THE MAXIMUM NUMBER OF THREE TERM ARITHMETIC PROGRESSIONS, AND TRIANGLES IN CAYLEY GRAPHS

ZACHARY CHASE

ABSTRACT. Let G be a finite Abelian group. For a subset $S \subseteq G$, let $T_3(S)$ denote the number of length three arithemtic progressions in S and $\text{Prob}[S] = \frac{1}{|S|^2}\sum_{x,y\in S} 1_S(x+y)$. For any $q \geq 1$ and $\alpha \in [0,1]$, and any $S \subseteq G$ with $|S| = \frac{|G|}{q+\alpha}$, we show $\frac{T_3(S)}{|S|^2}$ and $\text{Prob}[S]$ are bounded above by $\max\left(\frac{q^2-\alpha q+\alpha^2}{q^2}, \frac{q^2+2\alpha q+4\alpha^2-6\alpha+3}{(q+1)^2}, \gamma_0\right)$, where $\gamma_0 < 1$ is an absolute constant. As a consequence, we verify a graph theoretic conjecture of Gan, Loh, and Sudakov for Cayley graphs.

1. Introduction

The study of arithmetic progressions in subsets of integers and general Abelian groups is a central topic in additive combinatorics and has led to the development of many fascinating areas of mathematics. A famous result on three term arithmetic progressions (3APs) is Roth's theorem, which, in its finitary form, says that for each $\lambda > 0$, for N large, any subset $S \subseteq \{1, \ldots, N\}$ of size $|S| \geq \lambda N$ contains a 3AP.

Once Roth's theorem ensures that all subsets of a given size have a 3AP, one can generate many 3APs. For example, Varnavides [4] proved that for each $\lambda > 0$, there is some $c > 0$ so that for all large N, every subset $S \subseteq \{1, \ldots, N\}$ with $|S| \geq \lambda N$ contains at least cN^2 3APs. A natural question is then how many 3APs a subset of $\{1, \ldots, N\}$ of a prescribed size can have. We look at this question in the group theoretic setting.

Fix $\lambda \in (0, 1)$. Let p be a large prime and consider subsets $S \subseteq \mathbb{Z}_p$ of size $|S| = \lfloor \lambda p \rfloor$. If $T_3(S)$ denotes the number of 3APs in S, namely, the number of $x, d \in \mathbb{Z}_p$ with $x, x + d, x + 2d \in S$, then Croot [1] showed that

$$\lim_{p\to\infty} \max_{\substack{S\subseteq\mathbb{Z}_p \\ |S|=\lfloor\lambda p\rfloor}} \frac{T_3(S)}{|S|^2}$$

exists, and then Green and Sisask [2] proved that the limit is in fact $\frac{1}{2}$, for all λ less than some absolute constant. In $\mathbb{Z}_n$, for n not prime, the situation is quite different, since subgroups have many 3APs relative to their size. In this paper, we nevertheless get an upper bound, useful when the size of S is "far" from dividing n.

Date: September 11th, 2018.

Theorem 1. *There is an absolute constant $\gamma_1 < 1$ so that for any finite Abelian group G of odd order, and for any $q \in \mathbb{N}, \alpha \in [0,1]$,*

$$\max_{\substack{S \subseteq G \\ |S| = \frac{|G|}{q+\alpha}}} \frac{T_3(S)}{|S|^2} \leq \max\left(\frac{q^2 - \alpha q + \alpha^2}{q^2}, \frac{q^2 + 2\alpha q + 4\alpha^2 - 6\alpha + 3}{(q+1)^2}, \gamma_1 \right).$$

Related to $\frac{T_3(S)}{|S|^2} = \frac{1}{|S|^2} \sum_{x,y \in S} 1_S(\frac{x+y}{2})$ is the quantity $\frac{1}{|S|^2} \sum_{x,y \in S} 1_S(x+y)$. This quantity, which we denote $\text{Prob}[S]$, arises in the expression for the number of triangles in a Cayley graph with generating set S. Precisely, let G be an additive group of size n and $S \subseteq G$ a symmetric set not containing 0. Connect $x, y \in G$ iff $x - y \in S$. We obtain an undirected graph on G with no self loops. The number of triangles in our graph is

$$\frac{1}{6} \sum_{a,b,c \in G} 1_S(a-b) 1_S(b-c) 1_S(a-c).$$

Let $x = a - b$ and $y = b - c$. Then ranging over c, b, a is equivalent to ranging over c, y, x and thus

$$|T| = \frac{1}{6} \sum_{x,y,c} 1_S(x) 1_S(y) 1_S(x+y) = \frac{1}{6} n \sum_{x,y \in S} 1_S(x+y) = \frac{1}{6} n |S|^2 \,\text{Prob}[S].$$

Quite recently, Gan, Loh, and Sudakov [3] resolved a conjecture of Engbers and Galvin regarding the maximum number of independent sets of size 3 that a graph with a given minimum degree and fixed size can have. Phrased in complementary graphs, they showed that given a maximum degree d and a positive integer $n \leq 2d+2$, the maximum number of triangles that a graph on n vertices with maximum degree d can have is $\binom{d+1}{3} + \binom{n-(d+1)}{3}$. This immediately raised the question of what the maximum is for $n > 2d+2$. They conjectured the following.

Conjecture (Gan-Loh-Sudakov)**.** *Fix $d \geq 2$. For any positive integer n, if we write $n = q(d+1) + r$ for $0 \leq r \leq d$, then the maximum number of triangles that a graph on n vertices with maximum degree d can have is $q\binom{d+1}{3} + \binom{r}{3}$.*

For each d, n, an example of a graph achieving $q\binom{d+1}{3} + \binom{r}{3}$ is simply a disjoint union of K_{d+1}'s and a K_r. The conjecture for a Cayley graph on an additive group G with generating set S, $|S| = \frac{|G|}{q+\alpha}$, takes the form $\text{Prob}[S] \leq \frac{q+\alpha^3}{q+\alpha}$, up to smaller order terms. We verify the conjecture for Cayley graphs when $q \geq 7$.

