

ON SUM-FREE SUBSEQUENCES

BY

S. L. G. CHOI, J. KOMLÓS AND E. SZEMERÉDI

ABSTRACT. A subsequence of a sequence of n distinct integers is said to be sum-free if no integer in it is the sum of distinct integers in it. Let $f(n)$ denote the largest quantity so that every sequence of n distinct integers has a sum-free subsequence consisting of $f(n)$ integers. In this paper we strengthen previous results by Erdős, Choi and Cantor by proving

$$(n \log n / \log \log n)^{1/2} \ll f(n) \ll n(\log n)^{-1}.$$

1. Given a sequence of n distinct integers, a subsequence is said to be sum-free if no integer in it is the sum of distinct integers of the same subsequence. Let $f(n)$ denote the largest quantity so that every sequence of n distinct integers has a sum-free subsequence consisting of $f(n)$ integers. Some lower and upper estimates for $f(n)$ are known (see [1]–[4]). In this paper we strengthen these previous bounds by proving

THEOREM. *We have*

$$(1.1) \quad (n \log n / \log \log n)^{1/2} \ll f(n) \ll n(\log n)^{-1}.$$

2. In this section we establish the upper estimate in (1.1). First we prove the following Lemma which is perhaps of some independent interest.

LEMMA. *Suppose $a_1 < \cdots < a_{t_1}$ is a sequence so that the number of distinct sums $a_i + a_j$ is $\leq t_1^{2-c}$. Then the number of integers x so that $a_i + a_j = x$ has at least t_1^α solutions, where $\alpha < c$, is at least $2t + O(t^{1-(c-\alpha)/2})$.*

PROOF. Denote $t_1^{1-(c-\alpha)/2}$ by M . Let S_1 consist of the first M integers in the sequence and S_2 the last M integers in the sequence. Let S_3 denote the rest of the sequence. We shall prove that in $S_2 + S_3$, the number N of integers x so that $a_i + a_j = x$ has at least t_1^α solutions is at least $t_1 + O(M)$, and that a similar conclusion holds for $S_1 + S_3$. As the treatment of $S_1 + S_3$ and that of $S_2 + S_3$ are almost identical, it suffices to establish the assertion for $S_2 + S_3$. On noting that for each integer x in $S_2 + S_3$ there are at most $|S_2| = M$ solutions for

Received by the editors August 13, 1974.

AMS (MOS) subject classifications (1970). Primary 10L10.

Key words and phrases. Sum-free, subsequence, integers.

Copyright © 1975, American Mathematical Society

$x = a_i + a_j, a_i \in S_2, a_j \in S_3$, and recalling that the number of distinct integers in $S_2 + S_3$ is at most t_1^{2-c} , we obtain

$$\begin{aligned} N &\geq (|S_3||S_2| - t_1^{2-c}t_1^\alpha)/|S_2| = t_1 - 2M - t_1^{2-c+\alpha}M^{-1} \\ &= t_1 - 3M = t_1 - 3t_1^{1-(c-\alpha)/2}. \end{aligned}$$

This completes the proof of the Lemma.

We now proceed to establish the upper estimate in (1).

We put

$$(2.1) \quad t = [n((\log n)/3)^{-1}]$$

and determine s so that

$$(2.2) \quad t(s+1) \leq n < t(s+2).$$

We define a set A of n integers as follows:⁽¹⁾

For each $i = 0, 1, \dots, s$, let

$$(2.3) \quad A_i = 2^i[t, 2t).$$

Let A_{s+1} consist of any $n - t(s+1)$ integers distinct from those in A_0, \dots, A_s . Finally, put $A = A_0 \cup \dots \cup A_{s+1}$. Suppose B is a sum-free set from A . We shall establish the theorem by showing that

$$(2.4) \quad |B| \leq n(\log n)^{-1}.$$

For each $i = 0, \dots, s$, we denote by B_i the intersection of A_i and B . Let B_h be the first set with at least $t^{9/10} = t^{1/2+2/5}$ integers. For $i = h, h+1, \dots$ we define B_i^*, B'_{i+1} inductively as follows:

Let B_h^* be a subset of B_h consisting of at least $|B_h|/2$ integers of the same parity. (Note that the pairwise sums of integers in B_h^* clearly lie in A_{h+1} .) Let B'_{h+1} consist of those integers x so that $a_i + a_j = x$ has at least $t^{2/5}$ solutions in $a_i, a_j \in B_h^*$.

