On a question of additive number theory

by

P. Erdős (Saskatoon) and P. Scherk (Saskatoon)

1. Let \( A = \{a\}, B = \{b\}, \ldots \) denote sets of non-negative integers containing the number zero;

\[
\sum_{i=1}^{k} A_i = \sum_{i=1}^{k} a_i, \quad (a_{i+1} \leq a_i, \quad a_i \leq a_{i-1} = \ldots = a_1),
\]

Thus \( \sum A_i \) consists of all the numbers \( a_1 + a_2 + \ldots + a_k \) where each \( a_i \) lies in the corresponding \( A_i \). For a given integer \( n \) let \( [A] \) denote the number of positive elements of \( A \) up to and including \( n \). \( A \) denotes the set of the integers \( \leq n \) which do not belong to \( A \).

It is well known and easy to see that \( n \sum A_i \) implies \( |A_1| + |B| \leq n - 1 \). The corresponding problem for three or more sets does not lead to anything new. For then

\[
(1) \quad n \sum_{i=1}^{k} A_i
\]

implies \( n \sum A_i + A_k \) and thus \( |A_1| + |A_2| \leq n - 1; \quad 1 \leq k < \mu \leq k \). Adding these \( 1/2(k-1) \) inequalities we readily obtain

\[
(2) \quad \sum_{i=1}^{k} |A_i| \leq 1/2 k(n-1).
\]

That (2) cannot be improved can be seen by taking \( A_1 = A_2 = \ldots = A_k = \) set of integers between \( [k+1] \) and \( n-1 \) together with 0.

This question becomes more interesting if we require \( n \) to be the smallest number not in \( \sum A_i \). For \( k = 3 \) and \( n < 15 \) one can show (*) that

\[
|A_1| + |A_2| + |A_3| \leq n - 1.
\]

(*) Written communication from Professor H. B. Mann.
However this estimate becomes false if \( n \geq 15 \).

Surprisingly enough, (2) is asymptotically correct. Put
\[
\phi(n) = \max \sum_{i=1}^{k} |A_i|
\]
where \( A_1, \ldots, A_k \) range through those sets which satisfy (1) and
\[
[1, 2, \ldots, n-1] \subseteq \bigcup_{j=1}^{k} A_j.
\]

Thus \( \phi(n) = n-1 \). In the present paper we shall prove the existence of two positive constants \( \alpha = \alpha_k \) and \( \gamma = \gamma_k \) such that
\[
\frac{1}{\alpha k n} - \alpha n^{k-1} \gamma < \phi(n) < \frac{1}{\gamma k n} - \beta n^{k-1} \gamma
\]
for every \( k > 2 \). The first half of (5) will be proved in \( \S \) 2, the second in \( \S \) 3.

It would be of interest to obtain an explicit formula for \( \phi(n) \) if \( k > 2 \).

In particular it may be true that
\[
\phi(n) = \frac{1}{\alpha k n} + (\beta + o(1)) n^{k-1} \gamma
\]
for some positive constant \( \beta = \beta_k \). But we are unable to prove (6), still less to determine \( \beta \).

2. Let \( B_k = \{ b \} \) denote the set of all integers requiring only the digits \( 0 \) and \( 2^k \) in the number system with the basis \( 2^k \), \( k = 0, 1, \ldots, k-1 \). Thus every integer \( x \) permits a unique representation
\[
x = \sum_{i=1}^{k} b_i 2^i.
\]

Suppose that \( n \) has the representation
\[
n = \sum_{i=1}^{k} b_i 2^i, \quad b_i \in B_k.
\]

Obviously one of the \( b_i \)'s must be greater than \( \frac{1}{2} n \). Relabeling the \( B_i \)'s if necessary, we may assume
\[
b_\xi > \frac{1}{2} n.
\]

We obtain the set \( C_k \) by omitting the number \( b_\xi \) from \( B_k \). Thus
\[
n \in C_k + \sum_{i=1}^{k-1} B_i
\]
and every number lies in \( C_k + \sum_{i=1}^{k-1} B_i \) except the numbers
\[
b_\xi + \sum_{i=1}^{k-1} b_i.
\]

We now define
\[
C_k = B_k + [b_1^* + b_2^* + \ldots + b_{k-1}^* + b_\xi], \quad b_k \neq b_\xi; \quad k = 1, 2, \ldots, k-1.
\]

