Fourier Analytic Methods in Additive Number Theory

Kyle E. Kinneberg

Professor Asuman Aksoy, Advisor

Professor Michael O'Neill, Reader

May, 2009

CLAREMONT McKENNA COLLEGE

Department of Mathematics

Abstract

In recent years, analytic methods have become prominent in additive number theory. In particular, finite Fourier analysis is well-suited to solve some problems that are too difficult for purely combinatorial techniques. Among these is Szemerédi's Theorem, a statement regarding the density of integral sets and the existence of arithmetic progressions in those sets.

In this thesis, we give a general introduction to classical Fourier analysis over $\mathbb{R}$ and discrete Fourier analysis over the group of integers modulo N. We then give a complete explanation of Timothy Gowers's 1998 proof of Szemerédi's Theorem for arithmetic progressions of length four. This proof relies entirely on finite Fourier analytic methods. As a result, our explanation of it provides readers with a thorough demonstration of how these techniques are useful in additive number theory. We also give a short description of other problems in this field for which analytic methods are helpful.

Acknowledgments

The author thanks Professor Michael O'Neill for his tremendous help and guidance throughout this thesis project. He also thanks those in the mathematics department at Claremont McKenna College, especially Professor Asuman Aksoy, for the instruction he has received during the past four years.

The author thanks his family and friends as well, for their love, support, and contribution to his education. He is especially grateful to his mother, father, and brother for the example they set in the pursuit of true happiness.

Most importantly, the author thanks his patient Father and most Blessed Mother for never leaving his side.

Contents

Abstract **ii**

Acknowledgments **iii**

1 An Introduction to Fourier Analysis **1**
1.1 Classical Fourier Analysis 2
1.2 Discrete Fourier Analysis 7

2 The Work of Roth and Gowers on Szemerédi's Theorem **12**
2.1 Roth's Work on Progressions of Length Three 14
2.2 Preliminaries for Gowers's Work on Progressions of Length Four 20

3 Gowers's Proof of Szemerédi's Theorem for Progressions of Length Four **24**
3.1 Quadratic Uniformity 26
3.2 Sets that Fail to be Quadratically Uniform 42
3.3 Some Combinatorial Considerations 46
3.4 Quadratic Fourier Sums 57
3.5 Putting Everything Together 66

4 Extensions of Gowers's Work **72**
4.1 Sumsets in Cyclic Groups of Prime Order 76

Bibliography **81**

Chapter 1

An Introduction to Fourier Analysis

As its name suggests, classical number theory is rooted in the study of integers. Questions regarding divisibility, prime factorizations, congruence equations, and Diophantine equations are all questions about integers (usually positive integers). The traditional methods used to study such topics are finite in nature, which makes sense since the problems themselves are finite in nature. For example, the Fundamental Theorem of Arithmetic – arguably the most important fact about the integers – is usually proved by simple discrete arguments: the existence of prime factorizations can be verified using strong induction, and uniqueness can be shown using elementary properties of divisibility.

Due to the fundamentally finite nature of number theory, it is interesting when methods that are not naturally number theoretic arise and are useful. In modern number theory, methods from combinatorics, graph theory, probability, ergodic theory, convex geometry, incidence geometry, and algebraic geometry are surprisingly helpful [30]. Some of the most important number theoretic problems, however, have been resolved (or are in the process of being resolved) with analytic methods.

One of the earliest significant number theoretic problems that was solved using analytic arguments was Dirichlet's Theorem. In 1837, Dirichlet proved that for any positive integers a and b with $\gcd(a,b) = 1$ (that is, a and b are coprime), the set $\{a + bk : k \in \mathbb{N}\}$ contains infinitely many primes. In fact, Dirichlet proved a stronger statement, namely that the sum $\sum_{p \in a+b\mathbb{N}} \frac{1}{p}$ diverges, where p runs over primes. Other well-known results and questions in number theory that have a particularly analytic flavor are the Prime Number Theorem (regarding the density of primes in the integers), finding methods for solving Diophantine equations, Goldbach's Conjecture (stating that any even integer larger than 2 is the sum of two primes), and of course, the Riemann Hypothesis and it's generalization to L-functions.

Interestingly, the analysis used in the proof of Dirichlet's Theorem is a type of discrete Fourier analysis. It has been only in the past 60 years, however, that Fourier analytic methods have become generally popular in number theory. One of the primary reasons for the increased popularity of these methods is their usefulness in solving problems in additive number theory. In particular, their ability to help prove Szemerédi's Theorem, and their subsequent extension to prove the Green-Tao Theorem (that the primes contain arbitrarily long arithmetic progressions) have showed the mathematical community how much these methods can say about integral sets.

In this work, we attempt to give a general introduction to discrete Fourier analysis and a description of its uses in additive number theory. This description will take the form of a detailed explanation of Gowers's Fourier analytic proof of Szemerédi's Theorem (or more accurately, a special case of Szemerédi's Theorem for which the solution uses nearly all the techniques found in the proof of the full result). Although Gowers's proof of this theorem was one of the most important results in the past decade (partially because Green and Tao were able to extend his techniques to prove their theorem), it is surprisingly self-contained. Thus, we believe that a brief introduction to discrete Fourier analysis will suffice for readers to understand the entirety of Gowers's proof.

1.1 Classical Fourier Analysis

Before introducing discrete Fourier analysis, it will be helpful to discuss classical Fourier analysis. Readers may be familiar with this field, but a short reminder is useful. It will also enable us to present some of the results in discrete Fourier analysis as analogues of results in classical Fourier analysis. Because this section is meant to be a review of Fourier analysis, we will confine ourselves to functions defined on $\mathbb{R}$, rather than discussing the more general theory for functions defined on $\mathbb{R}^n$.

