Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

Counting k-arcs in $\mathbb{F}_q^2$

O. Roche-Newton, A. Warren

RICAM-Report 2020-42

Counting k-arcs in $\mathbb{F}_q^2$

Oliver Roche-Newton and Audie Warren

September 29, 2020

Abstract

An arc is a subset of $\mathbb{F}_q^2$ which does not contain any collinear triples. Let $A(q,k)$ denote the number of arcs in $\mathbb{F}_q^2$ with cardinality k. This paper is concerned with estimating the size of $A(q,k)$ when k is relatively large, namely $k = q^t$ for some $t > 0$. Trivial estimates tell us that

$$\binom{q}{k} \leq A(q,k) \leq \binom{q^2}{k}.$$

We show that the behaviour of $A(q,k)$ changes significantly close to $t = 1/2$. Below this threshold an elementary argument is used to prove that the trivial upper bound above cannot be improved significantly.

On the other hand, for $t \geq 1/2 + \delta$, we use the theory of hypergraph containers to get an improved upper bound

$$A(q,k) \leq \binom{q^{2-t+2\delta}}{k}.$$

This technique is also used to prove a finite field analogue of a result of Balogh and Solymosi [2], with a better exponent: there exists a subset $P \subset \mathbb{F}_q^2$ which does not contain any collinear quadruples, but with the property that for every $P' \subset P$ with $|P'| \geq |P|^{3/4+o(1)}$, P' contains a collinear triple.

1 Introduction

1.1 Basics

Let $\mathbb{F}_q$ be the finite field of order $q = p^r$ for some prime p. An arc in $\mathbb{F}_q^2$ is a subset of $\mathbb{F}_q^2$ with no three points collinear. A k-arc is an arc with cardinality k. Define $A(q,k)$ to be the number of such k-arcs. In this paper, we are interested in estimating $A(q,k)$, particularly in the case when k is relatively large with respect to q.

Previous work on this problem has focused on what happens for small values of k, and exact formulas for $A(q,k)$ are known for $2 \leq k \leq 9$, see [6] and the references within. For slightly larger

values of k, Kaipa [7] provided an upper bound for $A(q,k)$. However, the result of Kaipa is only effective when $k = O(\log q)$.

Our focus is on bounding $A(q,k)$ in the case $k = q^t$, for some $t > 0$. To provide some context, let us observe some trivial bounds for $A(q,k)$. Firstly, note that any subset of an arc is also an arc. Since the set

$$C = \{(x, x^2) : x \in \mathbb{F}_q\}$$

is an arc of cardinality q, all of the $\binom{q}{k}$ subsets of C of size k are arcs. Hence $A(q,k) \geq \binom{q}{k}$

A trivial upper bound for $A(q,k)$ is given by the number of subsets of $\mathbb{F}_q^2$ of cardinality k. To summarise, we have the following trivial bounds for $A(q,k)$:

$$\binom{q}{k} \leq A(q,k) \leq \binom{q^2}{k}. \tag{1}$$

1.2 Main results

The main purpose of this paper is to show that a threshold occurs at $t = 1/2$ (recall that we set $k = q^t$), at which point the behaviour of $A(q,k)$ appears to change considerably. For $t \leq 1/2$, an elementary probabilistic argument gives a rather precise description of $A(q,k)$.

Theorem 1. *Suppose $k \leq \frac{q^{1/2}}{1+\delta}$ for some $\delta > 0$. Then there exist an absolute constant $c > 0$ and a constant $C = C(\delta) > 0$ such that*

$$\binom{q^2}{k} e^{\frac{-Ck^3}{q}} \leq A(q,k) \leq \binom{q^2}{k} e^{\frac{-ck^3}{q}}.$$

Note that this value is rather close to the trivial upper bound given in (1). For example,

$$\binom{q^2}{k} e^{\frac{-Ck^3}{q}} \geq \binom{c'q^2}{k}$$

for some constant $c' > 0$ (depending on C), provided that q is sufficiently large with respect to C.

The argument leading to Theorem 1 fails for $t > 1/2$, and instead the machinery of hypergraph containers is used to give a much stronger upper bound. Our main result is the following, which deals with this case.

Theorem 2. *Let $\delta > 0$, and suppose that q is a sufficiently large (with respect to δ) prime power. Then for all $t \geq \frac{1}{2} + \delta$*

$$A(q,k) \leq \binom{2q^{2-t+3\delta/2}}{k}. \tag{2}$$

1.3 Two thresholds

These results show that there is a sudden change in the behaviour of $A(q,k)$, with the exponent in the top part of the binomial coefficient suddenly dropping from almost 2 to at most 3/2. For much

larger values of t (i.e. for $t \to 1$) the upper bound given by Theorem 2 approaches the trivial lower bound from (1).

