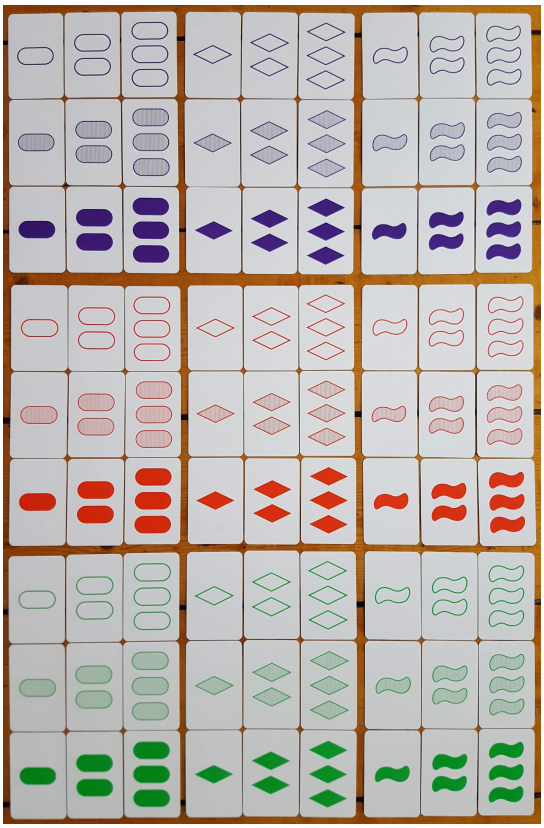
Utrecht University

Bachelor Thesis
The Cap Set Problem

Wytske Deutekom (5665698)
Supervisor: Damaris Schindler

January 11, 2019



Abstract

In this paper we will take a look at bounds that are found on the cap set problem. We will especially discuss results from the article 'Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small' by Croot, Lev and Pach and elaborate some of the proofs in this article. Before these specific results, we will extensively explain the idea of the cap set problem and show that the biggest cap set in the game Set with two attributes has size four. Besides that, we will give some background information on arithmetic progressions in different group settings. Moreover we will discuss entropy functions, the pigeonhole principle and the polynomial method, which play an important role in the work of Croot, Lev and Pach and the work of Ellenberg and Gijswijt, where they solve the classical cap set problem. Lastly we will take a look at the reasoning behind an improved construction of progression-free sets, where Elkin improves the best lower bound that was found in 1946 by Behrend.

Contents

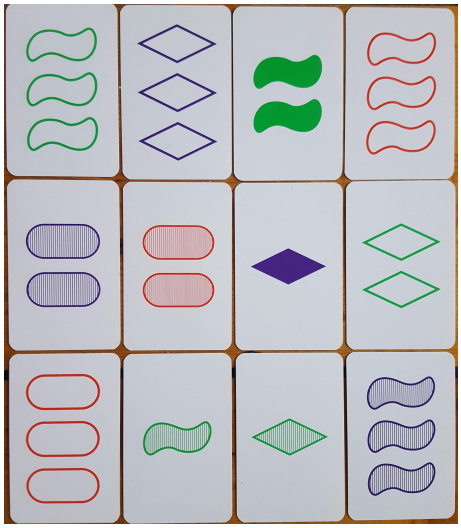


Figure 1: Possible state of the game Set.

1 The cap set problem

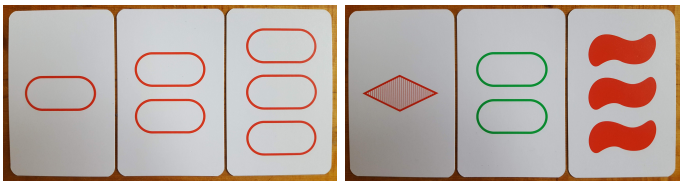
1.1 Explanation of the game Set

Set is a card game with a simple goal: find special triples within a deck of 81 cards (Klarreich, 2016 [15]). These special triples are called 'sets' and we will refer to these special triples as Sets. In this game, each card has a different design, where four attributes are used (as can be seen on the front page):

1. Color: red (0), purple (1) or green (2);
2. Shape: oval (0), diamond (1) or squiggle (2);
3. Shading: solid (0), striped (1) or outlined (2);
4. Number: one (0), two (1) or three (2).

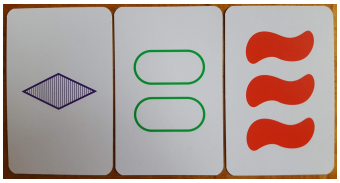
In this game, we place 12 cards face-up on the table and the players try to find a Set. A Set consists of three cards, whose designs must be either all the same or all different for each attribute (see example 1.1.1). A possible state of the game can be seen in figure 1. If there is no Set in these 12 cards, we place three extra cards on the table and we keep on doing this until there is a Set on the table. An interesting question to ask will then be: What is the size of the biggest collection of cards that contains no Set? This question leads us to the cap set problem.

Example 1.1.1. If we look at some examples of possible formations of cards (figure 2), we can easily create a table to find out whether the formation is a Set or not (see table 1). If the attribute is all the same, we will write down 'all the same' in the table, if the attribute is all different, we will write down 'all different' and if the attribute is not all the same nor all different, we will write down 'x'. If this last situation occurs, it is not possible to have a Set.

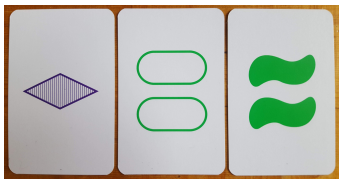


(a) Example 1

(b) Example 2



(c) Example 3



(d) Example 4

Figure 2: Examples of possible formations.

Example	Color	Shape	Shading	Number	Conclusion
1 (figure 2a)	all the same	all the same	all the same	all different	Set
2 (figure 2b)	x	all different	all different	all different	No Set
3 (figure 2c)	all different	all different	all different	all different	Set
4 (figure 2d)	x	all different	all different	x	No Set

Table 1: Example 1.1.1.

We can conclude that examples 1 and 3 are both Sets, since in those cases all attributes are either all the same or all different. In example 2 we have two cards with the same color, so the color isn't all the same nor all different for the three cards, which means this cannot be a Set. Lastly in example 4 we have two cards in green, therefore the color is again not all the same nor all different for the three cards, but also we have two cards with number two and the last card has number one, in consequence also the number isn't all the same nor all different in these three cards. Hence this can also not be a Set.

1.2 The problem

In the game Set, an answer to the question of section 1.1 was found by the Italian mathematician Giuseppe Pellegrino [19]. He proved that the biggest collection of cards without a Set would be 20. This answer wasn't sufficient for a lot of mathematicians, since we could also expand this game to more than four attributes. We could say that for every integer n, we could think of a version of Set with n attributes and three choices per attribute, this would give a total of 3^n different cards.

Definition 1.2.1. If a collection of cards contains no Set, we name that collection a cap set.

For certain amounts of attributes, it is possible to find the exact size of the biggest cap set, but for an enormous amount of attributes this is not possible yet. However, it is possible to find an upper bound on how big such a cap set can be. In this case we would look for the number of cards where it is guaranteed that it holds at least one Set.

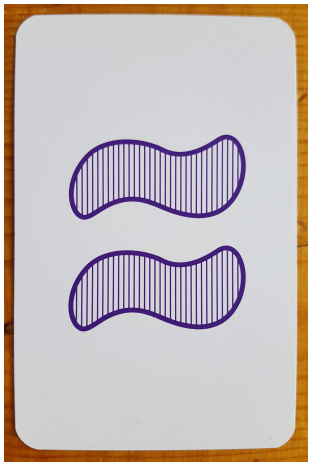


Figure 3: A specific card of the game Set.

1.3 The game Set in geometry

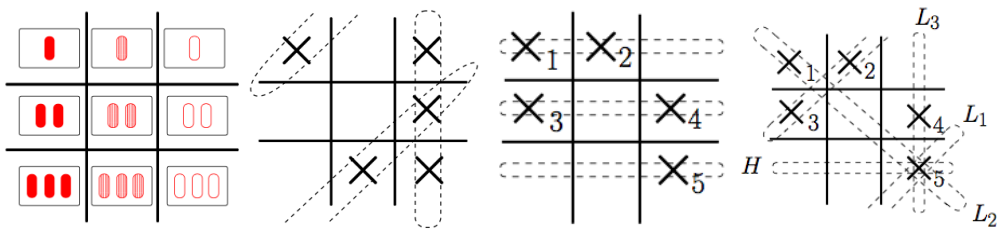
If we want to find an upper bound on the size of cap sets, we have to translate the Set game into geometry. We begin with a field $\mathbb{F}_3$ with three elements. Next we consider the vector space $\mathbb{F}_3^4$. The cards are represented as points of $\mathbb{F}_3^4$, therefore these points have four coordinates. Here each coordinate can take the value $0, 1$ or 2. For example if we look at the attribute color (which is the first coordinate in the point), we will get a '0' if the card is red, a '1' if it is purple and a '2' if it is green (see enumeration on page 3).

Example 1.3.1. The card in figure 3 would correspond to point $(1, 2, 1, 1)$ since the color is purple, the shape is a squiggle, the shading is striped and the number is two (according to the instructions in section 1.1).

When we consider the cards as points of $\mathbb{F}_3^4$, we can formulate the property of being a Set in the following way [8]: iff three points of $\mathbb{F}_3^4$ are collinear, then they form a Set. Imagine that we have three elements of $\mathbb{F}_3$: α, β and γ. Then we can see that $\alpha + \beta + \gamma = 0$ if and only if $\alpha = \beta = \gamma$ or if $\{\alpha, \beta, \gamma\} = \{0, 1, 2\}$. This shows that if we look at vectors a, b and c in $\mathbb{F}_3^4$, that these vectors will be all the same or all different with respect to each coordinate precisely when $a+b+c = 0$. This expression means that $a-b = b-c$, which is an arithmetic progression (section 5). Since this reasoning works for every value of n in $\mathbb{F}_3^n$, we can also define a version of Set with n attributes instead of four. In this case we would have points with n coordinates instead of four. To translate the rules of Set into geometry, we will use the resulting $n-$dimensional space. In this space, every line contains exactly three points. If three points lie on the same line (arithmetic progression, section 5), they form a Set. Therefore we can define a cap set as:

Definition 1.3.1. If a collection of points (subset of $\mathbb{F}_3^n$) contains no complete lines, we name that collection a cap set.

Proposition 1.3.1. The biggest cap set of $\mathbb{F}_3^2$ has size 4 (Davis & Maclagan 2003, [8]).



(a) Possible cards (b) Lines of Set (c) Representation of $\mathbb{F}_3^2$ (d) Four lines containing X_5

Figure 4: Set with two attributes.

Proof Proposition 1.3.1
In this case, we take a look at $n = 2$, which means that we only have two attributes. We could for example only look at the red ovals. In this way, we only keep the attributes 'number' and 'shading'. Since both attributes have three choices, we get a total of nine cards (figure 4a).

We could display these nine cards in a tic-tac-toe board, which then represents the vector space $\mathbb{F}_3^2$. In this tic-tac-toe board, a Set is represented by a line. These lines can be horizontal, vertical or diagonal. We see that some of these lines correspond to winning tic-tac-toes, but Set lines can also meet an edge and loop around to the other side of the board. In figure 4b you can see two examples of lines that represent a Set.

