 33, NUMBER 148
OCTOBER 1979, PAGES 1353–1359

Sets of Integers With No Long Arithmetic Progressions Generated by the Greedy Algorithm

By Joseph L. Gerver and L. Thomas Ramsey

Abstract. Let S_k be the set of positive integers containing no arithmetic progression of k terms, generated by the greedy algorithm. A heuristic formula, supported by computational evidence, is derived for the asymptotic density of S_k in the case where k is composite. This formula, with a couple of additional assumptions, is shown to imply that the greedy algorithm would not maximize $\Sigma_{n\in S} 1/n$ over all S with no arithmetic progression of k terms. Finally it is proved, without relying on any conjecture, that for all $\epsilon > 0$, the number of elements of S_k which are less than n is greater than $(1-\epsilon)\sqrt{2n}$ for sufficiently large n.

Szekeres [1] conjectured that if S is a set of positive integers such that $\Sigma_{n\in S} 1/n$ diverges, then S contains arithmetic progressions with an arbitrary (finite) number of terms. As Erdös has pointed out, this conjecture would imply that for each integer $k \geqslant 3$, there exists

$$A_k = \sup_{S\in \mathcal{S}_k} \sum_{n\in S} 1/n,$$

where $\mathcal{S}_k = \{S \subset Z^+ : S \text{ contains no arithmetic progression of } k \text{ terms}\}$.

Gerver [2] showed that for every integer $k \geqslant 3$, there exists a set $S_k \in \mathcal{S}_k$ such that

$$\sum_{n\in S_k} 1/n = [1 + o(1)]\, k \log k$$

for large k. In the case where k is prime, these S_k are generated recursively by the greedy algorithm; i.e., $n \in S_k$ if and only if $\{m \in S_k: m \leqslant n-1\} \cup \{n\}$ contains no arithmetic progression of k terms.

For the rest of this paper we let S_k be the set of positive integers with no arithmetic progression of k terms, generated by the greedy algorithm, regardless of whether k is prime or composite. We let $s_k(n)$ be the nth element of S_k, and let $\sigma_k(n)$ be the number of elements of S_k less than or equal to n.

We will investigate here the sets S_k in the case where k is composite. We derive heuristically a formula for the asymptotic density of such S_k, and show that this formula implies that for large k

$$\sum_{n\in S_k} 1/n = [1 + o(1)]\, k.$$

Received September 5, 1978.
AMS (MOS) subject classifications (1970). Primary 10L10, 10–04.

In other words, in the case where k is composite, we conjecture that the greedy algorithm does not maximize $\Sigma\ 1/n$. We then present some computational evidence in support of this formula. Finally, we prove, without relying on any conjecture, that for all $k \geq 3$, and all $\epsilon > 0$,

$$\sigma_k(n) > (1 - \epsilon)\sqrt{2n}$$

for sufficiently large n.

Now when k is prime, S_k has a great deal of structure. In fact, if you subtract one from each element of S_k, you end up with the set of all nonnegative integers which do not contain the digit $k - 1$ when written in base k. This follows easily from the Chinese remainder theorem. On the other hand, there is no obvious reason, when k is composite, that S_k should exhibit any particular structure.

To investigate this matter, we computed S_4 up to 2^{16} and S_6 up to 25000. In both cases, at first glance, the elements appear to be distributed randomly. For example, the elements of S_4 below 100 are

1, 2, 3, 5, 6, 8, 9, 10, 15, 16, 17, 19, 26, 27, 29, 30,

31, 34, 37, 49, 50, 51, 53, 54, 56, 57, 58, 63, 65, 66,

67, 80, 87, 88, 89, 91, 94, 99,

and between 20000 and 20100 are

20011, 20012, 20020, 20021, 20023, 20050, 20063, 20072, 20084,

while the elements of S_6 below 100 are

1, 2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 14, 15, 17, 18, 19,

20, 22, 23, 24, 25, 26, 33, 34, 35, 36, 37, 39, 43, 44,

45, 46, 47, 49, 50, 51, 52, 59, 60, 62, 63, 64, 65, 66,

68, 69, 71, 73, 77, 85, 87, 88, 89, 90, 91, 93, 96, 97,

98, 99,

and between 20000 and 20100 are

20010, 20011, 20017, 20025, 20028, 20034, 20038, 20052,

20058, 20060, 20061, 20069, 20079, 20080, 20082, 20085,

20093, 20095, 20098.

