Tinkering with Lattices: A New Take on the Erdős Distance Problem

Elena Kim[1], Fernando Trejos Suárez[2]

1. Pomona College, 2. Yale University

1. elena.kim@pomona.edu, 2. fernando.trejos@yale.edu

Advisors: Eyvindur Palsson (palsson@vt.edu) and Steven J Miller (sjm1@williams.edu)

SMALL REU at Williams College

2020 Québec-Maine Number Theory Conference

Joint work with Jason Zhao

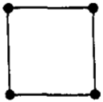
Erdős distinct distances problem

Question [Erdős, 1946]

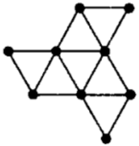
Given n points in a plane, what is the minimum number of distinct distances $f(n)$ that they determine?

Some Examples:

3 points; 1 distance



4 points; 2 distances



9 points; 4 distances

First Estimates

Theorem (Erdős, 1946)

Let $[P_n]$ be the class of subsets of the plane with n points, and let $f(n)$ be the minimum number of distinct distances determined by an element $P_n \in [P_n]$. Then,

$$(n - 3/4)^{1/2} - 1/2 \ \leq \ f(n) \ \leq \ cn/\sqrt{\log n}.$$

Upper Bound: Upper bound for distinct distances of the $\sqrt{n} \times \sqrt{n}$ integer lattice.

Lower Bound (the hard part): Work with the convex hull of an arbitrary point set P_n.

Erdős Distinct Distances Problem: Bounds

Upper bounds (unimproved since Erdős!):

- $\Delta(n) = O(\frac{n}{\sqrt{\log n}})$ (Erdős, 1946)

Lower bounds:

- $\Delta(n) = \Omega(n^{1/2})$ (Erdős, 1946)
- $\Delta(n) = \Omega(n^{4/5}/\log n)$ (Chung, Szemeredi Trotter, 1992)
- $\Delta(n) = \Omega(n^{4/5})$ (Szekely, 1993)
- $\Delta(n) = \Omega(\frac{n}{\log n})$ (Guth + Katz, 2015)

A set with $O(\frac{n}{\sqrt{\log n}})$ distinct distances is *near-optimal*. The integer lattice is a near-optimal set.

Lattice Distance Distribution

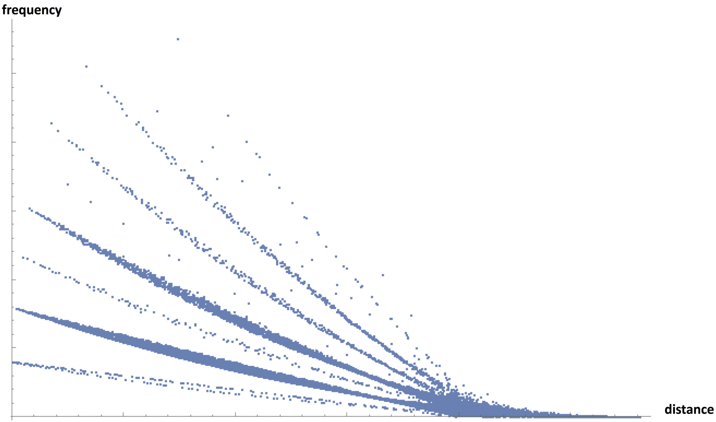
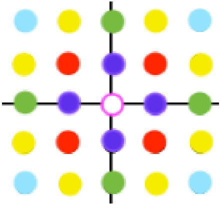


Figure: Distance distribution for 200×200 integer lattice

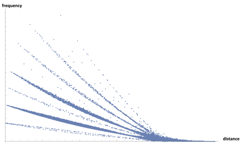
Repeating Distances



How often do distances on the integer lattice repeat?

- 4 points at a distance 1 from the origin.
- 4 points at a distance $\sqrt{2} = \sqrt{1^2 + 1^2}$ from the origin.
- 8 points at a distance $\sqrt{5} = \sqrt{2^2 + 1^2} = \sqrt{1^2 + 2^2}$.

Calculating Distance Frequency



What is the frequency of a distance $\sqrt{d}$ on a $N \times N$ lattice?

- Find all the decompositions of d into $d = a^2 + b^2$, where $N - 1 \geq a \geq b \geq 0$. If there are m ordered pairs (a, b) with $a^2 + b^2 = d$, $\sqrt{d}$ is on the m-th curve!
- If $b = 0$ or $a = b$, then the frequency of that particular decomposition is $2(N - a)(N - b)$. If $a > b$ then the frequency of that particular decomposition is $4(N - a)(N - b)$.
- Add all the frequencies together.

