Imperial College London

Department of Mathematics

Arithmetic Progressions within the Primes

Author:

Andrew Sawkins

Supervisor:

Dr Robert Barham

Submitted in partial fulfillment of the requirements for the MSc Pure Mathematics of Imperial College London

September 2021

Abstract

This dissertation is an expository account of one of the most celebrated theorems in number theory of the last 20 years, namely the Green-Tao theorem. The theorem states that the prime numbers contain arbitrarily long arithmetic progressions. This paper gives a complete proof and includes several of the simplifications that have been made in the years since the original publication.

Acknowledgments

I would like to thank my supervisor, Dr Robert Barham, for agreeing to supervise this dissertation and being unwavering in his support and patience through a particularly difficult year.

My family have always encouraged me to pursue my interests and given me the confidence to find my own path in education and beyond. For this and so much more, I am forever grateful.

Notation

Some notation used in this project:

- When $A \subseteq \mathbb{N}$, we say $A(N) = |\{a \in A : a \leq N\}|$.
- Let $\mathbb{P}$ be the set of prime numbers.
- $\pi(N) = |\{p \in \mathbb{P} : p \leq N\}|$. That is the size of the set of all primes less than or equal to N.
- Throughout the report, log will be the natural logarithm and a subscript will indicate the number of iterated logarithms. For example $\log_{[3]} N = \log \log \log N$.
- We will write $\mathbb{Z}_N$ to mean the abelian group $\mathbb{Z}/N\mathbb{Z}$.

Contents

Chapter 1

Introduction

1.1 Beginnings

Questions about prime numbers are always going to be attractive to mathematicians. There is something special about these atoms of the integers that give them a certain mystique. Unfortunately though many of these questions are incredibly complex, think about the Goldbach Conjecture and the Twin Prime Conjecture. Something that these problems share which adds immensely to their complexity is the combination of additive properties with multiplicative subsets. So another problem that easily fits into this category would be trying to find arithmetic progressions within the primes. Arithmetic progressions within the primes are very simple to find for a low number of terms. In fact, it is trivial for progressions of one or two terms and the first arithmetic progression with 3 terms is $3, 5, 7$. With a little more thought one will be able to find 4 and 5 term arithmetic progressions, however the problem of finding longer progressions quickly becomes incredibly challenging. How long would it take you to find the first 7-term arithmetic progression $7, 157, 307, 457, 607, 757, 907$? And in fact, the longest arithmetic progression of primes found so far has 27 primes. It was found by Rob Gahan and PrimeGrid on 23 September 2019:

$$224584605939537911 + 81292139 \cdot 23n \text{ for } n = 0, \ldots, 26$$

An immediate question that comes to mind when looking at this is whether there is a limit on the length of arithmetic progressions of the primes. Now, it is easy to show that there cannot be an infinite arithmetic progression of primes. (Suppose your progression is of the form $a + nd$, if this is infinite, then we must have that n takes the value a or a multiple of a at some point and clearly $a + ad$ is not prime.) So the best possible answer to our question is the following result of Ben Green and Terrence Tao:

Theorem 1 (Green-Tao, 2008 [13])**.** *There are arbitrarily long arithmetic progressions within the primes.*

Since the original proof of this theorem, there have been several simplifications, notably by Thomas Bloom [2] and more recently by David Conlon, Jacob Fox, and Yufei Zhao [4]. My aim is to give a full proof of this theorem in a more approachable form.

When looking at arithmetic progressions and in particular long arithmetic progressions, the first place we must look is the very deep result of Szemerédi:

Theorem 2 (Szemerédi, 1975 [19])**.** *Any subset of* $\mathbb{N}$ *with positive upper density contains arbitrarily long arithmetic progressions.*

There are many proofs of Szemerédi's theorem, and these proofs have brought in new ideas in areas as varied as combinatorics [19], ergodic theory

[9], and Fourier analysis [11]. The ideas found in these proofs have sparked new areas of research in all of their respective fields. Unfortunately, none of the proofs of Szemerédi's theorem are straightforward, and as such we shall be required to use it without proof and show how to prove the Green-Tao theorem assuming it.

The first theorem talking about non-trivial arithmetic progressions within the primes actually came far earlier than Szemerédi's result. In 1939 van der Corput proved the a result for 3-term arithmetic progressions of primes using the so called circle method.

Theorem 3 (van der Corput, 1939 [6]). *There are infinitely many 3-term arithmetic progressions consisting of all primes.*

In the decades after this result, despite the immense work done by mathematicians in finding results about sets with positive upper density no further results were found about the primes. The next result was in 1981 by Heath-Brown when he was able to prove a partial result for 4-term arithmetic progressions.

Theorem 4 (Heath-Brown, 1981 [15]). *There are infinitely many 4-term arithmetic progressions consisting of three primes and one almost prime.*

In this theorem an almost prime is a number with at most two prime factors when counted with multiplicity. This result was also obtained using the circle method, and it seemed as though this method had some barrier to be able to prove anything much stronger, and as such some new ideas were needed in order to prove Theorem 1.

1.2 Basic Definitions

Definition 1. *A **k-term arithmetic progression** (k-A.P) is a sequence of integers*

$$a, a+d, \ldots, a+(k-1)d$$

*with $a, d, k \in \mathbb{N}$. An A.P is considered **trivial** when $d = 0$.*

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

$$\limsup_{N \to \infty^+} \frac{A(N)}{N}$$

Definition 3. *Now if $A \subseteq \mathbb{P}$, then we define the **relative upper density of** $A \in \mathbb{P}$ to be*

$$\limsup_{N \to \infty^+} \frac{A(N)}{\pi(N)}$$

Definition 4. *Let $f : \mathbb{Z} \to \mathbb{C}$ be a function with finite support. Then define*

$$\Lambda(f) = \sum_{\substack{x+y=2z \\ x,y,z \in \mathbb{Z}}} f(x)f(y)f(z)$$

Remark 1. *In particular, we see that if $A \subseteq \mathbb{Z}$ (finite), then $\Lambda(1_A)$ is the number of 3-A.P's (including the trivial 3-A.P's) contained in A.*

Definition 5. *The number of non-trivial 3-A.P's in $A \subseteq \mathbb{Z}$ will be given by*

$$T(A) = \sum_{\substack{x+y=2z \\ x,y,z \in \mathbb{Z} \\ x \neq y \neq z}} 1_A(x)1_A(y)1_A(z).$$

Definition 6. *For any integer $N \geq 1$ we take*

$$r_3(N) = sup\{|A| : A \subseteq [1, N], T(A) = 0\}.$$

That is, $r_3(N)$ is the cardinality of any maximal subset of A not containing non-trivial 3-A.P's.

Chapter 2

Roth's Theorem

As incredible as the Szemerédi theorem is, it was not the first result of this type. About 20 years earlier Klaus Roth proved the first non-trivial theorem for 3-A.P's. Any subset of $\mathbb{N}$ with positive upper density contains a 3-A.P. It can be written precisely as follows.

Theorem 5 (Roth, 1953 [16])**.** *If $A \subseteq [1, N]$ and A does not contain any non-trivial 3-A.P's then*

$$\frac{|A|}{N} \ll \frac{1}{\log \log N} = \frac{1}{\log_{[2]} N}. \tag{2.1}$$

There is an equivalent statement to Roth's Theorem:

Theorem 6. *If $\delta > 0$, then for any $N > exp(exp(\frac{c}{\delta}))$ and any $A \subseteq [1, N]$ with $|A| \geq \delta N$, there is a non-trivial 3-A.P in A.*

Theorem 7 (Behrend, 1946 [1])**.** *There is a subset $A \subseteq [1, N]$ with*

$$\frac{|A|}{N} \geq exp(-c\sqrt{\log N}) \tag{2.2}$$

such that A does not contain any non-trivial 3-A.P's (for all N large enough).

Proof. Let $M \geq 1$, and $n \geq 1$. Let $x \in [1, M]^n$, so $x = (x_1, x_2, \ldots, x_n)$, where each x_i is an integer. Then,

$$\|x\|^2 = x_1^2 + x_2^2 + \ldots + x_n^2$$

which satisfies

$$n \leq \|x\|^2 \leq nM^2.$$

Thus the map $x \mapsto \|x\|^2$ maps the set $[1, M]^n$ with M^n elements to a set with $nM^2 - n + 1$ elements. Therefore by the pigeonhole principle there is an integer r such that at least $\frac{M^n}{nM^2-n+1}$ integer points $x \in [1, M]^n$ satisfy $x_1^2 + \ldots + x_n^2 = r$.

Let S be the set of such integer points. That is

$$S = \{x \in [1, M]^n : x_1^2 + \ldots + x_n^2 = r\}.$$

So we see that

$$|S| \geq \frac{M^n}{nM^2 - n + 1} \geq \frac{M^n}{nM^2} \geq \frac{M^{n-2}}{n}.$$

Consider the injective map $f : S \to \mathbb{Z}$

$$x \mapsto \sum_{1 \leq i \leq n} x_i (2M)^{i-1}$$

We see that f is the restriction of an $\mathbb{R}$-linear map. Also, for $x \in (-2M, 2M)^n$ we have that $f(x) = 0 \iff x = 0$. This implies that for any $x, y, z \in S$,

$$x + y = 2z \iff f(x) + f(y) = 2f(z).$$

Since S is a sphere, and a straight line can intersect at sphere in $[1, M]^n$ in at most 2 points, it does not contain any non-trivial solutions to $x + y = 2z$, so it follows that $f(S)$ also does not contain such a solution. That is $f(S)$ contains no non-trivial 3-A.P's.

Now,

$$|S| \geq \frac{M^{n-2}}{n} \Rightarrow |f(S)| \geq \frac{M^{n-2}}{n}.$$

And we have that,

$$f(x) = \sum_{1 \leq i \leq n} x_i (2M)^{i-1}, 1 \leq x_i \leq M.$$

Thus,

$$1 \leq f(x) \leq M \frac{(2M)^n - 1}{2M - 1} \leq (2M)^n$$

Now taking $N = (2M)^n$, $M = \frac{N^{\frac{1}{n}}}{2}$, and $n = \sqrt{\log N}$ the result follows. □

The function r_3 was first studied by Paul Erdős and Paul Turán [8]. In their paper they calculated the values of $r_3(n)$ for $n \leq 23$ and for $n = 41$. However, a paper by Arun Sharma [18] showed that the value of $r_3(20)$ given by Erdős and Turán was incorrect and he calculated the values of $r_3(n)$ for $n \leq 27$ and for $n = 41, 42, 43$. Subsequently, with the aid of computers the exact value of $r_3(n)$ has been calculated for all $n \leq 123$ [7].

These more recent papers have also been able to deal with some more cases of a conjecture of Szekeres:

Conjecture 1 (Szekeres)**.** *For $k \geq 1$ we have $r_3(\frac{1}{2}(3^k + 1)) = 2^k$.*

The cases $k = 1, 2, 3$ were dealt with by Erdős and Turán [8], $k = 4$ by Arun Sharma [18], and $k = 5$ by Janusz Dybizbański [7].

Although we are currently unable to prove or find a counterexample to Szekeres conjecture, there are still many more interesting properties of the function r_3 that can be studied.

Proposition 1. *For all $m, n \geq 1$ we have*

$$r_3(m+n) \leq r_3(n) + r_3(m).$$

That is, the function r_3 is a subadditive function.