We confirmed this initial impression by subjecting S_4 to a number of tests for randomness. For example, let X_i be the number of elements in S_4 between 60000 + $50(i - 1) + 1$ and $60000 + 50i$ inclusive for $1 \leq i \leq 100$. If S_4 is pseudorandom, X_i should have approximately a Poisson distribution. Below we compare the number of times that $X_i = r$ with the probability that $X_i = r$ assuming a Poisson distribution with $\lambda = 2.5$ (the sample mean $\bar{X}$ is 2.49).

r	0	1	2	3	4	5	6	7	$\geqslant 8$
$N(X_i = r)$	10	20	26	19	12	8	3	2	0
$100P(X_i = r)$	8.2	20.5	25.7	21.4	13.3	6.7	2.8	1.0	0.4

This result is in sharp contrast to the case of S_k where k is prime. In that case X_i would generally have a bimodal distribution whose shape would be quite sensitive to our arbitrary choice of the parameters 60000, 50, and 100.

Likewise the distribution of gaps $s_4(n) - s_4(n-1)$ is relatively smooth for $s_4(n) < 2^{15}$, viz.:

gap	1	2	3	4	5	6	7	8	9	10 . . .	138	>138
frequency	220	196	154	181	138	121	129	103	104	95 . . .	1	0

On the other hand, if k is prime, $s_k(n) - s_k(n-1)$ must be equal to $(k^m - 1)/(k-1) + 1$ for some nonnegative integer m.

Finally, the elements of S_4 appear to be randomly distributed among the congruence classes mod m for $m \leqslant 8$. We list below the number of elements of S_4 less than 2^{11} which are congruent to c mod m.

$c \backslash m$	2	3	4	5	6	7	8
0	174	115	91	62	56	46	49
1	177	113	96	71	62	54	52
2		123	82	63	67	47	42
3			82	68	59	56	38
4				87	51	58	42
5					56	55	44
6						35	40
7							44

Again this is in sharp contrast to the case where k is prime, and elements of S_k are never divisible by k.

We now derive heuristically an asymptotic formula for $\sigma_k(n)$, on the assumption that the elements of S_k are suitably "random",

Let $f_k(n)$ be the characteristic function of S_k. It will be helpful in what follows to think of $f_k(n)$ as the probability that $n \in S_k$. Now consider an arithmetic progression of positive integers whose kth term is n. Such a progression must be of the form $\{n-(k-1)r, n-(k-2)r, \ldots, n-r, n\}$, where r is a positive integer less than or equal to $(n-1)/(k-1)$. The probability that the first $k-1$ terms of this progression are all in S_k is

$$\prod_{i=1}^{k-1} f_k(n-ir).$$

Now $n \in S_k$ if and only if there exists no positive integer $r \leqslant (n-1)/(k-1)$ such that $n-(k-1)r, n-(k-2)r, \ldots, n-r$ are all in S_k. Therefore, f_k satisfies the functional equation

$$f_k(n) = \prod_{r=1}^{[(n-1)/(k-1)]} \left[1 - \prod_{i=1}^{k-1} f_k(n-ir)\right].$$

Equivalently, if we allow operations with $-\infty$, we have

$$\log f_k(n) = \sum_{r=1}^{[(n-1)/(k-1)]} \log\left[1 - \prod_{i=1}^{k-1} f_k(n-ir)\right].$$

We conjecture that when k is composite, $f_k(n)$ can be approximated by a continuous function $\varphi_k(x)$ which satisfies a nearly identical equation, viz.

$$\log \varphi_k(x) = \int_0^{x/(k-1)} \log\left[1 - \prod_{i=1}^{k-1} \varphi_k(x-ir)\right] dr. \tag{1}$$

Specifically, we conjecture that $\sigma_k(n) \sim \int_0^n \varphi_k(x)\,dx$, where φ_k is the unique function satisfying (1). The justification for this conjecture is that, since the elements of S_k are distributed practically at random on a local scale, it should be possible to smooth out f_k and interpret it literally as a probability, without altering the large scale behavior of σ_k.

We can find an asymptotic formula for $\varphi_k(x)$ if we assume that there exists a real number p, with $-1 < p < 0$, such that for all $\epsilon > 0$, $x^{p-\epsilon} < \varphi_k(x) < x^{p+\epsilon}$ for sufficiently large x. Then

$$\log \varphi_k(x) \sim -\int_0^{x/(k-1)} \prod_{i=1}^{k-1} \varphi_k(x-ir)\,dr$$

$$= -\varphi_k(x)^{k-1} \int_0^{x/(k-1)} \prod_{i=1}^{k-1} \frac{\varphi_k(x-ir)}{\varphi_k(x)}\,dr$$

$$\sim -\varphi_k(x)^{k-1} \int_0^{x/(k-1)} \prod_{i=1}^{k-1} \left(\frac{x-ir}{x}\right)^p dr$$

$$= -x\varphi_k(x)^{k-1} \int_0^{1/(k-1)} \prod_{i=1}^{k-1} (1-it)^p\,dt.$$

It follows that $p = -1/(k-1)$ and

$$\varphi_k(x) \sim x^{-1/(k-1)}(\log x)^{1/(k-1)} \cdot \left[\int_0^{1/(k-1)} \prod_{i=1}^{k-1} (1-it)^{-1/(k-1)}\,dt\right]^{-1/(k-1)} (k-1)^{-1/(k-1)}.$$

