NEW AND OLD RESULTS IN RAMSEY THEORY AND ADDITIVE COMBINATORICS

ANDREW LOTT

Contents

1. Introduction 1
2. Preliminary definitions 2
3. Ramsey's theorem 3
4. Schur's theorem and related results 4
5. SAT computations 7
6. Additive combinatorics 9
7. Roth's theorem for a general linear equation 10
8. Finitary density regularity vs. density regularity 21
9. Some cutting-edge results related to Roth's theorem 22
10. The transference principle 23
11. Acknowledgment 26
References 26

1. Introduction

The mathematical community has studied Ramsey theory, additive combinatorics, and closely related ancestors for over a century [15], [16]. These subjects are so rich that they have attracted the attention of notable mathematicians including Paul Erdős and Terence Tao.

Roughly speaking, Ramsey theory is the study of patterns in sets that are indestructible under finite partition. For instance, consider the set $\{1, 2, 3, 4, 5\}$. One may notice that it admits several solutions to the equation $a + b = c$, say $1 + 1 = 2$ or $1 + 2 = 3$. Assign each element to one of the colors red or blue (this is a "finite partition"). One can check that no matter how we choose the coloring, there will be either a red solution to $a + b = c$ or a blue solution to $a + b = c$. As such, the solutions to $a + b = c$ are "indestructible" under colorings of $\{1, 2, 3, 4, 5\}$ in red and blue. This is the simplest nontrivial case of Schur's theorem [2].

Questions like these gave birth to a branch of additive combinatorics concerning patterns which must be contained in sufficiently dense subsets of $\mathbb{N} = \{1, 2, 3, \ldots\}$. By a sufficiently dense subset of $\mathbb{N}$, we mean a set containing a positive proportion of $\mathbb{N}$, in the appropriate sense. For example, we will prove that a sufficiently dense subset of $\mathbb{N}$ must contain a 3-term arithmetic progression (3AP), which consists of 3 distinct numbers that are evenly spaced apart (e.g., $\{1,2,3\}$ or $\{4,9,14\}$ or $\{12,14,16\}$). This is an informal way of stating Roth's theorem [7].

We introduce some of the core results of Ramsey Theory: Schur's theorem [2], Rado's theorem [12], Ramsey's theorem [3], and Van Der Waerden's theorem [14]. We present an original extension of Schur's theorem in higher dimensions [13], and we provide a framework for finding exact answers to Ramsey type problems via SAT solver, a powerful computational tool [13],[1]. In addition, we include original SAT computations and examples related to Schur's theorem in higher dimensions.

We also explore some of the foundational ideas of additive combinatorics, starting with a formal statement of Szemerédi's theorem [8],[9]. We then provide a proof of Roth's theorem for a general linear equation [7] via the Hardy-Littlewood circle method, which includes explicit dependence on the coefficients of the linear equation which has not previously appeared in the literature. We also cite recent papers which give improved bounds and extensions of Roth's theorem.

Finally, we introduce the transference principle as it relates to density problems like the Green-Tao theorem [11],[10] and coloring problems like the Pythagorean triples conjecture [5]. We include proofs of several elementary propositions related to transference which, to our knowledge, have not previously appeared.

2. Preliminary definitions

In order to discuss Ramsey theory, we must first introduce several definitions.

Definition 1. For $r \in \mathbb{N}$ and a set A, an assignment of each element of A to one of r color classes is called an *r-coloring* of A. More formally, this is a map $\Delta : A \to \{1, 2, \ldots, r\}$.

Definition 2. Let A be a set. Given an r-coloring of A, a subset of A is called *monochromatic* if it is contained within a single color class.

Definition 3. If $\mathcal{F}$ is a set of subsets of A, we call $\mathcal{F}$ a *family.* The elements of $\mathcal{F}$ are called *members* of $\mathcal{F}$.

Definition 4. For $N \in \mathbb{N}$, let $[N] = \{1, 2, \ldots, N\}$.

Given a set A and a family $\mathcal{F}$ of subsets of A, we might ask the following question. Does every r-coloring of A contain a monochromatic member of $\mathcal{F}$? Many problems in Ramsey theory are of this form. Sometimes it is useful to state these theorems in terms of *partition regularity.*

Definition 5. Let $\mathcal{F}$ be a family of finite subsets of $\mathbb{N}$. We say that $\mathcal{F}$ *r-partition regular* if every r-coloring of $\mathbb{N}$ yields a monochromatic member of $\mathcal{F}$.

Definition 6. We call a family $\mathcal{F}$ *partition regular* if $\mathcal{F}$ is r-partition regular for every $r \in \mathbb{N}$.

For a family $\mathcal{F}$ of finite subsets of $\mathbb{N}$, One may check that the following two statements are equivalent

(1) $\mathcal{F}$ is partition regular.

(2) For every $r \in \mathbb{N}$, there exists an $N \in \mathbb{N}$ such that every r-coloring of $[N]$ yields a monochromatic member of $\mathcal{F}$.

We often state a theorem concisely in terms of (1) but prove it by showing (2).

3. Ramsey's theorem

Our goal in section 4 is to prove Schur's theorem, but our argument will rely on Ramsey's theorem, which is a generalization of what is often referred to as the party problem. How many people must attend a party in order to guarantee there are three party goers each of whom knows the other two or three party goers each of whom does not know the other two?

We may translate this scenario into a 2-coloring of the edges of a complete graph, which is a collection of vertices where each pair of vertices is connected by an edge. Each vertex represents a party goer and the edge between each pair of vertices is blue when the two know each other and red otherwise. The problem at this point is to find how many vertices the complete graph must have in order to guarantee that any 2-coloring yields a red triangle or a blue triangle. It turns out the answer is 6. We encourage the reader to prove this as an exercise (hint: Consider a 2-coloring of K_6 in red and blue. For a given vertex u, at least 3 of its incident edges must be the same color by the pigeonhole principle).

We prove Ramsey's theorem in its general form below. Recall that for each $n \in \mathbb{N}$, K_n denotes the complete graph on n vertices.

Theorem 1 (Ramsey's theorem)**.** *Suppose $r, w_1, ..., w_r \in \mathbb{N}$. Then, there exists $N \in \mathbb{N}$ such that every r-coloring of K_N yields a monochromatic K_{w_i} in color i for some $1 \leq i \leq r$. In particular, there is a smallest such N, denoted by $R_r(w_1, ..., w_r)$.*

Proof. Fix $r \in \mathbb{N}$. We proceed by induction on the sum $w_1 + \cdots + w_r$.

The base case is

$$w_1 + \cdots + w_r = r,$$

which is achieved only when $w_1 = \cdots = w_r = 1$. But $R_r(1, ..., 1) = 1$, so this case is handled.

For the inductive hypothesis, fix $t \in \mathbb{N}$, and assume $R_r(w_1, ..., w_r)$ exists for every $w_1, ..., w_r \in \mathbb{N}$ satisfying $w_1 + \cdots + w_r = t$.

Suppose $w_1, ..., w_r \in \mathbb{N}$ and $w_1 + \cdots + w_r = t + 1$. By the inductive hypothesis, we know $R_r(w_1, ..., w_i - 1, ..., w_r)$ exists for each $1 \leq i \leq r$. Thus, we may take

$$M = \max\{R_r(w_1 - 1, ..., w_r), R_r(w_1, w_2 - 1, ..., w_r), ..., R_r(w_1, ..., w_r - 1)\}.$$

Set $N = r(M - 1) + 2$. We claim any r-coloring of K_N yields a monochromatic K_{w_i} in color i for some $1 \leq i \leq r$.

Fix an r-coloring of K_N, and fix a vertex v. There are $N - 1$ edges connecting v to the other $N - 1$ vertices, and each of these edges falls under one of r color categories. By the pigeonhole principle, there exist

$$\begin{aligned}\left\lceil \frac{N-1}{r} \right\rceil &= \left\lceil \frac{r(M-1)+1}{r} \right\rceil \\ &= \left\lceil M-1+\frac{1}{r} \right\rceil \\ &= M.\end{aligned}$$

of these edges which are the same color, say color i. These M edges connect v to M vertices. Consider the complete subgraph induced by these M vertices. Call this subgraph G.

Observe that $M \geq R_r(w_1, ..., w_i - 1, ..., w_r)$. Thus, either G yields a monochromatic K_{w_j} in color j for some $j \neq i$ or G yields a monochromatic $K_{w_i - 1}$ in color i. In the former case, we are done, and in the latter case, the complete subgraph formed by v and the monochromatic K_{w_i-1} in color i is a monochromatic K_{w_i} in color i. Therefore, N satisfies the desired condition. Thus, $R_r(w_1, ..., w_r)$ exists and is at most N.

□

4. Schur's theorem and related results

Earlier we introduced the simplest nontrivial case of Schur's theorem. In particular, we claimed the family of solutions to $a + b = c$ is 2-partition regular. But what about solutions to $a + b + c = d$ or $a + b + c + d = e$? And what if we consider 3-colorings or 4-colorings of $\mathbb{N}$? It turns out Schur's theorem accounts for every family of equations of this form and every number of colors. We will prove this using Ramsey's theorem.

Definition 7. We call $\{x_1, \ldots, x_k\} \in \mathbb{N}$ a *Schur k-tuple* if $x_1 + \cdots + x_{k-1} = x_k$.

Definition 8. For $r, k \in \mathbb{N}$, let $R_r(k) = R_r(k, \ldots, k)$.

Theorem 2 (Schur's theorem)**.** *For every $k \in \mathbb{N}$, the family of Schur k-tuples is partition regular.*

Proof. Take $k, r \in \mathbb{N}$, and set $N = R_r(k) - 1$. Fix an r-coloring Δ of $[N]$. We claim Δ admits a monochromatic Schur k-tuple.