Theorem 2. *There is an absolute constant $\gamma_0 < 1$ so that the following holds. Let G be a finite Abelian group and take $q \in \mathbb{N}, \alpha \in [0,1]$. Then for any symmetric subset $S \subseteq G$ with $|S| = \frac{|G|}{q+\alpha}$,*

$$\frac{1}{|S|^2} \sum_{x,y \in S} 1_S(x+y) \leq \max\left(\frac{q^2 - \alpha q + \alpha^2}{q^2}, \frac{q^2 + 2\alpha q + 4\alpha^2 - 6\alpha + 3}{(q+1)^2}, \gamma_0 \right).$$

Consequently, the Gan-Loh-Sudakov conjecture holds for Cayley graphs with generating set $|S| \leq \frac{n}{7}$.

We give a fourier analytic proof of Theorems 1 and 2. Here is a quick high-level overview of the argument. We express the relevant "probability" (either $\frac{1}{|S|^2}\sum_{x,y\in S} 1_S(\frac{x+y}{2})$ or $\frac{1}{|S|^2}\sum_{x,y\in S} 1_S(x+y)$) in terms of the fourier coefficients of 1_S. If the probability is large, then some nonzero fourier coefficient must be large. We deduce that (a dilate of) the residues of S of a certain modulus concentrate near 0. Since there won't be "wraparound" near 0, this allows us to transfer the problem to $\mathbb{Z}$, which is a setting where it's easier to bound the relevant probabilities. We can show from the result in $\mathbb{Z}$ that we in fact must have many residues be 0. This allows us to conclude that S is very close to a subgroup. Induction and a purely combinatorial argument finish the job from there.

Here is an outline of the paper. We first set our notation for Fourier analysis on $\mathbb{Z}_n$. Then we give the proof of Theorems 1 and 2, modulo two Lemmas, which we prove afterwards. After, we show the calculations deducing the Gan-Loh-Sudakov conjecture from our main theorem. Finally, we prove Theorems 1 and 2 when $q = 1$.

2. Fourier Analysis on $\mathbb{Z}_n$

In this section, we briefly fix our notation for fourier analysis on $\mathbb{Z}_n$ and obtain the fourier representation of the relevant quantities in the proofs to be given below. For a function $f : \mathbb{Z}_n \to \mathbb{C}$, define its (finite) fourier transform $\widehat{f} : \mathbb{Z}_n \to \mathbb{C}$ by

$$\widehat{f}(m) := \frac{1}{n}\sum_{x\in\mathbb{Z}_n} f(x)e^{-2\pi i\frac{xm}{n}}.$$

The following well-known equalities are straightforward.

$$\sum_{m\in\mathbb{Z}_n} |\widehat{f}(m)|^2 = \frac{1}{n}\sum_{x\in\mathbb{Z}_n} |f(x)|^2$$

$$f(x) = \sum_{m\in\mathbb{Z}_n} \widehat{f}(m)e^{2\pi i\frac{xm}{n}}.$$

Let S be a symmetric subset of $\mathbb{Z}_n$. Then, $\frac{1}{|S|^2}\sum_{x,y\in S} 1_S(x+y) =$

$$\frac{1}{|S|^2}\sum_{x,y\in\mathbb{Z}_n}\left[\sum_{m_1\in\mathbb{Z}_n}\widehat{1_S}(m_1)e^{2\pi i\frac{xm_1}{n}}\right]\left[\sum_{m_2\in\mathbb{Z}_n}\widehat{1_S}(m_2)e^{2\pi i\frac{ym_2}{n}}\right]\left[\sum_{m_3\in\mathbb{Z}_n}\widehat{1_S}(m_3)e^{2\pi i\frac{(x+y)m_3}{n}}\right]$$

$$= \frac{1}{|S|^2}\sum_{m_1,m_2,m_3\in\mathbb{Z}_n}\widehat{1_S}(m_1)\widehat{1_S}(m_2)\widehat{1_S}(m_3)\left[\sum_{x\in\mathbb{Z}_n} e^{2\pi i\frac{x(m_1+m_3)}{n}}\right]\left[\sum_{y\in\mathbb{Z}_n} e^{2\pi i\frac{y(m_2+m_3)}{n}}\right],$$

and using

$$\sum_{x\in\mathbb{Z}_n} e^{2\pi i\frac{xk}{n}} = \begin{cases} n & k\equiv 0 \pmod n \\ 0 & k\not\equiv 0 \pmod n\end{cases},$$

we obtain

$$\frac{1}{|S|^2}\sum_{x,y\in S}1_S(x+y)=\frac{n^2}{|S|^2}\sum_{m\in\mathbb{Z}_n}\widehat{1_S}(-m)\widehat{1_S}(-m)\widehat{1_S}(m).$$

However, the symmetry of S implies that $\widehat{1_S}(m)=\widehat{1_S}(-m)$ for each $m\in\mathbb{Z}_n$. Therefore,

$$\text{Prob}[S]=\frac{1}{|S|^2}\sum_{x,y\in S}1_S(x+y)=\frac{n^2}{|S|^2}\sum_{m\in\mathbb{Z}_n}\widehat{1_S}(m)^3.$$

Similarly, for any subset $S\subseteq\mathbb{Z}_n$,

$$\frac{1}{|S|^2}\sum_{x,y\in S}1_S(\frac{x+y}{2})=\frac{n^2}{|S|^2}\sum_{m\in\mathbb{Z}_n}\widehat{1_S}(m)^2\widehat{1_S}(-2m).$$

3. Proof of Theorems 1 and 2

We induct on q. We discuss the base case $q=1$ in section 6. Take some $q\geq 2$ and $\alpha\in[0,1]$. Let $S\subseteq\mathbb{Z}_n$ be a symmetric[1] subset with $|S|=\frac{n}{q+\alpha}$.