Let \( x \neq n \); cf. (1) and (2). If \( b_k \neq b_\xi^* \),
\[
x \in C_k + \sum_{i=1}^{k-1} B_i \subseteq C_k + \sum_{i=1}^{k-1} C_i = \sum_{i=1}^{k-1} C_i.
\]

If \( b_k = b_\xi^* \), there is an \( h \geq 1 \) such that
\[
x = \sum_{i=1}^{k} b_i^* + \sum_{i=1}^{h} b_i, \quad b_h \neq b_\xi^*.
\]

Hence
\[
x \in C_h + \sum_{i=1}^{k-h} B_i \subseteq C_h + \sum_{i=1}^{k-h} C_i \subseteq \sum_{i=1}^{k-h} C_i.
\]

Thus every number \( \neq n \) lies in \( \sum_{i=1}^{k-1} C_i \).

We next show
\[
n \in \sum_{i=1}^{k-1} C_i.
\]

Suppose
\[
n = \sum_{i=1}^{k-1} c_i, \quad c_i \in C_i.
\]

Then for each \( h > 0 \) either \( c_h = b_h \in B_k \) or
\[
c_h = \sum_{i=1}^{k} b_i + b_h, \quad b_h \neq b_\xi^*.
\]

Since the representation (2) of \( n \) was unique and since \( b_\xi \in C_k \), the first alternative cannot occur for all \( h > 0 \). On the other hand (3) shows that (7) cannot occur more than once. Thus (7) will hold for exactly one index \( h > 0 \). This leads to
\[
n = \sum_{i=1}^{k-1} b_i + \left( \sum_{i=1}^{k} b_i + b_h \right) + \sum_{i=1}^{k-h} b_i, \quad b_h \neq b_\xi^*.
\]
Comparing (8) with (2) we obtain

\[ \sum_{k} b_k = \sum_{k} b_k \neq b_k. \]

The representation of the number (9) being unique, we obtain in particular \( b_k = b_k \), a contradiction. This proves (5).

Define

\[ D_h = \sum_{0 \leq k \leq h} C_k, \quad h = 0, 1, \ldots, k-1 \]

and let \( A_h \) be the union of \( C_1 \) with the set of all the numbers

\[ n - D_h - \frac{1}{k} n \cup D_h. \]

Then

\[ n \notin \sum_{0 \leq h \leq k-1} A_h. \]

Thus \( n \) remains the only number not in \( \sum_{0 \leq h \leq k-1} A_h \).

It remains to estimate \( \sum_{0 \leq h \leq k-1} A_h \).

Let \( 2^{m+1} < n \leq 2^{m+2} \). Then

\[ B_0 = 2^{m+1} \leq 2^{m+2} < 2^{m+2}, \quad \lambda = 0, 1, \ldots, k-1. \]

Therefore

\[ C_0 = 2^{m+1}; \quad C_1 = 4^{m+1} \quad \text{if} \quad 0 < \lambda \leq k-1. \]

Thus

\[ \sum_{0 \leq h \leq k-1} C_0 = \prod_{0 \leq h \leq k-1} [C_0] < 4^{-m-1} n^{m-1} n^{m-1} \]

and

\[ \sum_{0 \leq h \leq k-1} C_0 = \prod_{0 \leq h \leq k-1} [C_0] < 4^{-m-1} n^{m-1} n^{m-1}, \quad h = 1, \ldots, k-1. \]

Hence

\[ [A_0] > \frac{1}{k} n - 4^{m-1} n^{m-1} \]

\[ [A_0] > \frac{1}{k} n - 4^{m-1} n^{m-1} \]

\[ [A_0] > \frac{1}{k} n - 4^{m-1} n^{m-1} \]

This proves the first part of our result with \( \alpha = (k+1)2^{m-1} \).
On a question of additive number theory

LEMMA 2.

\begin{equation}
\frac{n}{2} + \frac{\gamma}{k-2} \gamma \eta^{(\theta-\eta)k} \quad \text{if} \quad i = 1,
\end{equation}

\begin{equation}
\frac{n}{2} + \frac{\gamma}{k-2} \gamma \eta^{(\theta-\eta)k} \quad \text{if} \quad 1 < i \leq k.
\end{equation}

Proof. \(B_i\) contains either \(A_i\) or \(A_i\). Thus the first estimate follows immediately from (7) with \(\lambda = 2\).