Construct an r-coloring of K_{N+1} as follows: denote the vertices of the graph by $x_1, ..., x_N$. For every $1 \leq i, j \leq N$, set $\Delta(|i-j|)$ as the color the edge between x_i and x_j. By Ramsey's Theorem, we know this r-coloring of K_{N+1} induces a monochromatic K_k. Say the vertices of this subgraph are $x_{y_1}, ..., x_{y_k}$, where $y_1 < y_2 < \cdots < y_k$. Transferring this configuration to Δ, we see $(y_2 - y_1), (y_3 - y_2), ..., (y_k - y_{k-1}), (y_k - y_1)$ is a monochromatic Schur k-tuple. Indeed, these all lie in the same color category by construction, and

$$(y_2 - y_1) + (y_3 - y_2) + ... + (y_k - y_{k-1}) = y_k - y_1,$$

where the sum on the left-hand side telescopes to the right-hand side.

□

Schur's theorem allows us to define Schur numbers:

Definition 9. For $r, k \in \mathbb{N}$, let $S(r, k)$ be the smallest natural number for which every r-coloring of $[S(r, k)]$ yields a monochromatic Schur k-tuple.

After seeing Schur's theorem, one might wonder whether it can be generalized to higher dimensional integer lattices. The natural approach is to define Schur k-tuples of points in $\mathbb{N}^d$ in the following way, via component by component vector addition.

Definition 10. Let $d \in \mathbb{N}$. We call $x_1, \dots, x_k \in \mathbb{N}^d$ a *Schur k-tuple* if $x_1 + \dots + x_{k-1} = x_k$.

However, with this definition, an extension of Schur's theorem is immediate and trivial.

Theorem 3. *Let $d, r, k \in \mathbb{N}$. Then there exists an $N \in \mathbb{N}$ such that every r-coloring of $[N]^d$ yields a monochromatic Schur k-tuple. In fact, $S(r, k)$ is the smallest such N.*

Proof. Let $N = S(r, k)$. Fix an r-coloring of $[N]^d$, and consider the main diagonal of $[N]^d$:

$$D = \{(x, \dots, x) \in [N]^d\}.$$

We have an induced r-coloring of D, which contains $N = S(r, k)$ elements. This r-coloring of D corresponds to an r-coloring of $[N]$. By Schur's theorem, this r-coloring of $[N]$ yields a monochromatic Schur k-tuple satisfying $x_1 + \dots + x_{k-1} = x_k$. Then the r-coloring of D yields a monochromatic schur k-tuple as well.

$$(x_1, \dots, x_1) + \dots + (x_{k-1}, \dots, x_{k-1}) = (x_k, \dots, x_k).$$

On the other hand, suppose $N < S(r, k)$. Then, there is an r-coloring Δ of $[N]$ with no monochromatic Schur k-tuple. Construct an r-coloring of $[N]^d$ in the following way. For every $l \in [N]$, let $\Delta(l)$ be the color every point in $[N]^d$ with first coordinate l. Then, this coloring of $[N]^d$ yields no monochromatic Schur k-tuple in the first coordinate, so it yields no monochromatic Schur k-tuple in $[N]^d$. $\square$

Thus, we have a theorem which is equivalent to Schur's theorem in $\mathbb{N}$. However, we may add a linear independence condition to make the problem more interesting.

Definition 11. Let $d \in \mathbb{N}$. We call $x_1, \dots, x_k \in \mathbb{N}^d$ a *nondegenerate Schur k-tuple* if $x_1 + \dots + x_{k-1} = x_k$ and $x_1, \dots, x_{k-1}$ are linearly independent.

Observe that the linear independence condition prevents the case where Schur tuple lies on the main diagonal. With this definition, the problem is tougher to tackle. This is one of the major original results of this thesis [13].

Theorem 4 (Schur's theorem in integer lattices)**.** *Let $r, d \in \mathbb{N}$. There exists $N \in \mathbb{N}$ such that every r-coloring of $[N]^d$ yields a monochromatic nondegenerate Schur $(d+1)$-tuple.*

Proof. Let $N = R_r(d+1)^d + 1$, and consider an r-coloring Δ of $[N]^d$. For each i in $[N]$, define $y_i = (i, i^2, ..., i^d)$. Construct a colored K_N with vertices $y_1, ..., y_N$ in the following way: for every $i > j$, set $\Delta(y_i - y_j)$ as the color of the edge connecting y_i and y_j. This is an r-coloring of a complete graph on $N = R_r(l)$ vertices, so it must yield a monochromatic K_{d+1}. Say the vertices of this monochromatic K_{d+1} are $y_{x_1}, ..., y_{x_{d+1}}$, where $x_1 < \dots < x_{d+1}$.

We claim

$$V = \left\{ y_{x_{i+1}} - y_{x_i} : i \in \mathbb{N}, 1 \leq i \leq d \right\} \cup \{ y_{x_{d+1}} - y_{x_1} \}.$$

is a monochromatic nondegenerate Schur k-tuple induced by Δ. Indeed, V is monochromatic by construction, and

$$(y_{x_2} - y_{x_1}) + (y_{x_3} - y_{x_2}) + \cdots + (y_{x_{d+1}} - y_{x_d}) = y_{x_{d+1}} - y_{x_1}.$$

In addition, V is linearly independent. Indeed, let $a_1, a_2, ..., a_d$ be scalars and suppose

$$\begin{bmatrix} x_2 - x_1 & x_2^2 - x_1^2 & \ldots & x_2^d - x_1^d \\ x_3 - x_2 & x_3^2 - x_2^2 & \ldots & x_3^d - x_2^d \\ \vdots & \vdots & \ddots & \vdots \\ x_{d+1} - x_d & x_{d+1}^2 - x_d^2 & \ldots & x_{d+1}^d - x_d^d \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

By taking the product above, we obtain

$$\begin{bmatrix} \sum_{j=1}^{d+1} \left(x_2^j - x_1^j \right) a_j \\ \sum_{j=1}^{d+1} \left(x_3^j - x_2^j \right) a_j \\ \vdots \\ \sum_{j=1}^{d+1} \left(x_{d+1}^j - x_d^j \right) a_j \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

By setting

$$p(x) = a_1 x + a_2 x^2 + \cdots + a_d x^d,$$

this becomes

$$\begin{bmatrix} p(x_2) - p(x_1) \\ p(x_3) - p(x_2) \\ \vdots \\ p(x_{d+1}) - p(x_d) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Hence,

$$\begin{gathered} p(x_2) = p(x_1) \\ p(x_3) = p(x_2) \\ \vdots \\ p(x_{d+1}) = p(x_d), \end{gathered}$$

so

$$r = p(x_1) = p(x_2) = p(x_3) = \cdots p(x_{d+1}),$$

for some constant r. Therefore, p achieves the same value at $d+1$ distinct points. However, the degree of p is at most d. Thus, by the fundamental theorem of algebra, $p(x) = 0$, so

$$a_1 = a_2 = \cdots = a_d = 0,$$

and the linear independence requirement is satisfied.

□

It is important to introduce Rado's theorem as it is a powerful generalization of Schur's theorem. Rewrite the Schur equation $a + b = c$ as $a + b - c = 0$. A subset $\{-1, 1\}$ of the coefficients of this linear equation sums to 0. It turns out that this condition is enough to guarantee partition regularity.

Theorem 5 (Rado). *Suppose $r \in \mathbb{N}$,$c_1, \ldots, c_k \in \mathbb{Z}$ and there is some subset of $\{c_1, \ldots, c_k\}$ that sums to 0. Then there is an $N \in \mathbb{N}$ such that every r-coloring $[N]$ yields a monochromatic solution to $c_1x_1 + \cdots + c_kx_k = 0$.*

We must also mention Van der Waerden's theorem here, as it is a direct predecessor of Szemerédi's theorem, which will be the focus of our discussion on additive combinatorics.

Theorem 6 (Van der Waerden's theorem). *For every $k \in \mathbb{N}$, the family of kAPs is partition regular.*

Notice that Rado's theorem implies Van der Waerden's theorm for $k = 3$. Indeed, a 3AP solution to the equation $x + z - 2y = 0$.

5. SAT computations

Given a family $\mathcal{F}$, we might have a clever way to prove that if a set A is large enough, then every r-coloring A yields a monochromatic member of $\mathcal{F}$, but determining exactly how large A must be usually requires immense comptutational power, and for this reason only a few of these values are known. The most powerful tool for finding these answers is a SAT solver.

In 2018, Marjin Heule conducted a multi-cpu year SAT computation [4] to determine $S(5, 3)$. In particular, he showed that for $n = 161$, every 5-coloring of $[n]$ yields a monochromatic solution to $a + b = c$, and he provided a counterexample to show this is not true for $n = 160$. To illustrate the magnitude of this computation, consider the fact that there are 5^{161} 5-colorings of [161]. This is far more than the number of atoms in the observable universe.

In a similar way, Heule [5] was able to show the family of Pythagorean triples (e.g., $\{3, 4, 5\}$ and $\{5, 12, 13\}$) is 2-partition regular by conducting a SAT computation to determine that every 2 coloring of [7825] yields a monochromatic Pythagorean triple. It is conjectured that the Pythagorean triples are partition regular.

