A UNIFORM SQUARE-ROOT LOWER BOUND FOR NON-UNIQUE SUMS AND A RESOLUTION OF ERDŐS PROBLEM 14

LIAM KRUER AND JENSEN KOHLMEYER

Abstract. For a set $A \subseteq \mathbb{N}_0$, let $E_A(N)$ count the positive integers at most N that do not have exactly one unordered representation as a sum of two elements of A, with equal summands allowed. We prove the uniform estimate $E_A(N) > \sqrt{N}/12000$ for every integer $N \geq 2 \cdot 10^{24}$. This answers both questions in Erdős problem 14: part I holds with exponent $1/2$, and the existence assertion in part II is false. The common finite estimate is $E_A(2T^4) > T^2/1000 - 1$ for every integer $T \geq 10^6$. Its proof combines a triple-sum moment identity, pair counting in an arithmetic progression, and two incompatible estimates for a finite generating polynomial. We give the shared argument once and describe the two accompanying Lean proof files and the scope of their reported local verification.

hexagon:2611.00761v1

1. Introduction and main results

For $A \subseteq \mathbb{N}_0 = \{0, 1, 2, \ldots\}$ and integers $n, N \geq 0$, define

$$\begin{aligned} r_A(n) &= |\{(a, b) \in A^2 : a \leq b,\ a + b = n\}|, \\ E_A(N) &= |\{1 \leq n \leq N : r_A(n) \neq 1\}|. \end{aligned} \tag{1.1}$$

The first function counts unordered pairs, including diagonal pairs (a, a). An exceptional integer has either no representation or more than one unordered representation. These conventions agree with the exceptional counting function in the Formal Conjectures formulation [2]. The set A may be finite or infinite.

Erdős discussed this counting function in connection with a problem of Erdős and Nathanson [1, pp. 43–44]. The two questions recorded as Erdős problem 14 ask whether

$$N^{1/2-\varepsilon} = O(E_A(N)) \quad (N \to \infty) \qquad \text{for every } A \text{ and every } \varepsilon > 0, \tag{1.2}$$

$$E_A(N) = o(\sqrt{N}) \quad (N \to \infty) \qquad \text{for some } A \subseteq \mathbb{N}_0. \tag{1.3}$$

All asymptotic statements here are taken along positive integers. We answer the first question affirmatively and the second negatively by proving one uniform lower bound.

Theorem 1.1. *For every $A \subseteq \mathbb{N}_0$ and every integer $N \geq 2 \cdot 10^{24}$,*

$$\sqrt{N} < 12000\, E_A(N). \tag{1.4}$$

Corollary 1.2 (Part I)**.** *For every $A \subseteq \mathbb{N}_0$ and every real $\varepsilon > 0$, one has $N^{1/2-\varepsilon} = O(E_A(N))$ as $N \to \infty$. The constant 12000 and threshold $2 \cdot 10^{24}$ work simultaneously for all A and ε.*

Corollary 1.3 (Part II)**.** *There is no set $A \subseteq \mathbb{N}_0$ for which $E_A(N) = o(\sqrt{N})$. More precisely, every $A \subseteq \mathbb{N}_0$ satisfies*

$$\liminf_{N \to \infty} \frac{E_A(N)}{\sqrt{N}} \geq \frac{1}{12000}. \tag{1.5}$$

Date: September 13, 2026.

2020 *Mathematics Subject Classification.* Primary 11B34; Secondary 11B13, 68V20.

Key words and phrases. Unique sums, additive representation functions, exceptional sets, Erdős problem 14, formal proof, Lean.

The numerical constants are chosen to simplify the estimates; no optimality is asserted. The main finite result is the following.

Theorem 1.4. *For every $A \subseteq \mathbb{N}_0$ and every integer $T \geq 10^6$,*

$$E_A(2T^4) > \frac{T^2}{1000} - 1. \tag{1.6}$$

Sections 2–5 prove Theorem 1.4. The structural step is a first-moment identity for triple sums: for targets whose relevant pair sums are uniquely represented, a unique weighted representation $n = a + 2b$ constrains $3b$ to a single congruence class while the prefix $A \cap [0, n]$ stays fixed. Pair counting then shows that almost every target has at least two weighted representations. This contradicts an upper bound for $F(z)F(z^2)$, where F is a finite generating polynomial. Section 6 passes to every sufficiently large scale and derives both answers.