By (9), \(n \neq A_i + B_i\). Hence \([B_i] < n - [A_i]\) and (7) also yields the second inequality.

LEMMA 3.

\begin{equation}
[B_i \cap A_j] < \frac{1}{k-2} \left(1 + \frac{k-\mu-3}{k-\mu+1}\right) \gamma \eta^{(\theta-\eta)k}, \quad \mu = 2, \ldots, k.
\end{equation}

Proof. If \(\lambda \neq \mu, \ A_i \subset B_i\). Thus \([B_i \cap A_j] = [B_i] - [A_i]\) and (11) is a corollary of Lemmas 1 and 2.

LEMMA 4.

\begin{equation}
[B_i \cup B_j \cup \cdots \cup B_k] < \frac{n}{2} + 3 \gamma \eta^{(\theta-\eta)k}.
\end{equation}

Proof. If \(x\) lies in \(B_1 \cup B_2 \cup \cdots \cup B_k\), \(x - A_i\) lies in \(B_i \cup \cdots \cup B_k\). Hence

\begin{equation}
[B_i \cup B_j \cup \cdots \cup B_k] = [A_i \cup A_j \cup \cdots \cup A_k] \leq [A_i] + [A_j] + \cdots + [A_k] + \sum_{j=1}^{k} [B_j \cap A_k],
\end{equation}

and

\begin{equation}
[B_i \cap A_j] < \frac{1}{k-2} \left(1 + \frac{k-\mu-4}{k-\mu+2}\right) \gamma \eta^{(\theta-\eta)k} \leq 2 \gamma \eta^{(\theta-\eta)k}, \quad \mu = 1, k.
\end{equation}

Now by (7) and (11)

\begin{equation}
[A_i] = n - [A_i] \leq n + \frac{2k-3}{k-2} \gamma \eta^{(\theta-\eta)k} \leq 2 \gamma \eta^{(\theta-\eta)k},
\end{equation}

\begin{equation}
[B_i \cap A_j] < \frac{2}{k-2} \gamma \eta^{(\theta-\eta)k} \leq 2 \gamma \eta^{(\theta-\eta)k},
\end{equation}

and

\begin{equation}
[B_i \cap A_j] < \frac{1}{k-2} \left(1 + \frac{2k-4}{2}\right) \gamma \eta^{(\theta-\eta)k} \leq 2 \gamma \eta^{(\theta-\eta)k},
\end{equation}

if \(2 \leq \mu \leq k-1\). Thus (13) yields (12).

Let \(C\) denote the set of those elements of \(\sum A_i\) which lie in none of the \(B_i\). Lemma 4 implies

\begin{equation}
[C] > \frac{n}{2} - 3 \gamma \eta^{(\theta-\eta)k}.
\end{equation}

For each \(c \in C\) we choose a canonical representation

\begin{equation}
c = \sum_{i=1}^{\lambda} a_i, \quad a_i \in \mathbb{A_i},
\end{equation}

in the following way: First \(a_1\) is chosen maximally among all the representations of \(c\). If \(a_1, \ldots, a_1\) have been fixed, \(a_{i+1}\) will be maximal among all the representations of \(c\) which use \(a_1 + a_2 + \cdots + a_i\).

LEMMA 6.

\begin{equation}
c' = \sum a_i' \in C, \quad a_i' \in \mathbb{A_i},
\end{equation}

be the canonical representation of \(c'\). Let

\begin{equation}
1 \leq \lambda_i < \lambda_i < \cdots < \lambda_i \leq k
\end{equation}

and suppose

\begin{equation}
\sum_{i=1}^{\lambda} a_{i} = \sum_{i=1}^{\lambda} a_{i}',
\end{equation}

Then

\begin{equation}
a_{i} = a_{i}', \quad \mu = 1, 2, \ldots, k.
\end{equation}

Proof. Substituting (17) in (15) we obtain another representation of \(c\). Since \(a_{i}\) was maximal, we have \(a_{i} = a_{i}').\) Similarly, (17) and (16) imply \(a_{i} = a_{i}').\) Thus \(a_{i} = a_{i}'),\) and (18) follows by induction.