Now we describe the general structure of the SAT expressions used to solve these problems. Consider a family $\mathcal{F}$ and an r-coloring of a finite set A given by

$$\Delta : A \to \{1, 2, ..., r\}.$$

Following the lead of Boza, Marín, Revuelta, and Sanz [1], write a logical expression in conjunctive normal form (cnf) which is true if and only if Δ yields no monochromatic member of $\mathcal{F}$. For each $p \in A$ and each $s \in \{1, 2, ..., r - 1\}$, define a boolean variable $\phi_s(p)$ by

$$\phi_s(p) = \begin{cases} \text{True} & \text{if } \Delta(p) = s \\ \text{False} & \text{otherwise} \end{cases}.$$

(Note that $\Delta(p) = r$ when $\phi_1(p), \phi_2(p), ..., \phi_{r-1}(p)$ are all false). In order to guarantee Δ assigns exactly one color to each point, our cnf expression must include

$$\mathcal{D} = \bigwedge_{p \in A} \bigwedge_{i < j \leq r-1} \big(\neg\phi_i(p) \vee \neg\phi_j(p)\big).$$

Observe that Δ yields no monochromatic member of $\mathcal{F}$ if and only if for each $m \in \mathcal{F}$ and for each color $i \in \{1, 2, ..., r\}$, Δ assigns at least one element of m to a color besides i. Thus, for each $m \in \mathcal{F}$ we include the expression

$$\mathcal{C}_m = \left(\bigwedge_{i \in [r-1]} \left(\bigvee_{p \in m} \neg \phi_i(p_j) \right) \right) \wedge \left(\bigvee_{i \in [r-1]} \bigvee_{p \in m} \phi_i(p_j) \right).$$

Set

$$\mathcal{C} = \bigwedge_{m \in \mathcal{F}} \mathcal{C}_m.$$

Then, Δ induces a coloring with no monochromatic member of $\mathcal{F}$ if and only if $\mathcal{D} \wedge \mathcal{C}$. When given this expression, a SAT solver will return "unsatisfiable", or it will print a satisfying assignment of the variables corresponding to an r-coloring with no monochromatic member of $\mathcal{F}$. Using a SAT solver, we computed several higher dimensional Schur numbers.

Definition 12. Let $r, d, k \in \mathbb{N}$. Define $S_r^d(k)$ to be the smallest natural number such that every r-coloring of $[S_r^d(k)]^d$ yields a monochromatic nondegenerate Schur k-tuple.

We determined $S_2^2(3) = 6, S_3^2(3) = 18$, and $S_4^2(3) \geq 49$. The counterexamples are displayed in the following figures.

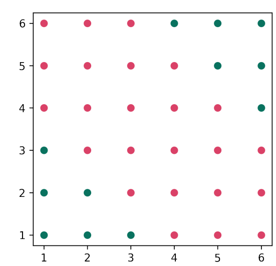

FIGURE 1. A 2-coloring of $[6]^2$ with no monochromatic nondegenerate Schur triples.

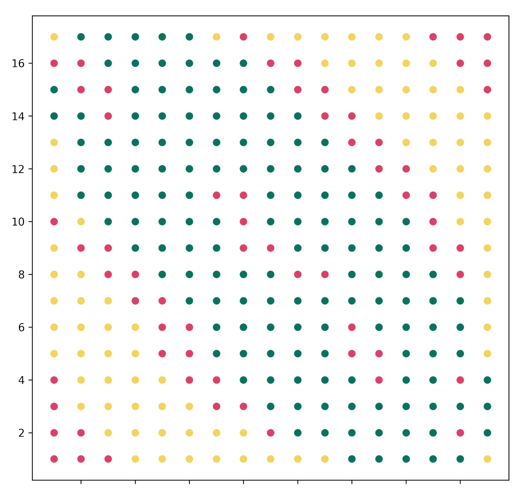

FIGURE 2. A 3-coloring of $[17]^2$ with no monochromatic nondegenerate Schur triples.

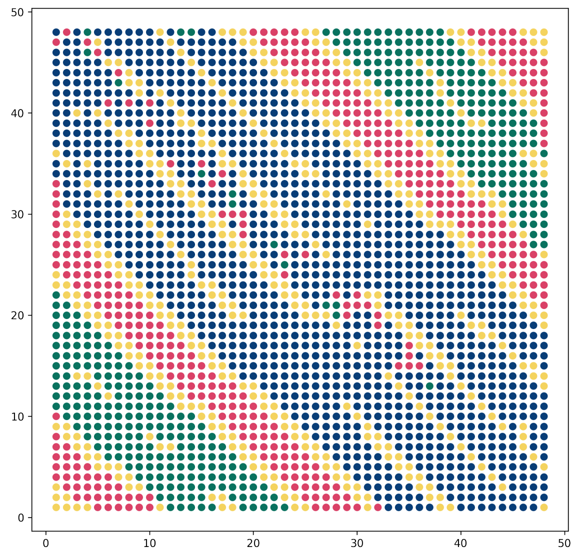

FIGURE 3. A 4-coloring of $[48]^2$ with no monochromatic nondegenerate Schur triples.

6. Additive combinatorics

In additive combinatorics, we might ask the following question. Given a family $\mathcal{F}$ in $\mathbb{N}$, does a sufficiently dense subset of $\mathbb{N}$ necessarily contain a member of $\mathcal{F}$? Of course, it is important to clarify what we mean by "sufficiently dense".

Definition 1. The *density* of $A \subset \mathbb{N}$ is given by

$$\delta(A) = \lim_{N\to\infty} \frac{|A \cap [N]|}{N},$$

if this limit exists.

Informally, we may think of density as the probability a given natural number lies in A. For instance, the even natural numbers have density $1/2$ while the multiples of three have density $1/3$.

We defined density as the limit of a sequence, but sequences do not always converge. For this reason, we define the upper density of a set A:

Definition 2. The *upper density* of $A \subseteq \mathbb{N}$ is given by

$$\overline{\delta}(A) = \limsup_{N\to\infty} \frac{|A \cap [N]|}{N}.$$

Note that $\overline{\delta}$ always exists since the terms of the sequence are bounded between 0 and 1. Then, instead of saying a A is "sufficiently dense", we say A has positive upper density.

In order to more clearly state the theorems in additive combinatorics, we introduce another definition.

Definition 3. Let $\mathcal{F}$ be a family in $\mathbb{N}$. We call $\mathcal{F}$ *density regular* if for every $A \subset \mathbb{N}$

$$\overline{\delta}(A) > 0 \implies A \text{ contains a member of } \mathcal{F}.$$

The terms in the sequence of definition 1 are nonnegative, so $\overline{\delta(A)} \geq 0$. Thus, if $\overline{\delta}(A) \not> 0$, then $\delta(A) = 0$. With this in mind, we write the contrapositve of definition 3.

Definition 3. (contrapositive) Let $\mathcal{F}$ be a family in $\mathbb{N}$. We call $\mathcal{F}$ density regular if for every $A \subset N$,

$$\text{A is } \mathcal{F}\text{-free} \implies \delta(A) = 0.$$

We call a set $\mathcal{F}$-free if it contains no member of $\mathcal{F}$. It is sometimes easier to prove theorems using this version of the definition.

7. Roth's theorem for a general linear equation

Szemerédi's theorem is a strengthening of van der Waerden's theorem and one of the foundational result of additive combinatorics.

Theorem 1 (Szemerédi's theorem)**.** *For every $k \in \mathbb{N}$, the family of kAP's is density regular.*

Proving Szemerédi's theorem [8], [9] in its entirety is enormously difficult and far beyond the scope of this paper, but we can handle the case $k = 3$, known as Roth's theorem [7]. Notice that if $x < y < z$ form a 3AP, then $x + y - 2z = 0$. This is a linear equation whose coefficients sum to 0. It turns if $c_1 + \cdots + c_k = 0$, then the family of solutions to the linear equation $c_1x_1 + \cdots + c_kx_k = 0$ is density regular. Note that this is a stronger condition than the one in Rado's theorem, which only requires that a subset of the coefficients sums to zero.

In order to prove Roth's theorem for a general linear equation, we adapt the traditional density increment argument. The fact that this adaptation is possible has been known in the mathematical community, but we write it out in careful detail, and we come to a quantitive bound with a multiplicative constant that is explicitly determined by the coefficients of the linear equation. We will rely on several Fourier analytic tools. The first of these is the orthogonality relation.

Proposition 1 (Orthogonality relation)**.**

$$\int_0^1 e^{2\pi i n\alpha} d\alpha = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{if } n \in \mathbb{Z} \setminus \{0\}. \end{cases}$$

Proof. Let $n \in \mathbb{Z}$. If $n = 0$, then

$$\int_0^1 e^{2\pi i n\alpha} d\alpha = \int_0^1 e^0 d\alpha = 1.$$

Otherwise, $n \neq 0$, and

$$\int_0^1 e^{2\pi i n\alpha} d\alpha = \frac{1}{2\pi i n}\left(e^{2\pi i n \cdot 1} - e^{2\pi i n \cdot 0}\right) = \frac{1}{2\pi i n}(1 - 1) = 0,$$

where we also used the fact that $n \in \mathbb{Z}$ and $e^{2\pi i n} = 1$. $\square$

We will use this to detect and count solutions to a given linear equation, but first we need to define the Fourier transform.

Definition 4. Let $\mathbb{T}$ be $\mathbb{R}/\mathbb{Z}$, the circle parametrized by $[0, 1]$ with 0 and 1 identified.

Definition 5. Suppose $F : \mathbb{Z} \to \mathbb{C}$ satisfies $F(n) = 0$ for all but finitely many n. Then, the fourier transform of F is $\widehat{F} : \mathbb{T} \to \mathbb{C}$, defined by

$$\widehat{F}(\alpha) = \sum_{n \in \mathbb{Z}} F(n) e^{-2\pi i n \alpha}.$$

Definition 6. For $A \subset \mathbb{Z}$, define $1_A : \mathbb{Z} \to \{0, 1\}$ by

$$1_A(n) = \begin{cases} 1 & \text{if } n \in A \\ 0 & \text{if } n \notin A \end{cases}.$$

We are ready to state the proposition which allows us to count solutions to

$$c_1 x_1 + \cdots c_k x_k = 0$$

via the orthogonality relation. Note that for a finite set S, we take $|S|$ to mean the number of elements in S.