More Facts About the Distance Distribution

Theorem (Fermat)

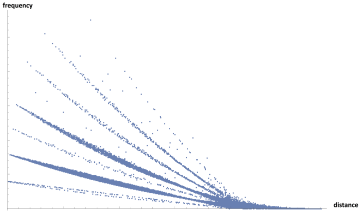
Suppose d has prime factorization $d = 2^f p_1^{g_1} \cdots p_m^{g_m} q_1^{h_1} \cdots q_n^{h_n}$, where $p_i \equiv 1 \pmod 4$, $q_i \equiv 3 \pmod 4$. Then there exist $r(d)$ ordered pairs $(a, b) \in \mathbb{Z}^2$ with $a^2 + b^2 = d$, where

$$r(d) = \begin{cases} 4\,(g_1+1)\cdots(g_m+1) & h_i \text{ is even for all } i, \\ 0 & \text{else.} \end{cases}$$

- The number of integers in the set $\{1, \ldots, 2n\}$ which can be written as the sum of two squares is of order $\frac{cn}{\sqrt{\log n}}$. (Source of Erdos's Upper Bound!)

What is the most common distance on the lattice?

- The first (left-most) distance on each curve has the highest frequency on that curve.
- Define n_k as the least positive integer such that there are k ordered pairs (a, b) with $a^2 + b^2 = n_k$, so that $\sqrt{n_k}$ is the first distance on the k-th curve. Then the sequence $n_1, n_2, \ldots$ will be a list of potential candidates for the most common distance on the lattice!



What is the most common distance on the lattice?

- The first distance on each curve has the highest frequency on that curve.
- Define n_k as the least positive integer such that there are k ordered pairs (a, b) with $a^2 + b^2 = n_k$, so that $\sqrt{n_k}$ is the left-most distance on the k-th curve. Then the sequence $n_1, n_2, \ldots$ will be a list of potential candidates for the most common integer on the lattice!

Lemma (SMALL 2020)

Let $k = q_1^{\alpha_1} q_2^{\alpha_2} \cdots q_n^{\alpha_n}$ be arbitrary, where $q_1 > q_2 > \ldots > q_n$, and let $5 = p_1 < p_2 < \cdots$ be the primes $\equiv 1 \pmod 4$, in increasing order. Then,

$$n_k = \Bigg(\underbrace{p_1 \cdots p_{\alpha_1}}_{\alpha_1 \text{ primes}}\Bigg)^{q_1 - 1} \Bigg(\underbrace{p_{\alpha_1+1} \cdots p_{\alpha_1+\alpha_2}}_{\alpha_2 \text{ primes}}\Bigg)^{q_2 - 1} \cdots \Bigg(\underbrace{p_{\alpha_1+\cdots+\alpha_{n-1}+1} \cdots p_{\alpha_1+\cdots+p_{\alpha_1+\cdots+\alpha_n}}}_{\alpha_n \text{ primes}}\Bigg)^{q_n - 1}$$

What is the most common distance on the lattice?

Although n_k is difficult to deal with, the extremal cases are simple:

- For k prime, $n_k = 5^{k-1}$.
- For $k = 2^m$, $n_k = p_1 \cdots p_m$ where $p_1 < \ldots < p_m$ are the first m primes such that $p_i \equiv 1 \pmod 4$.
- Adapting previous asymptotic results on the product of the first k primes,

$$n_k \approx e^{\frac{1}{2}(1+c)\log_2 2k \log\log_2 2k}.$$

We arrive at the following upper bound for the frequency of $\sqrt{n_k}$:

$$2kN\left(N - e^{\frac{1}{4}(1+c)\log_2 2k \log\log_2 2k}\right).$$

Error introduction

We want to compare the distance distribution of the integer lattice with those of its subsets.

Why do we care about this?

The integer lattice is a near-optimal set, *however* its subsets can have distance distributions with a wide range of behavior.

Basically, we are trying to solve the Erdős distance problem on subsets of the lattice.

Calculating error

How do we compare the distance distributions of subsets of the lattice with the distance distribution of the lattice?