LEMMA 7. Let \(1 \leq i \leq k\). The number of elements \(b_i\) occurring in the representation of elements \(c = a_i + b_i\) of \(C\) is less than

\begin{equation}
2 \frac{k-1}{k-2} \gamma \eta^{(\theta-\eta)k} \leq 4 \gamma \eta^{(\theta-\eta)k}.
\end{equation}

This remark is obvious. If \(b_i\) occurs in the representation of numbers of \(C\), \(b_i\) cannot occur in any \(A_i\) with \(i \neq i\). Hence the number of these \(b_i's\) is \(\leq [B_i \cap A_i]\). Choosing \(\mu = 1\) if \(i > 1\) and \(\mu\) arbitrarily if \(i = 1\), we obtain our estimate from (11).
We now construct a sequence of subsets

\[ C = D_k \supseteq D_{k-1} \supseteq D_{k-2} \supseteq \ldots \supseteq D_0 \]

of \( C \) in the following fashion: Let \( \delta > 0 \) be given. \( D_0 \) consists of those elements

\[ e^* = \sum_{i=1}^k a_i^* = b_i^* + a_i^* \quad (a_i^* \in A_i, \; i = 1, \ldots, k) \]

of \( D_{k-1} \) such that for every \( i \geq h \) there are not less than \( \delta 2^{-h} n^{3/8} \) elements of \( D_{k-1} \) of the form \( b_i^* + a_i^* \) (\( h = 1, \ldots, k \)).

**Lemma 6.**

\[ |D_k \cap D_i| < 4(5k-1)\delta n. \]

Proof. Let \( G_i \) denote the set of those numbers (19) of \( D_0 \) such that there are fewer than \( 4\delta n^{3/8} \) elements of \( D_k \) of the form \( b_i^* + a_i \) (\( i = 2, \ldots, k \)). Thus

\[ D_k \cap D_i = \bigcup_{i=1}^k G_i. \]

Let \( 1 \leq i \leq k \) be fixed. By Lemma 7 there are less than \( 4\gamma n^{3/8-\delta n} \) numbers \( b_i \) occurring in the representation of elements \( e = a_i^* \cdot b_i \) of \( C \). In particular there are fewer than \( 4\gamma n^{3/8-\delta n} \) numbers \( b_i^* \). Each of them occurs in fewer than \( \delta n^{3/8} \) elements of \( C_i \) and each \( e \cdot G_i \) has a representation \( e^* = b_i^* + a_i^* \). Hence

\[ |G_i| < 4\gamma n^{3/8-\delta n} \cdot \delta n^{3/8} = 4\gamma \delta n \]

and

\[ |D_k \cap D_i| < \sum_{i=1}^k |G_i| < 4(5k-1)\delta n. \]

**Lemma 9.**

\[ |D_k \cap D_{k+i}| < (k-h-1) |D_{k-1} \cap D_h|, \quad h = 1, 2, \ldots, k-2. \]

Proof. Let \( G_i \) denote the set of those elements (19) of \( D_k \cap D_{k+i} \) such that there are fewer than \( \delta 2^{-h} n^{3/8} \) elements of \( D_k \) of the form \( b_i^* + a_i \) (\( i = k+2, \ldots, k \)). Thus

\[ D_k \cap D_{k+i} = \bigcup_{i=1}^k G_i. \]

Let \( i \) be fixed; \( i+h-1 \leq i \leq k \). If \( b_i^* \) occurs in the representation of some \( e \cdot G_i \), there are not less than \( \delta 2^{-h} n^{3/8} \) elements of \( D_{k-1} \) of the form \( b_i^* + a_i \) while fewer than \( \delta 2^{-h} n^{3/8} \) of them belong to \( D_k \). Hence more than \( \delta 2^{-h} n^{3/8} \) of them will lie in \( D_{k-1} \cap D_h \). The number of these \( b_i^* \) is therefore less than

\[ |D_k \cap D_{k+i}| / (12^{-h} n^{3/8}). \]

Each of these \( b_i^* \) gives rise to less than \( \delta 2^{-h} n^{3/8} \) elements of \( C_i \). Conversely each element of \( C_i \) has a representation \( e^* = b_i^* + a_i^* \). Hence

\[ |G_i| < \delta 2^{-h} n^{3/8} (|D_{k-1} \cap D_h| / (12^{-h} n^{3/8})) = |D_{k-1} \cap D_h|. \]

This yields (21).

