On Incomplete Distance Sets in $\mathbb{Z}_p \times \mathbb{Z}_p$

Adam Scrivener

April 2016

1 Introduction

In this paper we discuss the Erdős-Falconer distance problem. The classical Erdős distance problem in $\mathbb{R}^d$, $d \geq 2$, asks for the smallest possible size of $\Delta(E) = \{|x-y| : x, y \in E\}$ with $E \subset \mathbb{R}^d$ a finite set.

An analogous problem is the Falconer distance problem which asks how large does the Hausdorff dimension of a compact set $E \subset \mathbb{R}^d$, $d \geq 2$, needs to be to ensure that the Lebesgue measure of $\Delta(E)$, defined as above, is positive.

The Erdős-Falconer distance problem in $\mathbb{Z}_p^2$ has features of both the Erdős and the Falconer distance problems. Let $E \subset \mathbb{Z}_p^2$ and define

$$\Delta(E) = \{||x-y|| : x, y \in E\},$$

where

$$||x|| = x_1^2 + x_2^2.$$

taken modulo p. The Erdős-Falconer distance problem is the following: how large does $|E|$ have to be to ensure that $\Delta(E) = \mathbb{Z}_p$?

In this paper, we argue that for any $u \in [0, \frac{1}{2})$, for all but finitely many primes p, there is a set of size at least $u \cdot p \log_2(p)$ which has an incomplete distance set, giving an explicit lower bound on the size of the largest such set.

2 Definitions

If $E \subset \mathbb{Z}_p \times \mathbb{Z}_p$, define $\Delta(E) = \{t | \exists x, y \in \mathbb{Z}_p \times \mathbb{Z}_p : d(x,y) = t\}$, where $d(x,y) = (x_1 - y_1)^2 + (x_2 - y_2)^2$ taken modulo p. Define $\mathscr{S}(p)$ to be the size of the largest such subset E which has distance set $\Delta(E) \subsetneq \mathbb{Z}_p$.

3 Previous Results

Iosevich and Rudnev showed that if $|E| > 2p^{\frac{3}{2}}$, then $\Delta(E) = \mathbb{Z}_p$. Here we sketch that proof.

Let E be a subset of $\mathbb{Z}_p \times \mathbb{Z}_p$. Let $1_A(\cdot)$ be the indicator function of a set A. Then, define

$$\nu(t) = \sum_{x,y \in \mathbb{Z}_p^2} 1_E(x) 1_E(y) 1_{S_t}(x-y),$$

where

$$S_t = \{x \in \mathbb{Z}^{p2} : ||x|| = t\}.$$

Theorem 1. *If $|E| > 0$, $\nu(t) > 0$ for every $t \in \mathbb{Z}_p$.*

Proof. In order to prove this theorem, we begin with a couple of lemmas.

Lemma 1. *Suppose that $p \equiv 1$(mod 4). Then*

$$|S_t| = p - 1.$$

If $p \equiv 3$(mod 4), then

$$|S_t| = p + 1.$$

Lemma 2. *Suppose that $m \neq$ (0,0). Then*

$$|\widehat{1}_{S_t}(m)| \leq 2p^{-\frac{3}{2}}.$$

Now, by Fourier Inversion, we have

$$\nu(t) = \sum_{x,y \in \mathbb{Z}_p^2} 1_E(x) 1_E(y) \sum_{m \in \mathbb{Z}_p^2} \widehat{1}_{S_t}(m) \chi((x-y) \cdot m).$$

Reversing the order of summation and using the definition of the Fourier transform, we obtain

$$p^4 \sum_{m \in \mathbb{Z}_p^2} |\widehat{1}_E(m)|^2 \widehat{1}_{S_t}(m)$$

$$= (|E|)^2 \cdot |S_t| \cdot p^{-2} + p^4 \sum_{m \neq (0,0)} |\widehat{1}_E(m)|^2 \widehat{1}_{S_t}(m). \tag{1}$$

With Lemma 1, the first term in equation (1) equals

$$(|E|)^2 p^{-1} \pm (|E|)^2 p^{-2}.$$

Assuming Lemma 2, the second term in equation (1) is bounded by

$$p^4 \cdot 2p^{-\frac{3}{2}} \cdot \sum_{m \in \mathbb{Z}_p^2} |\widehat{1}_E(m)|^2$$

$$= 2p^{\frac{1}{2}} \cdot |E|$$

by Plancherel.