Suppose B_k^*, B'_{k+1} have been defined. We let B_{k+1}^* be a subset of $B'_{k+1} \cup B_{k+1}$ consisting of at least $(|B'_{k+1}| \cup |B_{k+1}|)/2$ integers of the same parity. Finally we let B'_{k+2} denote the set of integers x so that $a_i + a_j = x$ has at least $t^{2/5}$ solutions in $a_i, a_j \in B_{k+1}^*$.

We shall now deduce (2.4) on the assumption that

(1) For real numbers $\alpha > \beta$ and integer y , we use $y[\alpha, \beta]$ to denote the set of integers ym , where m runs through all integers in $[\alpha, \beta]$.

$$(2.5) \quad \mathcal{B}'_i \cap \mathcal{B}_i = \emptyset, \quad i = h+1, h+2, \dots,$$

which we shall prove later.

We shall apply the Lemma with $a_1 < \dots < a_{t_1}$ consisting of the integers in \mathcal{B}_h^* so that

$$t_1 = |\mathcal{B}_h^*| = |\mathcal{B}_h|/2 \geq t^{9/10}/2.$$

Therefore, since $\mathcal{B}_h^* \subset \mathcal{B}_h \subset 2^h[t, 2t)$ and since all the integers in \mathcal{B}_h^* are of the same parity, we may conclude that the number of distinct sums $a_i + a_j$ is $\leq t \leq (2|\mathcal{B}_h^*|)^{10/9} \leq |\mathcal{B}_h^*|^{2^{3/20}} = |\mathcal{B}_h^*|^{2^{-17/20}}$ provided t is sufficiently large which we may clearly assume. Next, again assuming t sufficiently large, we have $|\mathcal{B}_h^*|^{2^{9/20}} \geq t^{2/5}$. Thus we may apply the Lemma with $c = 17/20$ and $\alpha = 9/20$ to obtain

$$|\mathcal{B}'_{h+1}| \geq 2|\mathcal{B}_h^*| + O(|\mathcal{B}_h^*|^{1-\frac{1}{2}((17-9)/20)}) = |\mathcal{B}_h| + O(t^{4/5}).$$

Similarly, $|\mathcal{B}'_{h+2}| \geq 2|\mathcal{B}_{h+1}^*| + O(|\mathcal{B}_{h+1}^*|^{4/5})$, which, since $\mathcal{B}'_{h+1} \cup \mathcal{B}_{h+1}$ is a disjoint union on using (2.5), is $\geq |\mathcal{B}'_{h+1}| + |\mathcal{B}_{h+1}| + O(t^{4/5})$. In general, we have

$$|\mathcal{B}'_k| \geq |\mathcal{B}'_{k-1}| + |\mathcal{B}_{k-1}| + O(t^{4/5}).$$

Thus

$$t \geq |\mathcal{B}'_{s+1}| \geq \sum_{i=h}^{s-1} |\mathcal{B}_i| + O(t^{4/5}s).$$

Hence

$$|\mathcal{B}| = \sum_{i=0}^{s+1} |\mathcal{B}_i| \leq t^{9/10}s + O(t^{4/5}s) + 2t = 2t + o(t),$$

and in view of (2.1) and (2.2) this proves (2.4).

It remains to show (2.5). We shall actually establish the stronger result that for each $i = 1, \dots, s-h$, any 2^{s-i} distinct integers in \mathcal{B}'_{h+i} have distinct representations; i.e. if x_k ($k = 1, \dots, 2^{s-i}$) are any 2^{s-i} distinct integers in \mathcal{B}'_{h+i} , then there exist $\sum_{k=1}^{2^{s-i}} N_k$ distinct integers $x_{k,j}$ ($k = 1, \dots, 2^{s-i}; j = 1, \dots, N_k$) belonging to \mathcal{B} such that $x_k = x_{k,1} + \dots + x_{k,N_k}$.