- The $N \times N$ lattice has $\frac{N^2(N^2-1)}{2} \approx \frac{N^4}{2}$ total distances. A subset with p points has $\frac{p(p-1)}{2} \approx \frac{p^2}{2}$ total distances.
- So we scale the distance distribution of the subset up by $\frac{N^4}{p^2}$.
- Then, for each unique distance we find the absolute difference between the scaled subset frequency and the lattice frequency.
- We then average these difference to find the error.

Configurations

What configuration of p points maximizes error?

Configurations

What configuration of p points maximizes error?

Figure: $p = 4$

Configurations

What configuration of p points maximizes error?

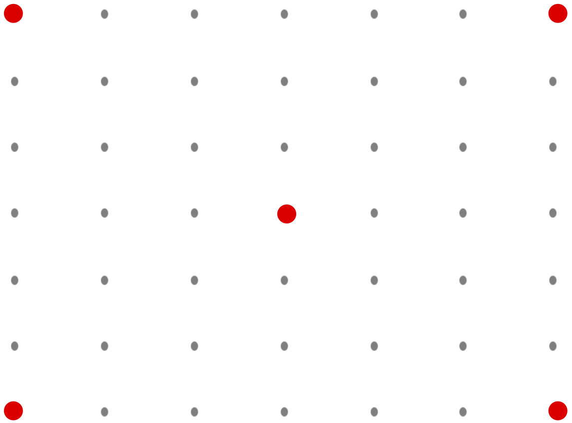


Figure: $p = 5$

Configurations

What configuration of p points maximizes error?

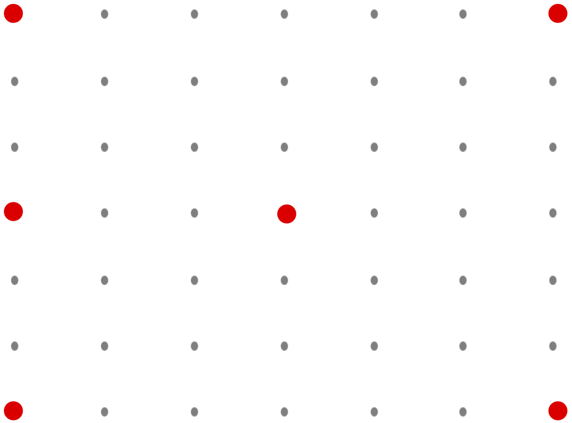


Figure: $p = 6$

Configurations

What configuration of p points maximizes error?

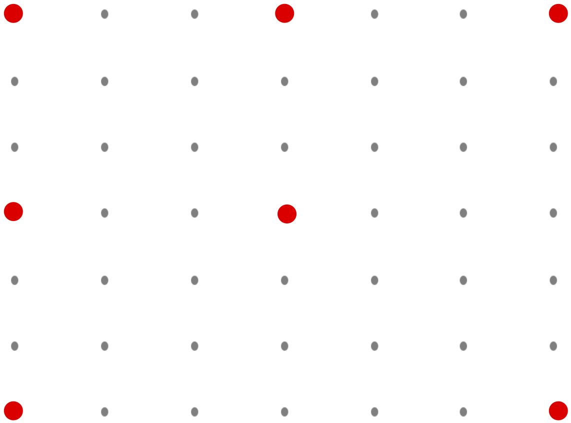


Figure: $p = 7$

Configurations

What configuration of p points maximizes error?

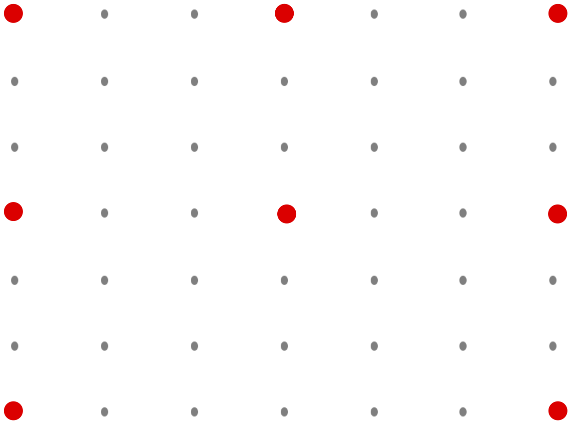


Figure: $p = 8$

Configurations

What configuration of p points maximizes error?