Let $\gamma=\max(\frac{q^2-\alpha q+\alpha^2}{q^2},\frac{q^2+2\alpha q+4\alpha^2-6\alpha+3}{(q+1)^2},\gamma_0)$. Assume, for the sake of contradiction, that $\text{Prob}[S]\geq\gamma$. Then, as explained in section 2,

$$\sum_m\widehat{1_S}(m)^3\geq\frac{d^2}{n^2}\gamma.$$

Note $\widehat{1_S}(0)^3=\frac{d^3}{n^3}$, so, since $\widehat{1_S}(m)$ is real for each m[2],

$$\gamma\frac{d^2}{n^2}-\frac{d^3}{n^3}\leq\sum_{m\neq 0}\widehat{1_S}(m)^3\leq\left(\sup_{m\neq 0}\widehat{1_S}(m)\right)\cdot\sum_{m\neq 0}\widehat{1_S}(m)^2=\left(\sup_{m\neq 0}\widehat{1_S}(m)\right)\cdot[\frac{d}{n}-\frac{d^2}{n^2}],$$

where we used Plancherel in the last step. Take $m_0\neq 0$ with

$$\widehat{1_S}(m_0)\geq\frac{d}{n}\frac{\gamma-\frac{d}{n}}{1-\frac{d}{n}}=:\frac{d}{n}\mu.$$

Then,

$$\mu\leq\frac{1}{d}\sum_{x\in S}e^{2\pi i\frac{m_0}{n}x}=\frac{1}{d}\sum_{x\in S}e^{2\pi i\frac{m_0/g}{n/g}x},$$

where $g:=\gcd(m_0,n)$. Let

$$A=\{x\in\mathbb{Z}_n:2\pi\frac{m_0/g}{n/g}x\in[-2\pi/3,2\pi/3]\pmod{2\pi}\}$$

[1] In the 3AP setting, we do not assume S is symmetric.

[2] In the 3AP setting, we instead do $\gamma\frac{d^2}{n^2}-\frac{d^3}{n^3}\leq\sup_{m\neq 0}|\widehat{1_S}(-2m)|\cdot[\frac{d}{n}-\frac{d^2}{n^2}]$. Then we take m_0 with $|\widehat{1_S}(m_0)|\geq\frac{d}{n}\mu$. Finally, we can translate S so that $\widehat{1_S}(m_0)$ is real and positive.

$$B = \mathbb{Z}_{n/g} \setminus A.^3$$

Then, since $\widehat{1_S}(m_0)$ is real,

$$d\mu \leq \sum_{x\in S} \cos(2\pi \frac{m_0/g}{n_0/g} x) \leq |A| + (d - |A|)(-\frac{1}{2}),$$

which implies

$$\frac{|A|}{d} \geq \frac{2\mu + 1}{3}.^4$$

For $z \in B$,

$$\#\{(x, y) \in S^2 : x + y = z\} \leq d$$

and for $z \in A$,

$$\#\{(x, y) \in B \times A : x + y = z\} \leq |B|$$
$$\#\{(x, y) \in S \times B : x + y = z\} \leq |B|$$
$$\#\{(x, y) \in A \times A : x + y = z\} =: C_z.^5$$

Therefore,,

$$d^2 \operatorname{Prob}[S] \leq d|B| + 2|A|\,|B| + \sum_{z\in A} C_z$$
$$= d(d - |A|) + 2|A|(d - |A|) + |A|^2 \operatorname{Prob}[A].$$

So, we must have

$$\operatorname{Prob}[A] \geq \frac{\gamma + 2\frac{|A|^2}{d^2} - \frac{|A|}{d} - 1}{\frac{|A|^2}{d^2}}.$$

If we let $f(x) = \frac{\gamma+2x^2-x-1}{x^2}$, then $f'(x) = -2\gamma x^{-3} + x^{-2} + 2x^{-3}$ is positive for $x > 0$. We've shown $\frac{|A|}{d} \geq \frac{2\mu+1}{3} =: v^6$, so we get that

$$\operatorname{Prob}[A] \geq \frac{\gamma + 2v^2 - v - 1}{v^2} =: \beta.$$

We now argue that the weight at 0 must be large. For each $i \in [-\frac{1}{3}\frac{n}{g}, \frac{1}{3}\frac{n}{g}]$, let $S_i = \{x \in S : x \equiv i \pmod{n/g}\}$. Let $a_i = |S_i|$. Note that for each $i, j \in [-\frac{1}{3}\frac{n}{g}, \frac{1}{3}\frac{n}{g}]$ such that $i + j \in [-\frac{1}{3}\frac{n}{g}, \frac{1}{3}\frac{n}{g}]$,

$$\#\{(x_i, y_j, z_{i+j}) \in S_i \times S_j \times S_{i+j} : x_i + y_j = z_{i+j}\} \leq \min(|S_i|\,|S_j|, |S_i|\,|S_{i+j}|, |S_j|\,|S_{i+j}|).^7$$

The uniqueness of 0 is that $0 + 0 = 0$, so that $\#\{(x_0, y_0, z_0) \in S_0^3 : x_0 + y_0 = z_0\}$ cannot be upper bounded by potentially smaller terms $|S_i|, i \neq 0$. Note that the

[3] In the 3AP setting, we let $A = \{x \in \mathbb{Z}_n : 2\pi \frac{m_0/g}{n/g} x \in [-\frac{\pi}{2}, \frac{\pi}{2}]\}$ and $B = \mathbb{Z}_{n/g} \setminus A$.

[4] In the 3AP setting, we get $d\mu \leq |A| + (d - |A|)0$ and thus $\frac{|A|}{d} \geq \mu$.

[5] In the 3AP setting, the sets will merely have $2z$ instead of z - the same estimates thus hold.

[6] In the 3AP setting, we have $\nu := \mu$.