It follows that the second term in equation (1) is smaller than the first term if $|E| > 2p^{\frac{3}{2}}$, which completes the proof since the positivity of the expression of equation (1) says that

$$\nu(t) = \sum_{||x-y||=t} E(x)E(y) > 0,$$

which means that for every $t \in \mathbb{Z}_p$ there exists $x, y \in E$ such that $||x - y|| = t$. That is to say that for every $t \in \mathbb{Z}_p$, there is a pair of points in E whose vector difference is in S_t.

Thus matters have been reduced to proving Lemma 2. We have

$$\widehat{1}_{S_t}(m) = p^{-2} \sum_{||x||=t} \chi(-x \cdot m),$$

which equals

$$p^{-3} \sum_{x\in\mathbb{Z}_p^2} \sum_{x\in\mathbb{Z}_p} \chi(s||x|| - t)\chi(-x \cdot m). \tag{2}$$

Then we have

$$sx_j^2 - x_j m_J = s\left(x_j^2 - \frac{m_j}{s}\right) = s\left(\left(x_j - \frac{m_j}{2s}\right)^2 - \frac{m_j^2}{4s^2}\right).$$

Plugging this back into equation (2) and making a change of variables

$$y_j = x_j - \frac{m_J}{2s},$$

we obtain

$$p^{-3} \sum_{x\neq 0} \chi\left(-st - \frac{||m||}{4s}\right) \sum_{y\in\mathbb{Z}_p^2} \chi(s||y||). \tag{3}$$

Then, we can write the inner sum of equation (3) in terms of a function $\psi(s)$ as

$$\sum_{y\in\mathbb{Z}_p^2} \chi(s||y||) = p \cdot \psi^2(s),$$

which $|\psi(s)| = 1$.

Plugging this into equation (3), we obtain

$$p^{-2} \sum_{x\neq 0} \chi\left(-st - \frac{||m||}{4s}\right) \psi^2(s).$$

To conclude, we use the following result due to Andre Weil, which we use as a black box.

Theorem 2. *With the notation above,*

$$\left|\sum_{x\neq 0}\chi\left(-st-\frac{||m||}{4s}\right)\psi^2(s)\right|\leq 2\sqrt{p}.$$

This gives us the bound

$$\left|\widehat{1}_{S_t}(m)\right|=p^{-2}\left|\sum_{x\neq 0}\chi\left(-st-\frac{||m||}{4s}\right)\psi^2(s)\right|\leq 2p^{-\frac{3}{2}},$$

thereby proving Lemma 2 and therefore Theorem 1. □

4 Computational Results

Whereas the aim of the last section was to give an upper bound on $\mathscr{S}(p)$, the goal of the following sections is to find a large subset which has an incomplete distance set, thereby giving a lower bound on $\mathscr{S}(p)$.

4.1 Brute force observations in $\mathbb{Z}_5$

Below is a visualization of $\mathbb{Z}_5 \times \mathbb{Z}_5$, with an example of a subset

$$E=\{(0,1),(1,1),(1,2),(2,3),(3,4),(4,4)\}:$$

$$\begin{array}{ccccc}
- & - & - & X & X \\
- & - & X & - & - \\
- & X & - & - & - \\
X & X & - & - & - \\
- & - & - & - & -
\end{array}$$

whose distance set is missing 4.

To get a handle on this problem, scripts were run for $\mathbb{Z}_5 \times \mathbb{Z}_5$. Interestingly, the following subsets of size 10 were found to have incomplete distance sets:

$$\begin{array}{ccccccccccccccccc}
- & - & - & X & X & & - & X & - & X & - & & X & - & X & - & - \\
- & - & X & X & - & & - & - & X & - & X & & X & - & X & - & - \\
- & X & X & - & - & & X & - & - & X & - & & X & - & X & - & - \\
X & X & - & - & - & & - & X & - & - & X & & X & - & X & - & - \\
X & - & - & - & X & & X & - & X & - & - & & X & - & X & - & -
\end{array}$$

$$\begin{array}{ccccccccccccccccc}
- & - & - & - & - & & - & X & X & - & - & & X & - & - & - & X \\
X & X & X & X & X & & - & - & X & X & - & & - & - & - & X & X \\
X & X & X & X & X & & - & - & - & X & X & & - & - & X & X & - \\
- & - & - & - & - & & X & - & - & - & X & & - & X & X & - & - \\
- & - & - & - & - & & X & X & - & - & - & & X & X & - & - & -
\end{array}$$