Proposition 2. *Let $A_1, \ldots A_k$ be finite subsets of $\mathbb{N}$, and let $c_1, \ldots, c_k \in \mathbb{Z}$. Then,*

$$|\{(x_1, \ldots, x_k) \in A_1 \times \cdots \times A_k : c_1 x_1 + \cdots c_k x_k = 0\}| = \int_0^1 \widehat{1_A}(c_1 \alpha) \cdots \widehat{1_A}(c_k \alpha) d\alpha.$$

Proof. By the definition of the Fourier transform,

$$\int_0^1 \widehat{1_A}(c_1 \alpha) \cdots \widehat{1_A}(c_k \alpha) d\alpha = \int_0^1 \sum_{x_1 \in A_1} e^{-2\pi i c_1 x_1 \alpha} \cdots \sum_{x_k \in A_k} e^{-2\pi i c_k x_k \alpha} d\alpha.$$

We may rewrite the product of sums in the integrand as follows.

$$\int_0^1 \sum_{x_1 \in A_1} e^{-2\pi i c_1 x_1 \alpha} \cdots \sum_{x_k \in A_k} e^{-2\pi i c_k x_k \alpha} d\alpha = \int_0^1 \sum_{(x_1, \ldots, x_k) \in A_1 \times \cdots \times A_k} e^{-2\pi i (c_1 x_1 + \cdots + c_k x_k) \alpha} d\alpha.$$

Next, we apply the integral sum property to obtain

$$\int_0^1 \sum_{(x_1, \ldots, x_k) \in A_1 \times \cdots \times A_k} e^{-2\pi i (c_1 x_1 + \cdots + c_k x_k) \alpha} d\alpha = \sum_{(x_1, \ldots, x_k) \in A_1 \times \cdots \times A_k} \int_0^1 e^{-2\pi i (c_1 x_1 + \cdots + c_k x_k) \alpha} d\alpha,$$

and we use the orthogonality relation to conclude

$$\begin{aligned} &\sum_{(x_1, \ldots, x_k) \in A_1 \times \cdots \times A_k} \int_0^1 e^{-2\pi i (c_1 x_1 + \cdots + c_k x_k) \alpha} d\alpha \\ =& \sum_{(x_1, \ldots, x_k) \in A_1 \times \cdots \times A_k} \begin{cases} 1 & \text{if } c_1 x_1 + \cdots + c_k x_k = 0 \\ 0 & \text{else} \end{cases} \\ =& |\{(x_1, \ldots, x_k) \in A_1 \times \cdots \times A_k : c_1 x_1 + \cdots c_k x_k = 0\}|. \end{aligned}$$

$\square$

Now we have the Fourier analytic tools necessary to approach Roth's theorem. Suppose the coefficients of a linear equation are $c_1, ..., c_k \in \mathbb{Z}$, where $c_1 + \cdots + c_k = 0$. Let

$$\mathcal{F} = \{\{x_1, \cdots, x_k\} \subset \mathbb{Z} : c_1x_1 + \cdots + c_kx_k = 0 \text{ and } x_i \neq x_1 \text{ for some } i\}.$$

We will come to an upper bound on the density of $\mathcal{F}$-free subsets of $\mathbb{N}$ which depends explicitly on the coefficients $c_1, \ldots, c_k$. In order to clearly state this dependence, we define the following parameters.

Without loss of generality, order $c_1, \ldots, c_k$ so $c_1, ..., c_j < 0$ and $c_{j+1}, ..., c_k \geq 0$. Set

$$\begin{aligned} J &= |c_1 + \cdots + c_j| \\ K &= c_{j+1} + \cdots + c_{k-1} \\ \beta &= \frac{K}{J}\left(1 - \frac{K}{J}\right), \text{ and} \\ \gamma &= 1 - \frac{K}{J}, \end{aligned} \tag{1}$$

Finally, define

$$r(c_1, \ldots, c_k) = \frac{(\gamma - \beta)^{k-1}}{c_k^{k-2} 2^k}. \tag{2}$$

Theorem 2 (Roth's Theorem for a general linear equation)**.** *Define $\mathcal{F}$ as above. If $N \in \mathbb{N}$ and $A \subset [N]$ is $\mathcal{F}$-free, then*

$$|A| \leq \frac{16N}{r \log\log N},$$

where $r = r(c_1, \ldots, c_k)$, as in equation (2)*.*

Our density increment argument can be summarized in the following 4 steps.

Step 1: (Fourier peak). Define the balanced function of $A \subset [N]$ by

$$f_A = 1_A - \delta 1_{[N]}.$$

We show that if the density of A satisfies several mild conditions and A contains no member of $\mathcal{F}$, then $\widehat{f_A}(\alpha)$, the Fourier transform of the balanced function of A, must be large for some $\alpha \in \mathbb{T}$.

Step 2: (Density increment). Suppose f_A achieves a Fourier peak as in step 1. In a certain sense this means A has a nonuniform distribution in $[N]$. Consequently, we may show that if A has density δ in $[N]$, then A has density $\delta + \epsilon/8$ in an arithmetic progression $P \subset [N]$, where ϵ is a positive constant determined by the Fourier peak.

Step 3: (Translate and scale). Suppose $A \subset [N]$ is $\mathcal{F}$-free and has density $\delta + \epsilon/8$ in an arithmetic progression $P \subset [N]$. Since $\mathcal{F}$ is invariant under scaling and translation, we may translate and scale P to $[N']$ and $A \cap P$ to $A' \subset [N']$ in such a way that A' is $\mathcal{F}$-free and has density $\delta + \epsilon/8$ in $[N']$.

Step 4: (Repeat). Given an $A \subset [N]$ which is $\mathcal{F}$-free, we apply steps 1 through 3 to obtain an incrementally denser $A_1 \subset [N_1]$ where A_1 is $\mathcal{F}$-free and $N_1 < N$. Repeat steps 1 through 3 again to obtain an even denser $\mathcal{F}$-free $A_2 \subset [N_2]$ where A_2 is $\mathcal{F}$-free and $N_2 < N_1$. We repeat this process for as long as possible. However, we cannot continue forever, or else we would come to some $A_k \subset [N_k]$ with density greater than 1, which is impossible. Based on the fact that this process must end, we extract an upper bound on the density of A.

Our summary hides many details and exceptions. This will become clear as we work through the proof.

Lemma 1 (Fourier peak)**.** *Let $N \in \mathbb{N}$, and let $A \subset [N]$ be $\mathcal{F}$-free with $|A| = \delta N$. If*

$$\delta \geq \left(rN^{k-2}\right)^{-1/(k-1)} \text{ and } |A \cap (\beta N, \gamma N]| \geq \frac{(\gamma - \beta)|A|}{2},$$

then

$$|\widehat{f_A}(\alpha)| \geq r\delta^2 N,$$

for some $\alpha \in [0, 1)$.

Proof. Let $N \in \mathbb{N}$ and suppose $A \subset [N]$ is $\mathcal{F}$-free. Suppose further

$$|A \cap (\beta N, \gamma N])| \geq \frac{(\gamma - \beta)|A|}{2}.$$

Let $x \in \mathbb{R}$. Then

$$c_1 x + c_2 x + \cdots + c_k x = x(c_1 + \cdots + c_k) = 0.$$

It follows that $\{(x, x, ..., x) \in \mathbb{N}^k\}$ is a set of trivial solutions to $c_1x_1 + \cdots + c_kx_k = 0$. Combining this with the fact that A is $\mathcal{F}$-free, we see the only solutions to $c_1x_1+\cdots+c_kx_k = 0$ in A are the trivial solutions, of which there are $|A| = \delta N$. Hence,

$$\delta N = |\{(x_1, \cdots, x_k) \in A^k : c_1x_1 + \cdots + c_kx_k = 0\}|.$$

Thus,

$$\delta N = \int_0^1 \widehat{1_A}(c_1\alpha) \cdots \widehat{1_A}(c_k\alpha) d\alpha,$$

by proposition 2.

Let f_A be the balanced function of A, defined by

$$f_A = 1_A - \delta 1_{[N]}.$$

Therefore,

$$\begin{aligned}
\delta N &= \int_0^1 \widehat{1_A}(c_1\alpha) \cdots \widehat{1_A}(c_{k-1}\alpha) \left(\widehat{f_A}(c_k\alpha) + \delta\widehat{1_{[N]}}(c_k\alpha\right)) d\alpha \\
&= \int_0^1 \widehat{1_A}(c_1\alpha) \cdots \widehat{1_A}(c_{k-1}\alpha)\widehat{f_A}(c_k\alpha) d\alpha \\
&+ \delta \int_0^1 \widehat{1_A}(c_1\alpha) \cdots \widehat{1_A}(c_{k-1}\alpha)\widehat{1_{[N]}}(c_k\alpha) d\alpha.
\end{aligned}$$

Apply proposition 2 once again to obtain

$$\begin{aligned}
&\int_0^1 \widehat{1_A}(c_1\alpha)\widehat{1_A}(c_2\alpha) \cdots \widehat{1_{[N]}}(c_k\alpha) d\alpha \\
=&|\{(x_1, ..., x_k) : \{x_1, ..., x_{k-1}\} \subset A, x_k \in [N], c_1x_1 + \cdots + c_kx_k = 0\}|.
\end{aligned}$$

Let

$$B = \{(x_1, ..., x_k) : \{x_1, ..., x_{k-1}\} \subset A, x_k \in [N], c_1x_1 + \cdots + c_kx_k = 0\}.$$

We have shown

$$\delta N = \int_0^1 \widehat{1_A}(c_1\alpha)\cdots\widehat{1_A}(c_{k-1}\alpha)\widehat{f_A}(c_k\alpha)d\alpha + \delta|B|. \tag{3}$$

Let $\{x_1, \ldots, x_{k-1}\} \subset A \cap (\beta N, \gamma N)$, $x_1 \equiv \cdots \equiv x_{k-1} \pmod{c_k}$, and

$$x_k = -\frac{c_1x_1 + \cdots + c_{k-1}x_{k-1}}{c_k}.$$

We claim $x_k \in [N]$. By (1),

$$KN\beta - JN\gamma < c_1x_1 + \cdots + c_{k-1}x_{k-1} < KN\gamma - JN\beta.$$

Since $\gamma = 1 - K/J$ and $K\beta \geq 0$,

$$K\beta - J\gamma = K\beta + K - J \geq K - J.$$

and it is not hard to check that

$$K\gamma - J\beta < 0.$$

Combining these facts gives

$$(K - J)N \leq c_1x_1 + \cdots + c_{k-1}x_{k-1} < 0. \tag{4}$$

We know

$$x_k = -\frac{c_1x_1 + \cdots + c_{k-1}x_{k-1}}{c_k}$$

and $K - J = -c_k$. Thus, (4) becomes $0 < x_k \leq N$. In addition

$$c_1x_1 + \cdots + c_{k-1}x_{k-1} \equiv x_1(c_1 + \cdots c_{k-1}) \equiv -x_1c_k \equiv 0 \pmod{c_k},$$

so x_k is an integer. In particular, $x_k \in [N]$.