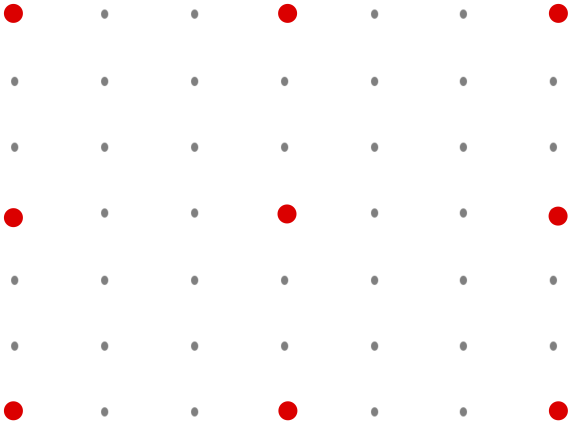


Figure: $p = 9$

Configurations

What configuration of p points maximizes error?

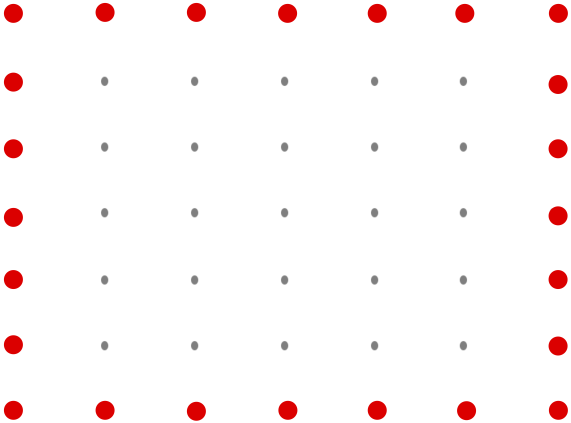


Figure: $p = 4(N-1)$

Configurations

What configuration of p points maximizes error?

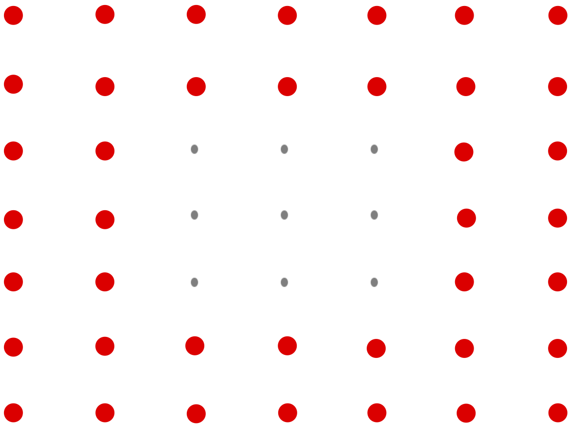


Figure: $p = 4(N-1) + 4(N-3)$

Configurations

What configuration of p points maximizes error?

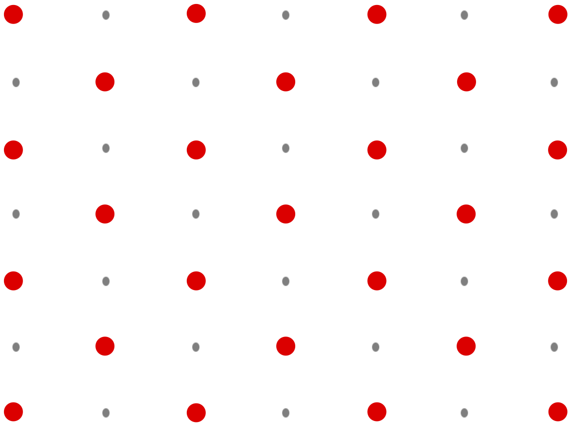
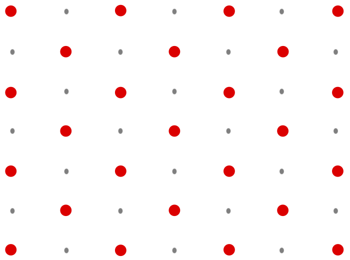


Figure: $p = \left\lceil \frac{N^2}{2} \right\rceil$

Error Calculations

How do we calculate the error for one of these configurations?

Ex: for $p = \left\lceil \frac{N^2}{2} \right\rceil$ we have a checkerboard lattice.



We simplify by looking at $\sqrt{a^2+b^2}$ instead of $\sqrt{d}$.

$\sqrt{a^2+b^2}$ only appears if a and b are both even or both odd.