To prove proposition 1.3.1, we will use a contradiction. We assume that there exists a cap set of size 5, so we have 5 points: X_1, X_2, X_3, X_4 and X_5. We can represent the vector space $\mathbb{F}_3^2$ as three horizontal lines as shown in figure 4c. Since we don't want a Set to exist, each line can only contain two of the points X_1, X_2, X_3, X_4 and X_5. We could have a distribution of points like this. Of course this distribution can also be different, but the bottom line is that we will always have two horizontal lines with two points and one horizontal line with only one point. In this case, this point is X_5, then without loss of generality we will prove it for this case.

When we look at figure 4c again, we see that there are four possible lines that contain the point X_5. We will denote these lines by H, L_1, L_2 and L_3 (figure 4d).

We see that line H only contains X_5 and none of X_1, X_2, X_3 or X_4. Since these four points have to be placed on the other three lines (L_1, L_2 or L_3), we find by the pigeonhole principle (section 6) that on a line L_i there must be at least two of these points. This shows that there exists a line L_i with these two points and also the point X_5. Subsequently this is a line with three points, which means that this is a Set. Which shows that this collection of five points is not a cap set. $\square$

The biggest cap set size is only known for values of n up to 5, when $n = 6$ it is known that the size is bound as follows: $112 \leq \text{size} \leq 114$. For bigger values of n mathematicians haven't been able to find such a precise bound. Yet there has been done a lot of research in this field, so we will look at some observations so far.

1.4 Bounds on the cap set problem so far

The first upper bound was found by Roth [20]. He looked at a collection $A \subseteq \{1, 2, \cdots, N\}$ of integers and if this set didn't contain three elements in arithmetic progression, then $|A| = o(N)$.

He also proved that as N grew, then indeed $|A| = O(\frac{N}{\log \log N})$. Since then, the problem has stimulated much research in estimating the largest possible size of a cap set. The current record in sets of integers was found by Bloom [4] as $|A| = O(\frac{N(\log \log N)^4}{\log N})$.

Roth's problem is equivalent to finding the largest possible size of a subset of the cyclic group $\mathbb{Z}_N$ with no three-term arithmetic progression. Because of this equivalence, several mathematicians have investigated other finite abelian groups (definition 2.1.1).

Definition 1.4.1. Let G be an abelian group and A a subset of G. We call A progression-free if there are no pairwise distinct $a, b, c \in A$, with $a + b = 2c$. The largest size of such a progression-free subset $A \subseteq G$ will be denoted by $r_3(G)$.

For abelian groups G of odd order, the same was proven by Brown and Buhler [5], but also independently by Frankl, Graham and Rödl [12]: $r_3(G) = o(|G|)$ as $|G|$ grows. Meshulam used the general lines of Roth's argument and discovered that $r_3(G) \leq \frac{2|G|}{\mathrm{rk}(G)}$ if G is an abelian group of odd order [17]. Here $\mathrm{rk}(G)$ stands for the rank of G. This discovery lead to the generalization: $r_3(\mathbb{Z}_m^n) \leq \frac{2m^n}{n}$.

It wasn't until many years later that Bateman and Katz proved that $r_3(\mathbb{Z}_3^n) = O(\frac{3^n}{n^{1+\varepsilon}})$ with an absolute constant $\varepsilon > 0$ [2].

Lev was the fist mathematician to consider abelian groups of even order. He continued on the Roth-Meshulam proof and showed that $r_3(G) < \frac{2|G|}{\mathrm{rk}(2G)}$ for any finite abelian group G [16]. Here it is used that $2G = \{2g : g \in G\}$. Sanders improved this result for homocyclic groups of exponent four: $r_3(\mathbb{Z}_4^n) = O(\frac{4^n}{n(\log\ n)^\varepsilon})$ where $\varepsilon > 0$ is again an absolute constant [23].

Croot, Lev and Pach used the polynomial method (section 7.1) to solve a closely related problem to the classical Set problem. Instead of three different options per attribute, they solved the problem for four different options. This problem is more tractable in the calculations than the original Set problem.

1.5 Guidance through the paragraphs

We will start with some prerequisites which are needed in later paragraphs. After that there will follow some paragraphs with some more elucidations on certain topics which are used in articles on the cap set problem (finite fields, progressions and the entropy function). Next there is a paragraph on the polynomial method since this plays an important role in the solution to the complicated cap set problem. Finally we will take a look at the work of some mathematicians who have tried to find lower and upper bounds for the progression free sets.

2 Prerequisites

In this section we will repeat some definitions to ensure that we are on the same page in terms of notation. For these definitions and examples, we will use the work of Paar & Pelzl [18], Chartrand, Polimeni & Zhang [6], Armstrong [1] and Igodt & Veys [14].

2.1 Algebraic structures

Definition 2.1.1. A(n) (abelian) group.

A set of elements G can be made into a group by adding an operation $\circ$ which combines two elements of G. A group has four important properties and a group can be abelian (property 5):

1. The group operation is closed:
   $\forall\ a, b \in G$: $a \circ b = c \in G$.
2. The group operation is associative:
   $\forall\ a, b, c \in G$: $a \circ (b \circ c) = (a \circ b) \circ c$.
3. There is a neutral element:
   $\exists\ 1 \in G$, (the neutral element/identity element), such that $\forall\ a \in G : a \circ 1 = 1 \circ a = a$.
4. There is an inverse element:
   $\exists\ a^{-1} \in G$, (the inverse of a), such that $\forall\ a \in G : a \circ a^{-1} = 1$.
5. A group G is abelian (or commutative) if, furthermore, $\forall\ a, b \in G : a \circ b = b \circ a$.

Example 2.1.1. Group.

An example of a group with neutral element 0, would be the set of integers $\mathbb{Z}_m = \{0, 1, \cdots, m-1\}$ with the operation addition modulo m. Since $a + (-a) = 0 \bmod m$, we see that every element a has inverse $-a$. We can also see that this set is a group with the operation multiplication if m is prime, because all elements a (except for $a = 0$) have an inverse such that $a \cdot a^{-1} = 1 \bmod m$.

Definition 2.1.2. Cardinality.

The cardinality of a group G (denoted by $|G|$), is also referred to as the order of a group. This represents the number of elements in that group. G is called a finite group if it's order is finite. If G has an infinite number of elements, then G is an infinite group.

Definition 2.1.3. Cyclic group.

G is called a cyclic group, if there exists an element x in G which generates all of G. We often write this as $\langle x \rangle = G$.

Example 2.1.2. Cyclic group.

Let n be a positive integer. The set $0, 1, 2, \cdots, n-1$ can be made into a group using addition modolu n. If x and y are members of this set, we can define the following:

$$x +_n y = \begin{cases} x + y & \text{if } 0 \leq x + y < n \\ x + y - n & \text{if } x + y \geq n \end{cases}$$

In $\mathbb{Z}_6$ we would have:
$\langle 0 \rangle = \{0\}$
$\langle 1 \rangle = \langle 5 \rangle = \mathbb{Z}_6$
$\langle 2 \rangle = \langle 4 \rangle = \{0, 2, 4\}$
$\langle 3 \rangle = \{0, 3\}$

For example we see that the elements of $\langle 5 \rangle$ are:
5
$5 +_6 5 = 4$
$5 +_6 5 +_6 5 = 3$
$5 +_6 5 +_6 5 +_6 5 = 2$
$5 +_6 5 +_6 5 +_6 5 +_6 5 = 1$
$5 +_6 5 +_6 5 +_6 5 +_6 5 +_6 5 = 0$

This means that $\langle 5 \rangle = \mathbb{Z}_6$, so we can conclude that there is an element in $\mathbb{Z}_6$ which generates all of $\mathbb{Z}_6$, thus $\mathbb{Z}_6$ is a cyclic group.

Definition 2.1.4. Vectorspace (linear space).

A real vectorspace (linear space) is written as $(\mathbb{R}, V, +)$ and consists of a cummutative (abelian) group $(V, +)$ together with an external operation; the scalar multiplication:
$\mathbb{R} \times V \to V : (\lambda, v) \mapsto \lambda v$. Furthermore, the following properties hold:

1. Distributivity-1:
   $\forall\ \lambda \in \mathbb{R}, \forall\ v, w \in V : \lambda(v + w) = \lambda v + \lambda w$.
2. Distributivity-2:
   $\forall\ \lambda_1, \lambda_2 \in \mathbb{R}, \forall\ v \in V : (\lambda_1 + \lambda_2)v = \lambda_1 v + \lambda_2 v$.
3. Mixed associativity:
   $\forall\ \lambda_1, \lambda_2 \in \mathbb{R}, \forall\ v \in V : \lambda_1(\lambda_2 v) = (\lambda_1 \lambda_2)v$.
4. Coefficient 1:
   $\forall\ v \in V : 1v = v$.

The numbers $\lambda \in \mathbb{R}$ are called the coefficients or scalars, the elements $v \in V$ are called the vectors and the neutral element 0 is called the zero vector.

2.2 Properties

Definition 2.2.1. Injective.

A function $f : A \to B$ is injective if for all $x, y \in A$ with $x \neq y$: $f(x) \neq f(y)$. This means that every two distinct elements of A, will also have distinct images in B. This shows that if a function $f : A \to B$ is not injective, then there must exist distinct elements w and z with the property that $f(w) = f(z)$. We can also formulate the contrapositive: A function $f : A \to B$ is injective, if it holds that if $f(x) = f(y)$, with $x, y \in A$, then $x = y$.

If A and B are finite sets, a function $f : A \to B$ can only be injective if every two elements in A have distinct images in B, which shows that we need at least as many elements in B as in A: $|A| \leq |B|$.

Definition 2.2.2. Surjective.

If every element of the codomain B is the image of some element of A ($f(A) = B$), then the function $f : A \to B$ is called surjective.

If A and B are finite sets, a function $f : A \to B$ can only be surjective if there are at least as many elements in A as in B. This shows that we need $|B| \leq |A|$.

Definition 2.2.3. Bijective.

A function $f : A \to B$ is called bijective or a one-to-one correspondence if it is both injective and surjective.
Hence if A and B are finite sets and there is a function $f : A \to B$ that is bijective, then $|A| = |B|$.

Definition 2.2.4. Isomorphic.

Two groups $(G, \star)$ and $(H, \circ)$ are called isomorphic if there exists a function $\phi : G \to H$ that is bijective and that satisfies the property:
$\forall\ a, b \in G : \phi(a \star b) = \phi(a) \circ \phi(b)$.
A function ϕ that meets these requirements is an isomorphism and is operation-preserving.

If $\phi : G \to H$ is an isomorphism, then ϕ is also a bijective function. This shows that ϕ has an inverse function $\phi^{-1} : H \to G$, which is again an isomorphism.
If $\phi : G \to H$ is also a linear map, then we can say that ϕ is a linear isomorphism.

Example 2.2.1. Isomorphic.

Define $\phi : \mathbb{R} \to \mathbb{R}^{pos}$ by $\phi(x) = e^x$. Then ϕ is a bijection and it satisfies the following property:
$\forall\ x, y \in \mathbb{R} : \phi(x + y) = e^{x+y} = e^x \cdot e^y = \phi(x) \cdot \phi(y)$. This shows (by definition 2.2.4) that $\mathbb{R}$ and $\mathbb{R}^{pos}$ are isomorphic groups. In this example, the group operation in $\mathbb{R}$ is addition, and the group operation in $\mathbb{R}^{pos}$ is multiplication.