[7] In the 3AP setting, we'll be looking at $[-\frac{1}{4}\frac{n}{g}, \frac{1}{4}\frac{n}{g}]$ instead. Also, we'll have $2z_{\frac{i+j}{2}} \in S_{\frac{i+j}{2}}$ instead of $z_{i+j} \in S_{i+j}$, and $|S_{\frac{i+j}{2}}|$ instead of $|S_{i+j}|$. This alters Lemma 1 not too significantly.

sets whose size we just bounded account for all the terms in the computation of Prob$[A]$, since, by our choice of A, there is no "wraparound".[8]

Take γ_0 so that $\beta > \frac{9}{10}$ (for any q, α). $\gamma_0 = .949$ works[9]. Then Lemma 1 applies and we obtain,

$$\frac{|S_0|}{|A|} \geq \text{Prob}[A] \geq \beta.$$

It should be noted that we already get a contradiction if $g \leq \beta\nu d$ since we clearly must have $|S_0| \leq g$. In any event, we argue that this large a weight at 0 forces S to be close enough to the subgroup $\{0, \frac{n}{g}, \frac{2n}{g}, \dots, \frac{(g-1)n}{g}\}$ for us to get a direct upper bound on Prob$[S]$. For ease, let

$$D = \{x \in S : x \equiv 0 \pmod{n/g}\}$$

$$E = S \setminus D.$$

Then,

$$\text{Prob}[S] = \frac{1}{d^2} \sum_{x,y \in S} 1_S(x+y)$$

$$= \frac{|D|^2}{d^2} \frac{1}{|D|^2} \sum_{x,y \in D} 1_S(x+y) + \frac{2}{d^2} \sum_{x \in D, y \in E} 1_S(x+y) + \frac{1}{d^2} \sum_{x,y \in E} 1_S(x+y).$$

Using that D is contained in a subgroup disjoint from E, we have the following (in)equalities

$$\sum_{x,y \in D} 1_S(x+y) = \sum_{x,y \in D} 1_D(x+y)$$

$$\sum_{x \in D, y \in E} 1_S(x+y) = \sum_{x \in D, y \in E} 1_E(x+y) = \sum_{y \in E} \sum_{x \in D} 1_{-y+E}(x) \leq \sum_{y \in E} |E|$$

$$\sum_{x,y \in E} 1_S(x+y) \leq |E|^2.\text{[10]}$$

Hence,

$$\text{Prob}[S] \leq \frac{|D|^2}{d^2} \text{Prob}[D] + \frac{3}{d^2}|E|^2.$$

Using a cheaper "approximation" argument, similar to the one used previously, that doesn't capitalize on the fact that D is contained in a subgroup disjoint from E will yield an upper bound for Prob$[S]$ larger than 1.

---

[8]In the 3AP setting, the lack of wraparound for $x, y \in [-\frac{1}{4}\frac{n}{g}, \frac{1}{4}\frac{n}{g}] \pmod{n/g}$ follows from the fact that either $x+y$ is even and then of course $\frac{x+y}{2} \in [-\frac{1}{4}\frac{n}{g}, \frac{1}{4}\frac{n}{g}]$, or it's odd and then $\frac{x+y}{2} = (x+y)\frac{n+1}{2} = \frac{x+y-1}{2} + \frac{g-1}{2}\frac{n}{g} + \frac{\frac{n}{g}+1}{2} = \frac{x+y-1}{2} + \frac{\frac{n}{g}+1}{2} \pmod{n/g}$; since $\frac{x+y-1}{2} \in [-\frac{1}{4}\frac{n}{g}, \frac{1}{4}\frac{n}{g}]$ we therefore see that $\frac{x+y}{2} \notin [-\frac{1}{4}\frac{n}{g}, \frac{1}{4}\frac{n}{g}] \pmod{n/g}$.

[9]In the 3AP setting, we get a larger value for γ_1, but of course, a value less than 1.

[10]In the 3AP setting, we replace $x+y$ with $\frac{x+y}{2}$. If $x, y \in D$, then $\frac{x+y}{2} \in D$. And if $x \in D, y \in E$, then $x+y$ can't be in $2^{-1}D = D$. The three analogous (in)equalities thus hold.

Note $\frac{|D|}{d} = \frac{|D|}{|A|}\frac{|A|}{d} \geq \beta\nu$. Let $\eta = \frac{|D|}{d}$, $k = \frac{n}{g} \in \mathbb{N}$, $q' = \lfloor \frac{g}{|D|} \rfloor$, and $\alpha' = \frac{g}{|D|} - q'$. Then by induction and the obvious observation that $\text{Prob}[D]$ is independent of whether the ambient group is $\mathbb{Z}_n$ or $\{0, \frac{n}{g}, \dots, (g-1)\frac{n}{g}\}$,

$$\text{Prob}[D] \leq \max\left(\frac{(q')^2 - \alpha' q' + (\alpha')^2}{(q')^2}, \frac{(q')^2 + 2\alpha' q' + 4(\alpha')^2 - 6\alpha' + 3}{(q'+1)^2}, \gamma_0\right);$$

hence,

$$\text{Prob}[S] \leq \eta^2 \max\left(\frac{(q')^2 - \alpha' q' + (\alpha')^2}{(q')^2}, \frac{(q')^2 + 2\alpha' q' + 4(\alpha')^2 - 6\alpha' + 3}{(q'+1)^2}, \gamma_0\right) + 3(1-\eta)^2.$$

Note that the induction is justified, as $q' = \lfloor \frac{g}{|D|} \rfloor \leq \frac{g}{|D|} < q$, since $\frac{g}{|D|} \leq \frac{n/2}{\beta v d} \leq \frac{n/2}{\frac{3}{4}d} = \frac{2}{3}(q+\alpha)$, where we used that $\beta v \geq \frac{3}{4}$, which holds for $q \geq 2$. We finish by appealing to Lemma 2, which indeed applies when $\beta\nu \geq \frac{3}{4}$.

