Conjectures.io

Formal proof verified through Conjectures

ERDŐS PROBLEM 653: ASYMPTOTICALLY MAXIMAL DIVERSITY OF PLANAR DISTANCE COUNTS

Liam Kruer and Jensen Kohlmeyer

Abstract. For an n-point subset P of the Euclidean plane, count the distinct distances from each point to the other points. Let $g(n)$ be the maximum number of different counts that can occur in such a configuration. The accepted submission by gus proves $g(n) = (1 - o(1))n$, answering Erdős Problem 653 affirmatively. Its construction uses an arithmetic grid to establish coarse distance-count levels, then geometric controls to separate the counts on almost all of a large witness family. We describe the finite construction interfaces, explain their quantitative assembly, and give the complete reduction from square-scale configurations to every sufficiently large cardinality. The full arithmetic, analytic, probabilistic and geometric proofs are retained in the public Lean file.

1. What is being counted

For a finite set $P \subset \mathbb{R}^2$ and $p \in P$, define

$$R_P(p) = \big|\{\|p - q\| : q \in P \setminus \{p\}\}\big|.$$

Its *distance-count spectrum* and *diversity* are

$$\mathcal{S}(P) = \{R_P(p) : p \in P\}, \qquad \operatorname{div}(P) = |\mathcal{S}(P)|.$$

The extremal function in the problem is

$$g(n) = \max_{P \subset \mathbb{R}^2,\ |P|=n} \operatorname{div}(P).$$

The maximum exists: the possible diversity values form a nonempty subset of the finite set $\{0, \dots, n\}$. In particular $g(n) \le n$.

This is diversity of *counts*, not the total number of distances determined by all pairs. For example, the vertices of a square each see two different distances, so its diversity is one. The theorem concerns the existence of configurations with high diversity, not all planar sets.

Theorem 1.1. *For every $\varepsilon > 0$, there exists N such that for every integer $n \ge N$ there is an n-point set $P \subset \mathbb{R}^2$ satisfying*

$$\operatorname{div}(P) \ge (1 - \varepsilon)n.$$

Consequently $g(n)/n \longrightarrow 1$ as $n \longrightarrow \infty$.

Date: 17 September 2026.

Key words and phrases. Discrete geometry, pinned distances, distance-count spectra, Gaussian integers, Lean.

Submitted by gus. Working proof exposition and source guide prepared by Conjectures with Codex assistance from the accepted Lean submission. This prose has not itself been formally verified and is not an author-approved journal publication. The complete proof, including the analytic and arithmetic construction lemmas summarised here, is the linked Lean source.

1.1. **Historical context.** The problem record cites Erdős's 1997 paper [1]. It reports earlier lower bounds of $3n/8$ by Erdős and Fishburn and $7n/10$ by Csizmadia, together with an upper bound of the form $n - cn^{2/3}$ [2]. The new asymptotic conclusion is compatible with such a sublinear deficit; it does not assert $g(n) = n$ for any finite n. The announcement's "29-year-old question" uses the calendar year difference $2026 - 1997$, not a certified first-proposal date.

2. The finite construction statement

The quantitative interface between the construction and the final limit is particularly simple. For an integer $s \geq 0$, put

$$t = 2(s+1), \qquad A_s = t^2 + 50t + 1, \qquad b_{s,\delta} = (1-\delta)(t-1)^2 - 1.$$

Proposition 2.1 (Square-scale configurations)**.** *For each $s \geq 0$ and $0 < \delta < 1/2$, there exists L_0 such that for every integer $L \geq L_0$ there is a finite planar set Q with*

$$|Q| \leq A_s L^2, \qquad \operatorname{div}(Q) \geq b_{s,\delta} L^2. \tag{1}$$

This is the theorem `FinalConstruction.squareFamilies` in the accepted source. Sections 3–5 describe the substantive formal lemmas used to construct these sets and derive the constants in (1). They are an exposition of the verified construction; the detailed sieve, projection-energy and finite selection proofs are in the linked Lean source rather than reproduced in full here. Sections 6–7 supply the elementary all-cardinalities deduction.