It is possible that this transition is even more sharp, and it is even conceivable that $A(q,k)$ is close to the trivial lower bound from (1) for all $t > 1/2$, but we were not able to prove or disprove this.

Furthermore, another sharp threshold is present in the statement of Theorem 1, occurring when $k \approx q^{1/3}$. The function $e^{\frac{-Ck^3}{q}}$ estimates the probability that a random set of points of size $k = q^t$ forms an arc. When $t < 1/3$ this probability tends to 1 as q grows, but when $t > 1/3$ it tends to zero. Some similar occurrences of such sharp probabilistic transitions in behaviour of combinatorial structures can be found in, for example, [4], [9], and [11].

Notation

Throughout the paper, the standard notation $\ll, \gg$ and respectively O and Ω is applied to positive quantities in the usual way. That is, $X \gg Y$, $Y \ll X$, $X = \Omega(Y)$ and $Y = O(X)$ all mean that $X \geq cY$, for some absolute constant $c > 0$. If both $X \ll Y$ and $Y \ll X$ hold we write $X \approx Y$, or equivalently $X = \Theta(Y)$. All logarithms are in base 2, unless stated otherwise.

1.4 Containers, supersaturation and the Balogh-Solymosi Theorem

The proof of Theorem 2 makes use of hypergraph containers. The theory of hypergraph containers was developed independently by Balogh, Morris and Samotij [1] and Saxton and Thomason [10]. We defer the full statement of the container theorem we use until Section 2.1. Roughly speaking, it says that if a hypergraph has a reasonably good edge distribution, we can obtain strong information about where the independent sets of the hypergraph may be found.

This new method has led to several significant breakthroughs in combinatorics in recent years, most notably in the field of extremal graph theory. Of more relevance to this paper is the work of Balogh and Solymosi [2], in which they prove the existence of point sets in $\mathbb{R}^2$ which do not contain collinear quadruples, but all large subsets contain a collinear triple.

Theorem 3 ([2], Theorem 2.1)**.** *For all $\delta > 0$, there exists $n_0 \in \mathbb{Z}$ such that for all $n \geq n_0$ there exists a set $P \subset \mathbb{R}^2$ with $|P| = n$ with the following properties. P does not contain any collinear quadruples, but for every subset $P' \subset P$ with $|P'| \geq n^{5/6+\delta}$, P' contains a collinear triple.*

A key step in [2] is the application of the container theorem to give detailed information about the subsets of $[n]^3$ which do not contain any collinear triples. It is an adaptation of this argument to the $\mathbb{F}_q^2$ case which is used in the proof of Theorem 2.

In most container applications, an important component is a "supersaturation lemma". In the context of collinear triples, we have already seen an example of a set $P \subset \mathbb{F}_q^2$ of size q which does not contain any collinear triples. On the other hand, a celebrated result of Segre [12] implies that any set of $q+2$ points in $\mathbb{F}_q^2$ contains at least one collinear triple.

A supersaturation lemma in this setting is a statement which says that as we increase the cardinality of our set beyond the threshold at which we are guaranteed at least one collinear triple, we find *many* such triples. The precise statement and its proof can be found in Section 2.2.

Our supersaturation lemma in $\mathbb{F}_q^2$ is optimal up to constants, whereas the corresponding result over $[n]^3$ in [2] is expected not to be. This allows us to give the following analogue of Theorem 3 in the $\mathbb{F}_q^2$ with better exponents.

Theorem 4. *For all $\delta > 0$ and for all sufficiently large (depending on δ) prime powers q, there exists a set $P \subset \mathbb{F}_q^2$ with $|P| \gg q^{2/3}$ having the following properties. P does not contain any collinear quadruples, but for every subset $P' \subset P$ with $|P'| \geq |P|^{3/4+\delta}$, P' contains a collinear triple.*

Given two integers $k < l$, define

$$f_{k,l}(n) := \min_{P} \max_{P' \subset P : P' \text{ does not contain a collinear } k\text{-tuple}} |P'|$$

where the minimum is taken over all $P \subset \mathbb{F}_q^2$ such that $|P| = n$ and P does not contain any collinear l-tuples. Theorem 4 says that $f_{3,4}(n) \leq n^{3/4+\delta}$.

It is not clear what the optimal exponent should be for $f_{3,4}(n)$. A simple greedy algorithm argument shows that any set of n points with no collinear quadruples contains a subset of size $\Omega(n^{1/2})$ with no collinear triples, that is, $f_{3,4}(n) \gg n^{1/2}$. A small improvement to this bound in the Euclidean setting was obtained by Füredi [5], using a deep graph theoretical result of Komlós, Pintz and Szemerédi [8].