2. Pair counting and the finite scale

All intervals in counting expressions are intersected with $\mathbb{N}_0$. For finite sets $U, V \subseteq \mathbb{N}_0$, write $U + V = \{u + v : u \in U,\ v \in V\}$. For a positive integer d, the notation $dU = \{du : u \in U\}$ denotes dilation. For finite $B \subseteq \mathbb{N}_0$, put

$$R_B(n) = |\{(a, b) \in B^2 : a + b = n\}|.$$

Unlike r_A, the function R_B counts ordered pairs.

Lemma 2.1. *Let $B, V, \mathcal{F}$ be finite sets of nonnegative integers, and let $m \geq 0$. Suppose that $B + B \subseteq V$, $|\mathcal{F}| \leq m$, and $R_B(n) \leq 2$ whenever $n \notin \mathcal{F}$. If $t = |B|$, then*

$$t^2 \leq 2|V| + mt. \tag{2.1}$$

If $L \geq 0$ and, in addition, $B \subseteq [0, L]$ lies in a single residue class modulo an integer $q > 0$, then

$$t^2 \leq 4\lfloor L/q \rfloor + 2 + mt. \tag{2.2}$$

Proof. For every n, the first coordinate of an ordered pair determines the second, so $R_B(n) \leq t$. Summing over $V \setminus \mathcal{F}$ and $V \cap \mathcal{F}$ gives

$$t^2 = \sum_{n \in V} R_B(n) \leq 2|V \setminus \mathcal{F}| + t|V \cap \mathcal{F}| \leq 2|V| + mt.$$

For the second assertion, choose the residue representative $0 \leq r < q$. Every $b \in B$ has the form $b = r + qj$ with $0 \leq j \leq \lfloor L/q \rfloor$. Thus

$$B + B \subseteq \{2r + qj : 0 \leq j \leq 2\lfloor L/q \rfloor\},$$

and the set on the right has $2\lfloor L/q \rfloor + 1$ elements. Apply (2.1) to this set of possible sums. □

We shall also use the elementary implication

$$t^2 \leq L^2 + mt \quad \Longrightarrow \quad t \leq m + L \qquad (t, m, L \geq 0). \tag{2.3}$$

Indeed, $t > m + L$ would give $t(t - m) \geq (t - m)^2 > L^2$, a contradiction.

Fix $A \subseteq \mathbb{N}_0$ and an integer $T \geq 10^6$, and set

$$\begin{aligned} X &= T^4, \qquad S = A \cap [0, X], \qquad k = |S|, \\ \mathcal{E} &= \{0 \leq n \leq 2X : r_A(n) \neq 1\}, \qquad m = |\mathcal{E}|. \end{aligned} \tag{2.4}$$

We shall prove $1000m > T^2$. Until the contradiction in Section 5, assume

$$1000m \leq T^2. \tag{2.5}$$

For every $n \in \mathbb{N}_0 \setminus \mathcal{E}$, one has $R_S(n) \le 2$. For $n \le 2X$ this follows from uniqueness of the unordered representation in A; for $n > 2X$ there are no representations from S. Lemma 2.1 therefore gives

$$k^2 \le 4X + 2 + mk \le (2T^2+1)^2 + mk.$$

By (2.3) and (2.5),

$$k \le m + 2T^2 + 1 \le 3T^2. \tag{2.6}$$

For integers $0 \le n \le X$, define the prefix and its statistics by

$$S_n = A \cap [0,n], \qquad K_n = |S_n|, \qquad H_n = \sum_{a \in S_n} a. \tag{2.7}$$

Every integer in $[0,n] \setminus \mathcal{E}$ is a sum of two members of S_n. Since there are K_n^2 possible ordered pairs and at most m exceptions,

$$n + 1 \le K_n^2 + m. \tag{2.8}$$

Lemma 2.2. *Under assumption* (2.5)*, every integer n with $X/1000 < n \le X$ satisfies*

$$K_n > 0, \qquad K_n + 1 \ge T^2/32. \tag{2.9}$$

Proof. If $K_n + 1 < T^2/32$, then (2.8) and (2.5) imply

$$\frac{T^4}{1000} < K_n^2 + m < \frac{T^4}{1024} + \frac{T^2}{1000}.$$

Hence $24T^4 < 1024T^2$, contrary to $T \ge 10^6$. Thus $K_n + 1 \ge T^2/32 > 1$, which also gives $K_n > 0$. □