For $i = 1$ this is clearly true since each integer in \mathcal{B}'_{h+1} is the sum of at least $t^{2/5}$ pairs of integers in \mathcal{B}_h^* . (Note that $t^{2/5} \geq (n \log^{-1} n)^{2/5} > n^{1/2} > 2^s$ in view of (2.1) and (2.2).) Now suppose the above result has been established for $i \leq k-1$. We shall prove that it holds for k also. So accordingly let x_1, \dots, x_Q , be $Q = 2^{s-k}$ distinct integers in \mathcal{B}'_{h+k} . Since each x_j ($j = 1, \dots, Q$) can be expressed as the sum of two integers in \mathcal{B}'_{h+k-1} in at least $t^{2/5}$ ways, we may thus have

$$\begin{aligned}
 x_1 &= y_1 + z_1 \\
 &\vdots \\
 x_Q &= y_Q + z_Q,
 \end{aligned}$$

where $y_1, z_1, \dots, y_Q, z_Q$ are $2Q = 2^{s-(k-1)}$ distinct integers in \mathcal{B}_{h+k-1}^* . Some of these may of course be in \mathcal{B}_{h+k-1} , but at any rate at most $2^{s-(k-1)}$ are in \mathcal{B}_{h+k-1}' . Now the induction hypothesis yields the result. This completes the proof of (2.5) and with it that of (2.4).

3. In this section we shall establish the lower estimate in (1.1). Suppose

$$(3.1) \quad a_1 < \dots < a_n$$

are any n positive integers. We shall proceed to select a sum-free subsequence with $\geq \sqrt{nw(n)}$ elements, where⁽²⁾

$$(3.2) \quad w = w(n) = (1/40)\sqrt{\log n (\log \log n)^{-1}}.$$

The semiopen intervals $B_s = [2^s, 2^{s+1})$ are called blocks, and the neighbourhood of the block B_s consists of all blocks B_j for which $|j - s| < \log \log n$.

On observing that any set, either consisting of integers all from the same block or consisting of integers each from a different nonadjacent block, forms a sum-free set, we may assume that we have k blocks, where

$$(3.3) \quad \sqrt{n}(2w)^{-1} \leq k \leq \sqrt{nw}$$

each consisting of at least $\sqrt{n}(2w)^{-1}$ and at most \sqrt{nw} of the numbers in (3.1). Put

$$(3.4) \quad l = \lfloor \sqrt{n}(2w \log \log n)^{-1} \rfloor; \quad t + 1 = \lfloor (\log n)/5 \rfloor.$$

We can choose l of these blocks say B_{i_1}, \dots, B_{i_l} such that

$$(3.5) \quad i_{j+1} - i_j \geq \log \log n, \quad j = 1, \dots, l-1.$$

By a theorem of Chvátal and Komlós [5, Theorem 4], in each block B_{i_j} there are at least $t + 1$ elements, $b_1 < \dots < b_{t+1}$ such that the t consecutive differences $b_2 - b_1, b_3 - b_2, \dots, b_{t+1} - b_t$ form an increasing or a decreasing sequence. Assume first that we have

$$(3.6) \quad M = \lfloor l/2 \rfloor$$

⁽²⁾ The functions $\log n, \log \log n$ will always be taken to the base 2.

blocks B_{i_j} ($j = j_1, \dots, j_M$) so that there exist in B_{i_j} $t + 1$ numbers $b_{j,1}, \dots, b_{j,t+1}$ such that the t pairs $b_{j,2}, b_{j,1}; b_{j,3}, b_{j,2}; \dots; b_{j,t+1}, b_{j,t}$ give rise to t differences $b_{j,2} - b_{j,1}; b_{j,3} - b_{j,2}; \dots; b_{j,t+1} - b_{j,t}$ forming an increasing sequence.

