The cardinality of a set containing the pairwise sums of four integers

Wouter van Doorn

Abstract

Choi, Erdős and Szemerédi proved the existence of a constant c such that for any set $A \subseteq \{1, 2, \ldots, 2n\}$ with $|A| \geq n + c$ there exist four integers b_1, b_2, b_3, b_4 with $b_i + b_j \in A$ for $1 \leq i < j \leq 4$. We show that one can take $c = 2044$.

1 Introduction

For an integer $k \geq 3$, let $g_k(n)$ be the smallest integer such that for all sets $A \subseteq \{1, 2, \ldots, 2n\}$ with $A \geq n+g_k(n)$, one can find distinct integers $b_1, b_2, \ldots, b_k$ (not necessarily contained in A) with $b_i + b_j \in A$ for $1 \leq i < j \leq n$. Choi, Erdős and Szemerédi introduced this function in [1], and proved many bounds on it. For example, they proved $g_3(n) = 2$, $g_4(n) \leq c$ for some constant c, $c_1 \log(n) < g_5(n) < c_2 \log(n)$ for some constants c_1, c_2 and $c_3\sqrt{n} < g_6(n) < c_4\sqrt{n}$ for some constants c_3, c_4. They did not necessarily attempt to optimize all these constants however, and estimating the value of $g_k(n)$ is listed as problem #866 at [2]. In this short note we will (without claiming much originality ourselves) revisit their proof that $g_4(n)$ is bounded, and find an explicit value for it.

2 An explicit upper bound

The goal is to find an explicit value of c in the theorem of Choi, Erdős and Szemerédi.

Theorem 1. *Let c be equal to 2044. If $A \subseteq \{1, 2, \ldots, 2n\}$ is any set with $|A| \geq n+c$ elements, then there exist distinct integers b_1, b_2, b_3, b_4 with $b_i+b_j \in A$ for $1 \leq i < j \leq 4$.*

In order to be able to prove this, we need a few lemmas, the first of which will essentially be a stronger version of the base case of Lemma A in [1].

Lemma 1. *Let $y_1 < y_2 < \ldots < y_t$ be a sequence of integers and set $y := y_t - y_1$. If $t \geq y^{1/2} + y^{1/4} + 1$, then there exist integers x_1, x_2, x_3 with $x_2 \neq x_3$ such that $x_1, x_1 + x_2, x_1 + x_3, x_1 + x_2 + x_3$ are all elements of the sequence. That is, the sequence contains all subset sums containing x_1.*

Proof. If a positive integer m and indices $i_1 < i_2$, $j_1 < j_2$ exist with $y_{j_1} - y_{i_1} = y_{j_2} - y_{i_2} = m$, then one can choose $x_1 = y_{i_2}, x_2 = y_{i_1} - y_{i_2}, x_3 = m$. However, if such an m and indices do not exist, then the sequence forms a so-called Sidon set, for which it is well-known (see e.g. [3]) that $t < y^{1/2} + y^{1/4} + 1$. $\square$

We will use the previous lemma to prove what is essentially the induction step of Lemma A in [1].

Lemma 2. *Let $y_1 < y_2 < \ldots < y_t$ be a sequence of integers and set $y := y_t - y_1$. If $t \geq 2y^{3/4} + 1.03y^{1/2} + 5$, then there exist integers x_1, x_2, x_3, x_4 with x_2, x_3, x_4 distinct such that the sequence contains all subset sums containing x_1.*

Proof. There are a total of $\frac{1}{2}t(t-1)$ differences $y_j - y_i$ in the interval $[1, y-1]$. Since one can check the inequality $\frac{1}{2}t(t-1) > 2y(y^{1/2} + y^{1/4} + 1)$, this implies there is an integer m that can be written in more than $t' := y^{1/2} + y^{1/4} + 1$ fully disjoint ways as a difference $y_{j_1} - y_{i_1} = y_{j_2} - y_{i_2} = \ldots = y_{j_{t'}} - y_{i_{t'}} = m$. Now we apply Lemma 1 to this sequence $y_{i_1}, y_{i_2} \ldots, y_{i_{t'}}$ to obtain x_1, x_2, x_3, and set $x_4 = m$. And x_4 is distinct from x_2 and x_3, because if, say, $x_2 = m$, then $x_1 + x_2 + x_4$ is both a y_{i_l} and a $y_{j_{l'}}$, contradicting the fact that these were disjoint. □

As in [1], the previous lemma can be used to deduce the existence of integers whose pairwise sums are contained in A, if A contains sufficiently many even integers.

Lemma 3. *Let $y_1 < y_2 < \ldots < y_t$ be a sequence of even integers and set $y := y_t - y_1$. If $t \geq 2y^{3/4} + 1.03y^{1/2} + 5$, then there exist distinct integers b_1, b_2, b_3, b_4 such that the sequence contains all pairwise sums $b_i + b_j$ with $1 \leq i < j \leq 4$.*

Proof. Apply Lemma 2 with $b_1 = \frac{1}{2}x_1$ and $b_i = \frac{1}{2}x_1 + x_i$ for $2 \leq i \leq 4$. □

We are now ready to prove our main result.

Proof of Theorem 1. With t the number of even elements of A, write $t = c + d$ for some $d \geq 0$. Let us first assume that there are no even elements of A contained in the interval $[4d+4, 2n-4d-4]$.

In that case, there are either at least $t' := \left\lceil \frac{1}{2}t \right\rceil \geq \frac{1}{2}c + \frac{1}{2}d$ even elements of A in $[2, 4d+2]$, or at least t' even elements of A in $[2n-4d-2, 2n]$. In either situation, with $y = 4d+2$ we have an interval of length y with at least t' even elements. If we can prove $t' \geq 2y^{3/4} + 1.03y^{1/2} + 5$, then we are done with this case by Lemma 3. But this follows from the inequality $t' \geq \frac{1}{2}c + \frac{1}{8}(y-2)$ and the fact that $\frac{1}{2}c + \frac{1}{8}(y-2) \geq 2y^{3/4} + 1.03y^{1/2} + 5$ holds for all $y \geq 1$.

We may therefore freely assume the existence of an even element $2m \in A$ with $2d + 2 \leq m \leq n - 2d - 2$. Now let (p, q) be a pair of odd integers in the interval $I := [m - 4d - 2, m + 4d + 2]$ with $p < q$ and $p + q = 2m$. With b_1, b_2 even integers such that $b_1 + b_2 = 2m$ and $m - 2 \leq b_1 < b_2 \leq m + 2$, we have $p + b_1 \geq m - 4d - 2 + m - 2 = 0$, which implies $p + b_1 \geq 1$ as this sum is odd. On the other hand, $q + b_2 \leq m + 4d + 2 + m + 2 = 2m + 4d + 4 \leq 2n$. Since we have $2d + 1$ such distinct pairs (p, q) in I, while A misses $t - c = d$ odd integers in the interval $[1, 2n]$, there must be a pair (p, q) in I such that $b_1 + p, b_2 + p, b_1 + q, b_2 + q$ are all in A. We can then take $b_3 = p, b_4 = q$. □

References

[1] S.L.G. Choi, P. Erdős, E. Szemerédi, *Some additive and multiplicative problems in number theory.* Acta Arithmetica, vol. 27, 37–50, 1975. Also available here.

[2] T. F. Bloom, https://www.erdosproblems.com.

[3] B. Lindström, *An inequality for B2-sequences*. Journal of Combinatorial Theory, Volume 6, Issue 2, 211–212, 1969. Also available here.