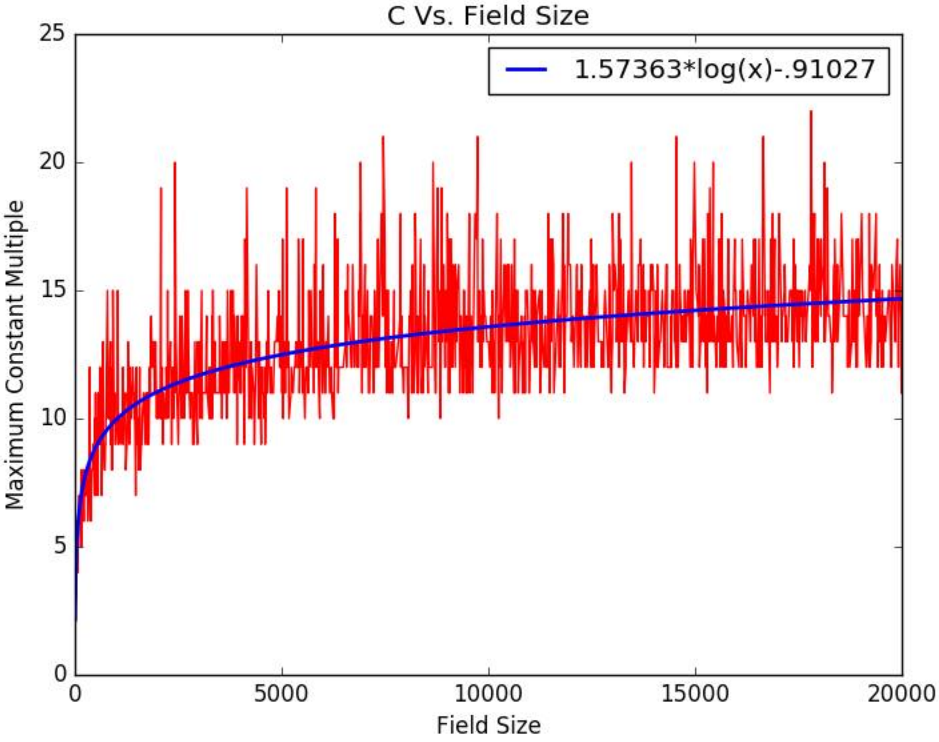
Observe that these subsets are pairs of parallel lines in $\mathbb{Z}_5 \times \mathbb{Z}_5$. In fact, by a brute force algorithm, one can check that every pair of parallel lines in $\mathbb{Z}_5 \times \mathbb{Z}_5$ has an incomplete distance set, and they are the only subsets of size 10 that enjoy this property. Furthermore, again by a simple brute force check, every subset of $\mathbb{Z}_5 \times \mathbb{Z}_5$ with exactly 11 points has a complete distance set.

This leads to a natural conjecture, namely that for any p, there exists a number of lines C such that the largest subsets of $\mathbb{Z}_p \times \mathbb{Z}_p$ with incomplete distance set are exactly the sets which are collections of C parallel lines. However, this can be shown to be incorrect, as one can algorithmically check that in $\mathbb{Z}_{19} \times \mathbb{Z}_{19}$, 4 vertical lines at indices 0,1,2, and 3, have an incomplete distance set, but 4 lines with column indices 0,1,2, and 4 have a complete distance set.

Another natural question is: how many lines can the set contain while still having an incomplete distance set?

4.2 Further simulations

To answer this question, a script was written which finds, for a given prime p, the largest number of adjacent vertical lines in $\mathbb{Z}_p \times \mathbb{Z}_p$ with an incomplete distance set. Note that it is easy to compute the distance set of adjacent vertical lines as we only need to compute the distances from every point to a fixed point in the line with the smallest column index. The script was run on every prime less than 20,000, and the number of adjacent lines (C) was plotted as a function of p, alongside a function which is $\Theta(log(x))$, computed using Mathematica's FindFit function.



This gives a strong indication that the largest number of adjacent lines with incomplete distance set grows with $\log(p)$. Note that this would be a lower bound on $\mathscr{S}(p)$, since there could possibly be larger subsets with an incomplete distance set which are not adjacent vertical lines.