The argument used in the proof of Theorem 4 works more effectively to give near-optimal bounds for $f_{3,k}(n)$ as k increases.

Theorem 5. *For all $\delta > 0$ and for all sufficiently large (depending on δ) prime powers q, there exists a set $P \subset \mathbb{F}_q^2$ with $|P| \gg q^{2-\frac{k}{k-1}}$ having the following properties. P does not contain any collinear k-tuples, but for every subset $P' \subset P$ with $|P'| \geq |P|^{\left(\frac{k-1}{k-2}\right)\left(\frac{1}{2}+\delta\right)}$, P' contains a collinear triple.*

Theorem 5 says that

$$f_{3,k}(n) \leq n^{\left(\frac{k-1}{k-2}\right)\left(\frac{1}{2}+\delta\right)} \tag{3}$$

for all $\delta > 0$. Taking k to be sufficiently large, the exponent in (3) gets arbitrarily close to $1/2$.

To find a lower bound for $f_{3,k}$, the greedy algorithm argument can be used to show that any set $P \subset \mathbb{F}_q^2$ of size n with no collinear k-tuples contains an arc $P' \subset P$ with $|P'| \gg n^{1/2}/k^{1/2}$. That is,

$$f_{3,k} \gg \frac{n^{1/2}}{k^{1/2}}.$$

Note that Theorem 5 implies Theorem 4.

1.5 Counting MDS codes

A linear code C of dimension k and length n over $\mathbb{F}_q$ is a linear subspace of dimension k in $\mathbb{F}_q^n$. The distance $d(c_1, c_2)$ between two elements of $c_1, c_2 \in C$ is the number of positions in which they differ. The minimum distance of C is the minimum value of $d(c_1, c_2)$ over all distinct $c_1, c_2 \in C$. A fundamental idea in coding theory is to construct codes with large minimum distance, since these codes have the greatest error detecting and correcting capabilities. The Singleton bound implies that the minimum distance of any linear code is at most $n-k+1$, and *maximum distance separable* (MDS) codes are those which attain this bound.

Let $B(q, n)$ denote the number of n-arcs in the projective plane $\mathbb{P}(\mathbb{F}_q^2)$. A connection between the number of projective arcs and MDS codes was established in [6]. Lemmas 1 and 2 therein imply that the number of MDS codes of dimension 3 and length n over $\mathbb{F}_q$ is

$$\frac{n!(q-1)^{n-2}}{q^3(q^2+q+1)(q+1)} B(q, n). \tag{4}$$

The arguments in this paper can also be framed in the projective setting to get similar bounds for $B(q, k)$. These modified versions of Theorems 1 and 2, combined with (4), give bounds for the number of $[n, 3]_q$ MDS codes. This is not the first time that containers have given applications in coding theory; Balogh, Treglown and Zsolt Wagner used the technique to count the number of t error correcting codes in [3].

2 Large arcs; $t > 1/2$

Throughout this section we assume $k = q^t$ for $t > 1/2$.

2.1 Statement of the container theorem

Before stating the theorem, we introduce some related quantities. For an $r-$uniform hypergraph $\mathcal{H} = (V, E)$ and $v \in V$, $d(v)$ denotes the degree of v. Let $d(\mathcal{H})$ denote the average degree of $\mathcal{H}$, so

$$d(\mathcal{H}) = \frac{1}{|V|} \sum_{v \in V} d(v) = \frac{r|E|}{|V|}.$$

We can also define the *co-degree* for a subset $S \subseteq \mathcal{H}$ of vertices as

$$d(S) = \{e \in E(\mathcal{H}) : S \subseteq e\}.$$

Using this definition we define the *k-maximum co-degree* Δ_k as

$$\Delta_k(\mathcal{H}) = \max_{\substack{S \subseteq \mathcal{H} \\ |S|=k}} d(S).$$

Since edges are subsets of size r, we see that $\Delta_r(\mathcal{H}) = 1$ (or $\mathcal{H}$ is empty), and $\Delta_k(\mathcal{H}) = 0$ for all $k > r$. For any $V' \subset V$, $\mathcal{H}[V']$ denotes the subgraph induced by V'.

We now state the container theorem we need, which is Corollary 3.6 in [10].