We shall consider these tM pairs of integers in these M blocks and prove

PROPOSITION P. *There exist $Q \geq tM/60$ pairs of integers contained in a number of blocks such that the neighbourhood of each block cannot contain the difference of a pair of integers in another block.*

We shall postpone the proof of Proposition P till later but shall first indicate how it leads to the existence of a sum-free subsequence of (3.1) with $\geq \sqrt{nw(n)}$ elements, where $w(n)$ satisfies (3.2).

We let c_1, \dots, c_Q be the greater numbers of the Q pairs of integers whose existence is asserted in Proposition P. We shall show that it is sum-free; we note that $Q \geq tM/60$ indeed implies $Q \geq w\sqrt{n}$ in view of (3.2), (3.4) and (3.6).

Accordingly let $e_1 < \dots < e_k$ be k of the integers c_1, \dots, c_Q , so that $e_k = e_{k-1} + \dots + e_1$. It is clear that e_k and e_{k-1} must belong to the same block. If $e_k = b_{j,i}$ and $d_{j,i} = b_{j,i} - b_{j,i-1}$ then

$$(3.7) \quad d_{j,i} \leq e_k - e_{k-1} \leq td_{j,i},$$

the right-hand inequality arising from the fact that $d_{j,i}, i = 1, \dots, t$, form an increasing sequence. On the other hand,

$$(3.8) \quad e_{k-2} \leq e_1 + \dots + e_{k-2} < (\log n)e_{k-2}/4,$$

since $t \leq (\log n)/5$ by (3.5), each block contains at most t integers of c_1, \dots, c_Q and (3.5) holds. Thus we obtain from (3.7), (3.8)

$$5e_{k-2}(\log n)^{-1} < d_{j,i} < (e_{k-2} \log n)/4,$$

and this means that $d_{j,i}$ lies in the neighbourhood of the block containing e_{k-2} . Since $d_{j,i} = b_{j,i} - b_{j,i-1}$ and $e_k = b_{j,i}$ we have the existence of a pair of integers, the greater of which is e_k , such that their difference is in the neighbourhood of the block containing e_{k-2} . As it is clear that the block containing e_k is different from the one containing e_{k-2} , we have obtained a contradiction to the assertion in Proposition P.

We remark that in the case where there are M blocks B_{i_j} ($j = j_1, \dots, j_M$) each with $t + 1$ integers giving rise to t consecutive differences forming a decreasing sequence, the proof regarding the existence of Q pairs satisfying Proposition P will be the same and the only modification occurs in the deduction from Proposition

P where we now take the smaller integers of the Q pairs and obtain in place of (3.7) the inequality

$$d_{j,i+1} \leq e_k - e_{k-1} \leq td_{j,i+1}, \quad \text{where } e_{k-1} = b_{j,i}.$$

It remains to prove Proposition P. We first introduce the following definition of a bad block.

A block is said to be bad if one of the following holds:

(i) it contains $\geq t/10$ pairs of integers and its neighbourhood contains their differences;

(ii) its neighbourhood contains the differences of $11t/10$ pairs of integers.

It is clear that we may assume there are $< M/2$ bad blocks, since if (i) holds for $\geq [M/4]$ blocks we already have Proposition P by simply taking the $[M/4]t$ pairs in these blocks, and if (ii) holds for $[M/4]$ blocks then we obtain $(11/10)[M/4]$ pairs of integers whose differences lie in their neighbourhood, so that after discarding those pairs which actually lie in these blocks we obtain $\geq (t/10)[M/4] > Mt/60$ pairs of integers satisfying the assertion in the proposition.

Hence we may take

$$(3.9) \quad M' = [M/2] + 1$$

blocks which are not bad. We may further assume that there are at most $M't/20$ pairs in these M' blocks with differences lying outside the neighbourhoods of these M' blocks. Thus we may assume there are M' blocks which are not bad and contain altogether at least $19M't/20$ pairs of integers whose differences actually lie in the neighbourhoods of these M' blocks. On this assumption we shall obtain Proposition P by proving that for each natural number $j \leq (M'/13) + 1$ there exist $[jt/2]$ pairs of integers and j blocks B_{p_1}, \dots, B_{p_j} such that all these pairs lie outside B_{p_1}, \dots, B_{p_j} but their differences lie in the neighbourhoods of these blocks. We shall prove this by induction.