3 Finite fields

Definition 3.0.1. (Finite) Field (Paar & Pelzl, 2010, p. 92) [18]

A field F is a set of elements with the following properties:

1. All elements of F form an additive group with the group operation "+" and the neutral element 0.
2. All elements of F except 0 form a multiplicative group with the group operation "×" and the neutral element 1.
3. When the two group operations are mixed, the distributivity law holds, i.e., $\forall\ a, b, c \in F$: $a(b + c) = (ab) + (ac)$.

Fields with a finite number of elements are called finite fields or Galois fields.

Example 3.0.1. Field (Paar & Pelzl, 2010, p. 92) [18].

An example of a field is the set $\mathbb{R}$. In this set, the additive group has neutral element 0 and the multiplicative group has neutral element 1. Every real number a has $-a$ as an additive inverse and every non-zero element a has $\frac{1}{a}$ as a multiplicative inverse. This shows that $\mathbb{R}$ is a field, but since the number of elements isn't finite, this is not a finite field.

In cryptography we are mostly interested in Galois fields (Paar and Pelzl, 2010), ergo fields with a finite number of elements. If the number of elements is finite, we call this number the order or cardinality of the field (definition 2.1.2). An interesting theorem about the order of a finite field is the following:

Theorem 3.0.1. A field with order m only exists if m is a prime power, in other words if $m = p^n$ for some positive integer n and prime integer p. In this case, p is called the characteristic of the finite field.

Example 3.0.2. We can have finite fields with 11 elements, since 11 is prime. We can also have finite fields with 81 elements, since $81 = 3^4$ and 3 is prime, or finite fields with 256 elements, since $256 = 2^8$ and 2 is prime. On the other hand, there is no finite field with 12 elements, since we can write 12 as $2^2 \cdot 3$, which shows that 12 is not a prime power.

The most simple examples of finite fields are the ones with prime order, therefore the fields where $n = 1$. Elements of these fields ($GF(p)$) can be represented by the integers $0, 1, \cdots, p-1$. The operations that are operated on this field are integer multiplication modulo p and modular integer addition.

Theorem 3.0.2. Let p be a prime. The integer ring $\mathbb{Z}_p$ is denoted as $GF(p)$ and is referred to as a prime field (or Galois field) with a prime order. All nonzero elements of $GF(p)$ have an inverse. Arithmetic in $GF(p)$ is done modulo p.

Example 3.0.3. Let's take a look at the finite field $GF(5) = \{0, 1, 2, 3, 4\}$. We can add and multiply (mod 5) as follows and we can also find the additive and multiplicative inverse of the elements:

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

Addition mod 5

$\times$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Multiplication mod 5

0	$-0 = 0$
1	$-1 = 4$
2	$-2 = 3$
3	$-3 = 2$
4	$-4 = 1$

Additive inverse

0	0^{-1} does not exist
1	$1^{-1} = 1$
2	$2^{-1} = 3$
3	$3^{-1} = 2$
4	$4^{-1} = 4$

Multiplicative inverse

In cryptography the finite field $GF(256) = GF(2^8)$ is used in AES encryption. They use this field since each element of the field can be represented by a byte. For this encryption, every byte of the data is treated as an element of $GF(2^8)$ and they perform arithmetic in the finite field to execute the operations.

4 Progressions

An arithmetic progression can be defined as a sequence of n numbers
$a_0 + [0, n) \cdot \alpha := \{a_0, a_0 + \alpha, \cdots, a_0 + (n-1)\alpha\}$ where the difference between successive terms is a constant α. Here $a_0, \alpha \in \mathbb{Z}$ and $n \in \mathbb{Z}^+$ (Tao & Vu, 2006 [25]).

$(a, b, c) \in \mathbb{Z}^3$ is a three-term progression if and only if $a + c = 2b$.
For a four-term progression, we find the following:

Lemma 4.0.1. $(a, b, c, d) \in \mathbb{Z}^4$ is a four-term progression if and only if $a+c = 2b$ and $b+d = 2c$.

Proof Lemma 4.0.1
$\Rightarrow$ If $(a, b, c, d) \in \mathbb{Z}^4$ is a four-term progression, we have the following equations (since the difference between successive terms is α):

$$b - a = \alpha \tag{1}$$

$$c - b = \alpha \tag{2}$$

$$d - c = \alpha \tag{3}$$

By combining (1) and (2), we get the equation $a + c = 2b$. By combining (2) and (3), we get the equation $b + d = 2c$.

$\Leftarrow$ If the equations $a + c = 2b$ and $b + d = 2c$ hold, we can rewrite these as:

$$a + c = b + b \tag{4}$$

$$b + d = c + c \tag{5}$$

Where we can write (4) as

$$b - a = c - b,$$

and (5) as

$$c - b = d - c.$$

Since $c-b$ is similar, it follows that $b-a = c-b = d-c$, which shows that the difference between the successive terms is a constant. Subsequently $(a, b, c, d) \in \mathbb{Z}^4$ is a four-term progression.

$\square$

For an n-term progression $(x_1, x_2, \cdots, x_n) \in \mathbb{Z}^n$ we get the following equations:

$$2x_2 = x_1 + x_3$$

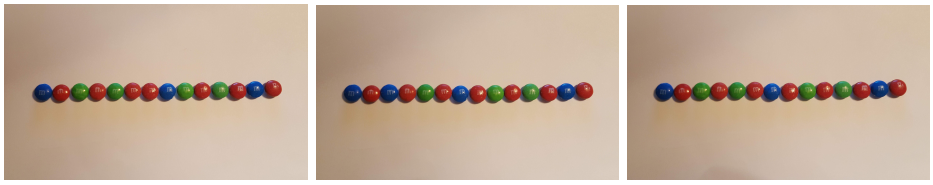
$$2x_3 = x_2 + x_4$$

$$\vdots$$

$$2x_{n-1} = x_{n-2} + x_n$$

This shows that for $2 \leq i \leq n-1$ we get:

$$2x_i = x_{i-1} + x_{i+1}$$



(a) Progression 1 (b) Progression 2 (c) Progression 3

Figure 5: Progressions

Example 4.0.1. Progressions

In this example we look at rows of M&Ms. We define an arithmetic progression in terms of color and we are looking for the longest n-term progression.

In figure 5 we see some M&Ms placed in a row. When we look at 5a, we see that there is an arithmetic progression in the color blue only when we 'ignore' one blue M&M. The distance between the first and second blue M&M is 7, while the distance between the second and third blue M&M is 5. This means that there exists just an arithmetic progression in blue of length 2 where $\alpha = 7$ (ignore the third blue M&M), $\alpha = 5$ (ignore the first blue M&M) or where $\alpha = 12$ (ignore the second blue M&M). In red we have two 3-term progressions, with $\alpha = 2$ (the first three red M&Ms and the last three red M&Ms), we also have a 4-term progression, with $\alpha = 4$ (we ignore the second, fourth and the sixth red M&M) and finally we also have a 3-term progression, with $\alpha = 5$ (if we ignore the second, third, fifth and last red M&M). In green we have two 2-term progressions with $\alpha = 2$. Hence in this row, the longest progression is has length 4 (in red).

If we switch the seventh and eighth M&Ms and change the third M&M from green to blue in the row that we had, we get figure 5b. Now we see that the distance between two red M&Ms is the constant number two. This means that we have an arithmetic progression. To determine n, we just count how many red M&Ms satisfy the requirement that the distance between the previous successive M&M and the current M&M is 2. In consequence here we have a 7-term progression in red, where $\alpha = 2$. This is the longest progression of this row. In blue and green we can also find progressions, but those have a maximum length of 3.

If we change the third M&M from blue to green again, we see in figure 5c that we still have the 7-term arithmetic progression with distance 2 (the red M&Ms), but now we also have another arithmetic progression where α is big: the blue M&Ms. The distance between two blue M&Ms is the constant number 6. Consequently now we also have a 3-term arithmetic progression. In green we have again two 2-term progressions with $\alpha = 2$. Hence here again the longest progression has length 7.

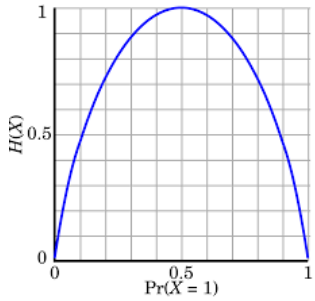


Figure 6: Binary entropy

5 Entropy function

The entropy of a random variable X, with probability mass function $p(x)$, is defined as

$$H(X) = -\sum_{x} p(x)\log(p(x))$$

If X is a binary random variable, the following can be said about the probabilities:

$$X = \begin{cases} 1 & \text{with probability } x \\ 0 & \text{with probability } 1-x \end{cases}$$

According to Roth (2016) [21], we can define the binary entropy function $H : [0,1] \to [0,1]$ by

$$H(x) = -x\ \log_2(x) - (1-x)\ \log_2(1-x).$$

The binary entropy function has some interesting characteristics (see figure 6). First of all $lim_{x\to 0}H(x) = lim_{x\to 1}H(x) = 0$, $H(x) > 0$ for $0 < x < 1$ and $H(x)$ is symmetric with respect to $x = 0.5$, hence it takes its maximum at that point: $H(0.5) = 1$. Moreover, $H(x)$ is $\cap-$concave since for every two points x_1 and $x_2 \in [0,1]$ the line segment that connects the points $(x_1, H(x_1))$ and $(x_2, H(x_2))$ lies entirely on or below the function curve in the real plane.

We are especially interested in $H(0.5-\varepsilon)$ and $H(2\varepsilon)$ over the interval $0 < \varepsilon < 0.25$ since we want to find the maximum of the average of these two functions. We need this in section 8. We plot both functions and their average (figure 7). We calculate the maximum of the average and we find 0.926, which we will define as γ:

$$\gamma := \max\{\frac{1}{2}(H(0.5-\varepsilon) + H(2\varepsilon)) : 0 < \varepsilon < 0.25\} \approx 0.926. \tag{6}$$

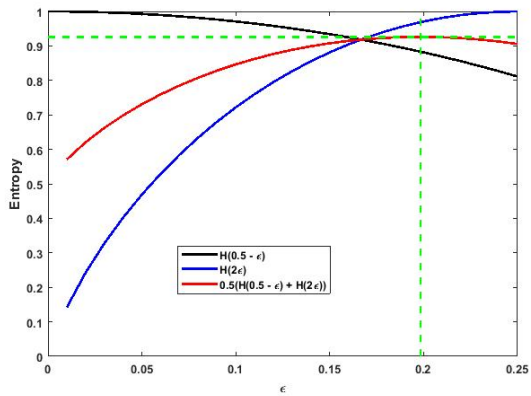


Figure 7: Entropy function: calculation of gamma

6 Pigeonhole principle

We start by considering a group of pigeons that are nestled in a set of n pigeonholes. It is easy to see that if there are n pigeons, then it it possible that each pigeon has his own pigeonhole where he can rest happily. This situation changes however if another pigeon arrives, making a total of more than n pigeons.

The arrival of this last pigeon causes that at least one of the pigeonholes will be filled with more than one pigeon.

This phenomenon is known as the pigeonhole principle:

Theorem 6.0.1. Pigeonhole principle (naive form):
If we place more than n objects into n boxes, then we know that at least one box will contain more than one object.

To illustrate how the pigeonhole principle can be used, we take a look at two small examples.