Proof. We have that $[1, m+n] = [1, m] \bigcup (m, m+n]$. Now for any $A \subseteq [1, m+n]$ we can write $A = (A \cap [1, m]) \bigcup (A \cap (m, m+n])$.
Now, if $T(A) = 0$, then we must have $T(A \cap [1, m]) = 0$. This implies that $|A \cap [1, m]| \leq r_3(m)$. Now, again since $T(A) = 0$ we must have that $T(A \cap (m, m+n]) = 0$. This now implies that

$$T(A \setminus m \cap [1, n]) = 0,$$

and thus

$$|A \cap (m, m+n]| \leq r_3(n).$$

Hence,

$$|A| \leq r_3(m) + r_3(n).$$

Finally, by passing to the supremum over all A with $T(A) = 0$ we get that

$$r_3(m+n) \leq r_3(m) + r_3(n).$$

□

Corollary 1. *The following limit exists:*

$$\lim_{n \to \infty^+} \frac{r_3(n)}{n}.$$

Proof. Let $k \geq 1$ be fixed. Then, by Euclid's algorithm, we can for any $n \geq 1$ write $n = qk + l$. Then by the proposition above we get

$$\begin{aligned} r_3(n) &\leq qr_3(k) + r_3(l) \\ \frac{r_3(n)}{n} &\leq \frac{qr_3(k)}{n} + \frac{r_3(l)}{n} \\ &\leq \frac{qr_3(k)}{qk} + \frac{r_3(l)}{n} \\ &\leq \frac{r_3(k)}{k} + \frac{r_3(l)}{n}. \end{aligned}$$

Hence,

$$\limsup_{n\to\infty} \frac{r_3(n)}{n} \leq \frac{r_3(k)}{k}.$$

This then implies that

$$\limsup_{n\to\infty} \frac{r_3(n)}{n} \leq \liminf_{k\to\infty} \frac{r_3(k)}{k}.$$

Therefore $\limsup = \liminf$ so the limit exists. □

From this study of r_3 we can formulate a version of Roth's theorem:

Theorem 8 (Weak Roth, 1952 [17])**.**

$$\lim_{n\to\infty^+} \frac{r_3(n)}{n} = 0.$$

Before we are able to prove this theorem, there is still some work to be done. There are definitions, a few lemmas and a basic understanding of Fourier transforms that are required. Therefore we have left the proof until Section 2.1.

In fact, if there is one 3-A.P, then there must be many. This was shown by a theorem of Varnavides. We shall just be assuming the result for the purposes of this paper, however the proof is not difficult to follow [20].

Theorem 9 (Varnavides, [20]). *Let $N \geq 1$ and suppose that $A \subseteq [1, N]$ with $T(A) \neq 0$. Then for all M, such that $1 \leq M \leq N$ we have*

$$T(A) \geq \left(\frac{|A|}{N} - \frac{(r_3(M)+1)}{M} \right) \frac{N^2}{4M^4}.$$

2.0.1 Fourier Facts

Let $f : \mathbb{Z} \to \mathbb{C}$ be of finite support. Then the Fourier transform of f is given by

$$\hat{f}(t) = \sum_{n \in \mathbb{Z}} f(n) e^{-2\pi i n t}. \tag{2.3}$$

The function $\hat{f}$ goes from the circle ($\mathbb{T} = \mathbb{R}/\mathbb{Z}$) to the complex plane. Then we have that if g is also of finite support, then the convolution of f and g is given by

$$\begin{aligned} f * g(n) &= \sum_{m \in \mathbb{Z}} f(m) g(n-m) \\ &= \sum_{m \in \mathbb{Z}} f(n-m) g(m) \\ &= g * f(n). \end{aligned}$$

We note that

$$\sum_{n \in \mathbb{Z}} f * g(n) = \left(\sum_{n \in \mathbb{Z}} f(n) \right) \left(\sum_{n \in \mathbb{Z}} g(n) \right).$$

$$f \hat{*} g(t) = \hat{f}(t) \hat{g}(t).$$

Then since

$$\int_0^1 e^{2\pi i n t} dt = \begin{cases} 1 & if\ m = 0 \\ 0 & if\ m \neq 0 \end{cases}.$$

we obtain

$$\int_0^1 |\hat{f}(t)|^2 dt = \sum_{n\in\mathbb{Z}} |f(n)|^2.$$

Then if $f(n) \geq 0$ for all $n \in \mathbb{Z}$ we get

$$\sup_{t\in\mathbb{T}} |\hat{f}(t)| \leq \sum_{n\in\mathbb{Z}} |f(n)| = \sum_{n\in\mathbb{Z}} f(n) = \hat{f}(0).$$

2.0.2 Bounding Bohr Sets

In our proof of Roth's theorem we use Bohr sets so here we put in a definition:

Definition 7. *Take $0 \leq \epsilon < \frac{1}{2}$, and $0 \leq \eta < 1$. Let $S = \{t : |\hat{f}(t)| > \eta N\}$. Then we define the **Bohr set** $B(S, \epsilon)$ as*

$$B(S, \epsilon) = \{n \in \mathbb{Z} : |n| \leq \epsilon N, \| nt \| \leq \epsilon \, \forall t \in S\}$$

Theorem 10. *Let $f : \mathbb{Z} \to \mathbb{C}$ be supported on $[1, N]$ and satisfy the following:*

- $\sum_{n\in\mathbb{Z}} |f(n)| \leq c_1 N$, *with* $c_1 \geq 1$.
- $\int_0^1 |\hat{f}(t)|^p dt \leq c_2 N^{p-1}$ *for some* $p \geq 2$.

Further, for any η, ϵ with $0 < \eta \leq 1, 0 < \epsilon \leq \frac{1}{2}$. Let

$$S = \{t \in \mathbb{T} : |\hat{f}(t)| \geq \eta N\}.$$

We then have

$$|B(S, \epsilon)| \geq \left(\frac{\epsilon}{4}\right)^{\frac{c_3(p)}{\eta^{p+1}}} N.$$

Proof. We will first show that there is an integer $r \geq 1$ and $\alpha_1, \ldots, \alpha_r \in S$ such that

$$r \leq \frac{c_4(p)}{\eta^{p+1}},$$

and

$$B(\{\alpha_1, \dots, \alpha_r\}, \frac{\epsilon}{2}) \subseteq B(S, \epsilon).$$

Indeed, let $r \geq 1$ be the largest integer such that there is a δ-spaced sequence $\alpha_1, \dots, \alpha_r$ in S, where $\delta = \frac{\eta}{4\pi c_1 N}$. That is $\|\alpha_i - \alpha_j\| \geq \delta$ for every $1 \leq i \neq j \leq r$. Then we find that every α in S satisfies $\|\alpha - \alpha_i\| \leq \delta$ for some $i, 1 \leq i \leq r$. Also, let $I_i := \{\alpha \in \mathbb{T} : \|\alpha - \alpha_i\| \leq \frac{\delta}{4}$ for each $i, 1 \leq i \leq r$. The $\{I_i\}_{1 \leq i \leq r}$ are mutually disjoint. In fact, if $\alpha \in I_i \cap I_j$ for $i \neq j$, then by the triangle inequality we get

$$\|\alpha_i - \alpha_j\| \leq \|\alpha - \alpha_i\| + \|\alpha - \alpha_j\| \leq \frac{\delta}{4} + \frac{\delta}{4} = \frac{\delta}{2},$$

contradicting that $\|\alpha_i - \alpha_j\| \geq \delta$.

On the other hand, for all $\alpha \in \bigcup_{1 \leq i \leq r} I_i$ we have $|\hat{f}(\alpha)| \geq \frac{\eta}{2} N$.

Say $\alpha \in I_i$. Then by the triangle inequality

$$|\hat{f}(\alpha)| \geq |\hat{f}(\alpha_i)| - |\hat{f}(\alpha) - \hat{f}(\alpha_i)|.$$

Now,

$$\begin{aligned}
|\hat{f}(\alpha) - \hat{f}(\alpha_i)| &\leq \sum_{n \in \mathbb{Z}} |f(n)||e(n\alpha) - e(n\alpha_i)| \\
&\leq \sum_{n \in \mathbb{Z}} |f(n)|(2\pi n\|\alpha - \alpha_i\| \\
&\leq c_1 N \cdot N \cdot \frac{\eta}{4\pi c_1 N} \\
&\leq \frac{\eta N}{2}.
\end{aligned}$$

Hence $|\hat{f}(\alpha)| \geq \eta N - \frac{\eta N}{2} = \frac{\eta N}{2}$. Next, we can use the Chebyshev bound to find an upper bound for the measure of this set. We have already seen that the Fourier transform on this set is large ($|\hat{f}(\alpha)| \geq \frac{\eta N}{2}$), and we know a bound on the L^p norm of the Fourier transform. Consequently, by the Chebyshev bound,

$$\mu\left(\bigcup_{1\leq i\leq r} I_i\right)\left(\frac{\eta N}{2}\right)^p \leq \int_0^1 |\hat{f}(t)|^p dt \leq N^{p-1}.$$

Also, because the I_i are disjoint, the measure of the union of the I_i is actually the sum of the measure of each I_i. That is,

$$\mu\left(\bigcup_{1\leq i\leq r} I_i\right) = \sum_{1\leq i\leq r} \mu(I_i) = \frac{\delta r}{2}.$$

Hence,

$$\frac{\delta r}{2} \leq c_2 N^{p-1}.$$

This then implies that

$$\frac{\eta r}{\pi c_1 N}\frac{\eta^p N^p}{2^p} \leq c_2 N^{p-1} \Rightarrow r \leq \frac{2^p \pi c_1 c_2}{\eta^{p+1}} = \frac{c_4(p)}{\eta^{p+1}}.$$

Now, we need to check that

$$B\left(\{\alpha_1, \ldots, \alpha_r\}, \frac{\epsilon}{2}\right) \subseteq B(S, \epsilon).$$

If we take $n \in B\left(\{\alpha_1, \ldots, \alpha_r\}, \frac{\epsilon}{2}\right)$, and $\alpha \in S$. Then, $n \leq \frac{\epsilon N}{2}$, and there is an α_i such that $\|\alpha - \alpha_i\| \leq \delta$. Hence,

$$\begin{aligned}\|n\alpha\| &\leq \|n\alpha_i\| + \|n(\alpha - \alpha_i)\| \\ &\leq \frac{\epsilon}{2} + n\|(\alpha - \alpha_i)\| \\ &\leq \frac{\epsilon}{2} + n\delta.\end{aligned}$$

Thus we get that,

$$\|n\alpha\| \leq \frac{\epsilon}{2} + \frac{\epsilon N}{2}\frac{\eta}{4\pi c_1 N} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

To obtain a lower bound for $|B\left(\{\alpha_1,\dots,\alpha_r\},\frac{\epsilon}{2}\right)|$ we split $\mathbb{T}^r$ into half-cubes (with half-cube meaning not going from $-T$ to T but 0 to T) of side length $\frac{1}{T}$ where $T=\left\lfloor\frac{2}{\epsilon}\right\rfloor+1\geq 1$. Now, by the pigeonhole principle we see that there are at least $\frac{\epsilon N}{2T^r}$ integers in the interval $[0,\frac{\epsilon N}{2}]$ such that the points $(n\alpha_1,\dots,n\alpha_r)$ all belong to the same half-cube. It then follows that there are at least $\frac{\epsilon N}{2T^r}$ integers $k\in\left[-\frac{\epsilon N}{2},\frac{\epsilon N}{2}\right]$ such that for all $1\leq i\leq r$,

$$\|k\alpha_i\|\leq\frac{1}{T}.$$

Finally, we get that

$$\begin{aligned}\frac{\epsilon N}{2T^r}&\geq\frac{N}{T^{r+1}}\\&\geq\left(\frac{\epsilon}{4}\right)^{r+1}N\\&\geq\left(\frac{\epsilon}{4}\right)^{\frac{c_4(p)+1}{\eta^{p+1}}}N.\end{aligned}$$

Then, if we set $c_3(p)=c_4(p)+1$ the result follows. $\square$

Lemma 1. *We have that*

$$\begin{aligned}\Lambda(f)&=\sum_{\substack{x+y=2z\\(x,y,z)\in\mathbb{Z}^3}}f(x)f(y)f(z)\\&=\int_0^1\hat{f}(t)^2\hat{f}(-2t)dt.\end{aligned}$$

Proof. By definition we have that $\hat{f}(t)=\sum_{n\in\mathbb{Z}}f(n)e^{-2\pi int}$. Therefore we see that

$$\begin{aligned}\hat{f}(t)^2\hat{f}(-2t)&=\sum_{n\in\mathbb{Z}}f(n)e^{-2\pi int}\sum_{m\in\mathbb{Z}}f(m)e^{-2\pi imt}\sum_{k\in\mathbb{Z}}f(k)e^{2\pi in(2k)t}\\&=\sum_{m,n,k}f(n)f(m)f(k)e^{-2\pi i(n+m-2k)t}\end{aligned}$$

Then, passing to the integral, we get that

$$\int_0^1 \hat{f}(t)^2\hat{f}(-2t) = \sum_{m,n,k} f(n)f(m)f(k)\int_0^1 e^{-2\pi i(n+m-2k)t}dt.$$

Since $\int_0^1 e^{-2\pi i(n+m-2k)t}dt = 1$ only when $n+m-2k=0$, and it will be 0 otherwise. $\square$

2.1 Proof of Weak Roth

We are now in a position to prove Theorem 8 the weak version of Roth's theorem that is $\lim_{N\to\infty}\frac{r_3(N)}{N} = 0$.