The above proof readily extends to an arbitrary finite Abelian group. Fix $r \geq 1$ and positive integers $n_1, \dots, n_r$. Let $n = n_1 \dots n_r$ and S be a subset of $\mathbb{Z}_{n_1} \times \dots \times \mathbb{Z}_{n_r}$ of size $|S| = \frac{n}{q+\alpha}$. Since $\widehat{1_S}(0, \dots, 0) = \frac{|S|}{n}$ and Plancherel holds, there is some $(m_1, \dots, m_r) \neq (0, \dots, 0)$ with

$$\frac{d}{n}\mu := \frac{d}{n}\frac{\gamma - \frac{d}{n}}{1 - \frac{d}{n}} \leq \widehat{1_S}(m_1, \dots, m_r) = \frac{1}{n}\sum_{(x_1, \dots, x_r) \in S} e^{2\pi i(\frac{m_1 x_1}{n_1} + \dots + \frac{m_r x_r}{n_r})}.$$

Analogous to before, letting $A = \{(x_1, \dots, x_r) \in S : 2\pi(\frac{m_1 x_1}{n_1} + \dots + \frac{m_r x_r}{n_r}) \in [\frac{-2\pi}{3}, \frac{2\pi}{3}] \pmod{2\pi}\}$, we must have $\frac{|A|}{d} \geq \frac{2\mu+1}{3}$. Let $S_j = \{(x_1, \dots, x_r) \in S : e^{2\pi i(\frac{m_1 x_1}{n_1} + \dots + \frac{m_r x_r}{n_r})} = e^{2\pi i \frac{j}{n}}\}$. Then, as before, we must have $\frac{|S_0|}{|A|} \geq \beta$. But S_0 is a subgroup of $\mathbb{Z}_{n_1} \times \dots \times \mathbb{Z}_{n_r}$, so the same inductive argument finishes the job. $\square$

4. Proof of Lemmas

Lemma 1. *Fix $d \geq 1$ and $\epsilon \in [0, \frac{1}{10})$. Let $\{a_j\}_{j \in \mathbb{Z}}$ be a collection of non-negative integers such that $\sum_{i \in \mathbb{Z}} a_i = d$ and $a_j = a_{-j}$ for each $j \in \mathbb{Z}$. Then if*

$$\sum_{i,j} \min(a_i a_j, a_i a_{i+j}, a_j a_{i+j}) \geq (1-\epsilon)d^2,$$

we must have that

$$a_0 \geq (1-\epsilon)d.$$

Proof. Define $\text{supp}(a_j) := \text{supp}((a_j)_{j \in \mathbb{Z}}) := \#\{n \geq 1 : a_n \neq 0\}$. We induct on $\text{supp}(a_j)$, with base case $\text{supp}(a_j) = 0$ obvious. Let $(a_j)_{j \in \mathbb{Z}}$ have $\text{supp}(a_j) =: N+1$. Let $n+1$ be the largest index j for which $a_j \neq 0$. First assume that $a_{n+1} \leq \frac{1}{10}d$.

Define $(b_j)_{j\in\mathbb{Z}}$ via $b_j = a_j$ if $|j| \le n$ and $b_j = 0$ if $|j| \ge n+1$. Then $b_j = b_{-j}$ for $j \in \mathbb{Z}$, $\operatorname{supp}(b_j) \le N$, and $\sum_{j\in\mathbb{Z}} b_j = d - 2a_{n+1}$. Note that

$$A_{n+1} := \sum_{i,j} \min(a_i a_j, a_i a_{i+j}, a_j a_{i+j})$$

$$\le \sum_{i,j} \min(b_i b_j, b_i b_{i+j}, b_j b_{i+j}) + 2\sum_{k=1}^{n} a_k a_{n+1} + 4\sum_{-n\le k\le -1} a_{n+1}a_k + 2a_{n+1}^2 + 4a_{n+1}^2$$

$$=: A_n + 6a_{n+1}(\frac{d - a_0 - 2a_{n+1}}{2}) + 6a_{n+1}^2.$$

Here we counted the number of ways $n+1$ or $-(n+1)$ can occur as $i+j$ for $i, j \ne 0$, then the number of ways $n+1$ or $-(n+1)$ can occur as i or j with no 0 as the other coordinate, and then accounted for the terms $(i,j) = (n+1, -(n+1)), (-(n+1), n+1)$, $(n+1, 0), (-(n+1), 0), (0, n+1)$, and $(0, -(n+1))$. If $A_{n+1} \ge (1-\epsilon)d^2$, then

$$(*) \qquad A_n \ge (1-\epsilon)d^2 - 3a_{n+1}(d - a_0).$$

We first show $3a_0 \ge (1+2\epsilon)d$. Bounding $a_0 \ge 0$ in (*) gives

$$A_n \ge \frac{(1-\epsilon)d^2 - 3a_{n+1}d}{(d-2a_{n+1})^2}(d - 2a_{n+1})^2.$$

To use the claim applied to $(b_j)_{j\in\mathbb{Z}}$ and total weight $d - 2a_{n+1}$, we must check that

$$1 - \frac{(1-\epsilon)d^2 - 3a_{n+1}d}{(d-2a_{n+1})^2} < \frac{1}{10}.$$

It suffices to show

$$1 - \frac{(1-\epsilon)d^2 - 3a_{n+1}d}{(d-2a_{n+1})^2} < \epsilon.$$

Rearranging gives

$$a_{n+1} < \frac{1-4\epsilon}{4(1-\epsilon)}d,$$

which is true for $\epsilon < 1/10$ and $a_{n+1} < \frac{d}{10}$. Hence, by induction,

$$3a_0 \ge 3\left[\frac{(1-\epsilon)d^2 - 3a_{n+1}d}{(d-2a_{n+1})^2}\right](d - 2a_{n+1}) = 3\frac{(1-\epsilon)d^2 - 3a_{n+1}d}{(d-2a_{n+1})}.$$

This is larger than $(1+2\epsilon)d$ iff

$$a_{n+1} < \frac{2-5\epsilon}{7-4\epsilon}d.$$

This is true for $\epsilon < 1/10$ and $a_{n+1} < d/10$.