Example 6.0.1. We have a box which contains three pairs of socks. They are coloured red, blue and white. Suppose we take out the socks without looking at them. How many socks should we take out of the box to be sure that we have a matching pair?

If we start by taking two or three socks, it might happen that they are all different. We could for example have taken out one red sock, one blue sock and one white sock. Of course if we take out another sock, we must have a matching pair of socks, since we all ready have every possible color of socks.

In this example, the four chosen socks can be seen as the 'objects' and the three colors are the 'boxes'. We find that the minimum number of socks that we have to take out, to be sure of a matching pair, is four.

Example 6.0.2. Given the following set $\{a_1, a_2, \cdots, a_n\}$, we want to proof that there exists a non-empty subset S whose sum is divisible by n.

We start by writing down the following n sums:

$$
\begin{aligned}
S_1 &= a_1 \\
S_2 &= a_1 + a_2 \\
&\vdots \\
S_n &= a_1 + \cdots + a_n.
\end{aligned}
$$

We are immediately done if one of these sums is divisible by n. So we assume that this is not the case. Since these sums are not divisible by n, we know that they all have a remainder when divided by n. Only $n-1$ distinct remainders are possible, in consequence by the pigeonhole principle we find that at least two of the sums must have the same remainder (since there were n sums).

To conclude this proof, we take two such sums S_i and S_j (with $j > i$) and subtract them:

$$S_j - S_i = a_{i+1} + \cdots + a_j.$$

which gives us the non-empty subset $\{a_{i+1}, \cdots, a_j\}$ whose sum is divisible by n.

7 Polynomial method

7.1 The method explained

In this section we follow the work of Tao [24] and Guth [13].

For a long time mathematicians have tried to use Fourier methods to solve the cap set problem. Eventually Croot, Lev and Pach made a major breakthrough in a problem quite similar to the cap set problem (where they used four choices per attribute), but they used a completely different method: the polynomial method. Later Jordan Ellenberg and Dion Gijswijt used the same polynomial method in their proof to solve the classic cap set problem.

Partly the idea of the polynomial method is borrowed from the philosophy of algebraic geometry. The interest of algebraic geometry often lies in special geometrical objects. If we look at a collection of one or more polynomials, these geometrical objects form the vanishing sets and they are often referred to as algebraic varieties. To understand more about the geometry of these objects, we look at the polynomials. The polynomial method is often used in combinatorical problems, where we often start with a certain field $\mathbb{F}$ (which is usually $\mathbb{R}$ or finite $\mathbb{F}_q$). In this field we often want to find a finite subset $S \subset \mathbb{F}^N$ that has an interesting property. To gain information about S, we can use multivariate polynomials over $\mathbb{F}$ with the property that they vanish on all points of S.

First we will take a look at the one-dimensional case, so we look at subsets of $\mathbb{F}$. We define S to be a finite subset of $\mathbb{F}$. There exists a polynomial of degree $|S|$ in $\mathbb{F}[X] \setminus \{0\}$, that vanishes on all points of S, namely:

$$\prod_{s \in S} (X - s).$$

This observation leads us to the well-known converse theorem:

Theorem 7.1.1. Factor theorem.
Let $\mathbb{F}$ be a field. Any polynomial in $\mathbb{F}[X] \setminus \{0\}$ that has degree d, will have at most d roots in $\mathbb{F}$.

This shows that if S is a finite subset of $\mathbb{F}$, the smallest possible degree of a polynomial that will vanish on all of S in $\mathbb{F}[X] \setminus \{0\}$, will be $|S|$.

Now we can expand this theorem to higher-dimensions. Again we start with a field $\mathbb{F}$, but now we have a multivariate polynomial $P \in \mathbb{F}[X_1, \cdots X_N]$ over $\mathbb{F}$. The highest power of X_i (which can occur in any monomial of P) will be the X_i-degree of P. This will be noted as $\deg_{X_i}(P)$. The degree in X of the polynomial $P(X_1, X_2, \cdots, X_N)$, is referred to as the total degree of P. Now we can formulate the higher-dimensional analogue of the factor theorem:

Lemma 7.1.1. Let $\mathbb{F}$ be a field and we have a non-trivial polynomial $P \in \mathbb{F}[X_1, \cdots, X_N] \setminus \{0\}$. We assume that $S_1, \cdots, S_N \subset \mathbb{F}$, where $|S_i| > \deg_{X_i}(P)$. Then it is not possible that P vanishes on all of $S_1 \times S_2 \times \cdots \times S_N$.

Proof Lemma 7.1.1.
We will prove this lemma by using induction on N. We know that for $N = 1$, the statement holds since that is exactly the Factor theorem (theorem 7.1.1). Now we will assume that it holds for $N - 1$ and prove that it also holds for N. To show that it holds for N, we need to show

that P will be the zero polynomial if we assume that $P = P(X_1, \cdots, X_N) \in \mathbb{F}[X_1, \cdots, X_N]$ vanishes on all points of $S_1 \times S_2 \times \cdots \times S_N$, with the property that for each $i \in \{1, \cdots, N\}$, $|S_i| > d_i := \deg_{X_i}(P)$.

We start by writing P as a polynomial in X_1:

$$P = P(X_1, \cdots, X_N) = \sum_{j=0}^{d_1} X_1^j P_j(X_2, \cdots, X_N).$$

Next we take a look at the polynomial

$$P(X_1, s_2, \cdots, s_N) = \sum_{j=1}^{d_1} X_1^j P_j(s_2, \cdots, s_N) \in \mathbb{F}[X_1].$$

We see that this polynomial vanishes on all of S_1 for any $(s_2, \cdots, s_N) \in S_2 \times \cdots \times S_N$. Furthermore, this polynomial has degree $d_1 < |S_1|$ (assumption), hence by the Factor theorem we can conclude that this must be the zero polynomial. This shows that for all $(s_2, \cdots, s_N) \in S_2 \times \cdots \times S_N$, we get $P_j(s_2, \cdots, s_N) = 0$. Next we can use the induction hypothesis to show that each P_j will be the zero polynomial, which gives us the desired result: P is the zero polynomial. $\square$

7.2 Why polynomials?

There are certain properties of polynomials that are used in mathematical proofs. First we will discuss some notations. Let $\mathbb{F}$ be a field. The space of polynomials over $\mathbb{F}$ with degree at most d and n variables is denoted by $\mathrm{Poly}_d(\mathbb{F}^n)$. This is then a vector space over $\mathbb{F}$. Later on we will need the dimension of the vector space $\mathrm{Poly}_d(\mathbb{F}^n)$, therefore we will discuss that first.

Proposition 7.2.1. The dimension of the vector space $\mathrm{Poly}_d(\mathbb{F}^n)$ is $\binom{d+n}{n} \geq \frac{d^n}{n!}$.

Proof of proposition 7.2.1.
A basis of this vector space is given by the monomials $x_1^{d_1}, \cdots, x_n^{d_n}$ where $d_1 + d_2 + \cdots + d_n \leq d$. In $n+1$ variables, the number of monomials of degree d is $\binom{d+n-1}{n-1} = \binom{d+n-1}{n}$. Since there is a one-to-one correspondence between these monomials and the monomials in n variables of degree at most d, we can substitute the extra variable by 1, which gives us $\binom{d+n}{n}$ monomials. This shows that $\mathrm{Poly}_d(\mathbb{F}^n)$ has dimemsion $\binom{d+n}{n}$. Next we want to show that $\binom{d+n}{n} \geq \frac{d^n}{n!}$. First we will rewrite this as follows:

$$\binom{d+n}{n} = \frac{(d+n)!}{n!(d+n-n)!} = \frac{(d+n)!}{n!\ d!}. \tag{7}$$

Then we can divide the numerator and denominator both by $d!$ since

$$(d+n)! = (d+n) \cdot (d+n-1) \cdots (d+1) \cdot d \cdot (d-1) \cdots 2 \cdot 1 = (d+n) \cdot (d+n-1) \cdots (d+1) \cdot d!.$$

When we do this, we can rewrite (7) as

$$\frac{(d+n)\cdot(d+n-1)\cdots(d+1)}{n!} \geq \frac{d^n}{n!}.$$

We find that indeed $\binom{d+n}{n} \geq \frac{d^n}{n!}$.

□

In the article of Croot, Lev and Pach (section 8), the polynomial method is used. In their proof, they use two important properties of polynomials, which we will discuss next.

Corollary 7.2.1. Parameter counting.
Let $S \subset \mathbb{F}^n$ be a finite set. Then there exists a non-zero polynomial P with degree at most $n|S|^{1/n}$ that vanishes on S.

Proof of corollary 7.2.1.
We will denote the vector space of functions $S \to \mathbb{F}$ as $\mathrm{Fcn}(S, \mathbb{F})$. Since we are looking at polynomials in S, we get the linear map $\mathrm{Poly}_d(\mathbb{F}^n) \to \mathrm{Fcn}(S, \mathbb{F})$. If and only if this (linear) map has a non-trivial kernel, there will be a non-zero polynomial vanishing on S of degree at most d. This map will not have a non-trivial kernel as long as the dimension of the range is smaller than the dimension of the domain. The dimension of the range is $|S|$ and we calculated the dimension of the domain in proposition 7.2.1: $\binom{d+n}{n}$. By using the bound of property 7.2.1, we can find that for $d \leq n|S|^{1/n}$, we will always get $\binom{d+n}{n} > |S|$, which is what we wanted to proof.

□

The second property of polynomials used is that a non-zero polynomial in one variable can never vanish on more points than its degree. We have already mentioned this in lemma 7.1.1, but we can rewrite this as the following more general lemma:

Lemma 7.2.1. Vanishing lemma.
Let L be a line in a vector space and let P be a polynomial of degree at most d. If P vanishes at $d+1$ points of L, then P will vanish on L.

Both corollary 7.2.1 and lemma 7.2.1 are used in section 8 to complete the proof of the theorem mentioned at the beginning of the article of Croot, Lev and Pach.

8 Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small

Definition 8.0.1. Let $m \in \mathbb{N}$. We call two integers a and b congruent modulo m if $m|a-b$. Therefore we often use the following notation: $a \equiv b \bmod m$. The set $\{b \in \mathbb{Z} | b \equiv a \bmod m\}$ is also known as the residue class ($a \bmod m$) and we denote the set of residue classes modulo m by $\mathbb{Z}/m\mathbb{Z} = \mathbb{Z}_m$. A group G is named cyclic if $G = \mathbb{Z}/m\mathbb{Z}$. Lastly, a direct product of one or several pairwise isomorphic cyclic groups is referred to as a homocyclic group.

We denote by H the binary entropy function (section 5).

Theorem 8.0.1. (Croot, Lev & Pach, 2016) [7]
Let $n \geq 1$ and take $A \subseteq \mathbb{Z}_4^n$ progression-free (section 4), if we define:

$\gamma := \max\{\frac{1}{2}(H(0.5-\varepsilon) + H(2\varepsilon)) : 0 < \varepsilon < 0.25\} \approx 0.926$,

then $|A| \leq 4^{\gamma n}$.

If we take a look at $G \cong Z_{m_1} \oplus \cdots \oplus Z_{m_k}$ (a finite abelian group), where $m_1 | \cdots | m_k$ are positive integers, we have some indices $i \in [1, k]$ with the property that $4|m_i$. We will denote the number of these indices by $rk_4(G)$. Now certain cosets of a supgroup isomorphic to $\mathbb{Z}_4^n$ will form G. We need precisely a union of $4^{-n}|G|$ of these cosets to form G, where we define n as $n := rk_4(G)$. Using Theorem 8.0.1, we can directly phrase the next corollary.

