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 $\sum_{n \in S} 1/n$ 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 $\sum_{n \in S} 1/n$ 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 $k \geq 3$, there exists

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

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

Gerver [2] showed that for every integer $k \geq 3$, there exists a set $S_k \in S_k$ 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 $\{m \in S_k: m \leq n - 1\} \cup \{n\}$ 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 $n$th 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

$$\sum_{n \in S_k} 1/n = [1 + o(1)]k.$$
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 \( e > 0 \),

\[
\sigma_k(n) > (1 - e)\sqrt{2n}
\]

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 \( 2^{5000} \). 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, 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).
INTEGERS WITH NO LONG ARITHMETIC PROGRESSIONS

<table>
<thead>
<tr>
<th>( r )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>( \geq 8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N(X_t = r) )</td>
<td>10</td>
<td>20</td>
<td>26</td>
<td>19</td>
<td>12</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>( 100P(X_t = r) )</td>
<td>8.2</td>
<td>20.5</td>
<td>25.7</td>
<td>21.4</td>
<td>13.3</td>
<td>6.7</td>
<td>2.8</td>
<td>1.0</td>
<td>0.4</td>
</tr>
</tbody>
</table>

This result is in sharp contrast to the case of \( S_k \) where \( k \) is prime. In that case \( X_t \) 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.:

<table>
<thead>
<tr>
<th>gap</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10 ...</th>
<th>138</th>
<th>&gt;138</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>220</td>
<td>196</td>
<td>154</td>
<td>181</td>
<td>138</td>
<td>121</td>
<td>129</td>
<td>103</td>
<td>104</td>
<td>95 ...</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

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 \( m \leq 8 \). We list below the number of elements of \( S_4 \) less than \( 2^{11} \) which are congruent to \( c \) mod \( m \).

<table>
<thead>
<tr>
<th>( m )</th>
<th>( 2 )</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>174</td>
<td>115</td>
<td>91</td>
<td>62</td>
<td>56</td>
<td>46</td>
<td>49</td>
</tr>
<tr>
<td>1</td>
<td>177</td>
<td>113</td>
<td>96</td>
<td>71</td>
<td>62</td>
<td>54</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>123</td>
<td>82</td>
<td>63</td>
<td>67</td>
<td>47</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>82</td>
<td>68</td>
<td>59</td>
<td>56</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>87</td>
<td>51</td>
<td>58</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>56</td>
<td>55</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>35</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 \( k \)th 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 \(r \leq (n - 1)/(k - 1)\) 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

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

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.

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

Specifically, we conjecture that \(\sigma_k(n) \sim \int_0^n \varphi_k(x) dx\), where \(\varphi_k\) 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 \(\sigma_k\).

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,
\begin{equation}
\sigma_k(n) \sim n^{(k-2)/(k-1)}(\log n)^{1/(k-1)}
\end{equation}
\begin{equation}
\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}.
\end{equation}