5 A Rephrasing in Terms of Quadratic Nonresidues

It turns out that finding a lower bound on the largest number of adjacent lines with an incomplete distance set can be phrased as a problem related to quadratic nonresidues. Explicitly,

Theorem 3. *Let p be prime, and $k \in [1, \dots, p]$. k adjacent lines have an incomplete distance set if and only if there is an element $y \in \mathbb{Z}_p$ such that each of $y, y-1^2, y-2^2, \dots, y-(k-1)^2$ are all quadratic nonresidues of $\mathbb{Z}_p$.*

Proof. First, define $A_D = \{y | x^2 + D^2 = y \text{ for some } x \in \mathbb{Z}_p\} \subset \mathbb{Z}_p$. Then, notice that the distance set of k vertical lines at column indices $l_1, \dots, l_k$ is

$$\bigcup_{i,j} A_{(l_i - l_j)}$$

This follows since if x and y are two points in lines l_i, l_j, the horizontal distance between them is $(l_i - l_j)^2$, and thus their distance is contained in $A_{(l_i-l_j)}$. Further, if $d \in \cup_{i,j} A_{(l_i-l_j)}$, there are lines at indices l_i, l_j where $d \in A_{(l_i-l_j)}$. So, $d = x^2 + (l_i - l_j)^2$ for some $x \in \mathbb{Z}_p$. This distance is achieved between points $(l_i, 0)$ and (l_j, x).

Then, note that $A_D = \{y | x^2 + D^2 = y \text{ for some } x \in \mathbb{Z}_p\} = \{y | x^2 = y - D^2 \text{ for some } x \in \mathbb{Z}_p\} = \{y | y - D^2 \text{ is a quadratic residue mod } p\}$.

We want to know when $p - |\cup_{i,j} A_{(l_i-l_j)}| > 0$. By convention, if $S \subset \mathbb{Z}_p$, let $\overline{S}$ be the elements in $\mathbb{Z}_p - S$. Then, we have that

$$\begin{aligned} p - \left| \bigcup_{i,j} A_{(l_i-l_j)} \right| &= \left| \overline{\bigcup_{i,j} A_{(l_i-l_j)}} \right| \\ &= \left| \bigcap_{i,j} \overline{A_{(l_i-l_j)}} \right| \\ &= \left| \bigcap_{i,j} \{y | y - (l_i - l_j)^2 \text{ is a quadratic nonresidue of } \mathbb{Z}_p\} \right| \\ &= \left| \bigcap_{i<j} \{y | y - (l_i - l_j)^2 \text{ is a quadratic nonresidue of } \mathbb{Z}_p\} \right| \end{aligned}$$

since $A_{(l_i-l_j)^2} = A_{(l_j-l_i)^2}$. Then, assuming our k lines are all adjacent, starting at column index 0, we have that $\{(l_i - l_j)^2 | 0 \le i, j < k\} = \{i^2 | 0 \le i < k\}$, since line l_j has horizontal distance j^2 from line l_0, and any two lines l_i, l_j (with $i > j$ without loss of generality) have horizontal distance $(l_i - l_j)^2 = i^2$ for some $i < k$.

And so we have

$$\begin{aligned} &\left| \bigcap_{i<j} \{y | y - (l_i - l_j)^2 \text{ is a quadratic nonresidue of } \mathbb{Z}_p\} \right| \\ = &\left| \bigcap_{0 \le i < k} \{y | y - i^2 \text{ is a quadratic nonresidue of } \mathbb{Z}_p\} \right| \end{aligned}$$

And thus, in particular, if each of $y, y - 1^2, y - 2^2, \ldots, y - (k-1)^2$ are quadratic nonresidues of $\mathbb{Z}_p$, this expression is positive, and therefore the distance set of k adjacent lines is incomplete. □

6 Distribution of Quadratic Nonresidues

Since there are $(p-1)/2$ quadratic nonresidues in $\mathbb{Z}_p$, it makes intuitive sense that there is about a $1/2^k$ probability that all of $y, y-1^2, \ldots, y-(k-1)^2$ are nonresidues, assuming that nonresidues behave essentially randomly. Then, it would stand to reason that if k is about $\log p$, then this probability is nonzero. The following argument formalizes this intuition by giving exact bounds on the likelihood that each of $y-s_1, y-s_2, \ldots, y-s_k$ are quadratic nonresidues, where the s_i *'s* are distinct elements in $\mathbb{Z}_p$.