Corollary 8.0.1. (Croot, Lev & Pach, 2016) [7]
We define G to be a finite abelian group, and we define n as before. Then we find the following bound: $r_3(G) \leq 4^{-(1-\gamma)n}|G|$, with $\gamma \approx 0.926$ the constant of Theorem 8.0.1.

Proof of Corollary 8.0.1.
We take a look at the finite abelian group G which is a union of $4^{-n}|G|$ cosets: $H + g_1, H + g_2, \cdots, H + g_{4^{-n}|G|}$. We take a look at B which is the largest progression free subset of G. Assume $|B| = r_3(G) > 4^{-(1-\gamma)n}|G|$. Since $4^{-n}|G|$ is the number of cosets of G, we find (by the pigeonhole principle (section 6) that there exists a coset $H + g_j$ where the number of elements of B in this coset is at least $4^{\gamma n}$. We find this bound by dividing the number of elements of B by the number of cosets: $\frac{|B|}{4^{-n}|G|} > \frac{4^{-(1-\gamma)n}|G|}{4^{-n}|G|} = 4^{\gamma n}$. We name this coset H_B where $(H + g_j) \cap B = H_B + g_j$. According to Theorem 8.0.1, this coset has a progression (since $|H_B| > 4^{\gamma n}$). If we take elements $h_1, h_2, h_3 \in H_B$, we can say that $h_1 + h_2 = 2h_3$. If we add $2g_j$ on both sides, we get the following equation: $(h_1 + g_j) + (h_2 + g_j) = 2(h_3 + g_j)$ where $h_1 + g_j, h_2 + g_j, h_3 + g_j \in B$. This shows that there exists a progression in B. But this is in contrast with the definition of B, since that was the largest progression-free subset of G. Subsequently we find that $|B| = r_3(G) \leq 4^{-(1-\gamma)n}|G|$.
□

To proof theorem 8.0.1, Croot, Lev and Pach use the following lemma:

Lemma 8.0.1. (Croot, Lev & Pach, 2016) [7])
We start with two integers $n \geq 1$ and $d \geq 0$ and we define P to be a multilinear polynomial in n variables with total degree at most d. Furthermore this polynomial is over a field $\mathbb{F}$ and we have a subset $A \subseteq \mathbb{F}^n$ where $|A| > 2\sum_{0 \leq i \leq d/2} \binom{n}{i}$. If for every distinct $a, b \in A$ $(a \neq b)$ $P(a-b) = 0$ holds, then also $P(0) = 0$.

To prove this lemma, we will take a look at a specific example and generalize this idea. For this lemma, we need to define the following parameters: $K_i \subseteq [n] = \{\emptyset, 1, 2, \cdots, n\}$, $|K_i| \leq \frac{d}{2}$, $\kappa = \{K_1, K_2, \cdots K_{n+1}\}$ and

$$m := \sum_{0 \leq i \leq d/2} \binom{n}{i}.$$

Example 8.0.1. Suppose $n = 3$ and $P(x_1, x_2, x_3) = x_1 + x_2x_3$. Then we see that $d = 2$ since the total degree is at most 2. Now we find that

$$m := \sum_{0 \leq i \leq 1} \binom{3}{i} = 4.$$

In order to apply lemma 8.0.1, we will assume that $|A| > 2m = 8$. Furthermore $K_i \subseteq [3] = \{1, 2, 3, \emptyset\}$ and $|K_i| \leq \frac{d}{2} = \frac{2}{2} = 1$, so $\kappa = \{K_1, \cdots, K_4\} = \{\{1\}, \{2\}, \{3\}, \emptyset\}$.

Now we have

$$\begin{aligned} P(x - y) &= (x_1 - y_1) + (x_2 - y_2)(x_3 - y_3) \\ &= x_1 - y_1 + x_2x_3 - x_2y_3 - x_3y_2 + y_2y_3 \end{aligned}$$

Since $x^I := \prod_{i \in I} x_i$, we can rewrite this as:

$$P(x - y) = x^{\{1\}} - y^{\{1\}} + x^{\{2,3\}} - x^{\{2\}}y^{\{3\}} - x^{\{3\}}y^{\{2\}} + y^{\{2,3\}} \tag{8}$$

Furthermore we will use that:

$$u_i(x) = x^{K_i}, \quad u_{m+i}(x) = \sum_{\substack{I \subseteq [3] \setminus K_i \\ 1 < |I| \leq 2 - |K_i|}} C_{I,K_i} x^I \tag{9}$$

and

$$v_i(y) = \sum_{\substack{J \subseteq [3] \setminus K_i \\ |J| \leq 2 - |K_i|}} C_{K_i,J} y^J, \quad v_{m+i}(y) = y^{K_i} \tag{10}$$

Now we can write the elements of (8) as follows:

$$x^{\{1\}} = x^{K_1} \sum_{\substack{J \subseteq \{2,3,\emptyset\} \\ |J| \leq 2 - |K_1|}} C_{K_1,J} y^J = u_1(x)v_1(y), \text{ where } C_{\{1\},\emptyset} = 1$$

$$-y^{\{1\}} + y^{\{2,3\}} = x^{K_4} \sum_{\substack{J \subseteq \{1,2,3\} \\ |J| \leq 2 - |K_4|}} C_{K_4,J} y^J = u_4(x)v_4(y), \text{ where } C_{\emptyset,\{1\}} = -1 \text{ and } C_{\emptyset,\{2,3\}} = 1$$

$$x^{\{2,3\}} = \sum_{\substack{I \subseteq \{1,2,3\} \\ 1 < |I| \leq 2 - |K_4|}} C_{I,K_4} x^I y^{K_4} = u_8(x)v_8(y), \text{ where } C_{\{2,3\},\emptyset} = 1$$

$$-x^{\{2\}}y^{\{3\}} = x^{K_2} \sum_{\substack{J\subseteq\{1,3,\emptyset\}\\ |J|\leq 2-|K_2|}} C_{K_2,J}y^J = u_2(x)v_2(y), \text{ where } C_{\{2\},\{3\}} = -1$$

$$-x^{\{3\}}y^{\{2\}} = x^{K_3} \sum_{\substack{J\subseteq\{1,2,\emptyset\}\\ |J|\leq 2-|K_3|}} C_{K_3,J}y^J = u_3(x)v_3(y), \text{ where } C_{\{3\},\{2\}} = -1$$

Now we can conclude that we can write (8) as:

$$\begin{aligned} P(x-y) &= \sum_{I\in\kappa} x^I \sum_{\substack{J\subseteq[3]\setminus I\\ |J|\leq 2-|I|}} C_{I,J}y^J + \sum_{J\in\kappa}\left(\sum_{\substack{I\subseteq[3]\setminus J\\ 1<|I|\leq 2-|J|}} C_{I,J}x^I\right) y^J \\ &= u_1(x)v_1(y) + u_2(x)v_2(y) + \cdots + u_8(x)v_8(y) \\ &= \langle u(x), v(x)\rangle \end{aligned}$$

This shows that we can interpret (8) as the scalar product of the vectors $u(x)$, $v(y) \in \mathbb{F}^8$ defined by (9) and (10).

To prove lemma 8.0.1, we will assume the contrary and work towards a contradiction: we assume that $P(a-b) = 0$, for all $a, b \in A$ with $a \neq b$, while $P(0) \neq 0$. This would imply the following:

Claim 8.0.1. The vectors $u(a)$ and $v(b)$ are orthogonal if and only if $a \neq b$.
Proof Claim 8.0.1:
$\Rightarrow$ If the vectors $u(a)$ and $v(b)$ are orthogonal, we know that $\langle u(a), v(b)\rangle = 0$, which implies that $P(a-b) = 0$. We assumed that $P(a-b) = 0$, for all $a, b \in A$ with $a \neq b$, while $P(0) \neq 0$, hence we see that this shows that $a \neq b$.
$\Leftarrow$ If $a \neq b$, we find that $P(a-b) = 0$ (assumption). Which implies that $\langle u(a), v(b)\rangle = 0$, subsequently we can conclude that $u(a)$ and $v(b)$ are orthogonal.

□

Claim 8.0.2. The vectors $u(a)$ ($\{u(a) : a \in A\} \subseteq \mathbb{F}^8$) are linearly independent.
Proof Claim 8.0.2:
Say $\sum_{a\in A} \lambda_a u(a) = 0$, where $\lambda_a \in \mathbb{F}$. After a scalar multiplication by $v(b)$ we have the following:

$$\begin{aligned} 0 &= \langle \sum_{a\in A} \lambda_a u(a), v(b)\rangle \\ &= \sum_{a\in A} \lambda_a \langle u(a), v(b)\rangle \\ &= \lambda_b \langle u(b), v(b)\rangle \\ &= \lambda_b P(b-b) \\ &= \lambda_b P(0), \end{aligned}$$

where we used that $\langle u(a), v(b)\rangle = 0$ when $a \neq b$ (Claim (8.0.1)) in line three. Which shows that $\lambda_b = 0$ for any $b \in A$ since $P(0) \neq 0$ (assumption). Therefore the vectors $u(a)$, ($\{u(a) : a \in A\} \subseteq \mathbb{F}^8$) are linearly independent.

□

As a result from claim (8.0.2) we find that $|A| \leq 8 = 2m$ which is in contrast with the assumptions that we took at the beginning of this lemma. Hence we can conclude that if $P(a-b) = 0$ for all $a, b \in A$ with $a \neq b$, then also $P(0) = 0$, which is lemma 8.0.1. We can do the same for bigger values of n. Where we defined

$$u_i(x) = x^{K_i}, \quad u_{m+i}(x) = \sum_{\substack{I \subseteq [n] \setminus K_i \\ d/2 < |I| \leq d - |K_i|}} C_{I,K_i} x^I$$

and

$$v_i(y) = \sum_{\substack{J \subseteq [n] \setminus K_i \\ |J| \leq d - |K_i|}} C_{K_i,J} y^J, \quad v_{m+i}(y) = y^{K_i}.$$

Next, Croot, Lev and Pach proof that the following estimate can be used for all integers $n \geq 1$ and real $0 < z \leq n/2$:

$$\sum_{0 \leq i \leq z} \binom{n}{i} < 2^{nH(z/n)} \tag{11}$$

In the proposition below, we will use that for $n \geq d \geq 0$, the dimension of the space of all multilinear polynomials in n variables of total degree of maximum d over the field $\mathbb{F}_2$ can be written as $\sum_{i=0}^{d} \binom{n}{i}$. We can also say that any non-zero multilinear polynomial actually represents a non-zero function.

The subgroup of $\mathbb{Z}_4^n$ generated by its involutions, will be denoted by F_n. Involutions are functions that are their own inverse. In other words, we can say $F_n \cong \mathbb{Z}_2^n$, since F_n is the kernal and the image of the doubling endomorphism of $\mathbb{Z}_4^n$. Where we can define the doubling endomorphism of $\mathbb{Z}_4^n$ as $g \to 2g$ $(g \in \mathbb{Z}_4^n)$. This leads us to the following proposition.