3. A triple-sum moment identity

Introduce the weighted representation function

$$g(n) = |\{(a,b) \in S^2 : a + 2b = n\}|. \tag{3.1}$$

Call a nonnegative integer n *clean* if $n \notin S + \mathcal{E}$. There are at most km integers which are not clean. If $n \le X$ is clean and $a \in S_n$, then $0 \le n - a \le X$ and $n - a \notin \mathcal{E}$. Otherwise $n = a + (n-a)$ would belong to $S + \mathcal{E}$. Consequently $n - a$ has a unique unordered representation as a sum of two elements of A, and both summands lie in S_n.

For $0 \le n \le X$, put

$$\mathcal{T}_n = \{(a,b,c) \in S_n^3 : a + b + c = n\}, \qquad B_n = \sum_{\substack{a,b \in S \\ a+2b=n}} b.$$

The symbol B_n denotes a sum of integers, not a set. All pairs counted by $g(n)$ belong to S_n^2, by nonnegativity.

Lemma 3.1 (Triple-sum identity)**.** *For every clean target $0 \le n \le X$,*

$$3H_n + 3B_n = n(K_n + g(n)). \tag{3.2}$$

Proof. Fix the first coordinate $a \in S_n$. The unique unordered pair representing $n - a$ contributes two ordered triples unless its entries coincide, in which case it contributes one. A diagonal pair is exactly a solution of $a + 2b = n$. Thus

$$|\mathcal{T}_n| = 2K_n - g(n), \qquad \sum_{(a,b,c) \in \mathcal{T}_n} a = 2H_n - \sum_{\substack{a,b \in S \\ a+2b=n}} a. \tag{3.3}$$

The set $\mathcal{T}_n$ is invariant under permutations of its three coordinates. The three coordinate sums are therefore equal, and their sum is $n|\mathcal{T}_n|$. Hence

$$3\sum_{(a,b,c)\in\mathcal{T}_n} a = n|\mathcal{T}_n|. \tag{3.4}$$

Also,

$$\sum_{\substack{a,b\in S\\ a+2b=n}} a = ng(n) - 2B_n.$$

Substitution in (3.3)–(3.4) gives

$$6H_n - 3ng(n) + 6B_n = 2nK_n - ng(n).$$

Rearranging and dividing by 2 proves (3.2). □

The diagonal correction in (3.3) is essential. The argument includes triples with all three coordinates equal and uses no distinctness assumption.

4. Few targets have at most one weighted representation

As n runs from 0 to X, the sets S_n are nested and take at most $k+1$ distinct values. Fix one such prefix, and consider clean targets $X/1000 < n \le X$ having this prefix. If there are no such targets, there is nothing to count. Otherwise write $K = K_n$, $H = H_n$, and $q = K + 1$ for their common statistics. By Lemma 2.2, $K > 0$ and $q \ge T^2/32$.

If $g(n) = 0$, then $B_n = 0$ and (3.2) becomes $3H = nK$. Since $K > 0$, at most one target with this prefix has $g(n) = 0$.

If $g(n) = 1$, let (a, b) be its unique weighted pair. Then $B_n = b$, so

$$3H + 3b = nq. \tag{4.1}$$

As $q > 0$, distinct targets yield distinct values of b. Moreover, all corresponding integers $3b$ are congruent to $-3H$ modulo q.

Lemma 4.1. *Under assumption* (2.5)*, for each fixed prefix there are at most* $m+20T$ *clean targets* $X/1000 < n \le X$ *with* $g(n) = 1$*.*

Proof. Let $B \subseteq S$ be the set of second coordinates of the unique weighted pairs for these targets, and let $t = |B|$. By the injectivity just proved, t equals the number of targets. The dilation $3B$ has cardinality t, is contained in $[0, 3X]$, and lies in a single residue class modulo q.