3. Coarse arithmetic levels

Fix s and hence $t = 2(s+1)$ and $K = t^2$. The construction starts with K blocks of L^2 points each. Each point is indexed by a block $b \in [K]$ and a grid coordinate $u \in \{0, \dots, L-1\}^2$. There is a subset of *good blocks* of cardinality

$$G = (t-1)^2.$$

Associated integers D_b, with $0 \leq D_b \leq K$, are distinct on the good blocks. They provide separated coarse levels for pinned counts.

The arithmetic selection uses distinct rational primes congruent to 3 modulo 4 and Gaussian-integer offsets. Such primes are inert in the Gaussian integers. The formal development controls collisions between squared-distance values through Gaussian norm and residue calculations, including estimates for high and low collision regimes. The selection theorem constructs the required primes and offsets simultaneously; they are not unproved hypotheses of the final theorem.

More precisely, `OffsetSelection.exists_admissible_family` produces a function $\epsilon_L \to 0$ and admissible data at every sufficiently large L. The actual initial point set P_L satisfies

$$|P_L| = KL^2, \qquad \left|R_{P_L}(p_{b,u}) - D_b L^2\right| \leq K\epsilon_L L^2. \tag{2}$$

This is `CoarseCounts.full_configuration`. Every point in (2) is a genuine point of the Euclidean plane.

The coarse estimate alone does not establish the theorem: many points in the same block could still have identical counts. The remaining construction separates almost all of those counts with controlled additional points.

4. Separating the counts

The numerical stage works with the integer initial counts in (2). It selects a common projection parameter, a bounded integer adjustment for each row, and a finer bounded adjustment for each individual grid index. These quantities are chosen before the final geometric extensions.

Write $T_{b,u}$ for the selected integer score after these adjustments. The formal selection theorem produces a set S of witness indices among the good blocks on which the scores are injective, with

$$|S| \geq (1-\delta)GL^2 - \lambda_L L^2, \qquad \lambda_L \geq 0, \qquad \lambda_L \longrightarrow 0. \tag{3}$$

The indices used for the corrections still range over the entire original configuration. Only the witness set is restricted; no unaccounted deletion of points changes the pinned counts.

4.1. **How the selection loses only a small fraction.** The development smooths projected grid counts and compares them to continuous projection densities. A common parameter satisfies the required energy and floor-discrepancy estimates. Independent bounded row adjustments distribute the integer scores into bins. A finite capacity selection discards overflow, and bounded individual corrections assign distinct target integers within the selected bins.

For fixed K and G, the formal error has the explicit form

$$\lambda_L = \sqrt{6G\omega_L} + \frac{2G(2E_L+1)}{H_L} + \frac{\sqrt{J_L(GL^2+2\omega_L L^2)}}{L^2}, \tag{4}$$

where $\omega_L \geq 0$ tends to zero, J_L is the number of reachable bins, and the schedules are

$$N_L = KL^2, \qquad H_L = \lfloor \sqrt{L} \rfloor,$$
$$B_L = \lfloor \sqrt{\log N_L} \rfloor, \qquad \ell_L = \left\lfloor \frac{\log N_L}{2\log 2} \right\rfloor, \qquad E_L = L^{1/10}.$$

The exact bin-window estimate gives

$$J_L \leq \frac{(K+8)L^2}{B_L} + 2.$$

Thus every term of (4) tends to zero: the first by $\omega_L \to 0$, the second by $E_L/H_L \to 0$, and the third by $J_L/L^2 \to 0$. This explains the vanishing loss in (3). The existence of the common parameters and the variance bounds are proved by the analytic and finite-probability lemmas leading to `FineSelection.exists_selection`.