Set

$$C = \{(x_1, ..., x_{k-1}) : \{x_1, ..., x_{k-1}\} \subset A \cap (\beta N, \gamma N), x_1 \equiv \cdots \equiv x_{k-1} \pmod{c_k}\}.$$

By the work done above,

$$|C| \leq |B|. \tag{5}$$

For each $i \in [c_k]$, set

$$p_i = \frac{|\{x \in A \cap (\beta N, \gamma N) : x \equiv i \pmod{c_k}\}|}{|A \cap (\beta N, \gamma N)|}$$

and

$$p = |C|/|A \cap (\beta N, \gamma N)|^{k-1}.$$

Then

$$p = p_0^{k-1} + \cdots + p_{c_k-1}^{k-1}.$$

Observe $p_0 + \cdots + p_{c_k-1} = 1$, so p is minimized when

$$p_0 = \cdots = p_{c_k-1} = 1/c_k.$$

Hence,

$$p \geq c_k(1/c_k^{k-1}) = 1/c_k^{k-2}.$$

Recall

$$|A \cap (\beta N, \gamma N)| \geq \frac{(\gamma - \beta)|A|}{2}.$$

Combining this with (5) gives

$$\begin{aligned}|B| &\geq |C| \\ &= p|A \cap (\beta N, \gamma N)|^{k-1} \\ &\geq \frac{|A \cap (\beta N, \gamma N)|^{k-1}}{c_k^{k-2}}. \\ &= \frac{(\gamma - \beta)|A|}{2^{k-1} c_k^{k-2}} \\ &= 2r|A|^{k-1}.\end{aligned}$$

In particular,

$$|B| \geq 2r|A|^{k-1} = 2r\delta^{k-1}N^{k-1}.$$

By applying (3) and triangle inequality, we observe

$$\begin{aligned}\int_0^1 |\widehat{1_A}(c_1\alpha)| \cdots |\widehat{1_A}(c_{k-1}\alpha)||\widehat{f_A}(c_k\alpha)| d\alpha &\geq \left| \int_0^1 \widehat{1_A}(c_1\alpha) \cdots \widehat{1_A}(c_{k-1}\alpha)\widehat{f_A}(c_k\alpha) d\alpha \right| \\ &\geq 2r\delta^k N^{k-1} - \delta N \\ &\geq \delta^k \left(2rN^{k-1} - \frac{N}{\delta^{k-1}} \right) \\ &\geq r\delta^k N^{k-1},\end{aligned}$$

where we accounted for the fact that $\delta \geq \left(rN^{k-2}\right)^{-1/(k-1)}$ and

$$|\widehat{1}_A(\alpha)| \leq \delta N \quad \forall \alpha \in \mathbb{T}.$$

Thus,

$$\begin{aligned}r\delta^k N^{k-1} &\leq \int_0^1 |\widehat{1_A}(c_1\alpha)| \cdots |\widehat{1_A}(c_{k-1}\alpha)||\widehat{f_A}(c_k\alpha)| d\alpha \\ &\leq \int_0^1 (\delta N)^{k-3} |\widehat{1_A}(c_1\alpha)||\widehat{1_A}(c_2\alpha)| \max_{\alpha \in \mathbb{T}} |\widehat{f_A}(\alpha)| d\alpha \\ &= \max_{\alpha \in \mathbb{T}} |\widehat{f_A}(\alpha)| \cdot \delta^{k-3} N^{k-3} \int_0^1 |\widehat{1_A}(c_1\alpha)||\widehat{1_A}(c_2\alpha)| d\alpha.\end{aligned}$$

By the Cauchy-Schwartz inequality,

$$\begin{aligned}&\max_{\alpha \in \mathbb{T}} |\widehat{f_A}(\alpha)| \cdot \delta^{k-3} N^{k-3} \int_0^1 |\widehat{1_A}(c_1\alpha)||\widehat{1_A}(c_2\alpha)| d\alpha \\ \leq &\max_{\alpha \in \mathbb{T}} |\widehat{f_A}(\alpha)| \cdot \delta^{k-3} N^{k-3} \left(\int_0^1 |\widehat{1_A}(c_1\alpha)|^2 d\alpha \right)^{1/2} \left(\int_0^1 |\widehat{1_A}(c_2\alpha)|^2 d\alpha \right)^{1/2}.\end{aligned}$$

By Plancherel's identity,

$$\begin{aligned}&\max_{\alpha \in \mathbb{T}} |\widehat{f_A}(\alpha)| \cdot \delta^{k-3} N^{k-3} \left(\int_0^1 |\widehat{1_A}(c_1\alpha)|^2 d\alpha \right)^{1/2} \left(\int_0^1 |\widehat{1_A}(c_2\alpha)|^2 d\alpha \right)^{1/2} \\ = &\max_{\alpha \in \mathbb{T}} |\widehat{f_A}(\alpha)| \cdot \delta^{k-3} N^{k-3} (\delta N)^{1/2} (\delta N)^{1/2}\end{aligned}$$

$$= \max_{\alpha \in \mathbb{T}} |\widehat{f_A}(\alpha)| \cdot \delta^{k-2} N^{k-2}.$$

In particular,

$$r\delta^k N^{k-1} \leq \max_{\alpha \in \mathbb{T}} |\widehat{f_A}(\alpha)| \cdot \delta^{k-2} N^{k-2}.$$

Dividing through by $\delta^{k-2}N^{k-2}$ gives

$$\max_{\alpha \in \mathbb{T}} |\widehat{f_A}(\alpha)| \geq r\delta^2 N.$$

$\square$

Lemma 2 (Density increment)**.** *Let $N \in \mathbb{N}$, and let $A \subset [N]$ be $\mathcal{F}$-free with $|A| = \delta N$. If $|\widehat{f_A}(\alpha)| \geq \epsilon N$ for some $\alpha \in \mathbb{T}$, then there exists an arithmetic progression $P \subset [N]$ of length $L = \lfloor (\epsilon N)^{1/2}/16 \rfloor$ satisfying $|A \cap P| \geq L(\delta + \epsilon/8)$.*

Proof. Let $N \in \mathbb{N}$ and suppose $A \subset [N]$, $|A| = \delta N$ and $\epsilon > 0$. Suppose further

$$|\widehat{f_A}(\alpha)| \geq \epsilon N. \tag{6}$$

for some $\alpha \in \mathbb{T}$, and let

$$L = \lfloor (\epsilon N)^{1/2}/16 \rfloor.$$

By the Dirichlet approximation theorem, there exists $q \in \mathbb{N}$, $q \leq 16L$ such that

$$||q\alpha||_{\mathbb{T}} \leq 1/(16L),$$

where $||q\alpha||_{\mathbb{T}}$ is the distance between $q\alpha$ and the nearest integer. Set $P_0 = \{-lq : 1 \leq l \leq L\}$. Let $\ell \in \mathbb{N}$. By the triangle inequality,

$$|e^{2\pi i \ell q\alpha}| \geq 1 - |1 - e^{2\pi i \ell q \alpha}|.$$

It follows that

$$\begin{aligned} |\widehat{1_{P_0}}(\alpha)| &= \sum_{\ell=1}^{L} |e^{2\pi i \ell q \alpha}| \\ &\geq L - \sum_{\ell=1}^{L} |1 - e^{2\pi i \ell q \alpha}|. \end{aligned}$$

For any $t \in \mathbb{R}$, we see

$$|1 - e^{2\pi i t}| \leq 2\pi ||t||_{\mathbb{T}},$$

by comparing straight line distance to arclength around the circle. Thus,

$$L - \sum_{\ell=1}^{L} |1 - e^{2\pi i \ell q \alpha}| \geq L - 2\pi \sum_{\ell=1}^{L} \ell ||q\alpha||_{\mathbb{T}}.$$

Now we use the fact that $||q\alpha||_{\mathbb{T}} \leq 1/(16L)$ along with the formula $\sum_{\ell=1}^{L} \ell = \frac{L(L+1)}{2}$:

$$L - 2\pi \sum_{\ell=1}^{L} \ell ||q\alpha||_{\mathbb{T}} \geq L - \frac{\pi(L+1)}{16} \geq L - \frac{\pi L}{8} \geq L - \frac{L}{2} = \frac{L}{2}.$$

Putting these facts together gives the following estimate.

$$|\widehat{1_{P_0}}(\alpha)| \geq L/2. \tag{7}$$

Next, we take the convolution of f_A and 1_{P_0}, defined by

$$\begin{aligned} f_A * 1_{P_0}(n) &= \sum_{m\in\mathbb{Z}} f_a(m) 1_{P_0}(n-m) \\ &= \sum_{m\in\mathbb{Z}} 1_A(m) 1_{P_0}(n-m) - \delta \sum_{m\in\mathbb{Z}} 1_{[N]}(m) 1_{P_0}(n-m). \end{aligned}$$

Observe that

$$1_A(m)1_{P_0}(n-m) = \begin{cases} 1 & m \in A \cap (n - P_0) \\ 0 & \text{otherwise} \end{cases}.$$