Definition 1. *A set S of random boolean variables are $\epsilon-$ independent if for every nonempty subset $T \subset S$, XOR_T, the probability that an odd number of elements of T are 1, is within ϵ of $\frac{1}{2}$.*

Theorem 4. *Let x be a random number in $\mathbb{Z}_p$. Let $S = \{s_1, \ldots, s_k\}$ be a subset of Z_p. Then define random variables A_i such that $A_i = 1$ if $x + s_i$ is a quadratic nonresidue modulo p and 0 otherwise. Then the A_i 's are ϵ-independent with $\epsilon = k(3+\sqrt{p})/2p$.*

Proof. Let $T = \{s_{i_1}, \ldots, s_{i_t}\}$ be any nonempty subset of S. Define

$$X_p(z) = \begin{cases} 1 & z \text{ is a nonzero quadratic residue modulo p} \\ 0 & z = 0 \\ -1 & z \text{ is a quadratic nonresidue modulo p} \end{cases}$$

Then, let $f_T(x) = \prod_{j=1}^{t}(x + s_{i_j})$. Further, let X^+ be the number of elements x in $\mathbb{Z}_p$ such that $X_p(f_T(x)) = 1$, and X^- be the number of elements x in $\mathbb{Z}_p$ such that $X_p(f_T(x)) = -1$. Note that $X_p(z) = 0$ if and only if $z = 0$, so there are exactly t elements in $\mathbb{Z}_p$ where $X_p(f_T(x)) = 0$, namely $\{-s_{i_1}, \ldots, -s_{i_t}\}$. Thus $X^+ + X^- = p - t$.

By the Weil bound, we have

$$t\sqrt{p} > \left|\sum_{x \in \mathbb{Z}_p} X_p\left(\prod_{j=1}^{t}(x + s_{i_j})\right)\right| = |X^+ - X^-| = |p - t - 2X^-|$$

Then, divide by $2p$ to get

$$\frac{t}{2\sqrt{p}} > \left|\frac{1}{2} - \frac{t}{2p} - \frac{X^-}{p}\right|$$

Then notice X^-/p is the probability that $f_T(x)$ is a quadratic nonresidue. Thus this probability is within $t/(2p) + t/(2\sqrt{p})$ of $\frac{1}{2}$. Also note that if $f_T(x) \neq 0$, then $f_T(x)$ is a nonresidue if and only if an odd number of A_{i_j} *'s* are 1, that is $XOR_T = 1$. Then we have that the probability that $f_T(x) = 0$ is t/p, so XOR_T deviates from $\frac{1}{2}$ by at most

$$\frac{t}{p} + \frac{t}{2p} + \frac{t}{2\sqrt{p}} = \frac{t(3+\sqrt{p})}{2p} \leq \frac{k(3+\sqrt{p})}{2p}$$

□

Lemma 3. *If S is a set of ϵ-independent random variables, then the joint distribution of all of the random variables deviates from the joint distribution of independent fair coins by no more than 2ϵ.*

Combining Lemma 3 and Theorem 4, the probability that the A_i $'s$ defined as in Theorem 1 are all equal to 1, that is that each of $x + s_i$ are nonresidues, is within $k(3 + \sqrt{p})/p$ of $(\frac{1}{2})^k$. Thus, in particular, if $p(\frac{1}{2})^k > k(3 + \sqrt{p})$, then the probability is positive, and thus there exists an $x \in \mathbb{Z}_p$ such that each of $x + s_i$ are quadratic nonresidues.

Putting this all together, we have our main theorem:

Theorem 5. *Take any $u \in [0, \frac{1}{2})$. Then, for all but finitely many primes p, $\mathscr{S}(p) \geq u \cdot p \log_2(p)$.*

Proof. Let p be prime, and let $u \in [0, \frac{1}{2})$. Then, let $S = \{0, -1^2, -2^2, \ldots, -(k-1)^2\}$, where k $= \lceil u \log_2(p) \rceil$. Note that these are all distinct since each nonzero value of $0, -1, \ldots, -(k-1)$ appear in the second half of $\mathbb{Z}_p$ $(\frac{1}{2}\log_2(p) < p/2)$, and thus no pair are additive inverses, and so their squares are all unique.