Now, let α be such that

$$(1-\epsilon)d^2 - 3a_{n+1}(d - 2a_{n+1} - a_0) - 6a_{n+1}^2 = (1-\alpha)(d - 2a_{n+1})^2.$$

Then, assuming $\alpha < \frac{1}{10}$, we can use induction to get that

$$a_0 \ge (1-\alpha)(d - 2a_{n+1}).$$

So to finish the induction, it suffices to show that

$$(1-\alpha)(d-2a_{n+1}) \geq (1-\epsilon)d,$$

which is equivalent to

$$\frac{(1-\epsilon)d^2 - 3a_{n+1}(d-a_0)}{d-2a_{n+1}} \geq (1-\epsilon)d,$$

which, after simplifying, is equivalent to

$$3a_0 > (1+2\epsilon)d,$$

which we have proven. Therefore, all we need to do is prove $\alpha < \frac{1}{10}$. It suffices to show $\alpha < \epsilon$. But, as we've just noted, $(1-\alpha)(d-2a_{n+1}) \geq (1-\epsilon)d$, so $\alpha \leq 1 - \frac{(1-\epsilon)d}{d-2a_{n+1}} \leq 1 - \frac{(1-\epsilon)d}{d} = \epsilon$, as desired.

We finish by arguing that we in fact must have $a_{n+1} < \frac{d}{10}$ for $\epsilon < \frac{1}{10}$. First note

$$\sum_{i,j} a_i a_j - \sum_{i,j} \min(a_i a_j, a_i a_{i+j}, a_j a_{i+j}) \geq 4 \sum_{1 \leq k \leq n} a_k a_{n+1} + 2a_{n+1}^2.$$

Therefore, we have that

$$d^2 \geq (1-\epsilon)d^2 + 4a_{n+1}(\frac{d-a_0-2a_{n+1}}{2}) + 2a_{n+1}^2$$

and hence,

$$2a_{n+1}^2 - 2a_{n+1}(d-a_0) + \epsilon d^2 \geq 0.$$

As one can verify, the proof given above (for $a_{n+1} < \frac{d}{10}$) works regardless of what a_{n+1} is, if $a_0 > (\frac{1+2\epsilon}{3})d$. Therefore, we may assume $a_0 \leq (\frac{1+2\epsilon}{3})d$ and get that we must have

$$2a_{n+1}^2 - 2a_{n+1}(\frac{2-2\epsilon}{3})d + \epsilon d^2 \geq 0.$$

So, $\frac{a_{n+1}}{d} < \frac{\frac{2-2\epsilon}{3} - \sqrt{(\frac{2-2\epsilon}{3})^2 - 2\epsilon}}{2}$ or $\frac{a_{n+1}}{d} > \frac{\frac{2-2\epsilon}{3} + \sqrt{(\frac{2-2\epsilon}{3})^2 - 2\epsilon}}{2}$. However, the first expression in ϵ is less than $\frac{1}{10}$ for $\epsilon < \frac{1}{10}$, and the second expression is greater than $\frac{1}{2}$ for $\epsilon < \frac{1}{10}$. Since we clearly can't have $a_{n+1} > \frac{d}{2}$, we're done. $\square$

Remark. It should be noted that the largest we can possibly take ϵ in the statement of Lemma 1 is $\epsilon = \frac{2}{9}$. Consider, for example, $a_0, a_{-1}, a_1 = \frac{d}{3}$. Extending Lemma 1 from $\epsilon < \frac{1}{10}$ to $\epsilon < \frac{2}{9}$ will just slightly lower the value of γ_0, and will not allow one to get all the way down to $q \leq 3$.

Remark. In the 3AP setting we may not necessarily have that $a_j = a_{-j}$ for each $j \in \mathbb{Z}$. However, a suitable adjustment of the given proof shows that, for ϵ small enough, $\sum_{i,j} \min(a_i a_j, a_i a_{\frac{i+j}{2}}, a_j a_{\frac{i+j}{2}}) \geq (1-\epsilon)d^2$ implies $a_j \geq (1-\epsilon)d$ for some j. We can then just translate S to assume $j = 0$.

Lemma 2. *For $q \in \mathbb{N}, \alpha \in [0,1]$, define*

$$F(q,\alpha) = \max\left(\frac{q^2 - \alpha q + \alpha^2}{q^2}, \frac{q^2 + 2\alpha q + 4\alpha^2 - 6\alpha + 3}{(q+1)^2}, \gamma_0\right).$$

For any $q \geq 2, \alpha \in [0,1], 1 \leq k \leq q, \eta \in (\frac{3}{4}, 1]$, if we let $q' = \lfloor \frac{q+\alpha}{k\eta} \rfloor$ and $\alpha' = \frac{q+\alpha}{k\eta} - q'$, then

$$\eta^2 F(q', \alpha') + 3(1-\eta)^2 < F(q,\alpha).$$

Proof. Fix any $q, k, q' \geq 1$ and $\alpha \in [0,1]$. Substitute $\eta = \frac{q+\alpha}{(q'+\alpha')k}$ and let

$$f(\alpha') := \frac{(q+\alpha)^2}{k^2} \frac{1}{(q'+\alpha')^2} F(q',\alpha') + 3(1 - \frac{q+\alpha}{(q'+\alpha')k})^2.$$

We show that $f(\alpha')$ attains its maximum at (one of) the extreme values of α'. Define

$$f_1(\alpha') := \frac{(q+\alpha)^2}{k^2} \frac{1}{(q'+\alpha')^2} \frac{(q')^2 - \alpha' q' + (\alpha')^2}{(q')^2} + 3(1 - \frac{q+\alpha}{(q'+\alpha')k})^2$$

$$f_2(\alpha') := \frac{(q+\alpha)^2}{k^2} \frac{1}{(q'+\alpha')^2} \frac{(q')^2 + 2\alpha' q' + 4(\alpha')^2 - 6\alpha' + 3}{(q+1)^2} + 3(1 - \frac{q+\alpha}{(q'+\alpha')k})^2.$$

A straightforward computation shows

$$f_1'(\alpha') = \frac{q+\alpha}{k^2} \frac{1}{(q'+\alpha')^3} \cdot$$

$$\left[(2\alpha' - q')(\alpha' + q')(q+\alpha) - 2((\alpha')^2 - 2q'\alpha' + (q')^2)(q+\alpha) + 6(k(\alpha'+q') - (q+\alpha))\right]$$

$$f_2'(\alpha') = \frac{q+\alpha}{k^2} \frac{1}{(q'+\alpha')^3} \cdot$$

$$\left[(\alpha'+q')(8\alpha'+2(q'-3))(q+\alpha) - 2(4(\alpha')^2 + 2(q'-3)\alpha' + (q')^2 + 3)(q+\alpha) + 6(k(\alpha'+q') - (q+\alpha))\right]$$