Theorem 6. *Let $\mathcal{H} = (V, E)$ be an $r-$uniform hypergraph on n vertices, and let $\epsilon, \tau \in (0, 1/2)$. Suppose that*

$$2^{\binom{r}{2}-1} \sum_{j=2}^{r} \frac{\Delta_j(\mathcal{H})}{d(\mathcal{H}) \cdot \tau^{j-1} \cdot 2^{\binom{j-1}{2}}} \leq \frac{\epsilon}{12r!} \tag{5}$$

and

$$\tau < \frac{1}{200r \cdot r!}. \tag{6}$$

Then there exists a set $\mathcal{C}$ of subsets of V such that

1. *if $A \subset V$ is an independent set then there exists $C \in \mathcal{C}$ such that $A \subset C$;*
2. *$|E(\mathcal{H}[C])| \leq \epsilon |E(\mathcal{H})|$ for all $C \in \mathcal{C}$;*
3. *$\log |C| \ll_r n\tau \log(\frac{1}{\epsilon}) \log(\frac{1}{\tau})$.*

The set $\mathcal{C}$ above is referred to as the set of *containers,* and a set $C \in \mathcal{C}$ is itself a container.

2.2 Supersaturation lemma

The second point in the statement of Theorem 6 above says that the number of edges in $\mathcal{H}(C)$ must be smaller than the total number of edges in $\mathcal{H}$. As in most applications of containers, we aim for this to ensure that the number of *vertices* in the containers is small. This connection is given by the following supersaturation lemma.

Lemma 1. *Let $P \subset \mathbb{F}_q^2$ with $|P| \geq 4q$. Then the number of collinear triples defined by P is $\Omega\left(\frac{|P|^3}{q}\right)$.*

Proof. Let $\mathcal{L}$ denote the set of all lines in $\mathbb{F}_q^2$. Let

$$\mathcal{L}' := \{l \in \mathcal{L} : |l \cap P| \geq 3\}.$$

Then

$$(q+1)|P| = \sum_{l \in \mathcal{L}} |l \cap P| = \sum_{l \in \mathcal{L}'} |l \cap P| + \sum_{l \in \mathcal{L} \setminus \mathcal{L}'} |l \cap P| \leq \sum_{l \in \mathcal{L}'} |l \cap P| + 2q(q+1).$$

Since $|P| \geq 4q$, it follows that

$$\frac{(q+1)|P|}{2} \leq \sum_{l \in \mathcal{L}'} |l \cap P|.$$

Applying the Hölder inequality gives

$$q|P| \ll (q^2+q)^{2/3} \left(\sum_{l \in \mathcal{L}'} |l \cap P|^3 \right)^{1/3} \ll q^{4/3} \left(\sum_{l \in \mathcal{L}'} |l \cap P|^3 \right)^{1/3} \ll q^{4/3} \left(\sum_{l \in \mathcal{L}'} \binom{|l \cap P|}{3} \right)^{1/3},$$

where the last inequality uses the fact that $|l \cap P| \geq 3$. A rearrangement gives

$$\#\text{collinear triples in } P = \sum_{l \in \mathcal{L}'} \binom{|l \cap P|}{3} \gg \frac{|P|^3}{q}.$$

□

Lemma 1 is optimal, up to the implied constant. To see this, first observe that $\mathbb{F}_q^2$ contains $\Omega(q^5)$ collinear triples; there are $q^2 + q$ lines, and each line contains $\binom{q}{3} \gg q^3$ collinear triples. Now define a random set $P \subset \mathbb{F}_q^2$, where each element $x \in \mathbb{F}_q^2$ belongs to P with probability $p = q^{-s}$. With high probability we have $|P| = \Theta(q^{2-s})$, and the number of collinear triples in P is $\Theta(q^{5-3s})$, since there are q^5 collinear triples to begin with, and each one survives the random selection process with probability q^{-3s}. Furthermore, the condition that $s < 1$ is necessary, since we already saw a simple algebraic construction of an arc with cardinality q. The corresponding supersaturation result for the grid $[n]^3$ was given in [2, Lemma 4.2]. It is tentatively believed that this result is not optimal. Any improvement would result in an improved exponent in Theorem 3.

2.3 Application of the container theorem

The hard work is done in the following lemma.