**Lemma 10.** Let \( 0 < h \leq k-1 \) be given,

\[ e^* = \sum_{i=1}^k a_i^* = b_i^* + a_i^* \cdot D_k. \]

Let \( i_1, \ldots, i_h \) be any \( h \)-tuple of distinct indices satisfying \( i_j > i_1 \geq 1, 2, \ldots, k \). Then there are at least

\[ g^{k-2(1/2)} n^{3/8} \]

numbers

\[ \left\{ e^* - \sum_{i=1}^h b_i^* + \sum_{i=1}^h a_i \cdot C \right\} \sum_{i=1}^k a_i^* \cdot C. \]

Proof. For \( h = 1 \) our assertion follows from the definition of \( D_1 \). Suppose it is proved for \( h-1 \) and assume (23). From the definition of \( D_1 \) there are at least \( \delta 2^{-h+1/2} n^{3/8} \) numbers \( a_i^* \) such that \( b_i^* + a_i^* \cdot D_{k-1} \). By induction assumption there are to each of them not less than

\[ g^{h-1-2(1/2)} n^{3/8} \]

numbers

\[ \left\{ b_i^* + a_i \cdot \sum_{i=1}^h a_i^* + \sum_{i=1}^h a_i \cdot e \right\} \sum_{i=1}^k a_i^* \cdot C. \]

Altogether we have at least

\[ (\delta 2^{-h+1/2}) n^{3/8} = g^{h-1-2(1/2)} n^{3/8} \]

numbers (23). By Lemma 6 they are mutually distinct.

**Lemma 11.** Let

\[ \delta = 2^{-12(1/2)} n^{3/8}. \]

Then \( D_{k-1} \) is empty.

Proof. The case \( h = k-1 \) of Lemma 10 yields: If there is a number

\[ e^* = \sum_{i=1}^k a_i^* \cdot D_{k-1} \quad \text{then there are at least} \quad g^{k-1-2(1/2)} n^{3/8} \]

elements \( a_i + b_i \). By Lemma 7 fewer than \( 4\gamma n^{3/8} \) numbers \( b_i \) can occur. Thus

\[ g^{k-1-2(1/2)} n^{3/8} < 4\gamma n^{3/8}. \]

This contradicts (24).
LEMMA 12. Let
\[ \gamma_n = \gamma = \frac{1}{2^{k-1}+1} \frac{1}{(k-1)!}. \]

Define \( \delta \) through (24). Then
\[ (1 - 8e(k-1)!\gamma_0) n^{1/\delta} > 6k \gamma \]
for every \( n \).

Proof. Since \( \frac{1}{4} \gamma \leq 1 \), we have
\[ 8e(k-1)!\gamma^2 + 6k \gamma < 8e(k-1)!2^{k-1} \gamma + 8(4 - e)(k-1)!2^{k-1} \gamma = 2^{k+3}(k-1)! \gamma < 1. \]
Hence
\[ (1 - 8e(k-1)!\gamma_0) n^{1/\delta} \geq 1 - 8e(k-1)!\gamma \delta > 6k \gamma. \]

We are now ready to show that the constant (25) satisfies (3).

Lemmas 8 and 9 imply by induction
\[ [D_1 \cap D_{k+1}] < 4 \cdot \frac{(k-1)!}{(k-h)(h-2)!} \gamma^h, \quad h = 0, 1, \ldots, k-2. \]

Thus by Lemmas 5 and 11,
\[ \frac{1}{2} n - 3kn^{1/\delta - 1/2} < [C] = \sum_{e=0}^{k-1} [D_e \cap D_{k+1}] \]
\[ < 4(k-1)! \gamma^h \sum_{e=0}^{k-1} \frac{1}{(k-h-2)!} \]
\[ < 4e(k-1)! \gamma^h. \]
Hence
\[ (1 - 8e(k-1)!\gamma_0) n^{1/\delta} < 6k \gamma. \]

Thus Lemma 12 shows that our assumption (5) leads to a contradiction

If \( n \) is a given integer and if \( S \) and \( C = [c] \) are sets of non-negative integers, the set \( S - C \) consists of all the integers \( x \geq 0 \) such that \( x \in S \) for every \( c \) with \( x + c \leq n \).

Let \( h > 1 \),
\[ n \in S, \quad 0 \in A_k \quad (k = 1, 2, \ldots, h) \]
and let
\[ S - \sum_{i=1}^{h} A_i = \{0\} \quad (\text{thus } \sum_{i=1}^{h} A_i \subseteq S). \]