Proof of Theorem 8. Let N be a sufficiently large integer, and let $A \subseteq [1,N]$ be a maximal subset not containing 3-A.P's (that is T(A) = 0). So, in particular $r_3(N) = |A|$ by definition. For the rest of this proof let us set $f = 1_A$.

Then, by Lemma 1 we have that $\Lambda(f) = \int_0^1 \hat{f}(t)^2\hat{f}(-2t)dt$. We wish to find a finitely supported function $g : \mathbb{Z} \to \mathbb{R}^+$ such that:

1. $\Lambda(f)$ is "close to" $\Lambda(g)$ (we define *close to* in Lemma 2),

2. $|\text{Supp}(g)| \geq \theta|A|$ where $\theta > 1$.

In view of Lemma 1 we can replace 1. by

3. $\hat{f}$ is "close to" $\hat{g}$.

We shall go into what "close to" means later on. Now, when looking for a function which in Fourier transform is close to a given function a convenient choice is to take the given function and convolve it with another function.

So we will look for $g = f * h$. The condition $\hat{f}$ is "close to" $\hat{g}$ forces upon us that $\hat{f}(0) \approx \hat{g}(0)$, and in fact we would like to find g such that $\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} g(n)$. in order to do this, the simplest choice for h will be the normalised characteristic function of some finite set, say $h = \frac{1_B}{B}$.

Now, the question is how do we choose B so that the support of g is much larger than the support of f. Since $\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} g(n)$, and f takes values up to 1, we want $g(n) < 1$.

Now, how do we choose B? Using the definitions, we find that

$$g(n) = \frac{1}{|B|} \sum_{m\in\mathbb{Z}} f(n-m) 1_B(m).$$

Since we want $g(n) < 1$, and $f = 1_A$, we need that no translate of B is contained in A. Therefore, we will choose B to be a 3-A.P since any translation of a 3-A.P is a 3-A.P. So, we set $B = \{0, b, 2b\}$ where b will be chosen later. Thus,

$$g(n) = \frac{1}{3}(f(n) + f(n-b) + f(n-2b)).$$

Hence, $g(n) = \frac{1}{3}$ or $\frac{2}{3}$ when $n \in \mathrm{Supp}(g)$. Now, we see from this that

$$\frac{2}{3}|\mathrm{Supp}(g)| \geq \sum_{n\in\mathbb{Z}} g(n) = \sum_{n\in\mathbb{Z}} g(n) = |A|.$$

Therefore we have what we desired, with $|\mathrm{Supp}(g)| \geq \frac{3}{2}|A|$.

Before moving forward with our proof, we need to say what we have meant by "close to". This involves a lemma of Browning and Prendiville called Configuration Control:

Lemma 2 (Configuration Control,[3]). *Suppose that for some p satisfying $2 \leq p < 3$ we have $\|\hat{f}\|_p << N^{1-\frac{1}{p}}$ and $\|\hat{g}\|_p << N^{1-\frac{1}{p}}$. Then,*

$$|\Lambda(f) - \Lambda(g)| \leq (\frac{\|\hat{f} - \hat{g}\|_\infty}{N})^{3-p} N^2.$$

For our proof of Roth's theorem, we only need to look at the case when $p = 2$.

Now, going back to our proof, how can we fix b such that $\|\hat{f} - \hat{g}\|$ is small? So we have that $\hat{g}(t) = \hat{f}(t)(\frac{1}{3}(1 + e(bt) + e(2bt)))$. Hence,

$$\begin{aligned}
|\hat{g}(t) - \hat{f}(t)| &= |\hat{f}(t)||\frac{(1 - e(bt)) + (1 - e(2bt))}{3}| \\
&\leq |\hat{f}(t)||\frac{2\pi\|bt\| + 2\pi\|2bt\|}{3}| \\
&\leq |\hat{f}(t)|2\pi\|bt\|.
\end{aligned}$$

We also know that $|\hat{f}(t)| \leq N$ and $|\hat{g}(t)| \leq |\hat{f}(t)| \leq N$.

Now, let η be such that $0 \leq \eta < 1$, and let $S = \{t : |\hat{f}(t)| \geq \eta N\}$. Suppose there exists a b, with $|b| \leq \epsilon N$ and $\|bt\| \leq \epsilon$ for all $t \in S$. Then we are done.

Now, take $\epsilon = \frac{4}{\log N}$. By Theorem 10 and the fact that for Roth's theorem we need $p = 2$, we have that

$$\begin{aligned}
|B(S, \epsilon)| &\geq \left(\frac{\epsilon}{4}\right)^{\frac{c}{\eta^3}} N \\
&= \left(\frac{1}{\log N}\right)^{\frac{c}{\eta^3}} N.
\end{aligned}$$

Now we want to take η such that $|B(S, \epsilon)| \geq 1$. So by taking $\eta = \left(\frac{2c \log\log N}{\log N}\right)^{\frac{1}{3}}$, then we have,

$$|B(S, \epsilon)| \geq \left(\frac{1}{\log N}\right)^{\frac{c}{\eta^3}} N \geq N^{\frac{1}{2}} N \geq N^{\frac{1}{2}} \geq 1.$$

Next, we have two cases, $t \in S$ or $t \notin S$. If $t \in S$, then $\|\hat{f}(t)\| \leq N$ and $\|bt\| < \epsilon$. Thus we have,

$$\|\hat{f} - \hat{g}\| \leq \|\hat{f}(t)\| \|bt\| \leq \epsilon N \leq \eta N.$$

In the other case, we have that $t \notin S$, then $\|\hat{f}(t)\| \leq \eta N$ and $\|bt\| \leq \frac{1}{2}$. So once again we get,

$$\|\hat{f} - \hat{g}\| \leq \|\hat{f}(t)\| \|bt\| \leq \frac{1}{2}\eta N \leq \eta N.$$

Now by Configuration Control, $|\Lambda(f) - \Lambda(g)| \leq \left(\frac{\|\hat{f}-\hat{g}\|_\infty}{N}\right) N^2 \leq \eta N^2$. There is however a small problem that must be looked in to, and that is that the support of g is not contained necessarily in $[1, N]$, but in $[-2\epsilon N, N+2\epsilon N]$. Now, instead of taking just the support of g, we take $S = \mathrm{Supp}(g) \cap [1, N]$. Therefore we have

$$|S| = |\mathrm{Supp}(g) \cap [1, N]| \geq |\mathrm{Supp}(g)| - 4\epsilon N.$$

Thus, $|S| \geq \frac{3}{2}|A| - 4\epsilon N$ and we would like this to be bigger than say $\frac{4}{3}|A|$. For this, we need $|A| \geq \frac{96N}{\log N}$.

Remark 2. *If $|A| < \frac{96N}{\log N}$ then we are done because we have proved that a set of maximal cardinality with no 3-A.P's is small.*

All we need to do now is apply Varnavides theorem to S.

So, we have that $\Lambda(S) \geq T(S)$, and $T(S) \neq 0$. Thus, by Varnavides,

$$\frac{|S|}{N} \leq \frac{r_3(M)+1}{M} + \frac{4M^4}{N^2}(\Lambda(A) - \Lambda(S)) + \frac{4\lambda(A)}{N^2}M^4.$$

Since $|S| \geq \frac{4}{3}|A|$, and $\Lambda(A) - \Lambda(S) \leq \eta N^2$, we get

$$\frac{4}{3}\frac{r_3(N)}{N} \leq \frac{r_3(M)+1}{M} + 4\eta M^4 + \frac{4\lambda(A)}{N^2}M^4.$$

We want to choose M such that every term on the right hand side goes to 0 as $N \to \infty$ other than than $\frac{r_3(M)+1}{M}$. So choose $M = \left(\frac{\log N}{2c \log\log N}\right)^{\frac{1}{13}}$.

With this choice, we have that as $N \to \infty$,

$$\frac{4M^4}{N^2}(\Lambda(A) - \Lambda(S)) \to 4M^4\eta \to 0.$$

and,

$$\frac{4\lambda(A)}{N^2}M^4 \to \frac{4M^4}{N} \to 0,$$

as required.

So we are left with

$$\frac{4}{3}\frac{r_3(N)}{N} \leq \frac{r_3(M)+1}{M}.$$

If we let $\lim_{M\to\infty} \frac{r_3(M)+1}{M} = \alpha$ we find that $\frac{4}{3}\alpha \leq \alpha$ and therfore $\alpha = 0$. Thus, we have proved the weak version of Roth's theorem. □

Chapter 3

Arithmetic Progressions and Szemerédi's Theorem

This chapter discusses more detail of counting arithmetic progressions and Szemerédi's theorem. We will also introduce the notion of pseudorandomness.

3.1 Counting arithmetic progressions

When looking for the existence of a structure in the natural numbers it can be more simple to look at the seemingly harder problem of counting how many such structures we expect to find in every finite subset of the natural numbers. Showing that this count is non-zero is then sufficient to solve our problem.

Another useful tool when looking for arithmetic progressions within subsets of the natural numbers is to pass from the subset $A \subseteq \mathbb{N}$ to its characteristic function $1_A : \mathbb{N} \to \{0, 1\}$. At first this may not seem to be very useful, but the following observation shows why this helps us count these progressions:

$$1_A(a)1_A(a+d)\dots 1_A(a+(k-1)d) = \begin{cases} 1, & a, a+d, \dots, a+(k-1)d \in A; \\ 0, & \text{otherwise.} \end{cases}$$

Therefore we can count all k-A.P's within $A \cap \mathbb{Z}_N$ by the sum

$$\sum_{a,d \in \mathbb{Z}_N} 1_A(a)1_A(a+d)\dots 1_A(a+(k-1)d), \tag{3.1}$$

and thus the existence of arithmetic progressions boils down to 3.1 being nonzero.

You will note that we have changed to considering arithmetic progressions in $\mathbb{Z}_N$ rather than in $[1, N]$. We do this to get around the issue that $a, d \in [1, N]$ does not mean that $a + (k-1)d \in [1, N]$. Now we have avoided one important problem, however we have added another, that being the wrap around problem. Working in $\mathbb{Z}_N$ allows APs to wrap around zero so for example, when $N = 7$ we would have $\{1, 6, 4\}$ as a 3-AP in $\mathbb{Z}_N$, but clearly not in $[1, N]$. This is encountered if and only if for some $1 \le i \le k-1$, the term $a + id$ is greater than N. We can get around this issue by simply embedding $[1, N]$ inside a slightly larger cyclic group, so no k-A.P's wrap around 0.

All of this helps motivate the following expected value notation.The definition was used throughout both the original paper [13] and the exposition [4] however this symbolic representation, I think, makes the rest of the paper easier to read.