5. Turning scores into actual distances

The final geometric step enlarges P_L through three extensions, retaining every original point. The source's exact realization theorem gives a set Q and a common integer C_* such that

$$R_Q(p_{b,u}) = T_{b,u} + C_* \tag{5}$$

on all original indices. Hence injectivity of T on S gives $\operatorname{div}(Q) \geq |S|$. This is why numerical score separation alone is not being substituted for a planar construction.

The three extensions have point budgets

$$C_0 \leq 50tL^2,$$
$$C_1 \leq 2KH_L L,$$
$$C_2 \leq 2B_L\left(\ell_L 2^{\ell_L} + K\left(\left\lfloor \frac{L^2}{\ell_L} \right\rfloor + 1\right)\right).$$

The rational control constructions encode prescribed midpoint coincidences and exclude unintended equalities. Their Euclidean realization establishes (5), including the common shift from all added points. The complete algebraic and geometric arguments are in the source modules `Dictionary`, `MarkerExistence`, `UnaryControls`, `RowControls`, `CorrectionGeometry`, and `PhysicalAssembly`.

For fixed K, the last two costs are $o(L^2)$. Indeed $C_1/L^2 \le 2KH_L/L \to 0$. Also $2^{\ell_L} \le \sqrt{N_L}$, $B_L = O(\sqrt{\log N_L})$, and ℓ_L is of order $\log N_L$. These estimates show separately that $B_L \ell_L 2^{\ell_L}/L^2 \to 0$ and $B_L K/\ell_L \to 0$. Therefore, eventually,

$$C_1 + C_2 \le L^2, \qquad \lambda_L \le 1.$$

Combining the point budget with (3) gives

$$|Q| \le (K + 50t + 1)L^2 = A_s L^2, \qquad \operatorname{div}(Q) \ge \big((1-\delta)(t-1)^2 - 1\big)L^2.$$

These are exactly the two conclusions of Proposition 2.1. No arithmetic, probabilistic or realization premise remains in the formal theorem `FinalConstruction.squareFamilies`.

6. Padding without losing diversity

The square-scale construction permits an upper bound on cardinality, whereas the original problem demands exactly n points. The following elementary observation bridges the difference.

Lemma 6.1 (Fresh-point insertion)**.** *For every finite planar set P and integer $m \ge |P|$, there is a set $Q \supseteq P$ with $|Q| = m$ and*

$$R_Q(p) = R_P(p) + (m - |P|) \quad (p \in P).$$

In particular $\operatorname{div}(Q) \ge \operatorname{div}(P)$.

Proof. For one insertion, choose a point sufficiently far from P that its distance from every old point exceeds the diameter of P. It contributes one new distance to each old point, so all their counts increase by exactly one. Their distinctions are unchanged. Repeat this operation, using the diameter of the current finite set at each step, until its cardinality is m. Each original count receives the same total shift. For an empty starting set the same construction begins with any point. □

Lemma 6.2 (From square scales to all sizes)**.** *Let A be a positive integer and b a real number. Suppose that for every sufficiently large integer L there is P_L with $|P_L| \le AL^2$ and $\operatorname{div}(P_L) \ge bL^2$. If $\eta > 0$ and $c = b/A - \eta > 0$, then for every sufficiently large n there is an n-point set Q_n with $\operatorname{div}(Q_n) \ge cn$.*

Proof. Take $L = \lfloor \sqrt{n/A} \rfloor$, so $AL^2 \le n < A(L+1)^2$. For sufficiently large n the required P_L exists and can be padded to n points by Lemma 6.1. As $L \to \infty$, $(L+1)^2/L^2 \to 1$; hence eventually $c(L+1)^2 \le (c+\eta)L^2$. Then

$$cn \le cA(L+1)^2 \le (c+\eta)AL^2 = bL^2 \le \operatorname{div}(P_L) \le \operatorname{div}(Q_n).$$

□

7. The asymptotic conclusion and exact formal target

For fixed δ, the coefficients in Proposition 2.1 satisfy

$$\frac{b_{s,\delta}}{A_s} = \frac{(1-\delta)(t-1)^2 - 1}{t^2 + 50t + 1} \longrightarrow 1 - \delta \qquad (s \to \infty).$$