Outside $3\mathcal{E}$, every integer has at most two ordered representations as a sum of two members of $3B$. Indeed, a represented integer is divisible by 3. For a represented integer $3u$, division by 3 identifies its representations from $3B$ with the representations of u from B. If $3u \notin 3\mathcal{E}$, then $u \notin \mathcal{E}$, and $R_B(u) \le R_S(u) \le 2$. Also $|3\mathcal{E}| = m$.

Apply (2.2) to $3B$, using $q \ge T^2/32$:

$$t^2 \le 4\lfloor 3X/q\rfloor + 2 + mt \le 384T^2 + 2 + mt \le (20T)^2 + mt.$$

Equation (2.3) gives $t \le m + 20T$. □

Working with $3B$ avoids inverting 3 modulo q. In particular, the argument remains valid when 3 divides $K + 1$.

Proposition 4.2. *Under assumption* (2.5)*, the number*

$$D = |\{0 \le n \le X : g(n) \le 1\}|$$

satisfies

$$D \le X/100. \tag{4.2}$$

Proof. There are $\lfloor X/1000 \rfloor + 1$ targets in the initial segment and at most km targets which are not clean. For each of the at most $k+1$ prefixes, there is at most one remaining target with $g(n) = 0$ and at most $m + 20T$ with $g(n) = 1$. Therefore

$$D \le \lfloor X/1000 \rfloor + 1 + km + (k+1)(1 + m + 20T). \tag{4.3}$$

Using $k \le 3T^2$ and $1000m \le T^2$, we obtain

$$\begin{aligned} 1000D &\le T^4 + 1000 + 3T^4 + (3T^2+1)(1000 + T^2 + 20000T) \\ &= 7T^4 + 60000T^3 + 3001T^2 + 20000T + 2000. \end{aligned}$$

For $T \ge 10^6$, each of $60000T^3$, $3001T^2$, and $20000T + 2000$ is at most T^4. Thus $1000D \le 10T^4 = 10X$. □

5. The generating polynomial contradiction

For $0 < z < 1$, let $F(z) = \sum_{a \in S} z^a$. All generating functions below are finite polynomials. Infinite geometric series are used only as upper bounds.

Lemma 5.1. *Under assumption* (2.5)*, for every real* $0 < z < 1$*,*

$$F(z)^2 \le \frac{2}{1-z} + mk, \qquad mk \le \frac{3X}{1000}. \tag{5.1}$$

At $z = 1 - 4/X$*, this implies*

$$F(z)F(z^2) < \frac{37X}{100}. \tag{5.2}$$

Proof. The coefficient identity $F(z)^2 = \sum_{n=0}^{2X} R_S(n) z^n$, the estimate $R_S(n) \le 2$ outside $\mathcal{E}$, and the general bound $R_S(n) \le k$ give

$$F(z)^2 \le 2 \sum_{n=0}^{2X} z^n + k \sum_{n \in \mathcal{E}} z^n \le \frac{2}{1-z} + mk.$$

The second inequality in (5.1) follows from (2.5) and (2.6).

Set $z = 1 - 4/X$. Since $X = T^4 \ge 10^{24}$, one has $0 < z < 1$. Evaluating (5.1) at z yields

$$F(z)^2 \le \frac{X}{2} + \frac{3X}{1000} \le \frac{51X}{100}. \tag{5.3}$$

Also,

$$\frac{2}{1-z^2} = \frac{X}{4} + \frac{X}{2(X-2)} \le \frac{X}{4} + 1.$$

Since $X \ge 1000$, evaluation at z^2 gives

$$F(z^2)^2 \le \frac{X}{4} + 1 + \frac{3X}{1000} \le \frac{26X}{100}. \tag{5.4}$$

Both polynomial values are nonnegative. As $51 \cdot 26 = 1326 < 1369 = 37^2$, multiplying (5.3) and (5.4) and taking square roots proves (5.2). □

Lemma 5.2. *Under assumption* (2.5)*, at* $z = 1 - 4/X$ *one has*

$$F(z)F(z^2) \ge \frac{19X}{50}. \tag{5.5}$$

Proof. The coefficient identity

$$F(z)F(z^2) = \sum_{n=0}^{3X} g(n)z^n$$

and nonnegativity give

$$F(z)F(z^2) \geq 2\sum_{n=0}^{X-1} z^n - 2D. \tag{5.6}$$

Indeed, $g(n) \geq 2$ except at the D targets counted in Proposition 4.2, and every omitted weight z^n is at most 1.