Therefore,

$$\sum_{m\in\mathbb{Z}} 1_A(m)1_{P_0}(n-m) = |A \cap (n-P_0)|,$$

and

$$\delta \sum_{m\in\mathbb{Z}} 1_{[N]}(m)1_{P_0}(n-m) = \delta|[N] \cap (n-P_0)|.$$

Hence,

$$f_A * 1_{P_0}(n) = |A \cap (n-P_0)| - \delta|[N] \cap (n-P_0)|. \tag{8}$$

By combining, (6) and (7), we have

$$\epsilon NL/2 \leq |\widehat{f_A}(\alpha)\widehat{1_{P_0}}(\alpha)| = |\widehat{f_A * 1_{P_0}}(\alpha)|,$$

where we also used the fact that the Fourier transform of a convolution is the product of the Fourier transforms. Now, we apply triangle inequality once again to obtain

$$|\widehat{f_A * 1_{P_0}}(\alpha)| = \left| \sum_{n\in\mathbb{Z}} f_A * 1_{P_0}(n) e^{-2\pi i n\alpha} \right| \leq \sum_{n\in\mathbb{Z}} |f_A * 1_{P_0}(n)|.$$

In particular,

$$\sum_{n\in\mathbb{Z}} |f_A * 1_{P_0}(n)| \geq \epsilon NL/2. \tag{9}$$

At this point, observe that

$$\sum_{n\in\mathbb{Z}} f_A(n) = \sum_{n\in\mathbb{Z}} 1_A(n) - \delta \sum_{n\in\mathbb{Z}} 1_{[N]}(n) = |A| - \delta N = 0.$$

It follows that

$$\sum_{n\in\mathbb{Z}} f_A * 1_{P_0}(n) = 0. \tag{10}$$

Take

$$\sum_{n\in\mathbb{Z}} (f_A * 1_{P_0}(n))_+$$

to be the sum of the positive terms in (10) and

$$\sum_{n\in\mathbb{Z}} (f_A * 1_{P_0}(n))_-$$

to be the sum of the negative terms. Since the sum in (10) equals 0, these positive and negative parts cancel each other out, and thus they have the same size in absolute value. Namely,

$$\sum_{n\in\mathbb{Z}} (f_A * 1_{P_0}(n))_+ = \sum_{n\in\mathbb{Z}} \left|(f_A * 1_{P_0}(n))_-\right|.$$

Therefore,

$$\sum_{n\in\mathbb{Z}} (f_A * 1_{P_0}(n))_+ = \frac{1}{2}\sum_{n\in\mathbb{Z}} |f_A * 1_{P_0}(n)| \geq \epsilon NL/4, \tag{11}$$

where we also used (9).

Finally, we look back to equation (8):

$$(f_A * 1_{P_0})(n) = |A \cap (n - P_0)| - \delta|[N] \cap (n - P_0)|.$$

Notice that if $n < -qL$ or $n > N$, then $f_A * 1_{P_0}(n) = 0$. In addition,

$$(f_A * 1_{P_0})(n) \leq |A \cap (n - P_0)| \leq L.$$

Therefore, by (11),

$$\sum_{n\in[0,N-qL]} (f_A * 1_{P_0})(n)_+ \geq \left(\sum_{n\in\mathbb{Z}} (f_A * 1_{P_0})(n)_+\right) - L(2qL) \geq \epsilon NL/4 - 32L^3.$$

Recall that $L \leq (\epsilon N)^{1/2}/16$, so $32L^3 \geq \epsilon NL/8$. Hence,

$$\sum_{n\in[0,N-qL]} (f_A * 1_{P_0}(n))_+ \geq \epsilon NL/8.$$

By the pigeonhole principle, there must be some value of $n \in [0, N - qL]$ for which

$$(f_A * 1_{P_0})(n) \geq \frac{\epsilon NL/8}{N - qL} \geq \epsilon L/8.$$

Combining this with (8) yields

$$|A \cap (n - P_0)| = \delta|[N] \cap (n - P_0)| + (f_a * 1_{P_0})(n) \geq \delta L + \epsilon L/8 = L(\delta + \epsilon/8).$$

At last, we conclude that $P = n - P_0$ is an arithmetic progression with the desired properties. $\square$

Corollary 1. *Suppose $N \in \mathbb{N}$ and $A \subset [N]$ with $|A| = \delta N$. If $\delta \geq \left(rN^{k-2}\right)^{-1/(k-1)}$ and A is $\mathcal{F}$-free, then there exists $N' \in \mathbb{N}$ and $A' \subset [N']$ such that A' is $\mathcal{F}$-free, $|N'| \geq r^{1/2}N^{1/2}\delta/32$, and $|A'| \geq N'(\delta + r^2\delta/8)$.*

Proof. We have 2 cases based on how many elements of A lie in $(\beta N, \gamma N]$.

Case 1: $|A \cap (\beta N, \gamma N]| < \frac{(\gamma-\beta)|A|}{2}$.

First suppose $k = 3$. Scaling $c_1x_1 + c_2x_2 + c_3x_3 = 0$ by -1 gives an equivalent problem. Thus, we may assume $c_1, c_2 > 0$ and $c_3 < 0$. It follows that $K = 0$, so $\beta = 0$ and $\gamma = 1$. Then

$$|A \cap (0, N]| = |A| > \frac{(\gamma - \beta)|A|}{2},$$

and the condition for case 1 cannot be met.

Suppose $k > 3$. By the pigeonhole principle,

$$(i)\ \ |A \cap (0, \beta N]| \geq |A|(1 + (\gamma - \beta)/4).$$

or

$$(ii)\ \ |A \cap (\gamma N, N]| \geq |A|(1 + (\gamma - \beta)/4).$$

Let (i) be true (the proof for (ii) is similar). Set $N' = \beta N$ and $A' = A \cap (0, \beta N]$. Since $k > 3$,

$$c_k^{(k-2)/2} \geq c_k.$$

Thus,

$$\begin{aligned} r^{1/2} N^{1/2} \delta/32 &\leq r^{1/2} N/32 \\ &= \frac{(\gamma - \beta)^{(k-1)/2} N}{32 c_k^{(k-2)/2} 2^{k/2}} \\ &\leq \frac{(\gamma - \beta)^{(k-1)/2} N}{32 c_k}. \end{aligned}$$

Recall that

$$\gamma - \beta = (1 - K/J)^2.$$

In addition, we may assume $K \neq 0$ by scaling $c_1 x_1 + \cdots c_k x_k = 0$ by -1. Then

$$\begin{aligned} \frac{(\gamma - \beta)^{(k-1)/2} N}{32} &= \frac{(1 - K/J)^{k-1} N}{32 c_k} \\ &= \frac{\frac{1}{J}(J - K)(1 - K/J)^{k-2} N}{32} \\ &\leq \frac{\frac{K}{J}(J - K)(1 - K/J)^{k-2} N}{32 c_k} \\ &= \frac{\beta (J - K)(1 - K/J)^{k-3} N}{32 c_k}. \end{aligned}$$

We also know $c_k = J - K$. In this way, we have

$$\begin{aligned} \frac{\beta (J - K)(1 - K/J)^{k-3} N}{32 c_k} &= \frac{\beta (1 - K/J)^{k-3} N}{32} \\ &\leq \beta N \\ &= N'. \end{aligned}$$

Putting these facts together, we obtain

$$N' \geq r^{1/2} N^{1/2} \delta/32.$$

Finally, we see

$$\begin{aligned} |A'| &\geq |A|(1 + (\gamma - \beta)/4) \\ &= N(\delta + \delta(\gamma - \beta)/4)) \\ &\geq N'(\delta + \delta(\gamma - \beta)/4). \end{aligned}$$

In addition, we know $k > 3$, $c_k \geq 1$, and $\gamma - \beta \leq 1$. Therefore,

$$\begin{aligned} N'(\delta + \delta(\gamma-\beta)/4) &\geq N'\left(\delta + \frac{\delta(\gamma-\beta)^{k-1}}{c_k^{k-2}2^k}\right) \\ &= N'(\delta + r^2\delta/8). \end{aligned}$$

as needed.

Case 2: $|A \cap (\beta N, \gamma N]| \geq \frac{(\gamma-\beta)|A|}{2}$.

Lemmas 1 and 2 provide an $\mathcal{F}$-free arithmetic progression $P = \{n + \ell q : 1 \leq \ell \leq L\} \subset [N]$ such that $|A \cap P| \geq L(\delta + r^2\delta/8)$. Set $N' = L$ and $A' = ((A \cap P) - n)/q$. $\mathcal{F}$ is invariant under scaling and translation. Therefore, since $A \cap P$ is $\mathcal{F}$ free, A' is also $\mathcal{F}$ free. Hence, A' and N' satisfy the desired conditions.

□

We are prepared to prove theorem 2.

