ON SEQUENCES WITH A PRESCRIBED NUMBER OF PAIRWISE SUMS

S. L. G. CHOI

ABSTRACT. In this paper we obtain an estimate, which is substantially best possible, for the number of distinct pairs of subsets \( \mathcal{A}, \mathcal{B} \) of \{1, \ldots, n\}, each consisting of \( k \) integers, so that the number of elements in \( \mathcal{A} + \mathcal{B} \) satisfies a prescribed upper bound.

1. Introduction. Let \( n \) be a positive integer \( \geq 2 \), \( k \) a positive integer not exceeding \( n \), and \( C \geq 2 \) a real number not exceeding \( k \). We define \( f(n; k; C) \) to be the number of distinct pairs \( (\mathcal{A}, \mathcal{B}) \) of subsets \( \mathcal{A}, \mathcal{B} \) of \{1, \ldots, n\} where \( |\mathcal{A}| = |\mathcal{B}| = k \) and \( |\mathcal{A} + \mathcal{B}| < Ck \). We define \( f_1(n; k; C) \) to be the number of subsets \( \mathcal{A} \) of \{1, \ldots, n\} where \( |\mathcal{A}| = k \) and \( |\mathcal{A} + \mathcal{A}| < Ck \).

Clearly we have \( f_1(n; k; C) \leq f(n; k; C) \) always. By letting \( k - \lfloor C - 2 \rfloor \) of the integers in \( \mathcal{A} \) to be the consecutive members of an arithmetic progression while allowing the remaining integers to be completely arbitrary, we see that \( f_1(n; k; C) \gtrsim n^{\lfloor C - 2 \rfloor} \), where the implied constant is absolute. In this note we establish an upper bound for \( f(n; k; C) \) which is substantially best possible.

Theorem. We have

\[
(1) \quad f(n; k; C) \leq k^8 k^3 C.
\]

It may be mentioned that whilst the exponent \( 3C \) of \( n \) in (1) is certainly not the most favourable attainable by our method, it is not clear how our method should be modified to achieve the best possible exponent.

2. Some lemmas and definitions. Throughout this section and the next we shall regard the integer \( n \) as fixed, \( k \leq n \), and \( k \geq C \geq 2 \).

Lemma 1. Suppose the integers \( x_1, \ldots, x_i \) and \( y_1, \ldots, y_j \) are given.
Then, with the exception of at most $ij^2$ choices for the integer $z$, the sums $z + y_q$ $(q = 1, \ldots, j)$ are all distinct from the sums $x_i + y_l$ $(1 \leq h \leq i; 1 \leq l \leq j)$.

**Proof.** There are at most $ij^2$ distinct integers of the form $x_i + y_l - y_q$ $(1 \leq h \leq i; 1 \leq l \leq j; 1 \leq q \leq j)$. Thus, unless $z$ coincides with one of these integers, $z + y_q$ $(1 \leq q \leq j)$ are clearly distinct from the integers $x_i + y_l$ $(1 \leq h \leq i; 1 \leq l \leq j)$.

Before stating the next lemma we introduce a number of definitions and notations.

Let $k$ be an integer $\geq 2$, and let $k^*$ be defined by

$$k^* = \lfloor k/2 \rfloor.$$ 

Suppose $(\mathcal{A}, \mathcal{B})$ is a pair of sets of integers where $\mathcal{A} = \{a_1 < \cdots < a_k\}$ and $\mathcal{B} = \{b_1 < \cdots < b_k\}$. We define the sets $\mathcal{A}_1, \mathcal{A}_2, \mathcal{B}_1, \mathcal{B}_2, \mathcal{A}_1', \mathcal{A}_2', \mathcal{B}_1', \mathcal{B}_2'$ as follows:

1. $\mathcal{A}_1 = \{a_1 < \cdots < a_{k^*}\}$
2. $\mathcal{A}_2 = \{a_{k^* + 1} < \cdots < a_k\}$
3. $\mathcal{B}_1 = \{b_1 < \cdots < b_{k^*}\}$
4. $\mathcal{B}_2 = \{b_{k^* + 1} < \cdots < b_k\}$
5. $\tilde{\mathcal{A}}_1$ is the subset of $\mathcal{A}_1$ consisting of those $a_i$ in $\mathcal{A}_1$ such that the sets \{a_i + b_j: b_j \geq a_i + (b_{k^* + 1} - a_{k^* + 1})\} and \{a_l + b_m: a_l > a_i; b_m > a_i + (b_{k^* + 1} - a_{k^* + 1})\} are disjoint.
6. $\tilde{\mathcal{A}}_2$ is the subset of $\mathcal{A}_2$ consisting of those $a_i$ in $\mathcal{A}_2$ such that the sets \{a_i + b_j: b_j \leq a_i + (b_{k^*} - a_{k^*})\} and \{a_l + b_m: a_l < a_i; b_m < a_i + (b_{k^*} - a_{k^*})\} are disjoint.
7. $\tilde{\mathcal{B}}_1$ is the subset of $\mathcal{B}_1$ consisting of those integers $b_i$ in $\mathcal{B}_1$ such that the sets \{b_i + a_j: a_j \geq b_i - (b_{k^* + 1} - a_{k^* + 1})\} and \{a_l + b_m: a_l \geq b_i - (b_{k^* + 1} - a_{k^* + 1})\} are disjoint.
8. $\tilde{\mathcal{B}}_2$ is the subset of $\mathcal{B}_2$ consisting of those integers $b_i$ in $\mathcal{B}_2$ such that the sets \{b_i + a_j: a_j \geq b_i - (b_{k^*} - a_{k^*})\} and \{a_l + b_m: a_l \leq b_i - (b_{k^*} - a_{k^*})\} are disjoint.

We further define $\mathcal{S}_1, \mathcal{S}_2, \mathcal{T}_1, \mathcal{T}_2$ by

$$\mathcal{S}_1 = \mathcal{S}_1(\mathcal{A}, \mathcal{B}) = \mathcal{A}_1 + \mathcal{B}_1,$$
\[
S_2 = S_2(\mathcal{A}, \mathcal{B}) = \mathcal{A}_2 + \mathcal{B}_2,
\]

\[
J_1 = (\mathcal{A}_1 + \tilde{\mathcal{B}}_2) \cup (\mathcal{B}_1 + \tilde{\mathcal{A}}_2),
\]

\[
J_2 = (\mathcal{A}_2 + \tilde{\mathcal{B}}_1) \cup (\mathcal{B}_2 + \tilde{\mathcal{A}}_1).
\]

Finally, for every nonnegative integer \( m \leq k + 1 \) we denote by \( \mathcal{C}_m = \mathcal{G}_m(k; C) \) the set consisting of all pairs \((\mathcal{A}, \mathcal{B})\) satisfying

(i) \( \mathcal{A} \) and \( \mathcal{B} \) are subsets of \( \{1, \ldots, n\} \) such that \( |\mathcal{A}| = |\mathcal{B}| = k \);
(ii) \( |\mathcal{A} + \mathcal{B}| < Ck \);
(iii) \( \max(|\mathcal{A}_1| + |\mathcal{B}_1|; |\tilde{\mathcal{A}}_2| + |\tilde{\mathcal{B}}_2|) = m \).

Lemma 2. We have

\[
|J_1| = (|\mathcal{B}_2| + |\tilde{\mathcal{A}}_2|)k^*,
\]

\[
|J_2| = (|\mathcal{B}_1| + |\tilde{\mathcal{A}}_1|)(k - k^*),
\]

and

\[
S_i \cap J_j = \emptyset \quad (i = 1, 2; j = 1, 2).
\]

Proof. It follows from the definition of \( \mathcal{B}_2 \) and \( \tilde{\mathcal{A}}_2 \) that whenever \( b, b^* \in \mathcal{B}_2 \) and \( a, a^* \in \tilde{\mathcal{A}}_2 \), where \( b \neq b^* \) and \( a \neq a^* \), then

\[
(b + \mathcal{A}_1) \cap (b^* + \tilde{\mathcal{A}}_1) = \emptyset, \quad (a + \mathcal{B}_1) \cap (a^* + \tilde{\mathcal{B}}_1) = \emptyset,
\]

\[
(b + \tilde{\mathcal{A}}_1) \cap (a + \mathcal{B}_1) = \emptyset.
\]

This yields (11). We obtain (12) similarly.

The cases \( i = j = 1 \) and \( i = j = 2 \) of (13) also follow immediately from the definitions of \( \mathcal{B}_2, \tilde{\mathcal{A}}_2 \) and \( \mathcal{B}_1, \tilde{\mathcal{A}}_1 \).