Definition 8. *Let $f : \mathbb{Z}_N \to \mathbb{R}^+$ be a non-negative function then define the*

normalised count of k-AP's in f *as:*

$$\begin{aligned}\Upsilon_k(f) :&= \mathbb{E}_{x,d\in\mathbb{Z}_N}[f(x)f(x+d)\dots f(x+(k-1)d)]\\ &= \frac{1}{N^2}\sum_{x,d\in\mathbb{Z}_N} f(x)f(x+d)\dots f(x+(k-1)d).\end{aligned}$$

Unfortunately, this definition also counts the trivial k-AP's where $d = 0$. This turns out not to be an issue for the proof as we are only taking estimates up to $o(1)$ errors and the trivial cases only add at most $\frac{1}{N}$ to Υ_k.

3.2 Relative Szemerédi Theorem

By changing from looking at AP's in $[1, N]$ to AP's in $\mathbb{Z}_N$ we may rewrite Szemerédi's Theorem as follows:

Theorem 11. *Let $\alpha > 0$ and $k \geq 3$ an integer. Then there exists a constant $c = c(\alpha, k) > 0$ such that every $f : \mathbb{Z}_N \to [0, 1]$ with density $\mathbb{E}(f) \geq \alpha$ satisfies*

$$\Upsilon_k(f) \geq c - o_{k,\alpha}(1).$$

It would be very useful if this theorem was enough to solve our problem, however this would require the set we are looking at to be dense within the natural numbers. If we look at the characteristic function of the primes $1_{\mathbb{P}}$, then by the prime number theorem we get that

$$\mathbb{E}(1_{\mathbb{P}}) \sim \frac{1}{\log N} \to 0 \text{ as } N \to \infty.$$

So an idea that would solve the issue of density would be to look not at the characteristic function $1_{\mathbb{P}}$, but to look at a weighted version of it, that being $f := \log N \cdot 1_{\mathbb{P}}$. Now we have satisfied the density hypothesis of Theorem 11, however the function f is no longer a $[0, 1]$ valued function. This leads to what is referred to as the Relative Szemerédi theorem which is the crux of the proof of the Green-Tao theorem. The Relative Szemerédi theorem weakens the hypotheses in Theorem 11 so that now our function f need only be bounded above by some pseudorandom function (a concept defined in the next section) and not 1.

Theorem 12 (Relative Szemerédi Theorem)**.** *Let $k \geq 3$ and $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ satisfy the k-linear forms condition. Now let $f : \mathbb{Z}_N \to \mathbb{R}$ be such that $0 \leq f \leq \nu$, then if there exists an $\alpha > 0$ such that the density $\mathbb{E}f \geq \alpha$, then there exists a constant $c' = c'(k, \alpha) > 0$ such that,*

$$\Upsilon_k(f) \geq c' - o_{k,\alpha}(1).$$

This theorem will be proven in Section 3.6.

3.3 Pseudorandomness

As promised in the previous section, we shall now define the notion of pseudorandomness which is crucial in the proof of the Relative Szemerédi Theorem and thus the whole of the Green-Tao theorem. In the original proof of Green and Tao [13] they used two pseudorandomness conditions called the **linear forms condition** and the **correlation condition**. Now these conditions were good enough to prove the theorem, however many mathematicians found them, particularly the correlation condition, unnatural. As such it has been a question since the original paper as to whether there were weaker and more natural hypotheses under which the Relative Szemerédi Theorem still holds. It was shown in [4] by Conlon, Fox, and Zhao that there are weaker conditions that suffice, and in fact just a weak linear forms condition is enough and we can remove the correlation condition entirely.

We have seen in Chapter 2 that proving the 3-term case of Szemerédi theorem (Roth's theorem) is not too difficult. However, proving the 4-term case is much more difficult, and the k-term case even more so. One of the brilliant pieces of insights from Green and Tao was this transference of Szemerédi's theorem to a sparse setting which unlike the full Szemerédi theorem has no more mathematical complexity in proving the k-AP case as the 3-AP case (of course the notation becomes significantly more cumbersome).

So we shall start by looking at the 3-term case and the notation used which shall help us understand the k-term case of the full Relative Szemerédi theorem.

Definition 9 (3-linear forms condition)**.** *Take a function $\nu : \mathbb{Z}_N \to \mathbb{R}^+$. We say ν satisfies the 3-linear forms condition if*

$$\begin{aligned}\mathbb{E}_{x,x',y,y',z,z'\in\mathbb{Z}_N}[&\nu(2x+y)\nu(2x'+y)\nu(2x+y')\nu(2x'+y')\nu(x-z)\nu(x'-z)\nu(x-z')\\&\cdot\nu(x'-z')\nu(-y-2z)\nu(-y'-2z)\nu(-y-2z')\nu(-y'-2z')] = 1+o(1)\end{aligned}$$

and also the above still holds when any of the twelve factors in the expectation are removed.

Theorem 13 (Relative Roth)**.** *Let $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ satisfy the 3-linear forms condition. Then for every $\alpha > 0$, there exists a constant $c = c(\alpha) > 0$ such that every function $f : \mathbb{Z}_N \to \mathbb{R}$ with $0 \le f \le \nu$ and $\mathbb{E}f \ge \alpha$ satisfies*

$$\Upsilon_3(f) \ge c - o_\alpha(1).$$

Now we shall look at the k-linear forms condition and see that it is a just an extension of the previous 3-linear forms condition.

Definition 10 (k-linear forms condition)**.** *Take a function $\nu : \mathbb{Z}_N \to \mathbb{R}^+$. We say ν satisfies the k-linear forms condition if*

$$\mathbb{E}_{x_1,x_1',\ldots,x_k,x_k' \in \mathbb{Z}_N} \left[\prod_{j=1}^{k} \prod_{\omega \in 0,1^{[1,k]\setminus j}} \nu \left(\sum_{i=1}^{k} (j-i) x_i^{(\omega_i)} \right)^{n_{j,\omega}} \right] = 1 + o(1)$$

for any choice of exponents $n_{j,\omega} \in \{0,1\}$.

3.4 Dense Model Theorem

The proof of the relative Szemerédi theorem comes down to two key parts now, this so called Dense Model Theorem, and then a counting lemma. In this section we are looking to come up with a Dense Model Theorem that lets us approximate our function f by another function $\tilde{f} : \mathbb{Z}_N \to [0,1]$ with respect to a cut norm.

Definition 11. *For any $r \in \mathbb{Z}^+$ and any function $f : \mathbb{Z}_N \to \mathbb{R}$ we define the* ***cut norm*** *of f by*

$$\|f\|_{\square,r} := sup|\mathbb{E}_{x_1,\ldots,x_r \in \mathbb{Z}_N} \left[f(x_1 + \ldots + x_r) \prod_{j=1}^{r} 1_{A_j}(x_{-j}) \right] |.$$

Where $x_{-j} := (x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_r) \in \mathbb{Z}_N^{r-1}$, and the supremum is taken over all $A_1, \ldots, A_r \subseteq \mathbb{Z}_N^{r-1}$.

Now we can make a similar definition for weighted hypergraphs. If we have a weighted hypergraph G on vertex set $X_1 \bigcup \cdots \bigcup X_r$ then we can define *the cut norm for G as*

$$\|G\|_{\square} := \sup |\mathbb{E}_{x_1 \in X_1, \ldots x_r \in X_r}[G(x_1, \ldots, x_r) 1_{A_1}(x_1) \cdots 1_{A_r}(x_r)]|.$$

Here we are taking the supremum over all subsets $A_i \subseteq X_{-i}$.

Now we shall state the dense model theorem which is taken mainly from the work of Zhao [22].

Theorem 14 (Dense Model)**.** *There exists an absolute constant C such that the following holds. Let $\epsilon > 0$, and suppose that $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ satisfies $\|\nu - 1\|_{\square,k-1} \leq \epsilon$. Then for every $f : \mathbb{Z}_N \to \mathbb{R}^+$ such that $0 \leq f \leq \nu$, there exists a function $\tilde{f} : \mathbb{Z}_N \to [0,1]$ such that $\|f - \tilde{f}\|_{\square,k-1} \leq \log^{-1/C}\left(\frac{1}{\epsilon}\right)$.*

In the proof of the dense model theorem it is important to use dual norms, and as such we shall define the general form of the dual norm.

Definition 12. *Suppose we have a norm $\|\cdot\|$ defined on $\mathbb{R}^N$, then we can define its* ***dual norm*** *as:*

$$\|f\|^* := sup\{|\langle f, g\rangle : \|g\| \leq 1\}.$$

Now, for the purposes of proving the Green-Tao theorem it shall be convenient to write $\mathbb{E}_{x,y}[f(x+y)1_A(x)1_B(y)]$ (which is what is within the supremum of $\|f\|_{\square,2}$) as $\langle f, \varphi\rangle = \mathbb{E}_x[f(x)\varphi(x)]$ for some $\varphi : \mathbb{Z}_N \to \mathbb{R}$. Now, if we define the convolution $f_1 * f_2(z) := \mathbb{E}_x[f_1(x)f_2(z-x)]$, then by a simple change of variables we can write,

$$\mathbb{E}_{x,y}[f(x+y)1_A(x)1_B(y)] = \mathbb{E}_{x,z}[f(z)1_A(x)1_B(z-x)] = \langle f, 1_A * 1_B\rangle.$$

What Conlon, Fox, and Zhao then did in [4] was find a nice way to look at the cut norm using convolutions which not only simplifies notation, but also

also allows the use of the Hahn-Banach theorem (which shall be introduced shortly after). If we let Φ_2 be the set of all convex convolutions $1_A * 1_B$ with $A, B \subseteq \mathbb{Z}_N$, then we may write

$$\|f\|_{\square,2} = \sup_{A,B\subseteq\mathbb{Z}_N} |\langle f, 1_A * 1_B\rangle| = \sup_{\varphi\in\Phi_2} |\langle f,\varphi\rangle|.$$

This convolution can then very easily be generalised for any given r,

Definition 13. *Take r functions $f_1, \ldots, f_r$ where each $f_i : \mathbb{Z}_N^{r-1} \to \mathbb{R}$, the* ***generalised convolution*** *$(f_1, \ldots, f_r)^* : \mathbb{Z}_N \to \mathbb{R}$ is defined as,*

$$(f_1, \ldots, f_r)^*(x) = \mathbb{E}_{\substack{y_1,\ldots,y_r\in\mathbb{Z}_N \\ y_1+\ldots+y_r=x}} [f_1(y_{-1})f_2(y_{-2})\ldots f_r(y_{-r})].$$

Then, if we take Φ_r to be the set of all generalized convolutions $(1_{A_1}, \ldots, 1_{A_r})^*$ with $A_1, \ldots, A_r \subseteq \mathbb{Z}_N^{r-1}$ we can write our general cut norm $\|f\|_{\square,r}$ in exactly the same manner to the case of $r = 2$ as,

$$\|f\|_{\square,r} = \sup_{A_1,\ldots,A_r\subseteq\mathbb{Z}_N^{r-1}} |\langle f, (1_{A_1}, \ldots, 1_{A_r})^*\rangle| = \sup_{\varphi\in\Phi_r} |\langle f,\varphi\rangle|.$$

Theorem 15 (Hahn-Banach theorem [12])**.** *Let K_1 and K_2 be closed convex subsets subsets of $\mathbb{R}^N$, with $0 \in K_1, K_2$, and suppose that $f \in \mathbb{R}^N$ cannot be written as a convex combination $c_1f_1 + c_2f_2$ with $f_1 \in K_1$ and $f_2 \in K_2$. Then there exists $\varphi \in \mathbb{R}^N$ such that $|\langle f,\varphi\rangle| > 1$ and $|\langle g,\varphi\rangle| \leq 1$ for every $g \in K_1 \bigcup K_2$.*

Now this is not the normal form of the Hahn-Banach theorem, however this formulation of it shall prove very useful.