In each $f_j'(\alpha')$, in the brackets, the quadratic term in α' vanishes. Therefore, in the brackers is a term linear in α'. In $f_1'(\alpha')$ the coefficient of α' is $q'(q+\alpha) + 4q'(q+\alpha) + 6k$, which is positive. Similarly, the coefficient of α' in $f_2'(\alpha')$ is $8q'(q+\alpha) + 2(q'-3)(q+\alpha) - 4(q'-3)(q+\alpha) + 6k = (6q'+6)(q+\alpha) + 6k$, which is positive. Hence, $f_1(\alpha'), f_2(\alpha')$ attain their maximum values only at the extreme values of α'. Since $f(\alpha') = \max(f_1'(\alpha'), f_2'(\alpha'))$[11], we see that $f(\alpha')$ attains its maximum at (one of) the extreme values of α'.

Suppose $\frac{q+\alpha}{(q'+\alpha')k} < 1$ for some $\alpha' \in (0,1)$. Then $\frac{q+\alpha}{(q'+1)k} < 1$. Note $\alpha' = 1 \implies F(q', \alpha') = 1$, and $\eta^2 + 3(1-\eta)^2$ is increasing for $\eta > \frac{3}{4}$. Since $\eta > \frac{3}{4}$ and since $\eta < 1$, we take $\eta = \frac{q+\alpha}{q+1}$ (since $q'k \in \mathbb{N}$). We obtain $\frac{q^2+2\alpha q+4\alpha^2-6\alpha+3}{(q+1)^2}$, which, of course, is at most $F(q,\alpha)$.

[11]Clearly $\eta^2\gamma_0 + 3(1-\eta)^2 \leq \gamma_0$ for $\eta \in (\frac{3}{4}, 1)$, since $\gamma_0 > \frac{3}{7}$. So, we assume $F(q', \alpha') \neq \gamma_0$.

If $\frac{q+\alpha}{q'k} < 1$, then we take $\alpha' = 0$ and argue as above. Otherwise, the extreme value of α' is the one making $\eta = 1$, namely $\alpha'_{crit} = \frac{q+\alpha}{k} - q'$. At $\eta = 1$, our desired inequality becomes $F(q', \alpha'_{crit}) \leq F(q, \alpha)$. Since $\alpha'_{crit} \in [0, 1]$ and $q' \in \mathbb{N}$, we have $q' = \lfloor \frac{q+\alpha}{k} \rfloor, \alpha'_{crit} = \{\frac{q+\alpha}{k}\}$, the fractional part. Therefore, it just suffices to show, generally, that

$$q, k \geq 1, \alpha \in [0, 1] \implies F(\lfloor \frac{q+\alpha}{k} \rfloor, \{\frac{q+\alpha}{k}\}) \leq F(q, \alpha).$$

Clearly, the inequality holds if $F(\lfloor \frac{q+\alpha}{k} \rfloor, \{\frac{q+\alpha}{k}\}) = \gamma_0$. If $q = 2$, then either $k = 1$ and the inequality is an equality, or $k = 2$ and $F(\lfloor \frac{q+\alpha}{k} \rfloor, \{\frac{q+\alpha}{k}\}) = F(1, \frac{\alpha}{2}) = 1 - \frac{\alpha}{2} + \frac{\alpha^2}{4}$, while $F(q, \alpha) \geq \frac{4-2\alpha+\alpha^2}{4} = 1 - \frac{\alpha}{2} + \frac{\alpha^2}{4}$. So, assume $q \geq 3$.

Note that $\frac{q^2-\alpha q+\alpha^2}{q^2} = 1 - \frac{\alpha}{q} + (\frac{\alpha}{q})^2$ is decreasing in $\frac{\alpha}{q}$ if $\frac{\alpha}{q} < \frac{1}{2}$. And for $q \geq 3$, $\frac{\alpha}{q}, \frac{\{\frac{q+\alpha}{k}\}}{\lfloor \frac{q+\alpha}{k} \rfloor} < \frac{1}{2}$. Therefore, to show that

$$\frac{\lfloor \frac{q+\alpha}{k} \rfloor^2 - \{\frac{q+\alpha}{k}\}\lfloor \frac{q+\alpha}{k} \rfloor + \{\frac{q+\alpha}{k}\}^2}{\lfloor \frac{q+\alpha}{k} \rfloor^2} \leq \frac{q^2 - \alpha q + \alpha^2}{q^2},$$

it suffices to show

$$\frac{\{\frac{q+\alpha}{k}\}}{\lfloor \frac{q+\alpha}{k} \rfloor} \geq \frac{\alpha}{q}.$$

But $q\{\frac{q+\alpha}{k}\} = q(\frac{q+\alpha}{k} - \lfloor \frac{q+\alpha}{k} \rfloor)$, so the inequality reduces to $\frac{q}{k} \geq \lfloor \frac{q+\alpha}{k} \rfloor$, which is true since $\lfloor \frac{q+\alpha}{k} \rfloor = \lfloor \frac{q}{k} \rfloor$, since if $\frac{q}{k} < m \in \mathbb{N}$, then $\frac{q}{k} \leq m - \frac{1}{k}$.