Proof. Set $N_0 = N$ and $A_0 = A$. By lemma 3, as long as $\delta_j \geq \left(rN_j^{k-2}\right)^{-1/(k-1)}$, there exist $N_{j+1} \in \mathbb{N}$ and $A_{j+1} \subset [N_{j+1}]$ with $A_{j+1} = \delta_{j+1}N_{j+1}$ where A_1 is $\mathcal{F}$ free, $\delta_1 \geq \delta + r\delta^2/8$ and $|N_{j+1}| \geq r^{1/2}N_j^{1/2}\delta/32$. But this process cannot continue forever. In fact, since the density increment at each step is $r\delta_j^2/8$, the δ_j's will exceed 2δ in at most $\delta/(r\delta^2/8) = 8/(r\delta)$ steps, at which point they will exceed 4δ in at most $4/(r\delta)$ steps, and so on. Successively doubling in this way, the δ_j's must eventually exceed 1, but this is absurd. Therefore, we must have $\delta_{j-1} < \left(rN_{j-1}^{k-2}\right)^{-1/(k-1)}$ for some

$$j < \frac{8}{r\delta} + \frac{4}{r\delta} + \frac{2}{r\delta} + \cdots = \frac{16}{r\delta}.$$

We use this inforation to create the following inequality chain.

$$\begin{aligned} \delta \leq \delta_{j-1} &\leq r^{-1/(k-1)}N_{j-1}^{-(k-2)/(k-1)} \\ &\leq r^{-1/(k-1)}N_{j-1}^{-1/2} \\ &\leq r^{-1/(k-1)}(32r^{-1/2}\delta^{-1})^{1/2}N_{j-2}^{-1/4} \\ &\leq r^{-1/(k-1)}(32r^{-1/2}\delta^{-1})^{3/4}N_{j-3}^{-1/8} \\ &\vdots \\ &\leq r^{-1/(k-1)}(32r^{-1/2}\delta^{-1})^{1-1/2^{j-1}}N^{-1/2^j} \\ &\leq r^{-1/(k-1)}(32r^{-1/2}\delta^{-1})N^{-1/2^{16/(r\delta)}}. \end{aligned}$$

Rearranging the extreme ends of the chain yields

$$N^{1/2^{16/(r\delta)}} \leq 32r^{-1/2-1/(k-1)}\delta^{-2}.$$

Take the natural logarithm of both sides to obtain

$$\log N \leq 2^{16/(r\delta)} \log\left(32r^{-1/2-1/(k-1)}\delta^{-2}\right).$$

Take the natural logarithm of both sides once more.

$$\log\log N \le \frac{16}{r\delta}\log 2 + \log\log\left(32r^{-1/2-1/(k-1)}\delta^{-2}\right)$$
$$\le \frac{16}{r\delta}\log 2 + \log\log\left(32r^{-2}\delta^{-2}\right).$$

Note for every $x > 0$, $\log x < x/2$. Thus,

$$\begin{aligned}\frac{16}{r\delta}\log 2 + \log\log\left(32r^{-2}\delta^{-2}\right) &< \frac{16}{r\delta}\log 2 + \log\left(16r^{-2}\delta^{-2}\right)\\ &= \frac{16}{r\delta}\log 2 + 2\log\left(4r^{-1}\delta^{-1}\right)\\ &\le \frac{16}{r\delta}\log 2 + \frac{4}{r\delta}\\ &\le \frac{16}{r\delta}.\end{aligned}$$

In total, these inequalities give the desired result:

$$|A| \le \frac{16N}{r\log\log N}.$$

□

8. Finitary density regularity vs. density regularity

Notice that our statement of Roth's theorem gives a quantitative upper bound on $\mathcal{F}$-free subsets of $\mathbb{N}$. By finding such an upper bound, we prove *finitary density regularity.*

Definition 1. We call $\mathcal{F}$ finitary density regular if for every $\delta > 0$, there exists $N_0 \in \mathbb{N}$ such that:

$$(A \subseteq [N] \text{ is } \mathcal{F}\text{-free} \implies |A| < \delta N) \text{ for all } N \ge N_0.$$

Notice that δ corresponds to the upper bound which grows arbitrarily small. In the case of Roth's theorem for a general linear equation, this is $16/\log\log N$. Even though proving density regularity is the goal in some problems, other times we may want to find improved bounds. As we will see in section 9, it is possible to do much better than $16/\log\log N$ in the case of Roth's theorem.

As expected, finitary density regularity implies density regularity.

Proposition 1. *If $\mathcal{F}$ is finitary density regular, then $\mathcal{F}$ is density regular.*

Proof. Let $A \subset \mathbb{N}$ be $\mathcal{F}$ free, and let $\epsilon > 0$. Since $\mathcal{F}$ is finitary density regular, there exists $N_0 \in \mathbb{N}$ such that

$$\frac{|A \cap [N]|}{N} < \epsilon$$

for every $N > N_0$. It follows that

$$\lim_{N\to\infty} \frac{|A \cap [N]|}{N} = 0.$$

Thus, $\delta(A) = 0$, and $\mathcal{F}$ is density regular, as needed. □

Interestingly, the converse is not necessarily true. We will construct an extreme example. Make the following definitions.

$$\alpha(\mathcal{F}) = \sup\{\bar{\delta}(A) : A \subset \mathbb{N} \text{ is } \mathcal{F}\text{-free}\}.$$

$$\beta(\mathcal{F}) = \sup\{\delta : \text{For infinitely many } N \in \mathbb{N}, \text{there exists an } \mathcal{F}\text{-free set } A \subset [N] \text{ with } |A| > \delta N\}.$$

Observe that $\mathcal{F}$ is density regular if and only if $\alpha(\mathcal{F}) = 0$ and $\mathcal{F}$ is finitary density regular if and only if $\beta(\mathcal{F}) = 0$. In sense, α measures how close $\mathcal{F}$ is to density regularity and β measures how close $\mathcal{F}$ is to finitary density regularity. Below, we construct a family that is density regular but not finitary density regular in the most extreme way possible ($\alpha(\mathcal{F}) = 0$ and $\beta(\mathcal{F}) = 1$).

Proposition 2. *Let*

$$\mathcal{F} = \{\{x, y\} \in \mathbb{N}^2 : y > x^2\}$$

Then $\alpha(\mathcal{F}) = 0$ and $\beta(\mathcal{F}) = 1$.

Proof. Let $A \subset \mathbb{N}$. Suppose $A \subset \mathbb{N}$ is infinite. Take $x \in A$. Since A is infinite, there exists some $y \in A$ such that $y > x^2$. Then A is not $\mathcal{F}$-free. Thus, if A is $\mathcal{F}$-free, then $\mathcal{F}$ must be finite, so $\delta(A) = 0$. It follows that $\alpha(\mathcal{F}) = 0$.

Let $0 < \delta < 1$, $N > \frac{1}{1-\delta}$. Notice $\{N, N+1, ..., N^2\} \subset [N^2]$ is $\mathcal{F}$-free, and

$$\begin{aligned}
|\{N, N+1, ..., N^2\}| &= N^2 - N + 1 \\
&> N^2 - N \\
&= N^2(1 - 1/N) \\
&> N^2(1 - (1 - \delta)) \\
&= \delta N^2.
\end{aligned}$$

It follows that $\beta(\mathcal{F}) = 1$.

□

9. Some cutting-edge results related to Roth's theorem

Mathematicians continue to refine and expand upon results related to Roth's theorem. Some have sought to improve density bounds while others have worked towards new generalizations. Recently, Bloom and Sisask [18] obtained a much anticipated, improved bound on the density of sets lacking 3AP's.

Theorem 3. *Let $N \in \mathbb{N}$. If $A \subset [N]$ contains no $3AP$, then*

$$|A| \leq CN/(\log N)^{1+\epsilon},$$

for some constants $C, \epsilon > 0$.

Note that $1/(\log N)^{1+\epsilon}$ shrinks far more quickly than $1/\log\log N$, so this is indeed an improved upper bound.

Schoen and Sisask [19] were able to find even better density bound for sets with no solutions to $x + y + z = 3w$.

Theorem 4. *If $N \geq 3$ and $A \subset [N]$ contains no solution to $x + y + z = 3w$ with x, y, z, w not all equal, then*

$$|A| \leq \frac{N}{\exp(c(\log N)^{1/7})}$$

for some positive constant c.

Notice that Roth's theorem for a general linear equation applies to $x + y + z = 3w$ (the coefficients sum to zero), but $1/\exp(c(\log N)^{1/7}$ shrinks much more quickly than $1/\log\log N$ or $1/(\log N)^{1+\epsilon}$.

Chapman and Chow very recently created a powerful extension of both Roth's theorem and Rado's theorem. Consider solutions to a polynomial equation of the form

$$a_1P(x_1) + \cdots + a_sP(x_s) = 0, \tag{1}$$

where P is an *intersective* polynomial with integer coefficients of degree d and $a_1, \ldots, a_s \in \mathbb{Z}$($P$ is intersective if $P(\mathbb{Z})$ contains a multiple of every integer). The equations in Schur's theorem, Rado's theorem, and Roth's theorem for a general linear equation are all of this form (with $P(x) = x$). As another example, Pythagorean triples satisfy such a polynomial equation ($P(x) = x^2$ and $a_1 = a_2 = 1, a_3 = -1$).

They showed that under the right conditions, the family of solutions to this polynomial equation are density regular if and only if the coefficients sum to zero (compare to Roth's theorem for a general linear equation) and partition regular if and only if a subset of the coefficients sums to zero (compare to Rado).

In particular, let $\mathcal{F}$ be a family of solutions to (1). Chapman and chow [20] proved the following result.

Theorem 5. *Let*

$$s_1(d) = \begin{cases} 5, & \text{if } d = 2 \\ 9, & \text{if } d = 3 \\ d^2 - d + 2\lfloor\sqrt{2d+2}\rfloor + 1, & \text{if } d \geq 4 \end{cases}.$$

If $s \geq s_1(d)$, then we have the following.

(1) $\mathcal{F}$ is partition regular if and only if there exists a non-empty subset of $\{a_1, \ldots, a_s\}$ which sums to 0.

(2) $\mathcal{F}$ is density regular if and only if $a_1 + \cdots + a_s = 0$.

If $s_1(2)$ were 3 instead of 5, we would have a solution to the Pythagorean triples problem!

10. The transference principle

We now introduce the idea of relative density.

Definition 1. For $A \subseteq B \subseteq \mathbb{N}$, define

$$\delta_B(A) = \lim_{N\to\infty} \frac{|A \cap B|}{|B \cap [N]|}.$$

Definition 2. For $B \subset \mathbb{N}$ we say a family $\mathcal{F}$ is *B-density regular* if every $\mathcal{F}$-free $A \subset B$ has $\delta_B(A) = 0$.

We think of $\delta_B(A)$ as the conditional probability a natural number lies in A given it lies in B. With these definitions, we are ready to state a suprising fact proven by Ben Green and Terrence Tao [11]. Let $\mathcal{P}$ denote the set of prime numbers.