We shall now prove (13) for \( i = 2, j = 1 \) (the case \( i = 1, j = 2 \) is similar). And for this purpose we shall only show that \((\tilde{\mathcal{A}}_2 + \mathcal{B}_1) \cap S_2 = \emptyset\), the proof of \((\mathcal{B}_2 + \tilde{\mathcal{A}}_1) \cap S_2 = \emptyset\) being similar. Accordingly let \( a_u \in \tilde{\mathcal{A}}_2 \); we proceed to show that

\[
(a_u + \mathcal{B}_1) \cap S_2 = \emptyset.
\]

Now by definition \( a_u + \mathcal{B}_1 \) are disjoint from \( \{a_v + b_w; v < u; b_w < a_u + (b_k^* - a_k^*)\} \). On the other hand, the integers of the form \( a_v + b_w \), where \( v \geq u \) and \( b_w \in \mathcal{B}_2 \) and the integers of the form \( a_v + b_w \), where \( v \in \tilde{\mathcal{A}}_2 \) and \( b_w \geq a_u + (b_k^* - a_k^*) \), are all greater than the integers \( a_u + \mathcal{B}_1 \). Thus the proof of (14) is complete.
Corollary. Suppose \((A, B)\) belongs to \(\mathcal{G}_m(k; C)\). Then one of the inequalities must hold:

\[ |S_1| < k^*(C - m/3), \quad |S_2| < (k - k^*)(C - m/3). \]

Proof. It follows from (11) and (12) and the definition of \(\mathcal{G}_m\) that

\[ \max(|S_1|, |S_2|) \geq mk^*. \]

Now from (13) and the fact that \(S_1 \cap S_2 = \emptyset\) (since every element in \(S_2\) is > every element in \(S_1\)), we have

\[ |S_1| + |S_2| + \max(|S_1|, |S_2|) \leq |A| + |B| < Ck, \]

and on making use of (15) the assertion of the corollary follows.

Lemma 3. The number of distinct pairs \((A, B)\), where \(A = \{a_1 < \cdots < a_k\}\) and \(B = \{b_1 < \cdots < b_k\}\) are subsets of \(\{1, \ldots, n\}\) such that \(A_1\) and \(B_1\) are prescribed and \(|A_2| + |B_2| = m\), is at most \(n^{m4k+1}k^3(k+1)\). Similarly the number of distinct pairs \((A, B)\), such that \(A_2, B_2\) are prescribed and \(|A_1| + |B_1| = m\), is at most \(n^{m4k+1}k^3(k+1)\).

Proof. We shall only prove the first half of the lemma (the other half can be proved in almost identical fashion). Let \(c_1 \leq \cdots \leq c_2(k-k^*)\) be the rearrangement in monotone increasing order of the integers \(a_1, \ldots, a_k\) and \(b_1, \ldots, b_k\). Then \(c_{k^*+1}, \ldots, c_k\) and \(b_{k^*+1} - (b_{k^*} - a_{k^*}), \ldots, b_k - (b_{k^*} - a_{k^*})\). (In the case where two integers \(a_u\) and \(b_v - (b_{k^*} - a_{k^*})\) are equal we adopt the convention that \(a_u\) appears first in the sequence \(c_1 \leq \cdots \leq c_2(k-k^*)\).

For a fixed subsequence \(q_1 < \cdots < q_{k-k^*}\) of \(1, 2, \ldots, 2(k - k^*)\), and a fixed subsequence \(t_1 < \cdots < t_m\) of \(1, 2, \ldots, 2(k - k^*)\) so that \(c_{q_1} < \cdots < c_{q_{k-k^*}}\) constitute \(A_2\) and \(c_{t_1} \leq \cdots \leq c_{t_m}\) constitute the set \(\tilde{A}_2 \cup \tilde{B}_2\), we have, by Lemma 1, on noting the number of choices for \(c_{1}, c_{2}, \ldots, c_{2(k-k^*)}\) in turn, that the number of choices for \((\tilde{A}_2, \tilde{B}_2)\) is \(\leq n^{m(k^3)2(2k-k^*)}\). The number of ways of choosing \(q_1 < \cdots < q_{k-k^*}\) and \(t_1 < \cdots < t_m\) is at most

\[ \binom{2(k-k^*)}{k-k^*} \binom{2(k-k^*)}{m} \leq 4^{2(k-k^*)}, \]

so that the total number of choices for the pair \((\tilde{A}_2, \tilde{B}_2)\) is \(\leq n^{m4k+1}k^3(k+1)\) as required.

Lemma 4. Suppose \(C < 2^{1/4}\). Then

\[ f(n, k; C) < n^{3C}k^8k. \]
Proof. The lemma clearly holds when \( k \leq 3n/2 \). We shall prove (16) by induction on \( k \). Accordingly let \( k > 3C/2 \) and suppose that (16) has been established with \( k' \) in place of \( k \) for all \( k' \leq [k/2] + 1 \).