Lemma 2. *Let $\delta > 0$ and suppose that q is a sufficiently large (with respect to δ) prime power. Let $1/2 + 2\delta \leq t \leq 1$. Then there exists a family $\mathcal{C}$ of subsets of $\mathbb{F}_q^2$ such that*

- $|\mathcal{C}| \leq 2^{c(\delta) q^{t-\delta} (\log q)^2}$,
- *For all $C \in \mathcal{C}$, $|C| \leq q^{2-t+3\delta}$,*
- *For every arc $P \subset \mathbb{F}_q^2$, there exists $C \in \mathcal{C}$ such that $P \subset C$.*

Proof. Define a 3−uniform hypergraph $\mathcal{H}$ with vertices corresponding to points in $\mathbb{F}_q^2$, with three points forming a (hyper)edge if they are collinear. We now employ an idea used in [2]; we will iteratively apply Theorem 6 to subsets of $\mathbb{F}_q^2$. We begin by applying it to the graph $\mathcal{H}$. As a result, we obtain a set $\mathcal{C}_1$ of containers. We iterate by considering each $A \in \mathcal{C}_1$. If A is not small enough, then we apply Theorem 6 to the graph $\mathcal{H}[A]$ to get a family of containers $\mathcal{C}_A$. If A is sufficiently small then we put this A into a final set $\mathcal{C}$ of containers (or to put it another way, we write $\mathcal{C}_A = A$).

Repeating this for all $A \in \mathcal{C}_1$ we obtain a new set of containers

$$\mathcal{C}_2 = \bigcup_{A \in \mathcal{C}_1} \mathcal{C}_A.$$

Note that $\mathcal{C}_2$ is a container set for $\mathcal{H}$. Indeed, suppose that X is an independent set in $\mathcal{H}$. Then there is some $A \in \mathcal{C}_1$ such that $X \subset A$. Also, X is an independent set in the hypergraph $\mathcal{H}[A]$, which implies that $X \subset A'$ for some $A' \in \mathcal{C}_A \subset \mathcal{C}_2$.

We then repeat this process, defining

$$\mathcal{C}_i = \bigcup_{A \in \mathcal{C}_{i-1}} \mathcal{C}_A.$$

By applying Lemma 1 and choosing the values of τ and ϵ appropriately, we can ensure that after relatively few steps we have all of the elements of $\mathcal{C}_m$ sufficiently small. We then declare $\mathcal{C} = \mathcal{C}_m$. It turns out that, because of m being reasonably small, $|\mathcal{C}|$ is also fairly small.

Now we give more precise details of how to run this argument. Let $A \in \mathcal{C}_j$, with $j \leq m$, and write $|A| = q^{2-s}$. If $s \geq t - 3\delta$ then do nothing. Otherwise, we will apply Theorem 6 to $\mathcal{H}[A]$.

Since $\mathcal{H}$ is $3-$uniform we have $\Delta_3(\mathcal{H}) = 1$. Furthermore, given a pair of points in the plane, the number of points which are collinear with the given pair is $q-2$, and thus $\Delta_2(\mathcal{H}) = q-2$. The subgraph $\mathcal{H}[A]$ therefore satisfies the bounds $\Delta_3(\mathcal{H}[A]) \leq 1$ and $\Delta_2(\mathcal{H}[A]) \leq q-2$. We also have the lower bound on the average degree of $\mathcal{H}[A]$, which follows from Lemma 1:

$$d(\mathcal{H}[A]) = \frac{1}{q^{2-s}} \sum_{v \in \mathcal{H}[A]} d(v) = \frac{3E(\mathcal{H}[A])}{q^{2-s}} \geq cq^{3-2s}.$$

Here $c > 0$ is the absolute constant coming from Lemma 1. We consider the two conditions (5) and (6) present in Theorem 6. Applying the above bounds, the condition (5) holds if

$$\frac{1}{cq^{2-2s}\tau} + \frac{1}{2cq^{3-2s}\tau^2} \leq \frac{\epsilon}{288}. \tag{7}$$

Let $\delta > 0$ be an absolute constant. We make the choices

$$\tau = q^{s+t-2-\delta}, \qquad \epsilon = q^{-\delta}.$$

Plugging these values into (7), this becomes

$$\frac{1}{cq^{t-s-\delta}} + \frac{1}{2cq^{2t-1-2\delta}} \leq \frac{q^{-\delta}}{288}.$$

It would suffice that

$$q^{t-s-2\delta} \geq c^{-1}1000, \quad \text{and} \quad q^{2t-1-3\delta} \geq c^{-1}1000.$$

Since $s < t - 3\delta$, the first of these inequalities holds for q sufficiently large (depending on δ). Similarly, since $t \geq \frac{1}{2} + 2\delta$, the second inequality holds for q sufficiently large. Thus (7) holds, and so does condition (5) of Theorem 6.