Proposition 8.0.1. (Croot, Lev & Pach, 2016) [7])
Consider $n \geq 1$ and $A \subseteq \mathbb{Z}_4^n$ a progression-free subset. The number of F_n- cosets with at least $2^{nH(0.5-\varepsilon)+1}$ elements of A (where $0 < \varepsilon < 0.25$), will be less than $2^{nH(2\varepsilon)}$.

In the proof of this proposition, they used the polynomial method (section 7.1), where they define a non-zero multilinear polynomial $P \in \mathbb{F}_2[x_1, \cdots x_n]$ of total dedree $P \leq d$, such that P vanishes on a certain set. Next they use lemma 8.0.1 to show that P vanishes on all of $\mathbb{F}_2^n$, so it turns out that P is the zero polynomial. This is contrary to the assumption that P was a non-zero polynomial. This shows that the number of F_n-cosets containing at least $2^{nH(0.5-\varepsilon)+1}$ elements of A is less than $2^{nH(2\varepsilon)}$.

Now we can use proposition 8.0.1 to prove theorem 8.0.1.

Proof Theorem 8.0.1.
For $x \geq 0$, the number of F_n- cosets with at least x elements of A, will be denoted by $N(x)$; this shows that $N(x) = 0$ for $x > 2^n$, hence we can say

$$|A| = \int_0^{2^{n+1}} N(x)\ dx = \int_0^{2^{nH(\frac{1}{4})+1}} N(x)\ dx + \int_{2^{nH(\frac{1}{4})+1}}^{2^{n+1}} N(x)\ dx. \tag{12}$$

For all $x \geq 0$, we can immediately say that $N(x) \leq 2^n$, then we can write the first integral as

$$\int_0^{2^{nH(\frac{1}{4})+1}} N(x)\ dx \leq \int_0^{2^{nH(\frac{1}{4})+1}} 2^n\ dx = [2^n x]_0^{2^{nH(\frac{1}{4})+1}} = 2^n \cdot 2^{nH(\frac{1}{4})+1} = 2 \cdot 2^{n(H(\frac{1}{4})+1)} \tag{13}$$

We can find the next upper bound:

$$\begin{aligned} H(\frac{1}{4}) + 1 &= -\frac{1}{4} \cdot \log_2(2^{-2}) - \frac{3}{4} \cdot \log_2(\frac{3}{4}) + 1 = -\frac{1}{4} \cdot -2 - \frac{3}{4} \cdot (\log_2(3) - \log_2(4)) + 1 \\ &= 1\frac{1}{2} - \frac{3}{4}\log_2(3) + \frac{3}{4}\log_2(2^2) = 1\frac{1}{2} - \frac{3}{4}\log_2(3) + \frac{3}{4} \cdot 2 = 3 - \frac{3}{4}\log_2(3) \\ &\approx 1.811 < 1.852 = 2\gamma. \end{aligned}$$

Which shows that we can rewrite (13) as

$$2 \cdot 2^{n(H(\frac{1}{4})+1)} < 2 \cdot 2^{2\gamma n} = 2 \cdot 4^{\gamma n}. \tag{14}$$

For the second integral, we substitute $x = 2^{nH(0.5-\varepsilon)+1}$ which gives us

$$\int_{2^{nH(\frac{1}{4})+1}}^{2^{n+1}} N(x)\ dx = \int_b^u N(2^{nH(0.5-\varepsilon)+1})\ d2^{nH(0.5-\varepsilon)+1} \tag{15}$$

To determine the boundaries of this integral, we equate the boundaries that we had to x. Thus for b we get

$$2^{nH(\frac{1}{4})+1} = x = 2^{nH(0.5-\varepsilon)+1},$$

which implies that $\varepsilon = 0.25$. For u we get

$$2^{n+1} = x = 2^{nH(0.5-\varepsilon)+1},$$

which implies that $\varepsilon = 0$.

Now we see that we can write (15) as

$$\int_{0.25}^{0} N(2^{nH(0.5-\varepsilon)+1})\ d2^{nH(0.5-\varepsilon)+1} = -\int_0^{0.25} N(2^{nH(0.5-\varepsilon)+1})\ d2^{nH(0.5-\varepsilon)+1} \tag{16}$$

To find the value of this integral, we need to determine $\frac{d(2^{nH(0.5-\varepsilon)+1})}{d\varepsilon}$ and therefore we need to know $\frac{d(H(0.5-\varepsilon))}{d\varepsilon}$.

$$\begin{aligned} \frac{d(H(0.5-\varepsilon))}{d\varepsilon} &= -1 \cdot H'(0.5-\varepsilon) = -1 \cdot \frac{d(H(x))}{dx}|_{x=0.5-\varepsilon} = -\log_2\left(\frac{1-x}{x}\right)|_{x=0.5-\varepsilon} \\ &= -\log_2\left(\frac{0.5+\varepsilon}{0.5-\varepsilon}\right) = -\log\left(\frac{0.5+\varepsilon}{0.5-\varepsilon}\right)/\log(2) \end{aligned}$$

Now we can use this to find the following

$$\begin{aligned}\frac{d(2^{nH(0.5-\varepsilon)+1})}{d\varepsilon} &= \log(2)\cdot 2^{nH(0.5-\varepsilon)+1}\cdot n\cdot \frac{d(H(0.5-\varepsilon))}{d\varepsilon}\\ &= \log(2)\cdot 2^{nH(0.5-\varepsilon)+1}\cdot n\cdot -\log\left(\frac{0.5+\varepsilon}{0.5-\varepsilon}\right)/\log(2)\end{aligned}$$

In consequence we find that

$$d(2^{nH(0.5-\varepsilon)+1}) = -n\cdot 2^{nH(0.5-\varepsilon)+1}\cdot \log\left(\frac{0.5+\varepsilon}{0.5-\varepsilon}\right)\ d\varepsilon$$

We can use this to rewrite (16) as

$$\begin{aligned}&n\int_0^{0.25} 2^{nH(0.5-\varepsilon)+1}N(2^{nH(0.5-\varepsilon)+1})\ \log\left(\frac{0.5+\varepsilon}{0.5-\varepsilon}\right)\ d\varepsilon\\ &< n\int_0^{0.25} 2\cdot 2^{nH(0.5-\varepsilon)}\cdot 2^{nH(2\varepsilon)}\ \log\left(\frac{0.5+\varepsilon}{0.5-\varepsilon}\right)\ d\varepsilon\end{aligned} \tag{17}$$

since $N(2^{nH(0.5-\varepsilon)+1}) < 2^{nH(2\varepsilon)}$ (Proposition 1).

Moreover we can rewrite (17) as

$$\begin{aligned}&2n\int_0^{0.25} 2^{n(H(0.5-\varepsilon)+H(2\varepsilon))}\ \log\left(\frac{0.5+\varepsilon}{0.5-\varepsilon}\right)\ d\varepsilon\\ &< 2n\cdot 1.5\int_0^{0.25} 2^{n(H(0.5-\varepsilon)+H(2\varepsilon))}\ d\varepsilon,\end{aligned} \tag{18}$$

where we used that $\log\left(\frac{0.5+0.25}{0.5-0.25}\right) = \log(3) < 1.5$.

Furthermore since $2\gamma = \max\{H(0.5-\varepsilon)+H(2\varepsilon) : 0<\varepsilon<0.25\}$, we can bound (18) as

$$< 3n\int_0^{0.25} 2^{n2\gamma}\ d\varepsilon = 3n\cdot[2^{n2\gamma}\varepsilon]_0^{0.25} = \frac{3}{4}n\cdot 2^{2n\gamma} < n\cdot 2^{2n\gamma} = n\cdot 4^{\gamma n} \tag{19}$$

Finally we can use (14) and (19) to bound (12) as

$$|A| < 2\cdot 4^{\gamma n} + n\cdot 4^{\gamma n} = (n+2)\cdot 4^{\gamma n}. \tag{20}$$

Now we use the tensor power trick to complete the proof. Since the subset $A\subseteq \mathbb{Z}_4^n$ is progression-free, we can also say that the set $A\times A\times\cdots\times A\subseteq \mathbb{Z}_4^{kn}$ is progression free. By using (20) we see that

$$|A|^k < (kn+2)\cdot 4^{\gamma kn},$$

which we can rewrite as

$$|A| < \sqrt[k]{(kn+2)\cdot 4^{\gamma kn}} = \sqrt[k]{kn+2}\cdot\sqrt[k]{4^{\gamma kn}} = (kn+2)^{\frac{1}{k}}\cdot 4^{\gamma n}$$

This expression holds for every integer $k \geq 1$, therfore we can find the limit from k to ∞:

$$|A| \leq \lim_{k\to\infty}(kn+2)^{\frac{1}{k}}\cdot 4^{\gamma n} = 1\cdot 4^{\gamma n} = 4^{\gamma n}.$$

$\square$

To that end, we indeed find the result of Theorem 8.0.1.

9 On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression

Jordan S. Ellenberg and Dion Gijswijt have shown that the ideas of Croot, Lev and Pach can be extended to vector spaces over a general finite field [10]. They worked on this problem independently and they developed the same idea simultaneously and decided to present them in a joint work.

In their paper, they start by generalizing lemma 8.0.1 of Croot, Lev and Pach. In their proposition they use some notations that we will discuss first. We define $\mathbb{F}_q$ to be a finite field and n a positive integer. We denote by M_n the set of monomials in $x_1, \cdots x_n$ with the requirement that their degree in each variable is at most $q-1$. The $\mathbb{F}_q-$vector space spanned by these monomials is called S_n.

We define M_n^d as the set of monomials in M_n of degree at most d, where d is any real number in $[0, 2n]$. Furthermore S_n^d is the subspace of S_n spanned by these monomials. For the dimension of S_n^d we will write m_d. Lastly we will speak of a "polynomial of degree at most d", when we speak of an element of S_n^d.

Proposition 9.0.1. We start with a finite field $\mathbb{F}_q$ and a subset $A \subseteq \mathbb{F}_q^n$. Next we need three elements of $\mathbb{F}_q$ that will sum to zero: α, β, γ. Assume that for every pair $a, b \in A$, with $a \neq b$, $P \in S_n^d$ satisfies the property that $P(\alpha a + \beta b) = 0$. Then the number of $a \in A$ which measure up to $P(-\gamma a) \neq 0$, will be at most $2m_{d/2}$.

The proof of this proposition is essentially the same as the proof of lemma 8.0.1 of Croot, Lev and Pach. The only essential addition is that P has to take vanish at a larger set of places. Again this is done by using the polynomial method (section 7.1).

Theorem 9.0.1. Again let α, β, γ be elements of $\mathbb{F}_q$, which sum up to zero. Take $A \subseteq \mathbb{F}_q^n$ a subset where apart from $a_1 = a_2 = a_3$, there are no solutions $(a_1, a_2, a_3) \in A^3$ for the equation

$$\alpha a_1 + \beta a_2 + \gamma a_3 = 0.$$

Then we find that $|A| \leq 3m_{(q-1)n/3}$.