Next, observe that

$$\frac{q^2 + 2\alpha q + 4\alpha^2 - 6\alpha + 3}{(q+1)^2} = \frac{(q+1)^2 - (2-2\alpha)(q+1) + (2-2\alpha)^2}{(q+1)^2},$$

so since $\frac{2-2\alpha}{q+1} \leq \frac{1}{2}$ for $q \geq 3$, as before it suffices to show that

$$\frac{2 - 2\{\frac{q+\alpha}{k}\}}{\lfloor \frac{q+\alpha}{k} \rfloor + 1} \geq \frac{2-2\alpha}{q+1}.$$

However, substituting $\{\frac{q+\alpha}{k}\} = \frac{q+\alpha}{k} - \lfloor \frac{q+\alpha}{k} \rfloor$, collecting terms with $q + \alpha$, and simplifying yields the equivalent

$$\lfloor \frac{q+\alpha}{k} \rfloor + 1 \geq \frac{q+1}{k}.$$

And this is clearly true. $\square$

5. Verifying the Gan-Loh-Sudakov Conjecture for Cayley Graphs

We verify that our bound implies the bound in the Gan-Loh-Sudakov conjecture when $q \geq 7$. Take a finite Abelian group G and a symmetric subset $S \subseteq G$ not containing 0. Let $n = |G|$, $S_0 = S \cup \{0\}$, $d = |S|$, $q = \lfloor \frac{n}{|S_0|} \rfloor$, and $\alpha = \frac{n}{|S_0|} - q$. The benefit of working with S_0 is that the graph-theoretic bound takes the simpler form

$$|T_{conj}| \leq q\binom{d+1}{3} + \binom{r}{3} = q\binom{|S_0|}{3} + \binom{\alpha|S_0|}{3}.$$

Note

$$\operatorname{Prob}[S_0] = \frac{1}{|S_0|^2} \sum_{x,y \in S_0} 1_{S_0}(x+y) = \frac{1}{|S_0|^2}\left[\sum_{x,y \in S} 1_{S_0}(x+y) + 2\sum_{y \in S} 1_{S_0}(y) + 1_{S_0}(0+0)\right].$$

Taking into account that for each $x \in S$ there is exactly one $y \in S$ for which $x + y = 0$, we see

$$\operatorname{Prob}[S] = \frac{|S_0|^2}{|S|^2}\left[\operatorname{Prob}[S_0] - \frac{3|S|+1}{|S_0|^2}\right].$$

The number of triangles in our Cayley graph is thus

$$\frac{1}{6}n|S|^2 \operatorname{Prob}[S] = \frac{1}{6}(q+\alpha)|S_0|^3 \left[\operatorname{Prob}[S_0] - \frac{3|S|+1}{|S_0|^2}\right].$$

For ease, let $M = \max\left(\frac{q^2-\alpha q+\alpha^2}{q^2}, \frac{q^2+2\alpha q+4\alpha^2-6\alpha+3}{(q+1)^2}, \gamma_0\right)$ so that, by Theorem 2 applied to S_0 (which is symmetric), we may bound the number of triangles by

$$\frac{1}{6}(q+\alpha)M|S_0|^3 - \frac{1}{6}(q+\alpha)|S_0|(3|S_0|-2).$$

As one may check, this is less than $q\binom{|S_0|}{3} + \binom{\alpha|S_0|}{3}$ iff

$$[(q+\alpha^3) - (q+\alpha)M]|S_0|^3 + [3\alpha - 3\alpha^2]|S_0|^2 \geq 0.$$

Therefore, it suffices to have $M \leq \frac{q+\alpha^3}{q+\alpha}$. We have $\gamma_0 \leq \frac{q+\alpha^3}{q+\alpha}$ for all $q \geq 7$ and any $\alpha \in [0,1]$. And, for any $q \geq 1, \alpha \in [0,1]$,

$$\frac{q+\alpha^3}{q+\alpha} - \frac{q^2-\alpha q+\alpha^2}{q^2} = \frac{\alpha^3(q^2-1)}{q^2(q+\alpha)},$$

$$\frac{q+\alpha^3}{q+\alpha} - \frac{q^2+2\alpha q+4\alpha^2-6\alpha+3}{(q+1)^2} = \frac{(1-\alpha)^2(q-1)((2+\alpha)q+3\alpha)}{(q+1)^2(q+\alpha)}$$

are non-negative.

6. Base Case $q = 1$

We finish by proving Theorems 1 and 2 when $|S| = \frac{n}{1+\alpha}$ for some $\alpha \in [0, 1]$. Note

$$\sum_{y\in S}\sum_{x\in G} 1_S(x+y) = \sum_{y\in S} |S| = |S|^2.$$

So,

$$\sum_{x,y\in S} 1_S(x+y) = |S|^2 - \sum_{x\notin S}\sum_{y\in S} 1_S(x+y) = |S|^2 - \sum_{x\notin S} |(-x+S)\cap S|.$$

By pigeonhole, $|(-x+S)\cap S| \geq 2|S| - n$, and thus,

$$|S|^2 \operatorname{Prob}[S] \leq |S|^2 - \sum_{x\notin S}(2|S| - n) = |S|^2(1-\alpha+\alpha^2).$$

As $1-\alpha+\alpha^2 = \frac{q^2-\alpha q+\alpha^2}{q^2}$ for $q = 1$, Theorem 2 is established. Replacing S with $2S$ in the appropriate places establishes Theorem 1 as well.

7. Acknowledgments

I would like to thank Po-Shen Loh for telling me the graph theoretic conjecture. I would also like to thank Adam Sheffer and Cosmin Pohoata for helpful comments.

References

[1] E. Croot, The minimal number of three-term arithmetic progressions modulo a prime converges to a limit, Canad. Math. Bull., 51 (2008), 47–56.

[2] B. Green and O. Sisask. On the maximal number of 3-term arithmetic progressions in subsets of $\mathbb{Z}/p\mathbb{Z}$. Bull. Lond. Math. Soc., 40(6):945–955, 2008.

[3] W. Gan, P. Loh, and B. Sudakov, Maximizing the number of independent sets of a fixed size, Combinatorics, Probability and Computing, 24 (2015), 521-527

[4] P. Varnavides On Certain Sets of Positive Density Journal London Math. Soc. 34 (1959), 358-360.

[5] J. Cutler and A. J. Radcliffe, The Maximum Number of Complete Subgraphs of Fixed Size in a Graph with Given Maximum Degree. J. Graph Theory, 84 (2017), 134-145

Department of Mathematics, California Institute of Technology, Pasadena, CA, 91125

E-mail address: `zchase@caltech.edu`