Because $X > 4$, one has $z^{-1} = (1 - 4/X)^{-1} \geq 1 + 4/X$. Bernoulli's inequality gives

$$z^{-X} \geq (1 + 4/X)^X \geq 1 + X(4/X) = 5,$$

so $z^X \leq 1/5$. Consequently,

$$2\sum_{n=0}^{X-1} z^n = \frac{2(1 - z^X)}{1 - z} = \frac{X}{2}(1 - z^X) \geq \frac{2X}{5}.$$

Insert this and $D \leq X/100$ in (5.6) to obtain

$$F(z)F(z^2) \geq \frac{2X}{5} - \frac{2X}{100} = \frac{19X}{50}.$$

□

Proof of Theorem 1.4. If $1000m \leq T^2$, Lemmas 5.1 and 5.2 would imply

$$\frac{38X}{100} \leq F(z)F(z^2) < \frac{37X}{100},$$

which is impossible. Hence $m > T^2/1000$. The count m includes the possible exception at 0, whereas $E_A(2T^4)$ counts only positive integers. Thus $m \leq E_A(2T^4) + 1$, and

$$E_A(2T^4) \geq m - 1 > \frac{T^2}{1000} - 1.$$

□

6. All large scales and the two answers

Proof of Theorem 1.1. Let $N \geq 2 \cdot 10^{24}$ and choose

$$T = \left\lfloor (N/2)^{1/4} \right\rfloor.$$

Then $T \geq 10^6$ and

$$2T^4 \leq N < 2(T + 1)^4 \leq 32T^4.$$

In particular, $\sqrt{N} < 6T^2$. The function E_A is nondecreasing because its defining sets are nested. By Theorem 1.4,

$$E_A(N) \geq E_A(2T^4) > \frac{T^2}{1000} - 1 \geq \frac{T^2}{2000} > \frac{\sqrt{N}}{12000}.$$

The middle inequality uses $T^2 \geq 2000$. □

Proof of Corollary 1.2. For $N \geq 1$ and $\varepsilon > 0$, one has $N^{1/2-\varepsilon} \leq \sqrt{N}$. Theorem 1.1 gives $N^{1/2-\varepsilon} \leq 12000E_A(N)$ for every $N \geq 2 \cdot 10^{24}$, which is the asserted big-O estimate. □

Proof of Corollary 1.3. Dividing (1.4) by $12000\sqrt{N}$ and taking the lower limit proves (1.5). Since $E_A(N) \geq 0$, the assertion $E_A(N) = o(\sqrt{N})$ would mean that for every $c > 0$ there is an integer N_0 such that $E_A(N) \leq c\sqrt{N}$ for all $N \geq N_0$. Taking $c = 1/24000$ contradicts (1.4) as soon as $N \geq \max\{N_0, 2 \cdot 10^{24}\}$. □

Remark 6.1 (Direct finite-scale proof of the negative answer). The nonexistence assertion also follows directly from Theorem 1.4, as in the accompanying Lean proof for part II. If $E_A(N) = o(\sqrt{N})$, choose N_0 with $E_A(N) \leq \sqrt{N}/4000$ for $N \geq N_0$. Choose an integer $T \geq 10^6$ such that $2T^4 \geq N_0$. Then

$$E_A(2T^4) \leq \frac{\sqrt{2T^4}}{4000} \leq \frac{T^2}{2000}, \qquad E_A(2T^4) > \frac{T^2}{1000} - 1 \geq \frac{T^2}{2000},$$

a contradiction. Thus both answers rest on the same finite obstruction, and the uniform theorem strengthens both asymptotic conclusions.

7. Lean formalization and verification scope

The two source manuscripts [5, 6] identify separate proof files: `Erdos14_PartI.lean` for part I and `Erdos14_PartII_Counterexample.lean` for the negative answer to part II. Copies with these names were available when preparing the unified manuscript. Their shared core proof bodies are identical. Each includes its helper lemmas in the namespace `Erdos14Verification`. The task wrapper supplies the imports of the Erdős 14 source and `TaskSupport`, together with the enclosing namespace `Bounty`.