Finally, we have, if our conjecture is true,

$$\sigma_k(n) \sim n^{(k-2)/(k-1)}(\log n)^{1/(k-1)} \tag{2}$$
$$\cdot\left[\int_0^{1/(k-1)} \prod_{i=1}^{k-1} (1-it)^{-1/(k-1)}\,dt\right]^{-1/(k-1)} (k-1)^{(k-2)/(k-1)}(k-2)^{-1}.$$

We now examine the behavior of $\varphi_k(x)$ as k tends to infinity. First, note that $\Pi_{i=1}^{k-1}(1-it)$ is a positive, monotonically decreasing function of t, and its derivative is monotonically increasing, over the interval $[0, 1/(k-1))$. At $t = 1/(k-1)$, the derivative of $\Pi_{i=1}^{k-1}(1-it)$ with respect to t is $-(k-2)!/(k-1)^{k-3}$. It follows that for $0 \leqslant t < 1/(k-1)$, we have

$$(k-2)!(k-1)^{-(k-3)}[(k-1)^{-1}-t] < \prod_{i=1}^{k-1}(1-it) < (k-1)[(k-1)^{-1}-t]$$

and, for large k,

$$[(k-1)^{-1}-t]^{-1/(k-1)} \lesssim \prod_{i=1}^{k-1}(1-it)^{-1/(k-1)} \lesssim e[(k-1)^{-1}-t]^{-1/(k-1)}.$$

Therefore, as k tends to infinity,

$$k \lesssim \int_0^{1/(k-1)} \prod_{i=1}^{k-1}(1-it)^{-1/(k-1)}\,dt \lesssim ek$$

and

$$\lim_{k\to\infty}\lim_{x\to\infty} \varphi_k(x)/x^{-1/(k-1)}(\log x)^{1/(k-1)} = 1.$$

This in itself tells us nothing about $\int_1^\infty \varphi_k(x)x^{-1}\,dx$. However, suppose that $\varphi_k(x)/x^{-1/(k-1)}(\log x)^{1/(k-1)}$ converges to 1 as k and x simultaneously tend to infinity; i.e., suppose that for all $\epsilon > 0$, there exists M such that if k and x are both greater than M, then $|1 - \varphi_k(x)/x^{-1/(k-1)}(\log x)^{1/(k-1)}| < \epsilon$. Then, since $0 < \varphi_k(x) \leqslant 1$ for all x and k, we would have

$$\lim_{k\to\infty} \frac{1}{k}\int_1^\infty \varphi_k(x)x^{-1}\,dx = 1.$$

Finally, if we are to evaluate $\Sigma_{n\in S_k} 1/n$, we must make an additional conjecture, namely that as n tends to infinity, $\sigma_k(n)^{-1}\int_0^n \varphi_k(x)\,dx$ converges to 1 uniformly for all composite k. This is reasonable, because if the elements of S_k were assigned at random, with the probability that $n \in S_k$ equal to $\varphi_k(n)$, we would have, with probability 1,

$$\sigma_k(n) = y + O(\sqrt{y \log y}),$$

where $y = \int_0^n \varphi_k(x)\,dx < n$. This conjecture, along with the others we have made, implies

$$\sum_{n\in S_k} 1/n = [1 + o(1)]\int_1^\infty \varphi_k(x)x^{-1}\,dx = [1 + o(1)]k.$$

We present below the values of $\sigma_4(n)$ and $\sigma_6(n)$ predicted by (2) and the actual computed values of these functions for n equal to all the powers of 2 from 2^6 to 2^{16}. We also include the sum of the reciprocals of the elements of S_4 (respectively S_6) up to n.