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 \( 0 < t < 1/(k - 1) \), we have
\begin{equation}
(k - 2)!(k - 1)^{-(k-3)}[(k - 1)^{-1} - t] < \prod_{i=1}^{k-1} (1 - it) < (k - 1)[(k - 1)^{-1} - t]
\end{equation}
and, for large \( k \),
\begin{equation}
[(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)}.
\end{equation}
Therefore, as \( k \) tends to infinity,
\begin{equation}
k \lesssim \int_0^{1/(k-1)} \prod_{i=1}^{k-1} (1 - it)^{-1/(k-1)} dt \lesssim ek
\end{equation}
and
\begin{equation}
\lim_{k \to \infty} \lim_{x \to \infty} \varphi_k(x)x^{-1/(k-1)}(\log x)^{1/(k-1)} = 1.
\end{equation}
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 \( 0 < \varphi_k(x) \leq 1 \) for all \( x \) and \( k \), we would have
\begin{equation}
\lim_{k \to \infty} \frac{1}{k} \int_1^\infty \varphi_k(x)x^{-1} dx = 1.
\end{equation}
Finally, if we are to evaluate \( \sum_{n \in S_k} 1/n \), 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,
\begin{equation}
\sigma_k(n) = y + O(\sqrt{y \log y}),
\end{equation}
where \( y = \int_0^n \varphi_k(x) dx < n \). This conjecture, along with the others we have made, implies
\begin{equation}
\sum_{n \in S_k} 1/n = [1 + o(1)] \int_1^\infty \varphi_k(x)x^{-1} dx = [1 + o(1)] k.
\end{equation}
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 \( n \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( 1.195 n^{2/3} (\log n)^{1/3} )</th>
<th>( \sigma_4(n) ) ratio</th>
<th>( \Sigma 1/n )</th>
<th>( 1.121 n^{4/5} (\log n)^{1/5} )</th>
<th>( \sigma_6(n) ) ratio</th>
<th>( \Sigma 1/n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 2^6 )</td>
<td>30.75</td>
<td>28</td>
<td>.911</td>
<td>3.175</td>
<td>41.53</td>
<td>42</td>
</tr>
<tr>
<td>( 2^7 )</td>
<td>51.38</td>
<td>46</td>
<td>.895</td>
<td>3.371</td>
<td>74.57</td>
<td>74</td>
</tr>
<tr>
<td>( 2^8 )</td>
<td>85.28</td>
<td>74</td>
<td>.868</td>
<td>3.525</td>
<td>133.3</td>
<td>131</td>
</tr>
<tr>
<td>( 2^9 )</td>
<td>140.8</td>
<td>125</td>
<td>.888</td>
<td>3.667</td>
<td>237.7</td>
<td>235</td>
</tr>
<tr>
<td>( 2^{10} )</td>
<td>231.5</td>
<td>211</td>
<td>.911</td>
<td>3.786</td>
<td>422.7</td>
<td>414</td>
</tr>
<tr>
<td>( 2^{11} )</td>
<td>379.3</td>
<td>351</td>
<td>.925</td>
<td>3.881</td>
<td>750.1</td>
<td>745</td>
</tr>
<tr>
<td>( 2^{12} )</td>
<td>619.8</td>
<td>574</td>
<td>.926</td>
<td>3.957</td>
<td>1329</td>
<td>1307</td>
</tr>
<tr>
<td>( 2^{13} )</td>
<td>1011</td>
<td>936</td>
<td>.926</td>
<td>4.019</td>
<td>2351</td>
<td>2318</td>
</tr>
<tr>
<td>( 2^{14} )</td>
<td>1644</td>
<td>1521</td>
<td>.925</td>
<td>4.070</td>
<td>4155</td>
<td>4070</td>
</tr>
<tr>
<td>( 2^{15} )</td>
<td>2670</td>
<td>2497</td>
<td>.935</td>
<td>4.112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 2^{16} )</td>
<td>4332</td>
<td>4077</td>
<td>.941</td>
<td>4.145</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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, \( \Sigma_{n \in S_5} 1/n = 3.007 \) and \( \Sigma_{n \in S_5} 1/n = 7.866 \). So \( S_4 \) may maximize \( \Sigma_{n \in S} 1/n \) for \( S \subsetneq S_4 \), but \( S_6 \) apparently does not maximize \( \Sigma_{n \in S} 1/n \) for \( S \subseteq S_6 \) (since \( S_5 \subseteq S_6 \)), nor presumably does \( S_k \) maximize \( \Sigma_{n \in S} 1/n \), \( S \subseteq S_k \), for any composite \( k \) greater than 6.

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

**Theorem.** For all integers \( k \geq 3 \) 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 \( \delta \). Now every positive integer which is not in \( S_k \) is the \( k \)th 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 \( k \)th term of an arithmetic progression is uniquely determined by the first two terms. Therefore, if \( \sigma_k(n) \leq (1 - \epsilon)\sqrt{2n} \) for some \( k \geq 3 \), and some sufficiently large \( n \), then

\[
(1 - \delta)n < \left( \frac{\sigma_k(n)}{2} \right)^2 = \frac{1}{2} [\sigma_k(n)^2 - \sigma_k(n)]
\]
for $\delta$ arbitrarily close to zero, and

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

for $\epsilon$ 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 $A = A^0_1$ be any set of $k$ integers, and let $\omega$ be the largest element of $A$. Let $A^x = \{ax + y : a \in A\}$, and let $S_A = \{S \subset Z^+ : \forall x \in Z^+, \forall y \in Z, A^x \setminus S\}$. 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” $A$,

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

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