Condition (6) in this instance becomes

$$q^{s+t-2-\delta} < \frac{1}{3600}.$$

Since $s + t - 2 - 2\delta < 2t - 2 - 2\delta < -2\delta$, this is satisfied as long as q is sufficiently large with respect to δ. The same is true for the condition $\epsilon < 1/2$. With these choices of τ and ϵ, Theorem 6 can be legitimately applied. We obtain a set of containers $\mathcal{C}_A$ with

$$|\mathcal{C}_A| \ll 2^{q^{t-\delta}(\log q)^2}.$$

We also know that for each $B \in \mathcal{C}_A$,

$$|E(\mathcal{H}[B])| \leq \epsilon |E(\mathcal{H}[A])| = q^{-\delta}|E(\mathcal{H}[A])|.$$

Therefore, at the ith level of this iterative procedure a container $B \in \mathcal{C}_i$ satisfies

$$|E(\mathcal{H}[B])| \leq q^{5-i\delta}.$$

Therefore, after $m(\delta)$ steps of this iteration, we can ensure that all of the containers B at this level satisfy

$$|E(\mathcal{H}[B])| \leq c' q^{5-3t}, \tag{8}$$

where $c' > 0$ is chosen to be sufficiently small so that Lemma 1 implies that $|B| \leq q^{2-t}$. That is, after $m(\delta)$ steps, this process will terminate. Choosing $m = \frac{4t}{\delta}$ will amply suffice, assuming once again that q is sufficiently large.

After the final iteration, we take the union of all containers we have found, and call this final set $\mathcal{C}$. We have

$$|\mathcal{C}| \ll 2^{\frac{4t}{\delta} q^{t-\delta} (\log q)^2}.$$

Each $C \in \mathcal{C}$ has size at most $q^{2-t+3\delta}$. Since independent sets in the hypergraph $\mathcal{H}$ correspond precisely with arcs in $\mathbb{F}_q^2$, the set $\mathcal{C}$ has the properties claimed in the statement of Lemma 2.

□

2.4 Proofs of Theorems 2 and 5

We can now bound the number of $k-$arcs.

Proof of Theorem 2. Fix $\delta > 0$ and recall that $k = q^t$ for $t \geq 1/2 + \delta$. Apply Lemma 2 with $\delta/2$. Then

$$A(q,k) \leq 2^{c(\delta) q^{t-\delta/2} (\log q)^2} \binom{q^{2-t+3\delta/2}}{k} \leq 2^k \binom{q^{2-t+3\delta/2}}{k} \leq \binom{2q^{2-t+3\delta/2}}{k}. \tag{9}$$

The second inequality above is valid provided that q is sufficiently large with respect to δ, while the final inequality is an instance of the bound $2^k \binom{a}{k} \leq \binom{2a}{k}$. Since $\delta > 0$ was arbitrary, the proof is complete.

□

We also deduce Theorem 5 from Lemma 2.

Proof of Theorem 5. Let $p = \frac{q^{\frac{-k}{k-1}}}{100}$ and construct a p-random subset $Q \subset \mathbb{F}_q^2$, each element of $x \in \mathbb{F}_q^2$ belonging to Q with probability p.

The expected number of collinear k-tuples in Q is at most $p^k q^{k+2}$, since there are less than q^{2+k} collinear k-tuples in $\mathbb{F}_q^2$ and each k-tuple survives the random selection process with probability p^k. The expected size of Q is pq^2. Therefore

$$\mathbb{E}[|Q| - \#\text{ collinear } k\text{-tuples in } Q\ ] \geq \left(\frac{1}{100} - \frac{1}{(100)^k}\right) q^{2-\frac{k}{k-1}}.$$

It follows that, with high probability,

$$\#\text{ collinear } k\text{-tuples in } Q \leq |Q|/2 \quad \text{and} \quad |Q| \gg q^{2-\frac{k}{k-1}}. \tag{10}$$

Now let $\delta > 0$ and apply Lemma 2 with this δ and $t = 1/2 + 2\delta$, giving a set of containers $\mathcal{C}$. The probability that Q contains an arc of size at least m is upper bounded by

$$|\mathcal{C}| \binom{q^{3/2+\delta}}{m} p^m. \tag{11}$$

This is because an arc of size m must be contained in some $C \in \mathcal{C}$, and each subset of size m belongs to the random subset Q with probability p^m. Every $C \in \mathcal{C}$ has size

$$|C| \leq q^{3/2+\delta},$$

and so the number of possible candidates for an arc of size m is at most

$$|\mathcal{C}| \binom{q^{3/2+\delta}}{m}.$$

An application of the union bound then gives (11).