We now look at the some crucial properties of Φ_r in the next lemma.

Lemma 3. *The set Φ_r has the following properties:*

$$1)\Phi_r \text{ is closed under multiplication;}$$
$$2)\Phi_r \text{ is the unit ball of } \|\cdot\|^*_{\square,r}.$$

Proof. The proof of 1) can be found as Lemma 5.2 in [4].
Now for the proof of 2) we first want to show that Φ_r is contained in the unit ball. We can see easily that each element of $\Phi_r \bigcup(-\Phi_r)$ is in the unit ball, then by application of the triangle inequality the same can be gathered for convex combinations. Now for the other direction, let us suppose that ψ is in the unit ball but not in $\Phi_r \bigcup(-\Phi_r)$. Then, by the Hahn-Banach theorem, there exists an f such that $|\langle f, \varphi\rangle| \leq 1$ for all $\varphi \in \Phi_r \bigcup(-\Phi_r)$ and $|\langle f, \psi\rangle| > 1$. However, the first of these inequalities implies that $\|f\|_{\square,r} \leq 1$, and thus the second inequality implies $\|\psi\|^*_{\square,r} > 1$ (we are allowed to do this since the definitions tell us that $|\langle f, \varphi\rangle| \leq \|f\|\|\varphi\|^*$), contradicting the assumption that ψ is in the unit ball. $\square$

So from this lemma we get that the unit ball of the dual norm is closed under multiplication, and therefore for every $\varphi, \psi : \mathbb{Z}_N \to \mathbb{R}$, we have that

$$\|\varphi\psi\|^*_{\square,r} \leq \|\varphi\|^*_{\square,r}\|\psi\|^*_{\square,r}.$$

Proof of Dense Model Theorem. For this proof we fix r and as such write $\|\cdot\|$ for $\|\cdot\|_{\square,r}$. We may assume that $\epsilon \leq \frac{1}{10}$. Now in order to prove the theorem, we need to show that there exists a $\tilde{f} : \mathbb{Z}_N \to \left[0, 1 + \frac{\epsilon}{2}\right]$ with $\|f - \tilde{f}\| \leq \frac{\epsilon}{2}$. Suppose, for contradiction, that there is no such function $\tilde{f}$. Next, let

$$K_1 := \{\tilde{f} : \mathbb{Z}_N \to \left[0, 1 + \frac{\epsilon}{2}\right]\}$$
$$K_2 := \{h : \mathbb{Z}_N \to \mathbb{R} | \|h\| \leq \frac{\epsilon}{2}\}.$$

It is seen that both K_1 and K_2 are closed convex sets in $\mathbb{R}^N$. By our assumption we have $f \notin K_1 + K_2 := \{\tilde{f} + h : \tilde{f} \in K_1, h \in K_2\}$. Since K_1 and K_2 are closed convex sets, we have that $K_1 + K_2$ is a convex set, and then by the Hahn-Banach theorem tells us that exists a function $\varphi : \mathbb{Z}_N \to \mathbb{R}$ such that:

$$(i)|\langle f, \varphi\rangle| > 1; \qquad (ii)|\langle g, \varphi\rangle| \leq 1 \text{ for every } g \in K_1 + K_2.$$

It is seen that $0 \in K_1, K_2$, and therefore $K_1, K_2 \subset K_1 + K_2$. Now if we take $g = (1 + \frac{\epsilon}{2})1_{\varphi>0} \in K_1$, then by (ii) we get that $\langle 1, \varphi_+\rangle \leq (1 + \frac{\epsilon}{2})^{-1}$. We define $\varphi_+ := \max\{0, \varphi\}$. Now ranging our function g over K_2, we get that $\|\varphi\|_\infty \leq \|\varphi\|^* \leq \frac{2}{\epsilon}$. This is easily verified as if $|\langle g, \varphi\rangle| \leq 1$ for all g with $|\|g\|| \leq \frac{\epsilon}{2}$, then $|\langle g, \varphi\rangle| \leq \frac{2}{\epsilon}$ for all g with $\|g\| \leq 1$.

By the Weierstrass approximation theorem (1885)[21], there exists a polynomial Q such that $|Q(x) - x_+| \leq \frac{\epsilon}{8}$ for all $x \in \left[-\frac{2}{\epsilon}, \frac{2}{\epsilon}\right]$. Define our polynomial Q as $Q(x) = \sum_{i=0}^{d} q_i x^i$ and take $R = \sum_{i=0}^{d} |q_i| \left(\frac{2}{\epsilon}\right)^i$.

Now, take $Q\varphi$ to be the function defined as $Q\varphi(x) = Q(\varphi(x))$. Then by the fact that $\|\varphi\|^* \leq \frac{2}{\epsilon}$ and use of the triangle inequality we get

$$\|Q\varphi\|^* \leq \sum_{i=0}^{d} |q_i| \|\varphi^i\|^* \leq \sum_{i=0}^{d} |q_i| (\|\varphi\|^*)^i \leq \sum_{i=0}^{d} |q_i| \left(\frac{2}{\epsilon}\right)^i = R.$$

We are making the assumption that $\|\nu - 1\| \leq \epsilon'$, and therefore we have

$$|\langle \nu - 1, Q\varphi\rangle| \leq \|\nu - 1\| \|Q\varphi\|^* \leq \epsilon' R.$$

Now, since we have $\|\varphi\|_\infty \leq \frac{2}{\epsilon}$, we get that $\|Q\varphi - \varphi_+\|_\infty \leq \frac{\epsilon}{8}$. Therefore,

$$|\langle \nu, Q\varphi\rangle| \leq |\langle 1, Q\varphi\rangle| + \epsilon' R \leq |\langle 1, \varphi_+\rangle| + \frac{\epsilon}{8} + \epsilon' R \leq (1 + \frac{\epsilon}{2})^{-1} + \frac{\epsilon}{8} + \epsilon' R.$$

Now, using the fact that $|\langle \nu - 1, 1\rangle| \leq \|\nu - 1\| \|1\|^*$, and $\|1\|^* = 1$, we get that $\|\nu\|_1 = |\langle \nu, 1\rangle| \leq \|\nu - 1\| + 1 \leq 1 + \epsilon'$. Therefore,

$$|\langle f, \varphi\rangle| \leq |\langle f, \varphi_+\rangle| \leq |\langle \nu, \varphi_+\rangle| \leq |\langle \nu, Q\varphi\rangle| + \|\nu\|_1 \|Q\varphi - \varphi_+\|_\infty \leq \left(1 + \frac{\epsilon}{2}\right)^{-1} + \frac{\epsilon}{8} + \epsilon' R + (1 + \epsilon')\frac{\epsilon}{8}.$$

Since we took $\epsilon \leq \frac{1}{10}$, the final term in our inequality can be no bigger than 1 when we take ϵ' small enough (take $\epsilon' = \frac{\epsilon}{8R}$), but this is a contradiction with (ii) from above. $\square$

3.5 Counting Lemma

The second section involved in proving the Relative Szemerédi theorem is the counting lemma which roughly states that the count of APs is preserved under the conditions of the dense model theorem from the previous section. We will look at this in terms of graphs, and in the case of graphs the counting lemma roughly states that if we take two weighted graphs that are "close" in cut norm then their triangle densities are similar.

Now let us consider a weighted tripartite graph G with vertex set $V_G = X \bigcup Y \bigcup Z$, with X, Y, Z all finite. There are three functions that define this graph

$$G_{XY} : X \times Y \to \mathbb{R},$$
$$G_{XZ} : X \times Z \to \mathbb{R},$$
$$G_{YZ} : Y \times Z \to \mathbb{R}.$$

Now we can write $\|G\|_\square = \max\{\|G_{XY}\|_\square, \|G_{XZ}\|_\square, \|G_{YZ}\|_\square\}$. We shall not write the subscripts where they are obvious from context.

So now we have done some of the set up we can start looking at the dense setting of our problem.

Lemma 4 (Triangle counting lemma, dense case)**.** *Let G and $\tilde{G}$ be two weighted tripartite graphs on the vertex set $X \bigcup Y \bigcup Z$ with weights in $[0, 1]$.*

If $\|G-\tilde{G}\|_\square \le \epsilon$, *then*

$$|\mathbb{E}_{x\in X,y\in Y,z\in Z}[G(x,y)G(x,z)G(y,z)-\tilde{G}(x,y)\tilde{G}(x,z)\tilde{G}(y,z)]|\le 3\epsilon.$$

Proof. Now if we take $f: X\to[0,1]$ and $g: Y\to[0,1]$ then by the definition of the cut norm we get that for every function f and g

$$|\mathbb{E}_{x\in X,y\in Y}[(G(x,y)-\tilde{G}(x,y))f(x)g(y)]|\le\epsilon.$$

Now if we were to fix a value for z then the following follows from above

$$|\mathbb{E}_{x\in X,y\in Y,z\in Z}[G(x,y)G(x,z)G(y,z)-\tilde{G}(x,y)G(x,z)G(y,z)]|\le\epsilon.$$

In a similar fashion we can see that the next two inequalities follow

$$|\mathbb{E}_{x\in X,y\in Y,z\in Z}[\tilde{G}(x,y)G(x,z)G(y,z)-\tilde{G}(x,y)\tilde{G}(x,z)G(y,z)]|\le\epsilon,$$

and

$$|\mathbb{E}_{x\in X,y\in Y,z\in Z}[\tilde{G}(x,y)\tilde{G}(x,z)G(y,z)-\tilde{G}(x,y)\tilde{G}(x,z)\tilde{G}(y,z)]|\le\epsilon.$$

Then by the triangle inequality the result follows. □

Unfortunately this is not enough to work in the case that G is unbounded, due to the fact we required that the functions f and g be bounded. Now, we need to find a triangle counting lemma in the sparse setting i.e when G is unbounded. For simplicity we are starting with the triangle counting case, towards the end of this section we will improve this to a simplex counting lemma on hypergraphs.

Theorem 16 (Triangle counting lemma, sparse case)**.** *Let* $\nu, G, \tilde{G}$ *be three weighted tripartite graphs with vertex set* $X\bigcup Y\bigcup Z$. *Now assume that* ν

satisfies the 3-linear forms condition, $0 \leq G \leq \nu$, *and* $0 \leq \tilde{G} \leq 1$. *Then, if* $\|G - \tilde{G}\|_{\square} = o(1)$, *then*

$$|\mathbb{E}_{x\in X,y\in Y,z\in Z}[G(x,y)G(x,z)G(y,z)]| = |\mathbb{E}_{x\in X,y\in Y,z\in Z}[\tilde{G}(x,y)\tilde{G}(x,z)\tilde{G}(y,z)]|+o(1).$$

The proof of this theorem requires a few classic techniques in the area such as repeated use of the Cauchy-Schwarz inequality, but also a new idea from Conlon, Fox, and Zhao which they called densification. The basic idea of densification is that after multiple applications of the Cauchy-Schwarz inequality it becomes useful to start looking at 4-cycles and their density given by $\mathbb{E}_{x\in X,y\in Y,z,z'\in Z}[G(x,z)G(x,z')G(y,z)G(y,z')]$. In order to start analysing these 4-cycles we shall bring in a new weighted graph $G' : X \times Y \to \mathbb{R}^+$, which is defined by $G'(x,y) := \mathbb{E}_{z'\in Z}[G'(x,z')G'(y,z')]$. This is useful to our problem as our 4-cycle density is now given by $\mathbb{E}_{x\in X,y\in Y,z\in Z}[G'(x,y)G(x,z)G(y,z)]$, and although this may look like just a rehash of what we had earlier, the graph G' in fact acts like a dense graph. So if we can do the same process for replacing the YZ and XZ parts of G with "dense" equivalents then we will get back to the dense case which we have already worked out how to deal with.