The final declarations in the two files have the respective target types

```
theorem target :
  fcTypeOfName% "Erdos14.erdos_14.parts.i"
```

and

```
theorem target :
  Not (fcTypeOfName% "Erdos14.erdos_14.parts.ii")
```

The second file writes `Not` with Lean's usual negation symbol. For the affirmative-answer target described in the original verification reports, the first source type expresses True if and only if the assertion in part I holds. The second expresses True if and only if the existence assertion in part II holds. Accordingly, the part I submission proves its target, and the part II submission proves the negation of its target. The part II result is the nonexistence statement of Corollary 1.3.

Table 1 records the correspondence between the mathematical argument and the principal declarations present in the files. The part I scale selection uses two integer square roots in place of an analytic fourth root. The part II file follows the direct finite-scale argument.

Table 1. Proof components and Lean declarations. The common helper namespace is omitted.

Mathematical component	Lean declaration
Lemma 2.1: pair and progression bounds	`pair_count_bound`, `progression_pair_count`
Lemma 3.1: triple-sum identity	`clean_triple_moment`
Proposition 4.2: low-target count	`low_targets_density`
Polynomial contradiction	`finite_obstruction`
Theorem 1.4: finite estimate for A	`set_scale_obstruction`
Theorem 1.1: uniform bound	`uniform_sqrt_bound`
Corollary 1.2: affirmative answer	`part_i_positive`
Corollary 1.3: nonexistence assertion	`part_ii_negative`

Reported verification. The source manuscripts report successful compilation with Lean 4.33.1 and mathlib [3] at revision

`0df444a360eaa60ab8c11dca51a86af692955474`.

They identify the audited Formal Conjectures revision as

`8432eac998110a563e03df65a28c117e97c8c142`,

and the validator [4] revision as

`076ead3c71c8b0bfc0641975076d95015d075ce3`.

The reported local checks compiled the trusted wrapper, compared the target types and referenced statement constants against the pinned challenge, checked axiom dependencies, and replayed the exported proof declarations into an empty Lean environment. According to those reports, each final theorem depends only on `propext`, `Classical.choice`, and `Quot.sound`, with no dependency on `sorryAx`, an unproved source conjecture, or a `native_decide` certificate.

Public verification records. The unified manuscript describes these checks as reported local verification. The prize submission [7] also links separate production verification records for part I [8] and part II [9], certified on September 16, 2026, and the pinned original-source edition and local verification reports [10]. The mathematical proof in Sections 2–6 is self-contained and independent of the submission infrastructure.

Acknowledgments

OpenAI's Codex assisted substantially in developing the argument, constructing and checking the Lean formalization described in the source manuscripts, and preparing the unified manuscript. The authors are responsible for the final mathematical content. We acknowledge the Lean and mathlib communities and the maintainers of Formal Conjectures and Conjectures.io for the libraries and verification infrastructure used in this work.

References

[1] P. Erdős, Some of my forgotten problems in number theory, *Hardy–Ramanujan J.* 15 (1992), 34–50. Journal PDF.

[2] The Formal Conjectures contributors, Formal Conjectures: Erdős problem 14, Lean source file `FormalConjectures/ErdosProblems/14.lean`. Repository. The reported revision is given in Section 7.

[3] The mathlib Community, The Lean mathematical library, *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020)*, ACM, 2020, 367–381. DOI.

[4] Conjectures.io, Conjectures validator, source repository. The reported revision is given in Section 7.

[5] L. Kruer and J. Kohlmeyer, A uniform square-root lower bound for non-unique sums, source manuscript, September 13, 2026, 7 pp., accompanying file `Erdos14_PartI.lean`.

[6] L. Kruer and J. Kohlmeyer, A negative answer to Erdős problem 14, part II, source manuscript, September 13, 2026, 7 pp., accompanying file `Erdos14_PartII_Counterexample.lean`.

[7] Justin Sun Prize submission JSP-000052, with unified manuscript PDF.

[8] Conjectures.io, Erdős problem 14, part i, production verification record.

[9] Conjectures.io, Erdős problem 14, part ii, production verification record.

[10] Pinned original-source edition. Local verification reports for part i and part ii.

Email address: `kruerl@purdue.edu`

Email address: `jenwin@purdue.edu`