In the proof of this theorem, we use proposition 9.0.1 to show that

$$m_d - q^n + |A| \leq 2m_{d/2}$$

Subsequently we get

$$|A| \leq 2m_{d/2} + (q^n - m_d),$$

where $q^n - m_d$ can be seen as the number of $q-$power-free monomials with degree greater than d. These monomials are in bijection with the monomials with degree less than $(q-1)n-d$. Since there are at most $m_{(q-1)n-d}$ of those monomials, we get $q^n - m_d \leq m_{(q-1)n-d} = m_{(q-1)n/3)}$, where $d = 2(q-1)n/3$. Now we can find the desired upper bound

$$|A| \leq 2m_{(q-1)n/3} + (q^n - m_{2(q-1)n/3}) \leq 3m_{(q-1)n/3}. \tag{21}$$

Next we define X to be a variable which takes values $0, 1, \cdots, q-1$. Each value has a probability of $\frac{1}{q}$ to be taken. From this it follows that we can consider $\frac{m_{(q-1)n/3}}{q^n}$ to be the probability that n independent copies of X have a maximum mean of $\frac{q-1}{3}$. Since this is a large deviation problem, we can use Cramér's theorem to find that

$$\lim_{n\to\infty} \frac{1}{n} \log\left(\frac{m_{(q-1)n/3}}{q^n}\right) = -I\left(\frac{q-1}{3}\right), \tag{22}$$

where $I(x)$ is the supremum of

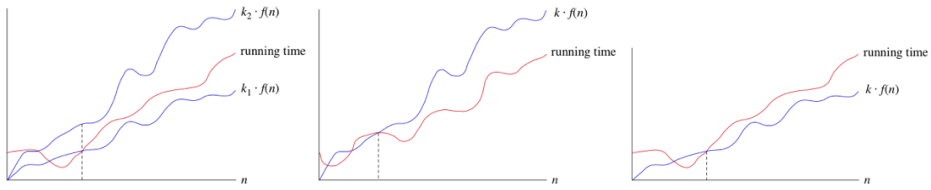
$$\theta x - \log\left(\frac{1 + e^{\theta} + \cdots + e^{(q-1)\theta}}{q}\right), \tag{23}$$

over all θ in $\mathbb{R}$.

Lastly we can find a new upper bound as follows:

Corollary 9.0.1. If A is a subset of $(\mathbb{Z}/3\mathbb{Z})^3$ with no three-term arithmetic progression, we can say the following about the number of elements of A:
$|A| = o(2.756^n)$.

We can take $q = 3$ and thus $x = 2/3$ to find that the supremum of (23) is attained when $e^{\theta} = (\sqrt{33} - 1)/8$. By using (21) and (22), we can now find the bound $3e^{-I(2/3)} < 2.756$. Lastly we can apply theorem 9.0.1 with $\alpha = \beta = \gamma = 1$, which directly gives us the desired result.



(a) Big−Θ (b) Big−O (c) Big−Ω

Figure 8: Asymptotic notations.

10 An improved construction of progression-free sets

In this paragraph we follow the work of Elkin [9] and we will use some asymptotic notations which we will explain first (these are also used in further paragraphs). The first asymptotic notation is $\Theta(f(n))$. This is referred to as the big−Theta notation and it asymptotically bounds the growth of a running time from above and from below (in our case, the running-time is the absolute value of $f(n)$). For n big enough, this notation tells us that for some constants k_1 and k_2, the running time will be at least $k_1 \cdot f(n)$ and at most $k_2 \cdot f(n)$ (figure 8a).
If we want to bound the running time just from above, we can use the big−O notation ($O(f(n))$). Here for n big enough and for k a constant, we can say that the running time will be at most $k \cdot f(n)$ (figure 8b).
Lastly if we want to bound the running time just from below, hence if we want to say that a running time is at least a certain amount of time, we can use the big−Omega notation ($\Omega(f(n))$). Here for n big enough and a constant k, the running time is at least $k \cdot f(n)$ (figure 8c).

The first mathematicians to look at the problem of finding subsets S of $\{1, 2, \cdots, n\}$ with no arithmetic three-term were Erdös and Turán in 1936. In their research they found a construction which consisted of $|S| = \Omega(n^{\log_3(2)})$ elements (Erdös & Túran, 1936 [11]). Later Salem and Spencer improved this result [22] and eventually Behrend was able to find a new lower bound in 1946: $|S| = \Omega\left(\frac{n}{2^{2\sqrt[2]{\log_2(n)}} \cdot \log^{1/4}(n)}\right)$ (Behrend, 1946 [3]). Since then there was no improvement in the result of Behrend until 2008: Michael Elkin discovered that Behrends construction wasn't optimal and he found the new lower bound: $|S| = \Omega\left(\frac{n}{2^{2\sqrt[2]{\log_2(n)}}} \cdot \log^{1/4}(n)\right)$.

Definition 10.0.1. A subset $S \subseteq \{1, 2, \cdots, n\}$ is called progression-free if it contains no three distinct elements $i, j, l \in S$ such that i is the arithmetic average of j and l: $i = \frac{j+l}{2}$. For a positive integer n, the largest size of a progression-free subset S (of $\{1, 2, \cdots, n\}$) will be denoted by $\nu(n)$. From now on, we will use $[\{n\}]$ when we mean $[\{1, 2, \cdots, n\}]$.

In 1946 Behrend used the pigeonhole principle (section 6) to show that $\nu(n) = \Omega\left(\frac{n}{2^{2\sqrt[2]{log(n)}} \cdot \log^{1/4}(n)}\right)$. For over sixty years, nobody was able to improve this upper bound. In a seminal paper, Roth found the first non-trivial upper bound $\nu(n) = O(\frac{n}{\text{loglog(n)}})$. Later Bourgain was able to improve this upperbound to the current best upper bound: $\nu(n) = O(n \cdot \frac{(\text{loglog(n)})^2}{\log^{2/3} n})$.

The past sixty years a lot of intensive research has been done in this area to improve Behrends lower bound. Though the recent found improvement by Elkin is not that large, it is interesting to show that Behrends construction isn't optimal.

10.1 Behrend construction

Behrend based his proof on the observation that a sphere cannot contain an arithmetic progression since a sphere is convex in any dimension (Behrend, 1946 [3]).

In Behrends proof we start with a sufficiently large positive integer n. After that, we set $y = \frac{n^{1/k}}{2}$, where we first assume that y is again an integer. Furthermore we will need a positive integer parameter k which we will determine later.

Now we consider random variables $Y_1, Y_2, \cdots, Y_k$ which are independent identically distributed, where for all $i \in [\{k\}]$, each Y_i is distributed uniformly over the set $[\{0, y-1\}]$. Next for all $i \in [\{k\}]$ we set $Z_i = Y_i^2$ and $Z = \sum_{i=1}^{k} Z_i$.

We can now use the expectation of Z_i to find the expectation of the random variable Z. We find that

$$\mu_Z = \mathbb{E}(Z) = \frac{k}{3}y^2 + \Theta(k \cdot y). \tag{24}$$

Then we can easily find the variance of Z_i as follows: $\mathrm{Var}(Z_i) = \mathbb{E}(Z_i^2) - \mathbb{E}(Z_i)^2$. We find that $\mathrm{Var}(Z) = k \cdot y^4 \cdot \frac{4}{45} \cdot (1 + O(\frac{1}{y}))$. Subsequently we see that the standard deviation of Z is the following

$$\sigma_Z = \sqrt{k} \cdot y^2 \cdot \frac{2}{3 \cdot \sqrt{5}} \cdot (1 + O(\frac{1}{y})). \tag{25}$$

We can now combine all these results together with the Chebyshev inequality to show that for any $a > 0$,

$$\mathbb{P}(|Z - \mu_Z| > a \cdot \sigma_Z) \leq \frac{1}{a^2}.$$

Next we can conclude that for a fixed value of $a > 0$, of all vectors v from the set $[\{0, y-1\}]$ we will have at least a $(1 - \frac{1}{a^2})$-fraction that have squared norm. Where squared norm satisfies $\mu_Z - a \cdot \sigma_Z \leq ||v||^2 \leq \mu_Z + a \cdot \sigma_Z$.

Thereafter we can use the Pigeonhole Principle to show that there is a value T that satisfies that at least $(1 - \frac{1}{a^2}) \cdot \frac{1}{2a \cdot \sigma_Z} \cdot y^k$ vectors in the set $[\{0, y-1\}]$ have squared norm T, if $\mu_Z - a \cdot \sigma_Z \leq T \leq \mu_Z + a \cdot \sigma_Z$. We will denote the set of these vectors by $\mathcal{S}$.

Next we can use (25) to bound the cardinality of $\mathcal{S}$:

$$|\mathcal{S}| \geq (1 - \frac{1}{a^2}) \cdot \frac{1}{2a \cdot \sqrt{k} \cdot y^2} \cdot \frac{3\sqrt{5}}{2} \cdot (1 - O(\frac{1}{y})) \cdot y^k = \frac{y^{k-2}}{\sqrt{k}} \cdot c,$$

where $c = c(a)$ is a fixed positive constant.

If we set $a = 2$, we have the universal constant $c = c(2)$ and it follows that $|\mathcal{S}| = \Omega\left(\frac{n^{k-2}}{2^k \sqrt{k}}\right)$.

Now we can maximize the right-hand-side by setting $k = \left\lceil \sqrt{2 \cdot \log_2(n)} \right\rceil$.

Subsequently we find that

$$|\mathcal{S}| = \Omega\left(\frac{n}{2^{2\sqrt[2]{log_2(n)}} \cdot \log^{1/4}(n)}\right).$$

For every three vectors $v, u, w \in \mathcal{S}$, we see that $v \neq \frac{u+w}{2}$ since all vectors in $\mathcal{S}$ have the same norm $\sqrt{T}$.

The coordinates of vectors from $\mathcal{S}$ that we will consider, need to be digits of a $2y-$ary representation. This means that we get $\hat{v} = \sum_{i=0}^{k-1} v_{i+1} \cdot (2y)^i$ for every vector $v = (v_1, v_2, \cdots, v_k) \in \mathcal{S}$. Using this representation, we can describe the set S as follows: $S = \{\hat{v} \mid v \in \mathcal{S}\}$. We will denote this mapping by $f(\cdot) : \mathcal{S} \to S$.

We see that for every $v \in \mathcal{S}$,

$$0 < \hat{v} \leq (2y)^k - 1 = n - 1$$

The mapping f is injective (definition 2.2.1) since $\mathcal{S} \subseteq [\{0, y-1\}]^k$. Hence for $u, v \in \mathcal{S}$, if $u \neq v$, then also $\hat{u} \neq \hat{v}$. Subsequently we find that

$$|S| = |\mathcal{S}| = \Omega\left(\frac{n}{2^{\sqrt[2\sqrt{2}]{log_2(n)}} \cdot \log^{1/4}(n)}\right)$$

Lastly we have to show that S is progression-free. We start by assuming the contradiction that for distinct numbers $\hat{u}, \hat{v}, \hat{w} \in S$ we have $\hat{v} = \frac{\hat{u}+\hat{w}}{2}$. The corresponding vectors in S are $u = (u_1, u_2, \cdots, u_k), v = (v_1, v_2, \cdots, v_k)$ and $w = (w_1, w_2, \cdots, w_k)$, which shows that

$$\hat{v} = \sum_{i=0}^{k-1} \frac{u_{i+1} + w_{i+1}}{2} \cdot (2y)^i = \sum_{i=0}^{k-1} v_{i+1} \cdot (2y)^i.$$

This shows that for every index $i \in [\{k\}]$ we have $v_i = \frac{u_i+w_i}{2}$ since all coordinates $u_1, u_2, \cdots, u_k, v_1, v_2, \cdots, v_k, w_1, w_2, \cdots, w_k \in [\{0, y-1\}]$. Since the equality holds for every index i, we can conclude that $v = \frac{u+w}{2}$, which contradicts the fact that $||u|| = ||v|| = ||w||$. In the end we find that S is progression-free and that S has size $\Omega\left(\frac{n}{2^{\sqrt[2\sqrt{2}]{log_2(n)}} \cdot \log^{1/4}(n)}\right)$.