Therefore, provided that q is sufficiently large, the probability that Q contains an arc of size $m = q^{1/2+\delta}$ is at most

$$2^{c(\delta)q^{t-\delta}(\log q)^2} \binom{q^{3/2+\delta}}{m} p^m \leq 2^{c(\delta)q^{1/2+\delta}(\log q)^2} \left(\frac{eq^{\frac{3}{2}+\delta}}{q^{\frac{k}{k-1}+\frac{1}{2}+\delta}}\right)^{q^{1/2+\delta}} \leq \left(\frac{2eq^{\frac{3}{2}+\delta}}{q^{\frac{k}{k-1}+\frac{1}{2}+\delta}}\right)^{q^{1/2+\delta}} = \left(\frac{2e}{q^{\frac{1}{k-1}}}\right)^{q^{1/2+\delta}}.$$

In the first inequality above, we have used the bound $\binom{a}{b} \leq \left(\frac{ea}{b}\right)^b$.

So, with probability tending to 1, Q does not contain an arc of size $q^{1/2+\delta}$. We also have (10) with high probability. In particular, there exists a set Q satisfying (10) and not containing any arcs of size m. We now assume that this is true for our random set Q.

We prune Q to find a subset $|Q'| \gg q^{2-\frac{k}{k-1}}$ with no collinear k-tuples. That is, for every collinear k-tuple in Q, remove one element to destroy the k-tuple. At the end of this process, we have deleted at most half of the points and have no collinear k-tuples. Since Q does not contain an arc of size m, Q' also does not.

Finally, the set Q' has the desired property.

□

3 Small k

Let $L(P)$ be the set of lines containing at least two points from P. Define the point set

$$\mathcal{L}_P := \bigcup_{\ell \in L(P)} \ell.$$

We begin with a simple lemma.

Lemma 3. *Let $P \subset \mathbb{F}_q^2$ be an arc with $|P| = k \leq \sqrt{q}$. Then*

$$\frac{q}{2}\binom{k}{2} \leq |\mathcal{L}_P| \leq q\binom{k}{2}.$$

Proof. For set P of k points in general position, $|L(P)| = \binom{k}{2}$. It follows that, for each $\ell \in L(P)$, $|\ell \cap (\mathcal{L}_P \setminus \ell)| \leq |L(P)| = \binom{k}{2} \leq \frac{q}{2}$. The last inequality uses the assumption that $k \leq \sqrt{q}$. Therefore,

$$|\mathcal{L}_P| \geq \sum_{\ell \in L(P)} (q - |\ell \cap (\mathcal{L}_P \setminus \ell)|) \geq \frac{q}{2}\binom{k}{2}.$$

The upper bound is trivial.

□

The following lemma is the key result of this section.

Lemma 4. *For any $k \leq \sqrt{q}$, we have*

$$\prod_{i=2}^{k-1}\left(1 - \frac{i^2}{q}\right) \leq \mathbb{P}(k \text{ points in } \mathbb{F}_q^2 \text{ form an arc}) \leq \prod_{i=1}^{k-2}\left(1 - \frac{i^2}{4q}\right)$$

Proof. We prove this result by induction on k. For the base step $k = 2$, two points form an arc with probability 1, and we have empty products on both sides of the desired inequality, so the base step is proved.

To perform the inductive step, notice that

$$\mathbb{P}(k \text{ points form an arc}) = \mathbb{P}(k{-}1 \text{ points form an arc}) \cdot \mathbb{P}(k\text{'th point is not collinear with two others}).$$

The first probability is bounded using the inductive hypothesis. The second can be bounded using Lemma 3.

Let P be an arc of cardinality $k-1$. The set of 'bad points' (meaning points which are collinear with two points of P) are precisely the elements of $\mathcal{L}_P$. Lemma 3 then gives

$$\mathbb{P}(\text{a point is bad}) = \frac{|\mathcal{L}_P| - (k-1)}{q^2 - (k-1)} \geq \frac{\frac{q}{2}\binom{k-1}{2} - k}{q^2} \geq \frac{(k-2)^2}{4q}.$$

On the other hand, also using Lemma 3,

$$\mathbb{P}(\text{a point is bad}) \leq \frac{q\binom{k-1}{2}}{q^2-(k-1)} \leq \frac{2\binom{k-1}{2}}{q} \leq \frac{(k-1)^2}{q}.$$

Therefore,

$$1-\frac{(k-1)^2}{q} \leq \mathbb{P}(\text{a point is good}) \leq 1-\frac{(k-2)^2}{4q}.$$

We conclude that

$$\mathbb{P}(k \text{ points form an arc}) \leq \prod_{i=1}^{k-3}\left(1-\frac{i^2}{4q}\right)\left(1-\frac{(k-2)^2}{4q}\right) = \prod_{i=1}^{k-2}\left(1-\frac{i^2}{4q}\right)$$

and

$$\mathbb{P}(k \text{ points form an arc}) \geq \prod_{i=2}^{k-2}\left(1-\frac{i^2}{q}\right)\left(1-\frac{(k-1)^2}{q}\right) = \prod_{i=2}^{k-1}\left(1-\frac{i^2}{q}\right),$$

as required. □

Proof of Theorem 1. In order to prove Theorem 1, the final task is to find a convenient approximation for the quantities in the statement of Lemma 4. All we are doing here is giving an approximation which is perhaps more easily digestible for the reader. There may be some loss in this last step, and so we emphasise that a more accurate statement is given by Lemma 4.