We note that \( |A + B| < Ck \) implies that \( (A, B) \) belongs to \( G_0(k, C) \), since otherwise the corollary to Lemma 2 shows that one of \( |S_1| < 2k^* - 1 \) and \( |S_2| < 2(k - k^*) - 1 \) must hold and this is clearly not possible. (Here we are using the trivial observation that for any set \( X \) of \( t \) integers, \( |X + X| > 2|X| - 1 \).

Thus, by the corollary to Lemma 2 again, we have

\[
|S_1| < Ck^*
\]

or

\[
|S_2| < C(k - k^*).
\]

The number of choices for \( (A_1, B_1) \), for which (17) holds is, by induction, at most \( n^3C(k/2)^8(k/2) \). The number of choices for \( (A, B) \) with \( (A_1, B_1) \) pre-
scribed is, by Lemma 3, \( \leq k^3(k+1)4^{k+1} \), since in this case, \( m = 0 \). Thus the total number of choices for \( (A, B) \) so that (17) holds is at most

\[
n^3C(k/2)^8(k/2)k^3(k+1)4^{k+1}.
\]

Similarly the estimate \( n^3C((k+1)/2)^8((k+1)/2)k^3(k+1)4^{k+1} \) holds for the number of choices for \( (A, B) \) so that (18) holds. Thus, altogether the number of choices for \( (A, B) \) is at most

\[
2n^3Ck^3(k+1)(k+1)4^{k+4}2^{-(4k+4)+(2k+2)} \leq n^3Ck^8k,
\]

and this completes the proof.

3. Proof of Theorem. For \( C < 2\sqrt{3} \) the Theorem is a consequence of Lemma 4. The proof will be by induction on \( C \). Accordingly we assume \( C \geq 2\sqrt{3} \) and that (1) has been established for every \( k \) with \( C' \) in place of \( C \) for all \( C' \leq C - 1/3 \). Furthermore, as the Theorem holds trivially for \( k \leq 3C/2 \) we shall assume that \( k > 3C/2 \) and (1) has been established for \( C \) and for all \( k' \leq [k/2] + 1 \) in place of \( k \).

We recall the definition (following Lemma 1 in §2) of \( G_m(k; C) \). Since \( m \leq k + 1 \) so that \( f(n; k; C) \leq \sum_{m=0}^{k+1} |G_m| \), it clearly suffices to show

\[
|G_m| \leq (k + 2)^{-1}k^{8k}n^{3C}; \quad 0 \leq m \leq k + 1.
\]

Accordingly suppose \( (A, B) \in G_m \). Then it follows from the corollary to Lemma 2 that either

\[
|S_1| < k^*(C - m/3)
\]
or

\[ |S_2| < (k - k^*)(C - m/3). \]

We estimate the number of \((\mathcal{A}, \mathcal{B})\) for which (20) holds. The number of choices for \((\mathcal{A}_1, \mathcal{B}_1)\) is, by induction, \(\leq (k/2)^{8(k/2)}n^{3(C-m/3)}\). (If \(m = 1\), we use the induction hypothesis on \(C\); and if \(m < 1\) we use the induction hypothesis on \(k\).) Corresponding to each choice of \((\mathcal{A}_1, \mathcal{B}_1)\) the number of choices for \((\mathcal{A}_2, \mathcal{B}_2)\) is, by Lemma 3, \(\leq n^m k^{3k+1}4^{k+1}\). We argue similarly if (21) holds. Thus the total number of elements in \(\mathcal{S}_m\) is

\[
\leq 2n^m k^{3(k+1)}4^{k+1}((k + 1)/2)^{8((k+1)/2)} \leq n^m k^{8k(k+1)-1},
\]

which yields (19). This completes the proof of the Theorem.

Concluding remarks. The function \(f(n; k; C)\), defined in \(\S\) 1, clearly also denotes the number of distinct pairs \((\mathcal{A}, \mathcal{B})\), where \(\mathcal{A}\) is a subset of \(k\) integers from a fixed set of \(n\) consecutive integers, and \(\mathcal{B}\) is a subset of \(k\) integers from another fixed set of \(n\) consecutive integers, so that \(|\mathcal{A} + \mathcal{B}| < Ck|\).

This observation enables us to obtain an upper bound, by means of the Theorem, for the function \(f_2(n; k; C)\) which denotes the number of subsets \(\mathcal{A}\) of \(\{1, \ldots, n\}\), where \(|\mathcal{A}| = k\) so that \(|\mathcal{A} - \mathcal{A}| < Ck|\).

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER, BRITISH COLUMBIA, CANADA