Since there are $\geq 19M't/20$ pairs of integers whose differences lie in the neighbourhoods of M' blocks, there exists one block, say B_{p_1} , whose neighbourhood contains the differences of $\geq 19t/20$ pairs of integers. Since B_{p_1} is not bad, it can contain at most $t/10$ of these pairs. Thus we may choose $[t/2]$ pairs of integers not lying in B_{p_1} but with differences lying in the neighbourhood of B_{p_1} . This establishes the assertion for $j = 1$. Suppose now $1 \leq j \leq M'/13$ and that the assertion has been proved for j . We shall proceed to establish it for $j + 1$. Accordingly let there exist j blocks B_{p_1}, \dots, B_{p_j} and $[jt/2]$ pairs of integers lying outside these blocks but with differences lying in the neighbourhoods of these blocks. Denote the set consisting of the $[jt/2]$ pairs by S_j . We let $B_1^{(j)}, \dots, B_{S_j}^{(j)}$ include all the blocks each containing at least $t/10$ elements of S_j . There can

be at most $(11t/10)6j$ pairs with differences lying in the neighbourhoods of $B_1^{(j)}, \dots, B_{5j}^{(j)}, B_{p_1}, \dots, B_{p_j}$, and at most jt pairs lying in the blocks B_{p_1}, \dots, B_{p_j} . Thus there are $\geq 19M't/20 - 76jt/10$ pairs lying outside B_{p_1}, \dots, B_{p_j} and whose differences lie outside the neighbourhoods of $B_1^{(j)}, \dots, B_{5j}^{(j)}, B_{p_1}, \dots, B_{p_j}$. Hence there is one block say $B_{p_{j+1}}$ which contains the differences of at least $(19M't/20 - 76jt/60)(M' - 6j)^{-1}$ pairs. In view of $j \leq M'/13$, this number is $\geq t/2 + t/5$. These pairs are clearly distinct from those in S_j . Furthermore $B_{p_{j+1}}$ contains at most $t/10$ of the pairs just chosen (since $B_{p_{j+1}}$ is not a bad block) and at most $t/10$ pairs from S_j (since $B_{p_{j+1}}$ is different from $B_1^{(j)}, \dots, B_{5j}^{(j)}$). Hence we may now assert that there are $\geq [jt/2] + (t/2 + t/5) - t/5 \geq [(j+1)t/2]$ pairs of integers lying outside $B_{p_1}, \dots, B_{p_{j+1}}$ and whose differences lie in their neighbourhoods. This concludes the induction argument and with it the proof of Proposition P. This completes the proof of the lower bound in (1.1).

We remark that it is possible to iterate the process for the lower bound and obtain $f(n) > \sqrt{nw^2}/2$ and by iterating k times we may obtain $f(n) > \sqrt{nw^k}/k!$ which grows up as high as $n^{1/2+\epsilon}$. As the proof of this involves rather messy details we have refrained from carrying this out in this paper. It is conceivable that $f(n) > n^{1-\epsilon}$ for every ϵ and $n \geq n_0(\epsilon)$.

REFERENCES

1. P. Erdős, *Extremal problems in number theory*, Proc. Sympos. Pure Math., vol. 8, Amer. Math. Soc., Providence, R. I., 1965, pp. 181–189. MR 30 #4740.
2. S. L. G. Choi, *The largest sum-free subsequence from a sequence of n numbers*, Proc. Amer. Math. Soc. 39 (1973), 42–44. MR 47 #1771.
3. D. Cantor (to appear).
4. S. L. G. Choi, *On sequences not containing a large sum-free subsequence*, Proc. Amer. Math. Soc. 41 (1973), 415–418. MR 48 #3910.
5. V. Chvátal and J. Komlós, *Some combinatorial theorems on monotonicity*, Canad. Math. Bull. 14 (1971), 151–157.

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

MATHEMATICS INSTITUTE, HUNGARIAN ACADEMY OF SCIENCES, BUDAPEST, HUNGARY

COMPUTER SCIENCE DEPARTMENT, STANFORD UNIVERSITY, STANFORD, CALIFORNIA 94305