Error Calculations

The error is:

$$\frac{4}{N+2}\left[\frac{3}{4}\left(4\left(\frac{N(5N-1)}{6}\right)-\frac{N(5N-1)}{6}\right)+\frac{1}{4}\left(\frac{N(5N-1)}{6}\right)\right]+$$

$$\frac{N-2}{N+2}\left[\frac{1}{2}\left(4\left(\frac{N(3N-1)}{3}\right)-\frac{N(3N-1)}{3}\right)+\frac{1}{2}\frac{N(3N-1)}{3}\right]$$

$$=2N^2-\frac{25N}{6}-\frac{121}{21(N+2)}+\frac{188}{21(3N-1)}+\frac{71}{6}$$

Lower Bounds

How do you calculate a lower bound for the error?

- Scale frequency down by $\frac{p^2}{N^4}$ and round frequency to nearest whole number
- We call this the optimal distance distribution for p points

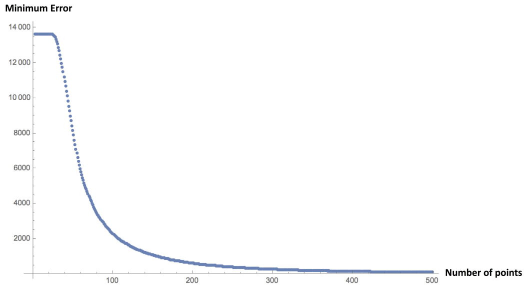


Figure: data for $N = 100$

Calculating Lower Bound

$$\text{Error} \geq \begin{cases} \frac{N^3}{N+2} + \frac{N^2}{N+2} - \frac{10N}{3(N+2)} & \text{if } \; p \leq \frac{\log_5(N)}{5}\left(11 - 2\sqrt{10}\right), \\ \frac{N^4}{8p^2} & \text{if } \; p \text{ sufficiently large.} \end{cases}$$

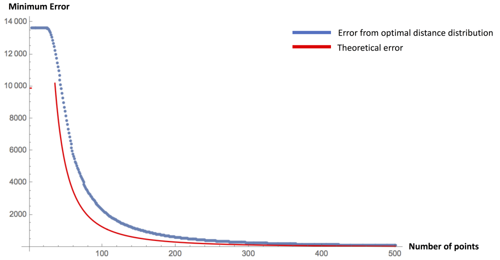


Figure: $N = 100$

Calculating Lower Bound

$$\text{Error} \geq \begin{cases} \frac{N^3}{N+2} + \frac{N^2}{N+2} - \frac{10N}{3(N+2)} & \text{if } p \leq \frac{\log_5(N)}{5}\left(11 - 2\sqrt{10}\right), \\ \frac{N^4}{8p^2} & \text{if } p \text{ sufficiently large.} \end{cases}$$

Optimal distance distribution is actually the empty distance distribution.

So error is the average frequency in the full lattice.

If the most frequent distance on the lattice is F, then p small enough that $N^4/p^2 > 2F$ will be sufficient. (Error contribution for adding any distance will result in strict increase in absolute difference).

$p \leq \log_5(N)(11 - 2\sqrt{10})/5$ ensures $N^4/p^2 > 2F$.

Lower Bound Formula

$$\text{Error} \;\geq\; \begin{cases} \frac{N^3}{N+2} + \frac{N^2}{N+2} - \frac{10N}{3(N+2)} & \text{if } \; p \leq \frac{\log_5(N)}{5}\left(11 - 2\sqrt{10}\right), \\ \frac{N^4}{8p^2} & \text{if } \; p \text{ sufficiently large.} \end{cases}$$

- Some intuition: the average error should be around $\frac{p^2}{4N^4}$
- *However*, for small p, many original frequencies are very close to 0, so average is smaller than $\frac{N^4}{4p^2}$

Further work

- Characterizing sets of maximum error.
- Characterizing sets of minimum error.
- Extending results to other lattice structures.

Acknowledgements

Thanks to

- Jason Zhao (Collaborator)
- Prof. Eyvindur Palsson (Mentor),
- Prof. Steven J. Miller (Mentor, NSF Grant DMS1561945),
- the SMALL REU program (NSF grant DMS1947438)
- Yale University
- The Quebec-Maine Conference organizers,
- and to you, for your attention today!

Questions?

Elena Kim
(elena.kim@pomona.edu)

Fernando Trejos Suárez
(fernando.trejos@yale.edu)