Given $\varepsilon > 0$, let $\delta = \min(\varepsilon, 1)/4$. Choose s large enough that $b_{s,\delta}/A_s > 1 - 2\delta$. Proposition 2.1 and Lemma 6.2, with $\eta = \delta$, then give all sufficiently large n an n-point configuration with diversity at least

$$\left(\frac{b_{s,\delta}}{A_s} - \delta\right) n > (1 - 3\delta)n \ge (1 - \varepsilon)n.$$

The coefficient is positive since $\delta \le 1/4$. Together with $g(n) \le n$, this proves Theorem 1.1.

7.1. **The self-distance convention.** The published formal helper includes the distance from a point to itself. For $p \in P$ this adds precisely the new value zero, so its count is $R_P(p) + 1$. Translation by one is injective on integers and leaves spectrum cardinality unchanged. The proof explicitly establishes this correspondence; it does not rely on a different geometric problem.

Set $o(n) = 1 - g(n)/n$ for positive n, with an arbitrary value at zero. The preceding theorem gives $o(n) \to 0$, and $(1 - o(n))n = g(n)$. This is the little-o form of the committed target:

$$\exists o : \mathbb{N} \to \mathbb{R}, \quad o(n) = o(1), \qquad (1 - o(n))n \le g(n) \quad \text{eventually.}$$

The terminal Lean theorem applies the verified square-family theorem to the verified bridge to this exact statement. Earlier conditional lemmas in the file do not leave a conditional final result.

8. Source guide, attribution and verification

The submission credited to **gus** was verified on 15 September 2026 and approved in review on 16 September [3]. Its public verification report records exact statement comparison, permitted-axiom checks and Lean kernel acceptance. The public proof downloaded for this exposition matches the retained accepted bytes. No fresh Lean replay was performed while preparing this PDF.

The source incorporates adapted files from *Prime Number Theorem and More*, organised by Alex Kontorovich and Terence Tao with the project's contributors [4]. Its opening notice specifies revision `a5154676af9aa3095150ee410cdda80555aa0642`, the six adapted files, licensing and modifications. That attribution is preserved in the accompanying source. The credit "submitted by gus" is not a claim that all supporting library material was authored by gus.

Accepted source component	Line
`OffsetSelection.exists_admissible_family`	17396
`ConstructionScales` schedules and budgets	22308
`FineSelection.exists_selection`	27476
`CoarseCounts.full_configuration`	27828
`PhysicalAssembly.exists_realization`	27963
`PhysicalAssembly.exists_configuration`	28033
`exists_padding_card`	28221
`official_pinnedCount_eq`	28245
`exact_sizes_of_square_family`	28352
`epsilon_of_squareFamilies`	28467
`FinalConstruction.squareFamilies`	28533
`Bounty.target`	28603

The complete accepted file contains 28,608 lines. This document supplies the construction's quantitative interfaces and final deductions, with a source map for the longer proofs; it is not a replacement for all 28,608 lines or an independent specialist audit of every auxiliary lemma. The accepted source SHA-256 is

`f3b51c0df8d3c8fb5ad2b71b01474345`
`fc3dece2d1bd37c9d60b45acc711fde4.`

References

[1] P. Erdős, *Some of my favourite unsolved problems*, Math. Japon. (1997), pp. 527–537. Bibliographic record.

[2] T. F. Bloom, *Erdős Problem 653*, https://www.erdosproblems.com/653, accessed 17 September 2026.

[3] Conjectures, *Erdős 653: accepted submission by gus*, 2026. Result and review; complete Lean proof; verification report.

[4] A. Kontorovich, T. Tao and project contributors, *Prime Number Theorem and More*, version 0.1.0, revision a5154676, 2026. https://github.com/AlexKontorovich/PrimeNumberTheoremAnd.