We use the approximation $1-x \leq e^{-x}$, which holds for all x. This gives

$$\mathbb{P}(k \text{ points form an arc}) \leq \prod_{i=1}^{k-2}\left(1-\frac{i^2}{4q}\right) \leq \prod_{i=1}^{k-2} e^{\frac{-ci^2}{q}} \leq e^{\frac{-c'k^3}{q}}$$

where c' is some positive absolute constant. Therefore

$$A(q,k) = \mathbb{P}(k \text{ points form an arc})\binom{q^2}{k} \leq e^{\frac{-c'k^3}{q}}\binom{q^2}{k},$$

as required.

A similar argument is used to bound $A(q,k)$ from below. By Lemma 4,

$$A(q,k) = \mathbb{P}(k \text{ points form an arc})\binom{q^2}{k} \geq \prod_{i=2}^{k-1}\left(1-\frac{i^2}{q}\right)\binom{q^2}{k}.$$

Assuming the condition $k \leq \frac{q^{1/2}}{1+\delta}$ for some $\delta > 0$, we can ensure that $\frac{i^2}{q} < \frac{1}{(1+\delta)^2}$. There then exists some constant $D = D(\delta) > 0$ such that the approximation $e^{-Dx} \leq 1-x$ is valid for all $0 \leq x \leq \frac{1}{(1+\delta)^2}$. This implies that

$$A(q,k) \geq \prod_{i=2}^{k-1} e^{-D\frac{i^2}{q}}\binom{q^2}{k} \geq e^{-C\frac{k^3}{q}}\binom{q^2}{k},$$

for some constant $C(\delta) > 0$. □

Acknowledgements

The authors were partially supported by the Austrian Science Fund FWF Project P 30405-N32. We are very grateful to Peter Allen, Nathan Kaplan and Cosmin Pohoata for helpful discussions and also for pointing out several useful references. We also thank Krishna Kaipa for clarifying some details concerning the connection between arcs and MDS codes.

References

[1] J. Balogh, R. Morris and W. Samotij, 'Independent sets in hypergraphs', *J. Amer. Math. Soc.* 28 (2015), no. 3, 669-709.

[2] J. Balogh and J. Solymosi, 'On the number of points in general position in the plane', *Discrete Anal.*, Paper No. 16, 20 pp.

[3] J. Balogh, A. Treglown and A. Zsolt Wagner, 'Applications of graph containers in the Boolean lattice.', *Random Structures Algorithms* 49 (2016), no. 4, 845-872.

[4] D. Conlon and W. T. Gowers, 'Combinatorial theorems in sparse random sets.', *Ann. of Math.* (2) 184 (2016), no. 2, 367-454.

[5] Z. Füredi, 'Maximal independent subsets in Steiner systems and in planar sets', *SIAM J. Discrete Math.* 4 (1991), no. 2, 196-199.

[6] A. V. Iampolskaia, A. N. Skorobogatov and E. A. Sorokin, 'Formula for the number of $[9, 3]$ MDS codes', *IEEE Trans. Inform. Theory* 41 (1995), no. 6, part 1, 1667-1671.

[7] K. V. Kaipa, 'An asymptotic formula in q for the number of $[n, k]$ q-ary MDS codes', *IEEE Trans. Inform. Theory* 60 (2014), no. 11, 7047-7057.

[8] J. Komlós, J. Pintz and E. Szemerédi, 'A lower bound for Heilbronn's problem', *J. London Math. Soc.* (2) 25 (1982), no. 1, 13-24.

[9] D. Osthus, H. Prömel and A. Taraz, 'For which densities are random triangle-free graphs almost surely bipartite?', *Combinatorica* 23 (2003), no. 1, 105-150.

[10] D. Saxton and A. Thomason, 'Hypergraph containers', *Invent. Math.* 201 (2015), no. 3, 925-992.

[11] M. Schacht, 'Extremal results for random discrete structures', *Ann. of Math.* (2) 184 (2016), no. 2, 333-365.

[12] B. Segre. 'Ovals in a finite projective plane', *Canadian Journal of Mathematics* 7 (1955), pp. 414416.