n	$1.195\, n^{2/3}(\log n)^{1/3}$	$\sigma_4(n)$	ratio	$\Sigma\, 1/n$	$1.121\, n^{4/5}(\log n)^{1/5}$	$\sigma_6(n)$	ratio	$\Sigma\, 1/n$
2^6	30.75	28	.911	3.175	41.53	42	1.011	3.927
2^7	51.38	46	.895	3.371	74.57	74	.992	4.263
2^8	85.28	74	.868	3.525	133.3	131	.983	4.578
2^9	140.8	125	.888	3.667	237.7	235	.992	4.859
2^{10}	231.5	211	.911	3.786	422.7	414	.979	5.102
2^{11}	379.3	351	.925	3.881	750.1	745	.993	5.325
2^{12}	619.8	574	.926	3.957	1329	1307	.983	5.517
2^{13}	1011	936	.926	4.019	2351	2318	.986	5.690
2^{14}	1644	1521	.925	4.070	4155	4070	.980	5.839
2^{15}	2670	2497	.935	4.112				
2^{16}	4332	4077	.941	4.145				

Extrapolating from the above figures, we can estimate $\Sigma_{n\in S_4} 1/n \approx 4.3$ and $\Sigma_{n\in S_6} 1/n \approx 6.9$. For comparison, $\Sigma_{n\in S_3} 1/n = 3.007$ and $\Sigma_{n\in S_5} 1/n = 7.866$. So S_4 may maximize $\Sigma_{n\in S} 1/n$ for $S \in \mathcal{S}_4$, but S_6 apparently does not maximize $\Sigma_{n\in S} 1/n$ for $S \in \mathcal{S}_6$ (since $S_5 \in \mathcal{S}_6$), nor presumably does S_k maximize $\Sigma_{n\in S} 1/n$, $S \in \mathcal{S}_k$, for any composite k greater than 6.

We now derive a lower bound for the asymptotic density of S_k.

THEOREM. *For all integers $k \geqslant 3$ and real $\epsilon > 0$, there exists an integer n_0 such that for all $n > n_0$,*

$$\sigma_k(n) > (1-\epsilon)\sqrt{2n}.$$

Proof. Suppose the contrary. Then there exist arbitrarily large n such that the number of positive integers less than n which are not in S_k is greater than $n - \sqrt{2n}$, which is greater than $(1-\delta)n$ for arbitrarily small δ. Now every positive integer which is not in S_k is the kth term of an arithmetic progression of which the first $k-1$ terms (and in particular the first two terms) are in S_k. But the kth term of an arithmetic progression is uniquely determined by the first two terms. Therefore, if $\sigma_k(n) \leqslant (1-\epsilon)\sqrt{2n}$ for some $k \geqslant 3$, and some sufficiently large n, then

$$(1-\delta)n < \binom{\sigma_k(n)}{2} = \tfrac{1}{2}[\sigma_k(n)^2 - \sigma_k(n)]$$

for δ arbitrarily close to zero, and

$$\sigma_k(n) > \tfrac{1}{2} + \sqrt{\tfrac{1}{4} + 2(1-\delta)n} > (1-\epsilon)\sqrt{2n}$$

for ϵ arbitrarily close to zero. This contradiction establishes the theorem.

Remark. We can generalize our conjecture about the asymptotic density of S_k as follows: Let $A = A_1^0$ be any set of k integers, and let ω be the largest element of A. Let $A_x^y = \{ax + y: a \in A\}$, and let $S_A = \{S \subset Z^+: \forall x \in Z^+, \forall y \in Z, A_x^y \not\subset S\}$. Thus, if $A = \{1, 2, \ldots, k\}$, then $S_A = S_k$, the set of all sets of positive integers containing no arithmetic progression of k terms. In general, the elements of S_A avoid containing a certain geometric pattern of integers. Let S_A be the element of S_A generated by the greedy algorithm, and let $\sigma_A(n)$ be the number of elements of S_A less than n. We conjecture that for "most" A,

$$\sigma_A(n) \sim n^{(k-2)/(k-1)}(\log n)^{1/(k-1)} \cdot \left(\int_0^{1/(k-1)} \prod_{i \in A} [1 - (\omega - i)t]^{-1/(k-1)} dt\right)^{-1/(k-1)} (k-1)^{(k-2)/(k-1)}(k-2)^{-1}.$$

It would be interesting to find examples of sets A for which the above is false other than arithmetic progressions with a prime number of terms.

Department of Mathematics
University of Hawaii
Honolulu, Hawaii 96822

1. P. ERDÖS & P. TURAN, "On certain sequences of integers," *J. London Math. Soc.*, v. 11, 1936, pp. 261–264.

2. J. GERVER, "The sum of the reciprocals of a set of integers with no arithmetic progression of k terms," *Proc. Amer. Math. Soc.*, v. 62, 1977, pp. 211–214.