Then, $p(\frac{1}{2})^k$ grows faster than $k(3 + \sqrt{p})$, so for all but finitely many p, $p(\frac{1}{2})^k > k(3 + \sqrt{p})$. Thus, from the argument above, for these p, there exists an $x \in \mathbb{Z}_p$ such that each of $x + s_i$ are quadratic nonresidues. Then, by Theorem 3, this means that k adjacent columns in $\mathbb{Z}_p \times \mathbb{Z}_p$ have an incomplete distance set (in fact, it must be missing distance x), and the size of the set containing these columns is $p\lceil u \log_2(p) \rceil \geq u \cdot p \log_2(p)$. □

7 Conclusion

Putting the results of this paper with the result from Iosevich and Rudnev, we have that for any $u \in [0, \frac{1}{2})$, for all but finitely many primes p,

$$u \cdot p \log_2(p) \leq \mathscr{S}(p) \leq 2p^{\frac{3}{2}}.$$

However, the author conjectures that $\mathscr{S}(p) \in O(p \log_2(p))$, as evidenced both by the data presented in the graphs in this paper and the further assumption that adjacent lines are near-optimal for creating subsets of $\mathbb{Z}_p \times \mathbb{Z}_p$ which have small distance sets.

The author also notes that if a subset of size k of $\mathbb{Z}_p$ can always be found which is near $\log_2(p)$ in size which has a squared distance set of size strictly less than k, then we can increase the lower bound proved in this paper. This will allow us to choose our k lines to have an even smaller distance set than k adjacent lines. However, to increase our lower bound, the size of the squared distance set would have to be smaller than k by a factor that grows faster than a constant, and the author conjectures that this is impossible.

Figure 1 below shows the accuracy of the lower bound produced in this paper as well as a potential (unproven) upper bound, as evidence of the $O(p \log_2(p))$ conjecture

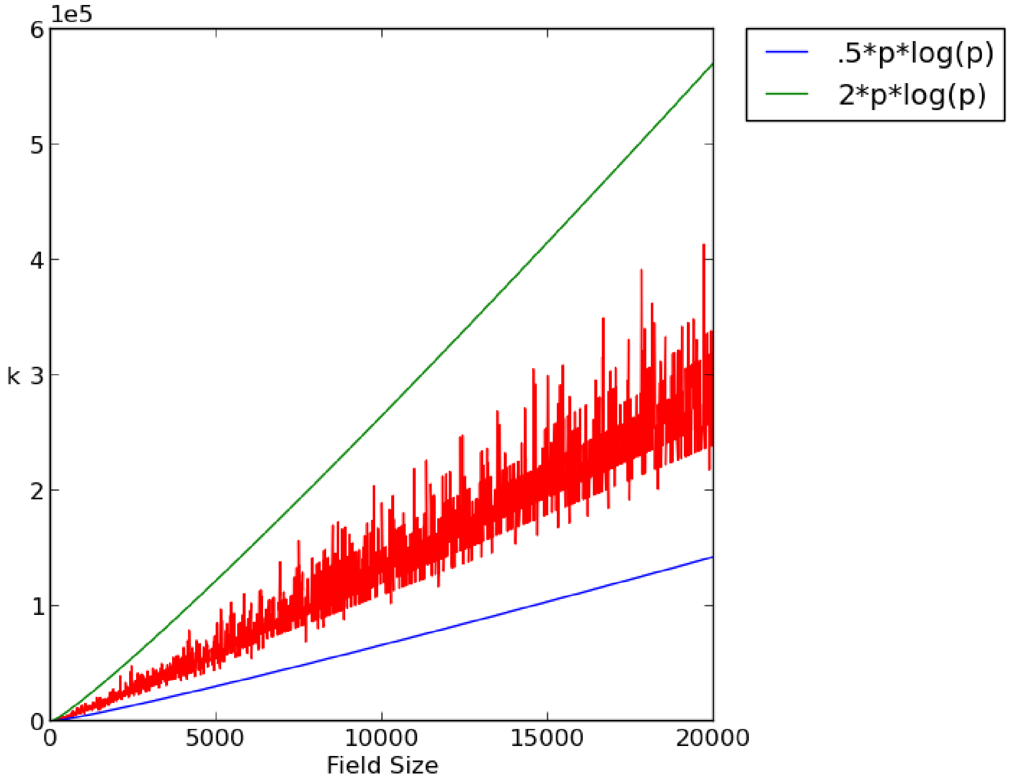


Figure 1: Size of largest set of adjacent lines with incomplete data set, k, plotted vs. p, the field size. Lower bound of $\mathscr{S}(p)$ proven in this paper in blue, and a function which is $O(p\log_2(p))$ in green.