Lemma 5. *For any* $\nu : X \times Y \to \mathbb{R}$,

$$\|\nu-1\|_{\square} \leq (\mathbb{E}_{x,x'\in X,y,y'\in Y}[(\nu(x,y)-1)(\nu(x',y)-1)(\nu(x,y')-1)(\nu(x',y')-1)])^{\frac{1}{4}}.$$

Proof. Take $A \subseteq X$ and $B \subseteq Y$, then by repeatedly using the Cauchy-

Schwarz inequality we get

$$
\begin{aligned}
|\mathbb{E}_{x\in X,y\in Y}[(\nu(x,y)-1)1_A(x)1_B(y)]|^4 &\le |\mathbb{E}_{x\in X}[(\mathbb{E}_{y\in Y}[(\nu(x,y)-1)1_B(y)])^2 1_A(x)]|^2 \\
&\le |\mathbb{E}_{x\in X}[(\mathbb{E}_{y\in Y}[(\nu(x,y)-1)1_B(y)])^2]|^2 \\
&= |\mathbb{E}_{x\in X,y,y'\in Y}[(\nu(x,y)-1)(\nu(x,y')-1)1_B(y)1_B(y')]|^2 \\
&\le \mathbb{E}_{y,y'\in Y}[(\mathbb{E}_{x\in X}[(\nu(x,y)-1)(\nu(x,y')-1)])^2 1_B(y)1_B(y')] \\
&\le \mathbb{E}_{y,y'\in Y}[(\mathbb{E}_{x\in X}[(\nu(x,y)-1)(\nu(x,y')-1)])^2]
\end{aligned}
$$

$$
= \mathbb{E}_{x,x'\in X,y,y'\in Y}[(\nu(x,y)-1)(\nu(x',y)-1)(\nu(x,y')-1)(\nu(x',y')-1)].
$$

Clearly, the lemma follows. $\square$

We need one more lemma before we are able to prove the sparse case of the triangle counting lemma. It utilises the 3-linear forms condition from earlier in the paper.

Lemma 6. *Let $\nu, G, \tilde{G}$ be three weighted tripartite graphs with vertex set $X\bigcup Y\bigcup Z$. Now assume that ν satisfies the 3-linear forms condition, $0\le G\le\nu$, and $0\le\tilde{G}\le 1$. Then*

$$
\mathbb{E}_{x\in X,y\in Y,z,z'\in Z}[(\nu(x,y)-1)G(x,z)G(x,z')G(y,z)G(y,z')] = o(1).
$$

This will still be true if any number of the four G factors are replaced by $\tilde{G}$.

Proof. First we will provide a proof when none of the G factors are replaced, and then give an example of the simple modification needed for the other cases. As before we start by using the Cauchy-Schwarz inequality

$$
\begin{aligned}
&|\mathbb{E}_{x\in X,y\in Y,z,z'\in Z}[(\nu(x,y)-1)G(x,z)G(x,z')G(y,z)G(y,z')]|^2 \\
&\le \mathbb{E}_{y\in Y,z,z'\in Z}[(\mathbb{E}_{x\in X}[(\nu(x,y)-1)G(x,z)G(x,z')])^2 G(y,z)G(y,z')]\mathbb{E}_{y\in Y,z,z'\in Z}[G(y,z)G(y,z')] \\
&\le \mathbb{E}_{y\in Y,z,z'\in Z}[(\mathbb{E}_{x\in X}[(\nu(x,y)-1)G(x,z)G(x,z')])^2 \nu(y,z)\nu(y,z')]\mathbb{E}_{y\in Y,z,z'\in Z}[\nu(y,z)\nu(y,z')]
\end{aligned}
$$

By the 3-linear forms condition we see that $\mathbb{E}_{y\in Y,z,z'\in Z}[\nu(y,z)\nu(y,z')] \leq 1 + o(1)$. Therefore all we need to look at now is the first factor, and once again we shall be using the Cauchy-Schwarz inequality

$$
\begin{aligned}
&|\mathbb{E}_{y\in Y,z,z'\in Z}[(\mathbb{E}_{x\in X}[(\nu(x,y)-1)G(x,z)G(x,z')])^2\nu(y,z)\nu(y,z')]|^2 \\
&= |\mathbb{E}_{x,x',y,z,z'}[(\nu(x,y)-1)(\nu(x',y)-1)G(x,z)G(x,z')G(x',z)G(x',z')\nu(y,z)\nu(y,z')]|^2 \\
&= |\mathbb{E}_{x,x',z,z'}[\mathbb{E}_y[(\nu(x,y)-1)(\nu(x',y)-1)\nu(y,z)\nu(y,z')]G(x,z)G(x,z')G(x',z)G(x',z')]|^2 \\
&\leq \mathbb{E}_{x,x',z,z'}[(\mathbb{E}_y[(\nu(x,y)-1)(\nu(x',y)-1)\nu(y,z)\nu(y,z')])^2G(x,z)G(x,z')G(x',z)G(x',z')] \\
&\cdot \mathbb{E}_{x,x',z,z'}[\nu(x,z)\nu(x,z')\nu(x',z)\nu(x',z')] \\
&\leq \mathbb{E}_{x,x',z,z'}[(\mathbb{E}_y[(\nu(x,y)-1)(\nu(x',y)-1)\nu(y,z)\nu(y,z')])^2\nu(x,z)\nu(x,z')\nu(x',z)\nu(x',z')] \\
&\cdot \mathbb{E}_{x,x',z,z'}[\nu(x,z)\nu(x,z')\nu(x',z)\nu(x',z')]
\end{aligned}
$$

Once again, by the 3-linear forms condition we get that the factor

$$\mathbb{E}_{x,x',z,z'}[\nu(x,z)\nu(x,z')\nu(x',z)\nu(x',z')]$$

is at most $1 + o(1)$. Finally, if we are to expand the first factor out we find that all terms are of the form $1 + o(1)$, but the parity means that the 1's cancel out and we are left with the first factor being $o(1)$. Thus the lemma follows. $\square$

Now we have the lemmas we need in order to prove the Triangle counting lemma, sparse case (Theorem 16). This is the last major hurdle in proving the Relative Szemerédi theorem as the main complexity in generalising Theorem 16 is notation.

Proof of Theorem 16. If we have that $\nu = 1$, then we are in the dense case and as such our theorem follows from Lemma 4. So now we shall prove the remaining case via induction on the number of $\nu_{XY}, \nu_{XZ}, \nu_{YZ}$ which are not 1. Since we are able to relabel if needed, we can assume without loss of generality that ν_{XY} is not 1. We can now apply the step introduced by Conlon, Fox, and Zhao known as densification [5]. The key insight in this method is that although our graphs ν and G are potentially unbounded, we can define new weighted graphs ν' and G' which shall behave as dense graphs. So define $\nu', G, G' : X \times Y \to \mathbb{R}^+$ as

$$\nu'(x,y) := \mathbb{E}_{z\in Z}[\nu(x,z)\nu(y,z)],$$
$$G'(x,y) := \mathbb{E}_{z\in Z}[G(x,z)G(y,z)],$$
$$\tilde{G}'(x,y) := \mathbb{E}_{z\in Z}[\tilde{G}(x,z)\tilde{G}(y,z)].$$

Unfortunately, the weights on ν' and G' are not necessarily bounded by 1, however this is not a difficult issue to deal with. We can bound the weights of ν' and G' by defining $\nu'_1 := \min\{\nu', 1\}$ and $G'_1 := \min\{G', 1\}$, and show that this bounding has an effect of order $o(1)$. Now we have

$$\begin{aligned}
&\mathbb{E}_{x\in X, y\in Y, z\in Z}[G(x,y)G(x,z)G(y,z) - \tilde{G}(x,y)\tilde{G}(x,z)\tilde{G}(y,z)] \\
&= \mathbb{E}_{(x,y)\in X\times Y}[GG' - \tilde{G}\tilde{G}'] \\
&= \mathbb{E}_{(x,y)\in X\times Y}[G(G' - \tilde{G}')] + \mathbb{E}_{(x,y)\in X\times Y}[(G - \tilde{G})\tilde{G}'].
\end{aligned}$$

Now we can see that the final term of the above $(\mathbb{E}_{(x,y)\in X\times Y}[(G - \tilde{G})\tilde{G}'])$ is equal to $\mathbb{E}_{(x,y)\in X\times Y}[(G(x,y) - \tilde{G}(x,y))\tilde{G}(x,z)\tilde{G}(y,z)]$. Then using the fact $0 \le \tilde{G} \le 1$ we see that the absolute value of $\mathbb{E}_{(x,y)\in X\times Y}[(G(x,y) - \tilde{G}(x,y))\tilde{G}(x,z)\tilde{G}(y,z)]$ is no greater than $\|G - \tilde{G}\|_\square = o(1)$. Therefore we are left to bound the term $\mathbb{E}_{(x,y)\in X\times Y}[G(G' - \tilde{G}')]$. Unsurprisingly, we shall be using the Cauchy-Schwarz inequality to help do this, we have

$$
\begin{aligned}
(\mathbb{E}_{(x,y)\in X\times Y}[G(G'-\tilde{G}')])^2 &\leq \mathbb{E}_{(x,y)\in X\times Y}[G(G'-\tilde{G}')^2]\mathbb{E}_{(x,y)\in X\times Y}[G] \\
&\leq \mathbb{E}_{(x,y)\in X\times Y}[\nu(G'-\tilde{G}')^2]\mathbb{E}_{(x,y)\in X\times Y}[\nu]
\end{aligned}
$$

$$= \mathbb{E}_{x\in X,y\in Y}[\nu(x,y)(\mathbb{E}_{z\in Z}[G(x,z)G(y,z)-\tilde{G}(x,z)\tilde{G}(y,z)])^2]\mathbb{E}_{x\in X,y\in Y}[\nu(x,y)].$$

Now by the 3-linear forms condition, the factor $\mathbb{E}_{x\in X,y\in Y}[\nu(x,y)] = 1 + o(1)$. Then we get that the first term $\mathbb{E}_{x\in X,y\in Y}[\nu(x,y)(\mathbb{E}_{z\in Z}[G(x,z)G(y,z) - \tilde{G}(x,z)\tilde{G}(y,z)])^2]$ differs from $\mathbb{E}_{x\in X,y\in Y}[(\mathbb{E}_{z\in Z}[G(x,z)G(y,z)-\tilde{G}(x,z)\tilde{G}(y,z)])^2] = \mathbb{E}_{(x,y)\in X\times Y}[(G'-\tilde{G}')^2]$ just by $o(1)$. This can be seen by Lemma 6 if you just expand out the square and then apply Lemma 6 to each term.

The 3-linear forms condition tells us that both $\mathbb{E}[\nu']$ and $\mathbb{E}[\nu'^2]$ are $1+o(1)$. Thus, by an application of the Cauchy-Schwarz inequality, we get

$$(\mathbb{E}[|\nu'-1|)])^2 \leq \mathbb{E}[(\nu'-1)^2] = o(1).$$

So now we would like to show that $\mathbb{E}_{(x,y)\in X\times Y}[(G'-\tilde{G}')^2] = o(1)$. Since $0 \leq G' \leq \nu'$, we get that

$$0 \leq G' - G'_1 = \max\{G'-1, 0\} \leq \max\{\nu'-1, 0\} \leq |\nu'-1|.$$

We also have that

$$\mathbb{E}_{(x,y)\in X\times Y}[(G'-\tilde{G}')^2] = \mathbb{E}[(G'-\tilde{G}')(G'-G'_1)] + \mathbb{E}[(G'-\tilde{G}')(G'_1-\tilde{G}')].$$

Now using the above we have that the first term on the right hand side $(\mathbb{E}[(G'-\tilde{G}')(G'-G'_1)])$ is no greater than

$$\mathbb{E}[(\nu'+1)|\nu'-1|] = \mathbb{E}[(\nu'-1)|\nu'-1|] + 2\mathbb{E}[|\nu'-1|] = o(1).$$

Now we want to check that $\|G'_1 - \tilde{G}'\|_\square = o(1)$.