We still have to take a look at the case when $y = \frac{n^{1/k}}{2}$ is not an integer. The construction is exactly the same except that we work with $\lfloor y \rfloor$ instead of y. In this case we have $n' = (2\lfloor y \rfloor)^k$. We can use the exact same arguments as in the integer case, to find that

$$|S| = |\mathcal{S}| = \Omega\left(\frac{n'}{2^{\sqrt[2\sqrt{2}]{log_2(n')}} \cdot \log^{1/4}(n')}\right) = \Omega\left(\frac{n'}{2^{\sqrt[2\sqrt{2}]{log_2(n)}} \cdot \log^{1/4}(n)}\right)$$

Lastly we can replace n' by n in this last expression since

$$\frac{n}{n'} \leq (\frac{y}{y-1})^k = 1 + \Theta(\frac{k}{y}) = 1 + \Theta\left(\frac{\sqrt{\log(n)}}{2^{(1/\sqrt{2})} \cdot \sqrt{\log(n)}}\right).$$

This shows that the resulting lower bound is at most a constant factor smaller than when y is an integer.

Therefore also when y is not an integer number, we find that $|S| = \Omega\left(\frac{n}{2^{\sqrt[2\sqrt{2}]{log_2(n)}} \cdot \log^{1/4}(n)}\right)$.

10.2 The improved construction by Michael Elkin

Theorem 10.2.1. There exist progression-free sets $S \subseteq [\{n\}]$ of at least $\Omega\left(\frac{n}{2^{2\sqrt{2}\sqrt[4]{log_2(n)}}} \cdot \log^{1/4}(n)\right)$ elements.

Proof idea:
In this proof, we will demonstrate that we can use a thin annulus, since such a thin annulus will contain a large convexly independent subset U that consists of integer points. We will show that between the width of the annulus and the size of U there exists an inherent trade off. We want to show that for certain widths of the annulus, U will contain at least a constant fraction of the integer points of the concerned annulus. We will use the largest width of the annulus for which we can proof this.

Again we set $k = \left\lceil \sqrt{2 \cdot \log_2(n)} \right\rceil$ and $y = \frac{n^{1/k}}{2}$.

First of all we bound y in the following way

$$\frac{2^{k/2}}{2\sqrt{2}} = \frac{1}{2\sqrt{2}} \cdot 2^{\frac{\sqrt{\log(n)}}{\sqrt{2}}} \leq y \leq \frac{1}{2} \cdot 2^{\frac{\sqrt{\log(n)}}{\sqrt{2}}} = \frac{2^{k/2}}{2}. \tag{26}$$

Again we start by assuming that y is an integer.

We begin by considering the $k-$dimensional ball which is centered at the origin and which has radius R'

$$R'^2 = \mu_Z = \frac{k}{3}y^2 + \Theta(ky). \tag{27}$$

We denote by C the discrete cube $[\{0, y-1\}]^k$ and we denote by $\hat{\mathcal{S}}$ the annulus of all vectors that have squared norm in $[R'^2 - 2 \cdot \sigma_Z, R'^2 + 2 \cdot \sigma_Z]$. Now we can again use Chebyshev inequality to show that $\hat{\mathcal{S}}$ contains at least $\frac{3}{4} \cdot y^k$ integer points of C.

Next we define $g = \epsilon \cdot k$, with $\epsilon > 0$ a universal constant to determine. We use g to partition the annulus $\hat{\mathcal{S}}$ into $l = \left\lceil \frac{4\sigma_Z}{g} \right\rceil$ annuli: $\hat{\mathcal{S}}_1, \hat{\mathcal{S}}_2, \cdots, \hat{\mathcal{S}}_l$. Here $\hat{\mathcal{S}}_i$ contains all vectors with squared norms in the range

$$\begin{cases} [R'^2 - 2\sigma_Z + (i-1) \cdot g, R'^2 - 2\sigma_Z + i \cdot g] & \text{for } i \in [\{l-1\}] \\ [R'^2 - 2\sigma_Z + (l-1) \cdot \sigma_Z, R'^2 + 2\sigma_Z] & \text{for } i = l \end{cases}$$

The sets of integer points in $\hat{\mathcal{S}}_i$ and $\hat{\mathcal{S}}_j$ are disjoint for distinct indices $i, j \in [\{l\}]$. Therefore we find that there exists an index $i \in [\{l\}]$ (pigeonhole principle) for which the annulus $\hat{\mathcal{S}}_i$ contains at least $\frac{3}{4l} \cdot y^k = \Omega(g \cdot \frac{y^{k-2}}{\sqrt{k}} = \Omega(\epsilon\sqrt{k} \cdot y^{k-2})$ integer points of $C \cap \hat{\mathcal{S}}$.

This means that for the annulus $\mathcal{S}$, with all vectors that have squared norm in $[R^2 - g, R^2]$, we can find a radius R where $R^2 \in [R'^2 - 2\sigma_Z, R'^2 + 2\sigma_Z]$, such that $\mathcal{S}$ contains at least $\Omega(\sqrt{k} \cdot y^{k-2})$ integer points of $C \cap \hat{\mathcal{S}}$.

We can now use (24), (25) and (27) to show that

$$R^2 \leq R'^2 + 2\sigma_Z \leq \frac{k}{3} \cdot y^2 + O(k \cdot y) + O(\sqrt{k} \cdot y^2) \leq \frac{k}{3} \cdot y^2 \left(1 + O\left(\frac{1}{\sqrt{k}}\right)\right).$$

We denote by $\tilde{\mathcal{S}}$ the set of all integer points of $C \cap \mathcal{S}$. Now $\tilde{\mathcal{S}}$ contains a subset $\check{\mathcal{S}}$ with at least $|\check{\mathcal{S}}| \geq \frac{|\tilde{\mathcal{S}}|}{2}$ integer points, which is convexly independent. Therefore and by using (26), we find

$$|\check{\mathcal{S}}| \geq \frac{|\tilde{S}|}{2} = \Omega(\sqrt{k} \cdot y^{k-2}) = \Omega\left(\log^{1/4}(n) \cdot \frac{n}{2^{2\sqrt[2]{\log_2(n)}}}\right).$$

We can again use the mapping of section 10.1 to define the set $\check{S} = f(\check{\mathcal{S}})$. We find that $|\check{S}| = |\check{\mathcal{S}}|$ since we took $\mathcal{S}$ as a convexly independent set. Furthermore, $\check{S}$ is progression-free, hence we can conclude that $|\check{S}| = \Omega\left(\log^{1/4}(n) \cdot \frac{n}{2^{2\sqrt[2]{\log_2(n)}}}\right)$.

11 Conclusion

We have seen that the cap set problem is a very interesting field for research, since there is always room for improvement in the bounds of biggest cap sets. Only precise sizes of cap sets are known for small values of n (not too many attributes). For bigger values of n there are only estimates of what the size of the biggest cap set will be. A lot of mathematicians have tried to find upper- and lower bounds for different types of the cap set problem. An important recent breakthrough was the one of Ellenberg and Gijswijt, who found a new upper bound on cap sets of $\mathbb{F}_q^n$ by using the polynomial method. Before them, Croot, Lev and Pach did important research in subsets of $\mathbb{F}_4^n$, but they only found an upper bound for the version with four choices per attribute. Furthermore we have seen that sometimes bounds can be improved after many years, like the lower bound that Behrend found in 1946, which was improved by Elkin in 2011. This shows that this problem is still very current and that bounds can always be improved.

12 References

[1] Armstrong, M. A. (1988). Groups and Symmetry. New York, U.S.A.: Springer.

[2] Bateman, M., & Katz, N. H. (2012). *New bounds on cap sets*, J. Amer. Math. Soc. 25, no. 2, 585-613.

[3] Behrend, F. (1946). *On sets of integers which contain no three terms in arithmetic progression.* Proceedings of the National Academy of Sciences of the United States of America, 32, 331-332.

[4] Bloom, T. F. (2014). *A quantitative improvement for Roths theorem on arithmetic progressions*, submitted. arXiv:1405.5800.

[5] Brown, T. C., & Buhler, J.C. (1982). *A density version of a geometric Ramsey theorem*, J. Combin. Theory, Ser. A 32, 20-34.

[6] Chartrand, G., Polimeni, A., & Zhang, P. (2013). Mathematical Proofs: A Transition to Advanced Mathematics (3e ed.). London, United Kingdom: Pearson Education Limited.

[7] Croot, E., Lev, V., & Pach, P. P. (2016). *Progression-free sets in Z_4^n are exponentially small.* Annals of Mathematics, 185 (2017), 331-337.

[8] Davis, B. L. & Maclagan, D. (2003). *The Mathematical Intelligencer* 25: 33.

[9] Elkin, M. (2011). *An improved construction of progression-free sets.* Israel Journal of Mathematics, 184, 93-128.

[10] Ellenberg, J., & Gijswijt, D. (2016). *On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression*, preprint. arXiv:1605.09223.

[11] Erdös, P., & Turán, P. (1936). *On some sequences of integers.* J. London Math. Society, 11: 261-264.

[12] Frankl, P., Graham, G., & Rödl, V. (1987). *On subsets of abelian groups with no 3-term arith- metic progression*, J. Combin. Theory, Ser. A 45, 157-161.

[13] Guth, L. (2013). *Unexpected Applications of Polynomials in Combinatorics.* The Mathematics of Paul Erds I, edited by Ronald L. Graham et al., Springer New York, pp. 493-522.

[14] Igodt, P., & Veys, W. (2015). Lineaire algebra (2nd ed.). Leuven, Belgium: University Press.

[15] Klarreich, E. (2016). Simple Set Game Proof Stuns Mathematicians, Quantamagazine.

[16] Lev, V. F. (2004). *Progression-free sets in finite abelian groups*, J. Number Theory 104, no. 1, 162-169.

[17] Meshulam, R. (1995). *On subsets of finite abelian groups with no 3-term arithmetic progressions*, J. Combin. Theory Ser. A 71, no. 1, 168-172.

[18] Paar, C., & Pelzl, J. (2010). Understanding Cryptography. Berlin Heidelberg, Germany: Springer.

[19] Pellegrino, G. (1983). On Pellegrino's 20-Caps in $S_{4,3}$, Volume 78, pages 433-447.

[20] Roth, K. (1953). *On certain sets of integers*, J. London Math. Soc. 28, 104-109.

[21] Roth, R. (2006). Introduction to Coding Theory. Cambridge: Cambridge University Press.

[22] Salem, R., & Spencer, D. (1942). *On sets of integers which contain no three in arithmetic progression.* Proceedings of the National Academy of Sciences of the USA, 28, 561-563.

[23] Sanders, T. (2011). *On Roths theorem on progressions*, Ann. of Math. (2) 174, no. 1, 619-636.

[24] Tao, T. (2014). *Algebraic combinatorial geometry: the polynomial method in arithmetic combinatorics, incidence combinatorics, and number theory.* arXiv:1310.6482.

[25] Tao, T., & Vu, V. (2006). Prologue. In Additive Combinatorics (Cambridge Studies in Advanced Mathematics). Cambridge: Cambridge University Press.