Theorem 6 (Green-Tao theorem)**.** *For every $k \in \mathbb{N}$, the family of kAP's is $\mathcal{P}$-density regular, where $\mathcal{P}$ is the set of primes.*

Interestingly, $\delta(\mathcal{P}) = 0$, so the Green-Tao theorem is not a consequence of Szemerédi's theorem. Nevertheless, this property of dense subsets of $\mathbb{N}$ "transfers" to a property of relatively dense subsets of the prime numbers. This is the transference principle.

Mathematicians often study the transference principle in the density setting, as in the Green-Tao theorem, but the same cannot be said of the transference principle in the coloring setting. Here, we provide some definitions and observations in this direction.

Definition 3. For $B \subseteq \mathbb{N}$, we say a family $\mathcal{F}$ is *B-partition regular* if for every $r \in \mathbb{N}$, every r-coloring of B yields a monochromatic member of $\mathcal{F}$.

Let S be the set of perfect squares. We may ask the following question: are the Schur triples S-partition regular? This is the unsolved Pythagorean triples problem.

Many results concerning partition regularity carry over to the relative setting. We illustrate this fact with a few exercises, after which our discussion is complete. The first provides a relationship between relative partition regularity and relative density regularity.

Proposition 1. *For any $B \subset \mathbb{N}$ and family $\mathcal{F}$ of finite subset of $\mathbb{N}$, B-density regularity of $\mathcal{F}$ implies B-partition regularity of $\mathcal{F}$.*

Proof. Suppose $B \subset \mathbb{N}$, $\mathcal{F}$ is a family of finite subsets of $\mathbb{N}$ such that $\mathcal{F}$ is not B-partition regular, and assume for the sake of contradiction that $\mathcal{F}$ is B-density regular. Then there exists a partition $\{C_1, C_2, ..., C_k\}$ of B which yields no monochromatic member of $\mathcal{F}$. Thus, for each C_i, we have $\delta_B(C_i) = 0$ because we have assumed $\mathcal{F}$ is B-density regular. Therefore,

$$\begin{aligned} 0 &= \sum_{i=1}^{k} \delta_B(C_i) \\ &= \sum_{i=1}^{k} \lim_{N\to\infty} \frac{|C_i \cap [N]|}{|B \cap [N]|} \\ &= \lim_{N\to\infty} \sum_{i=1}^{k} \frac{|C_i \cap [N]|}{|B \cap [N]|}, \end{aligned}$$

by the limit sum property. We also know $\{C_1, C_2, ..., C_k\}$ is a partition of B, so the the C_i's are pairwise disjoint and $\bigcup_{i=1}^k C_i = B$. Thus,

$$\lim_{N\to\infty} \sum_{i=1}^{k} \frac{|C_i \cap [N]|}{|B \cap [N]|} = \lim_{N\to\infty} \frac{\left|\bigcup_{i=1}^{k} (C_i \cap [N])\right|}{|B \cap [N]|}$$

$$
\begin{aligned}
&= \lim_{N\to\infty} \frac{\left|[N] \cap \bigcup_{i=1}^{k} C_i\right|}{|B \cap [N]|} \\
&= \lim_{N\to\infty} \frac{|B \cap [N]|}{|B \cap [N]|} \\
&= 1.
\end{aligned}
$$

Putting everything together, we have $0 = 1$, a contradiction. Hence, $\mathcal{F}$ is not B-density regular.

□

By setting $B = \mathbb{N}$, we see that density regularity implies partition regularity. For instance, Szemerédi's theorem implies Van der Waerden's theorem. However, the converse is not necessarily true. For instance, the Schur triples are partition regular but not density regular. Indeed, the set of odd numbers contains no Schur triple even though it has density $1/2$.

Here is an interesting consequence of partition regularity: if $\mathcal{F}$ is partition regular, then every r-coloring of $\mathcal{F}$ yields infinitely many monochromatic members of $\mathcal{F}$. It turns out that this is true in the relative setting as well.

Proposition 2. *Suppose $\mathcal{F}$ is a family of finite subsets of $\mathbb{N}$, each of which contains at least two elements, $B \subset \mathbb{N}$ and $r \in \mathbb{N}$, and further suppose $\mathcal{F}$ is B-partition regular. Then every r-coloring of B yields infinitely many monochromatic members of $\mathcal{F}$.*

Proof. Let $C_1, ..., C_r$ be an arbitrary r-coloring of B. Assume $C_1, ..., C_r$ yields finitely many monochromatic members of $\mathcal{F}$, whose elements form the set $M = \{m_1, ..., m_j\}$. Then

$$\{m_1\}, ..., \{m_j\}, C_1 - M, C_2 - M, ..., C_r - M.$$

is a finite coloring of B with no monochromatic member of $\mathcal{F}$. Thus, $\mathcal{F}$ is not B-partition regular. □

Proposition 3. *Suppose $c_1, ..., c_k \in \mathbb{Z}$ and $c_1 + \cdots + c_k \neq 0$. Let $\mathcal{F}$ be the family of solutions to $c_1x_1 + \cdots + c_kx_k = 0$. If $\mathcal{F}$ is B-partition regular, then B contains infinitely many multiples of every natural number.*

Proof. Suppose $c_1x_1 + \cdots + c_kx_k$ is B-partition regular. Let $c_1 + \cdots + c_k = c$, and let $m \in \mathbb{N}$. We focus on the following partition of B:

$$\{B_1, B_2, ..., B_{cm}, \},$$

where

$$B_k = \{r \in B : r \equiv k \pmod{cm}\}.$$

By combining the fact that $\mathcal{F}$ is B-partition regular with proposition 2, we know there is some B_l which contains infinitely members of $\mathcal{F}$. Say $\{x_1, ..., x_k\} \subset B_l$ is a member of $\mathcal{F}$. Then

$$0 = c_1x_1 + \cdots c_kx_k \equiv lc \pmod{cm}.$$

Then $cm \mid lc$, so $m \mid l$. It follows that every element of B_l is divisible by m, and we also know B_l contains infinitely many elements because it contains infinitely many members of $\mathcal{F}$. Thus, B contains infinitely many multiples of m.

□

Proposition 4. *Let $B \subset \mathbb{N}$ contain $q\mathbb{N}$. Let $\mathcal{F}$ be a partition regular family in $\mathbb{N}$, and suppose that if $r \in \mathbb{N}$ and $m \in \mathcal{F}$, then $rm \in \mathcal{F}$. Then, $\mathcal{F}$ is B-density regular.*

Proof. Let $C_1, .., C_r$ be a partition of $q\mathbb{N}$. For each i, set $D_i = C_i/q$. Then $D_1, ..., D_r$ is a partition of $\mathbb{N}$. Because $\mathcal{F}$ is partition regular, we know that for some l, there is a monochromatic member of $\mathcal{F}$, say m, in D_l. Then qm is a monochromatic member of $\mathcal{F}$ in C_l, and we are done. It follows that $\mathcal{F}$ is $q\mathbb{N}$ partition regular, so $\mathcal{F}$ is B-partition regular, as needed.

□

11. Acknowledgment

I want to thank my advisor Alex Rice for being with me every step of the way. Without him, this project would not have been possible.

References

[1] L. Boza, J.M. Marín, M.P. Revuelta, M. I. Sanz, 3-*color Schur numbers*, Discret. Appl. Math. 263 (2019), 59-68.
[2] I. Schur, *Über die kongruenz $x^m + y^m = z^m$ (mod p)*, Jahresber. Deutsche Math.-Verein., vol. 25 (1916), 114-116.
[3] F. P. Ramsey, *On a problem of formal logic*, Proc. London Math. Soc. 30 (1929), 264-286.
[4] M.J.H. Heule *Schur number five*, Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (2018), 6598-6606.
[5] Heule, Kullman, Marek *Solving and Verifying the boolean Pythagorean Triples problem via Cube-and-Conquer*, SAT 2016, LNCS 9710, pages 228-245.
[6] B. M. Landman and A. Robertson, *Ramsey theory on the integers*, Student Mathematical Library, 24. American Mathematical Society, Providence, RI, 2004.
[7] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. 28 (1953), 104-109.
[8] E. Szmerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. 27 (1975), 199-245.
[9] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. 11 (2001), no. 3, 465-588.
[10] B. Green, *Roth's theorem in the primes*, Ann. of Math. 161 (2005), no. 3, 1609-1636.
[11] B. Green, T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. 167 (2008), no. 2, 481-547.
[12] R. Rado, *Studien zur kombinatorik*, Math. Z. 36 (1933), no. 1, 424-470.
[13] V. Balaji, A. Lott, A. Rice, *Schur's theorem in integer lattices*, INTEGERS Vol. 22.
[14] B. L. van der Waerden, *Beweis einer baudetschen Vermutung*, Nieuw Arch. Wisk. 15 (1927), 212-216.
[15] P. Erdös, P. Turán, *On some sequences of integers*, Journal London Math. Vol. 11 (1936)
[16] D. Hilbert, *Uber die Irreducibilitat ganzer rationaler Functionen mit ganzzahligen Coefficienten*, J. reine angew. Math. 110 (1892) 104–129.
[17] A. Blondal, V. Jungić, *Proof of van der Waerden's theorem in nine figures*, Arnold Math. Journal (2018) 4: 161-168
[18] T. Bloom, O. Sisask, *Breaking the logarithmic barrier in Roth's theorem on arithmetic progressions*, preprint (2020), https://arxiv.org/abs/2007.03528
[19] T. Schoen, O. Sisask, *Roth's theorem in four variables and additivie structures in sums of sparse sets*, Forum of Mathematics, Sigma, vol. 4 (2016).
[20] J. Chapman, S. Chow, *Generalised Roth and Rado criteria*, preprint (2022), https://arxiv.org/abs/2210.04347