For any $A \subseteq X, B \subseteq Y$, we get that

$$\begin{aligned}\mathbb{E}_{x\in X,y\in Y}[(G_1' - \tilde{G}')(x,y)1_A(x)1_B(y)] &= \mathbb{E}_{x\in X,y\in Y}[(G_1' - \tilde{G}')1_{A\times B}] \\ &= \mathbb{E}_{x\in X,y\in Y}[(G_1' - G')1_{A\times B}] + \mathbb{E}_{x\in X,y\in Y}[(G' - \tilde{G}')1_{A\times B}]\end{aligned}$$

The first term of the last line can be seen to be at most $\mathbb{E}[|\nu' - 1|] = o(1)$. Now we need to look at the second term. It can be rewritten as $\mathbb{E}_{x\in X,y\in Y,z\in Z}[1_{A\times B}(x,y)G(x,z)G(y,z) - 1_{A\times B}(x,y)\tilde{G}(x,z)\tilde{G}(y'z)]$. Now if we substitute $\nu_{XY}, G_{XY}, \tilde{G}_{XY}$ by $1, 1_{A\times B}, 1_{A\times B}$ respectively this will increase how many of $\nu_{XY}, \nu_{XZ}, \nu_{YZ}$ are exactly 1, and thus by the inductive hypothesis $\mathbb{E}_{x\in X,y\in Y,z\in Z}[1_{A\times B}(x,y)G(x,z)G(y,z) - 1_{A\times B}(x,y)\tilde{G}(x,z)\tilde{G}(y'z)]$ is $o(1)$, thus showing that $\|G_1' - \tilde{G}'\|_\square = o(1)$ as we wanted.

We are now able to expand $\mathbb{E}[(G' - \tilde{G}')(G_1' - \tilde{G}')]$ as

$$\mathbb{E}[G'G_1'] - \mathbb{E}[G'\tilde{G}'] - \mathbb{E}[\tilde{G}'G_1'] + \mathbb{E}[(\tilde{G}')^2].$$

Now we claim that each of the terms from above are actually $\mathbb{E}[(\tilde{G}')^2] + o(1)$. We see that

$$\mathbb{E}[G'G_1'] - \mathbb{E}[(\tilde{G}')^2] = \mathbb{E}_{x\in X,y\in Y,z\in Z}[G_1'(x,y)G(x,z)G(y,z) - \tilde{G}'(x,y)\tilde{G}(x,z)\tilde{G}(y,z)].$$

Again by the induction hypothesis we get that this is $o(1)$. This is seen by substituting $\nu_{XY}, G_{XY}, \tilde{G}_{XY}$ by $1, G_1', \tilde{G}'$ respectively, which satisfies $\|G_1' - \tilde{G}'\|_\square = o(1)$ by above, and this will increase how many of $\nu_{XY}, \nu_{XZ}, \nu_{YZ}$ are exactly 1. The same method also shows that the rest of the other terms are also $\mathbb{E}[(\tilde{G}')^2] + o(1)$. Therefore, $\mathbb{E}[(G' - \tilde{G}')(G_1' - \tilde{G}')]$ is $o(1)$ and as such the theorem follows. $\square$

So we have now proved the 3-term case in, however clearly we need to do a little work to move to the k-term case. Fortunately, the main issue with doing this is the notation so there are no major new ideas needed. So now instead of the triangle counting lemma in weighted tripartite graphs, we want to have a simplex counting lemma in weighted k-partite hypergraphs. Denote the vertex sets as $X_1, \dots, X_k$, and as as before for the purposes of the Green-Tao theorem each of the X_i will be $\mathbb{Z}_N$. Similarly to before we shall define $X_{-i} := X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots X_k$. Now if we have a weighted hypergraph G it is defined by functions $f_{-i} : X_{-i} \to \mathbb{R}$ for all $i \in \{1, \dots, k\}$. Now, if we take $\|f_{-i}\|_\square$ to be the cut norm of f_{-i} then we can write $\|G\|_\square = \max\{\|f_{-1}\|_\square, \dots, \|f_{-k}\|_\square$.

Definition 14. *Let* $x_{-j}^{(\omega)} := (x_1^{(\omega_1)}, \dots, x_{j-1}^{(\omega_{j-1})}, x_{j+1}^{(\omega_{j+1})}, \dots, x_k^{(\omega_k)}) \in X_{-j}$. *Then the* ***k-linear forms condition*** *is satisfied by a weighted hypergraph* ν *if*

$$\mathbb{E}_{x_1, x_1' \in X_1, \dots, x_k, x_k' \in X_k} \left[\prod_{j=1}^{k} \prod_{\omega \in 0,1^{[1,k] \setminus j}} \nu\left(x_{-j}^{(\omega)}\right) \right] = 1 + o(1)$$

and the same holds if any of the $k2^{k-1}$ ν *factors are removed.*

We now have the definitions needed to state our simplex counting lemma which is our generalisation of Theorem 16.

Theorem 17 (Simplex counting lemma)**.** *Take* $\nu, G, \tilde{G}$ *be three weighted k-partite hypergraphs with vertex set* $X_1 \bigcup \cdots \bigcup X_k$. *Now assume that* ν *satisfies the k-linear forms condition,* $0 \le G \le \nu$ *and* $0 \le \tilde{G} \le 1$. *Then, if* $\|G - \tilde{G}\|_\square = o(1)$, *we have*

$$|\mathbb{E}_{x_1, x_1' \in X_1, \dots, x_k, x_k' \in X_k}[G(x_1) \dots G(x_k) - \tilde{G}(x_1) \dots \tilde{G}(x_k)]| = o(1).$$

3.6 Proof of Relative Szemerédi Theorem

Now we are in a position to use the dense model theorem and counting lemma to prove the relative Szemerédi theorem.

Proof of Theorem 12. Utilising the Lemma 5 and the k-linear forms condition we see that $\|\nu - 1\|_{\square,k-1} = o(1)$. By Theorem 14, the so called Dense Model theorem, there exists a function $\tilde{f} : \mathbb{Z}_N \to [0,1]$ such that for every $f : \mathbb{Z}_N \to \mathbb{R}^+$ with $0 \leq f \leq \nu$ we have $\|f - \tilde{f}\|_{\square,k-1} = o(1)$.

Now if we set each of $X_1, \ldots, X_k$ to be $\mathbb{Z}_N$, then for every $j \in \{1, \ldots, k\}$ define the linear maps $\psi_j : X_{-j} \to \mathbb{Z}_N$ by

$$\psi_j(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k) := \sum_{i \in [1,k] \setminus j} (j - i)x_i.$$

This is useful as it now enables us to construct k-partite weighted hypergraphs $G, \tilde{G}, \nu$ on vertex set $X_1 \bigcup \cdots \bigcup X_k$. This is done by defining

$$\begin{aligned}
G_j(x_{-j}) &:= f(\psi_j(x_{-j})),\\
\tilde{G}_{-j}(x_{-j}) &:= \tilde{f}(\psi_j(x_{-j}))\\
\nu_{-j}(x_{-j}) &:= \nu(\psi_j(x_{-j})).
\end{aligned}$$

It should be noted that in the final line of the above definition the ν_{-j} on the left hand side is concerning the weighted hypergraph, whereas the ν on the right hand side is speaking of the function defined on $\mathbb{Z}_N$. Now we want to show that for every j the following hold

$$\begin{aligned}
\|\nu_{-j} - 1\|_{\square} &= \|\nu - 1\|_{\square,k-1}, \text{ and}\\
\|G_{-j} - \tilde{G}_{-j}\|_{\square} &= \|f - \tilde{f}\|_{\square,k-1}.
\end{aligned}$$

Once again we must make a note, the cut norm on the left hand side is the weighted hypergraph cut norm, and the right hand side is the cut norm for functions on $\mathbb{Z}_N$. We shall just show how to prove the second of these claims and the first is just a straightforward variation of the same idea. Take the left hand side $\|G_{-j} - \tilde{G}_{-j}\|_\square$, by expanding out we find that this is equal to

$$\sup_{A_1,\ldots,A_{k-1}\subseteq\mathbb{Z}_N^{k-2}} |\mathbb{E}_{x_1,\ldots,x_k\in\mathbb{Z}_N}[(f-\tilde{f})(\sum_{j=1}^{k-1}(k-j)x_j)1_{A_1}(x_2,x_3,\ldots,x_{k-1})\cdots 1_{A_{k-1}}(x_1,\ldots,x_{k-2})]|.$$

Now taking the right hand side $\|f-\tilde{f}\|_{\square,k-1}$ we find that it is equal to

$$\sup_{B_1,\ldots,B_{k-1}\subseteq\mathbb{Z}_N^{k-2}} |\mathbb{E}_{x_1,\ldots,x_k\in\mathbb{Z}_N}[(f-\tilde{f})(\sum_{j=1}^{k-1}x_j)1_{B_1}(x_2,x_3,\ldots,x_{k-1})\cdots 1_{B_{k-1}}(x_1,\ldots,x_{k-2})]|.$$

Now we see that simply by a change of variables (e.g. $(k-1)x_1 \leftrightarrow x_1$) the right and left hand sides are equal.

Thus we have that $\|G_{-j} - \tilde{G}_{-j}\|_\square = \|f-\tilde{f}\|_{\square,k-1} = o(1)$. Now from our simplex counting lemma (Theorem 17) we get that

$$\mathbb{E}_{x_1,\ldots,x_k\in\mathbb{Z}_N^k}[G_{-1}(x_{-1})\cdots G_{-k}(x_{-k})] = \mathbb{E}_{x_1,\ldots,x_k\in\mathbb{Z}_N^k}[\tilde{G}_{-1}(x_{-1})\cdots\tilde{G}_{-k}(x_{-k})]+o(1).$$

Looking at the left hand side we get that it is equal to

$$\mathbb{E}_{x_1,\ldots,x_k\in\mathbb{Z}_N^k}[f(\psi_1(x_{-1}))\cdots f(\psi_k(x_{-k}))].$$

Next, if we set $x = \psi_1(x_{-1})$ and $d = x_1+\cdots+x_k$, then $\psi_j(x_{-j}) = x+(j-1)d$. Therefore we get that the above is equal to $\Upsilon(f)$. By exactly the same reasoning we get find a similar statement for the right hand side and as such we have that

$$\Upsilon(f) = \Upsilon(\tilde{f}) + o(1),$$

and by Theorem 11 this is at least $c(k,\alpha) - o_{k,\alpha}(1)$ as required. $\square$

Chapter 4

Arithmetic Progressions in the Primes

4.1 Counting the Primes

Now the immediate counting function for the primes is the indicator function $1_{\mathbb{P}}$, however this has the same problems that stopped us from just applying the standard Szemerédi Theorem to prove Theorem 1, the primes have density 0. Fortunately, there is a standard method for dealing with this difficulty, we shall not be using the prime indicator function, but a weighted function. This allows us to weight the primes not by 1, but by $\log p$ using the von Mangoldt function Λ.

Definition 15. *The **von Mangoldt function** Λ is defined by*

$$\Lambda(n) = \begin{cases} \log p, & \textit{if } n = p^k \textit{ for } p \in \mathbb{P}, k \in \mathbb{N}; \\ 0, & \textit{otherwise.} \end{cases}$$

Now the von Mangoldt function can be seen as a natural set of weights to study in this case since by the prime number theorem and Golomb[10] we have that

$$\mathbb{E}(\Lambda) = 1 + o(1).$$

Therefore, this density is positive (for sufficiently large N).

It is immediately apparent that this is not a perfect set of weights as the von Mangoldt function is skewed to some residue classes (e.g. the only even prime is 2) and since pseudorandom sets behave in a similar manner to random sets, they should have a uniform distribution over all the residue classes as N goes to infinity. Therefore, pseudorandomness statements can not be made unless we can get rid of these biases. Green and Tao were able to come up with a method to deal with this issue called the **W-trick**.

So if we let $w = w(N)$ to be a sufficiently large integer depending only on N, then define

$$W := \prod_{p \leq w} p$$

to be the product of the primes up to w. Now the trick to avoid our bias mod p is to restrict ourselves to consider just the primes $p \equiv 1(\mod W)$.

Now incorporating this W-trick with the von Mangoldt function, we can define a new prime counting function which we shall prove to be suitable for use with the Relative Szemerédi Theorem.

Definition 16 (Prime Counting Function)**.**

$$\tilde{\Lambda}(n) = \begin{cases} \frac{\phi(W)}{W} \log(Wn+1), & \textit{when } Wn+1 \in \mathbb{P}; \\ 0, & \textit{otherwise.} \end{cases}$$

The factor $\frac{\phi(W)}{W}$ is necessary due to the fact that precisely $\phi(W)$ of the residue classes $\mod W$ have infinitely many primes. Now in order to see that $\tilde{\Lambda}$ is applicable for the Relative Szemerédi Theorem, we shall use a theorem of analytic number theory know as Dirichlet's Theorem to show that the density is bounded.

Theorem 18 (Dirichlet's Theorem)**.**

$$\sum_{\substack{p\leq x\\ p\equiv a(\mod b)}} \log p = \frac{x}{\phi(b)}(1+o(1)).$$

This theorem can then be used to show that the density of the prime counting function $\tilde{\Lambda}$ satisfies the hypotheses for the Relative Szemerédi Theorem.

Lemma 7. *For w growing sufficiently slowly with N, the function $\tilde{\Lambda}$ has positive density:*

$$\mathbb{E}_{x\in\mathbb{Z}_N}[\tilde{\Lambda}(x)] > 0.$$

Proof. All that is required is to expand the expectation notation, and then apply Theorem 18.

$$\begin{aligned}
\mathbb{E}_{x\in\mathbb{Z}_N}[\tilde{\Lambda}(x)] &= \frac{\phi(W)}{WN}\sum_{\substack{\frac{N}{2}\leq n<N\\ Wn+1\in\mathbb{P}}} \log(Wn+1)\\
&= \frac{\phi(W)}{WN}\sum_{\substack{\frac{WN}{2}\leq p<WN\\ p\equiv 1(\mod W)}} \log p + o(1)\\
&= \frac{\phi(W)}{WN}(\frac{WN}{\phi(W)} - \frac{WN}{2\phi(W)})(1+o(1))\\
&= \frac{1}{2} + o(1).
\end{aligned}$$

This is positive for sufficiently large N and the result follows. □

4.2 Constructing the Majorising Measure

We are finally at the point where we are going to be able to use the Relative Szemerédi Theorem to prove the Green-Tao theorem, which has been our goal from the start. In order to successfully do this we need to construct a majorising measure for our prime counting function $\tilde{\Lambda}$ that satisfies our k-linear forms condition.

Proposition 2. *For every $k \geq 3$, there exists $\alpha_k > 0$ such that for every N, sufficiently large, there exists a function $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ which satisfies the k-linear forms condition and $\nu(n) \geq \alpha_k \tilde{\Lambda}(n)$ for all $\frac{N}{2} \leq n < N$.*

Assuming we have this majorising measure ν for $\tilde{\lambda}$ along with the Relative Szemerédi Theorem, we get the Green-Tao theorem.

Proof of Theorem 1 assuming Proposition 2. Define a function $f : \mathbb{Z}_N \to \mathbb{R}^+$ by

$$f(n) = \begin{cases} \alpha_k \tilde{\Lambda}(n) & \text{if } \frac{N}{2} \leq n < N; \\ 0, & \text{otherwise.} \end{cases}$$

By Theorem 18,

$$\sum_{\frac{N}{2} \leq n < N} f(n) = (\frac{1}{2} + o(1))\alpha_k N.$$

Thus, for large enough N we have $\mathbb{E}[f] \geq \frac{\alpha_k}{3}$. Now, since we have assumed that ν satisfies the k-linear forms condition and $0 \leq f \leq \nu$, it follows from the Relative Szemerédi Theorem (Theorem 12) that

$$\Upsilon_k(f) \geq c(k, \frac{\alpha_k}{3}) - o_{k,\alpha}(1).$$

Thus, for N large enough, we have that, for some $\frac{N}{2} \leq x < N$ and $d \neq 0$,

$$f(x)f(x+d)\cdots f(x+(k-1)d) > 0.$$

We have that f has support on $[\frac{N}{2}, N)$, we get that $x, x+d, x+2d, \ldots, x+(k-1)d$ is a k-A.P. Due to the support, this is not only a k-A.P in $\mathbb{Z}_N$ but also in $\mathbb{Z}$. Therefore, for $i \in \{0, \ldots, k-1\}$ we have that

$$(x+id)W+1$$

is a k-A.P of primes. $\square$

We have now been able to prove that if we can construct a majorising measure for $\tilde{\Lambda}$ then we are done.

If we let $w(n)$ be the number of prime factors of n, then define the **Möbius function** μ as

$$\mu(n) := \begin{cases} (-1)^{w(n)} & \text{if } n \text{ is squarefree;} \\ 0, & \text{otherwise.} \end{cases}$$

There is an basic identity know as the Möbius inversion formula that says

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(\frac{n}{d}).$$

In the original proof by Green and Tao [13], they used a truncated version of Λ to construct the majorising measure. This was needed to also take into account the numbers with few prime factors (we can consider these our "almost primes"). For an integer n with $w(n)$ large, most of the divisors of n will be small in comparison to n itself. Thus, if we truncate the sum in the Möbius inversion formula to only sum over divisors less than or equal to some R, we will be left with a function that approximates Λ well for the numbers n with $w(n)$ large i.e. many prime factors, but differs from Λ at our set of "almost primes". Thus, for $R > 0$ define

$$\Lambda_R(n) := \sum_{\substack{d|n \\ d \leq R}} \mu(d) \log(\frac{R}{d}).$$

A simplification of the proof was then found by Green and Tao [14], and they were able to replace Λ_R by the following:

Definition 17 (Modified Truncated Divisor Sum)**.** *Let χ be a smooth, bounded, compactly supported function with*

$$\chi(0) = 1,\ \chi(x) = 0 \textit{ for } |x| \geq 1.$$

Then define

$$\tilde{\Lambda}_R(n) := \sum_{d|n} \mu(d) \chi(\frac{\log d}{\log R}).$$

Now we are in a position to construct our majorising measure ν and then show that Proposition 2 follows from it. Unfortunately I am unable to produce a proof of this Proposition 3 , however a full proof of it can be found in [4] Section 9.

Proposition 3 (Pseudorandom Majorising Measure)**.** *Let $k \geq 3$ and $R := N^{k^{-1}2^{-k-3}}$. Then as earlier, assume that w grows slowly with N and take $W := \prod_{p \leq w} p$. Also, let $c_\chi := \int_0^\infty |\chi'(x)|^2\, dx$. Define $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ by*

$$\nu(n) := \begin{cases} \frac{\phi(W)}{W} \frac{\tilde{\Lambda}(Wn+1)^2}{c_\chi} & if \frac{n}{2} \leq n < N; \\ 1, & otherwise. \end{cases}$$

Then ν satisfies the k-linear forms condition.

We are now in a position to prove Proposition 2.

Proof of Proposition 2 using Proposition 3. Let $\alpha_k = k^{-1}2^{-k-4}c_\chi^{-1}$. We need to prove that for N large enough we have $\nu(n) \geq \alpha_k \tilde{\Lambda}(n)$ for all $n \in [\frac{N}{2}, N)$. We have that if $Wn+1 \notin \mathbb{P}$ then $\tilde{\Lambda}(n) = 0$, therefore it suffices to look at the inequality when $Wn+1$ is a prime. From the hypothesis in Proposition 3 we get that

$$\begin{aligned}
\log R &= k^{-1}2^{-k-3}\log N \\
&\geq k^{-1}2^{-k-4}\log(WN+1) \\
&= c_\chi \alpha_k \log(WN+1),
\end{aligned}$$

where this inequality holds provided that w grows slowly enough with N and N sufficiently large. In the case that $Wn+1$ is a prime it is seen that $\tilde{\Lambda}_R(Wn+1) = 1$, thus

$$\begin{aligned}
\alpha_k \tilde{\Lambda}(n) &= \alpha_k \frac{\phi(W)}{W}\log(Wn+1) \\
&\leq \alpha_k \frac{\phi(W)}{W}\log(WN+1) \\
&\leq \frac{\phi(W)}{W}\frac{1}{c_\chi} \\
&= \nu(n),
\end{aligned}$$

as required. $\square$

Bibliography

[1] F.A. Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proceedings of the National Academy of Sciences of the United States of America*, 32(12):331—332, December 1946.

[2] T. Bloom. The green-tao theorem on arithmetic progressions within the primes. 08 2010.

[3] T. D. Browning and S. M. Prendiville. A Transference Approach to a Roth-Type Theorem in the Squares. *International Mathematics Research Notices*, 2017(7):2219–2248, 06 2016.

[4] D. Conlon, J. Fox, and Y. Zhao. The green-tao theorem: An exposition. *arXiv: Number Theory*, 1:249–282, 2014.

[5] D. Conlon, J. Fox, and Y. Zhao. A relative szemerédi theorem. *Geometric and Functional Analysis*, 25(3):733–762, Mar 2015.

[6] J.G. van der Corput. Über summen von primzahlen und primzahlquadraten. *Mathematische Annalen*, 116:1–50, 1939.

[7] J. Dybizbanski. Sequences containing no 3-term arithmetic progressions. *Electron. J. Comb.*, 19:P15, 2012.

[8] P. Erdös and P. Turán. On some sequences of integers. *Journal of the London Mathematical Society*, s1-11(4):261–264, 1936.

[9] H. Furstenberg, Y. Katznelson, and D. Ornstein. The ergodic theoretical proof of Szemerédi's theorem. *Bulletin (New Series) of the American Mathematical Society*, 7(3):527 – 552, 1982.

[10] S. W. Golomb. The lambda method in prime number theory. *Journal of Number Theory*, 2(2):193–198, 1970.

[11] W. T. Gowers. A new proof of szemerédi's theorem. *Geometric & Functional Analysis GAFA*, 11:465–588, 2001.

[12] W. T. Gowers. Decompositions, approximate structure, transference, and the hahn-banach theorem. *Bulletin of the London Mathematical Society*, 42(4):573–606, Jun 2010.

[13] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, 167:481–547, 2008.

[14] B. Green and T. Tao. Linear equations in primes. *Annals of mathematics*, 171(3):1753–1850, 2010.

[15] D.R. Heath-Brown. Three Primes and an Almost-Prime in Arithmetic Progression. *Journal of the London Mathematical Society*, s2-23(3):396–414, 06 1981.

[16] K. F. Roth. On certain sets of integers. *Journal of the London Mathematical Society*, s1-28(1):104–109, 1953.

[17] K.F. Roth. Sur quelques ensembles d'entiers. *C. R. Acad. Sci. Paris*, 234:388–390, 1952.

[18] A. Sharma. Sequences of integers avoiding 3-term arithmetic progressions. *Electron. J. Comb.*, 19:P27, 2012.

[19] E. Szemerédi. On sets of integers containing k elements in arithmetic progression. *Acta Arithmetica*, 27(1):199–245, 1975.

[20] P. Varnavides. On certain sets of positive density. *Journal of the London Mathematical Society*, s1-34(3):358–360, 1959.

[21] K. Weierstrass. Über die analytische darstellbarkeit sogenannter willkürlicher functionen einer reellen veränderlichen. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften zu Berlin*, 2:633–639, 1885.

[22] Y. Zhao. An arithmetic transference proof of a relative szemerédi theorem. *Mathematical Proceedings of the Cambridge Philosophical Society*, 156(2):255–261, Nov 2013.