One of the first necessities in Fourier analysis is the identification of the unit circle $\mathbb{T}$ in the complex plane with the interval $(-\pi, \pi]$ in $\mathbb{R}$. As we shall see, the fundamental concepts in Fourier analysis are defined by integrating a function $f : \mathbb{T} \to \mathbb{C}$ over its entire domain $\mathbb{T}$. In order to keep this review simple, we wish to use Riemann integration rather than Lebesgue integration. As a result, we identify $\mathbb{T}$ with $(-\pi, \pi]$ by $e^{i\theta} \leftrightarrow \theta$. We then set $f(-\pi) = f(\pi)$ so that f is defined on $[-\pi, \pi]$ and we can integrate f over this closed interval. Hence, from this point forward, we use the closed interval $[-\pi, \pi]$ and the unit circle $\mathbb{T}$ interchangeably, remembering that $-\pi$ and π are essentially the same point (functions that we consider will always have $f(-\pi) = f(\pi)$). We also use the term ″integrable″ to refer to Riemann integrable functions.

Observe that the set of functions defined on $\mathbb{T}$ can be identified with the set of functions on $\mathbb{R}$ that have period 2π. Indeed, any function on $\mathbb{T}$ can be periodically extended to all of $\mathbb{R}$, and any function on $\mathbb{R}$ with period 2π can be restricted to $\mathbb{T}$. Hence, it is natural to consider functions of the form $e_n(x) = e^{inx}$. We can easily check that the collection $\{e_n : n \in \mathbb{Z}\}$ of such functions is an orthonormal set with regard to the inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx.$$

(Recall that this inner product is defined for the complex vector space of integrable functions on $\mathbb{T}$.) In other words,

$$\langle e_n, e_m \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)x} dx = \begin{cases} 0 & \text{if } n \neq m \\ 1 & \text{if } n = m. \end{cases}$$

Hence, we would like to say that the set $\{e_n : n \in \mathbb{Z}\}$ is an orthonormal basis for

$$\operatorname{span}(e_n : n \in \mathbb{Z}) = \left\{ \sum_{n \in \mathbb{Z}} a_n e_n : a_n \in \mathbb{C} \right\}.$$

The problem here, however, is with convergence of these infinite sums. Certainly, we must restrict the set $\{\sum_{n\in\mathbb{Z}} a_n e_n : a_n \in \mathbb{C}\}$ to sums that converge in some sense; perhaps uniform convergence, perhaps pointwise convergence, perhaps convergence in some norm.

Suppose we decide on uniform convergence, and suppose additionally that the series $\sum_{n\in\mathbb{Z}} a_n e^{inx}$ converges (uniformly) to an integrable function f. Then, for each $n \in \mathbb{Z}$, we have

$$\begin{aligned} a_n &= \sum_{m\in\mathbb{Z}} a_m \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m-n)x} dx \right) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\sum_{m\in\mathbb{Z}} a_m e^{imx} \right) e^{-inx} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx}. \end{aligned}$$

Thus, although we have not settled any questions regarding convergence, it seems natural to define our coefficients a_n in this manner. This leads us to the following definitions, and with it, the foundation of classical Fourier analysis.

Definition 1. Let $f : \mathbb{T} \to \mathbb{C}$ be integrable on $[-\pi, \pi]$. Then for $n \in \mathbb{Z}$, we say that

$$\hat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx$$

is the n-th Fourier coefficient of f.

Note that in particular, $\hat{f}(0)$ is simply the average value of f over $\mathbb{T}$.

Definition 2. If $f : \mathbb{T} \to \mathbb{C}$ is integrable on $[-\pi, \pi]$, then we say that the function

$$S(\xi) = \sum_{n\in\mathbb{Z}} \hat{f}(n) e^{in\xi}$$

is the Fourier series of f.

Note that S is "defined" for any $x \in \mathbb{R}$, and it takes values in $\mathbb{C}$. But as we discussed above, S may not be well-defined for some x because the sum may not converge. Questions regarding convergence of these series formed the foundation of the early study of Fourier analysis. Tied closely with these questions of convergence are questions concerning the relationship between S and f. In order to discuss convergence, though, it is convenient to use the following definition.

Definition 3. If $f : \mathbb{T} \to \mathbb{C}$ is integrable on $[-\pi, \pi]$, then we call

$$S_N(\xi) = \sum_{n=-N}^{N} \hat{f}(n) e^{in\xi}$$

the N-th partial sum of the Fourier series of f.

As we shall now see, the convergence of the Fourier series of f can depend on how "nice" f is. We shall also see, however, that there are examples of fairly nice functions for which the corresponding Fourier series do not come close to converging, even pointwise, on some (perhaps infinite) subset of $\mathbb{T}$.

The following theorems are results regarding the convergence of the Fourier series under different assumptions about f. We arrange the theorems from weakest to strongest, and since this section serves only as a review, we omit proofs. Justification for these results can be found in most texts on Fourier analysis, such as [14] and [23].

Theorem 1.1. *Let $f : \mathbb{T} \to \mathbb{C}$ be continuous, and suppose that its Fourier series converges absolutely; that is, $\sum_{n \in \mathbb{Z}} |\hat{f}(n)| < \infty$. Then the partial sums S_N converge absolutely and uniformly to f.*

Certainly, the conclusions of this theorem are strong, but since the hypotheses are also strong (namely, absolute convergence of the Fourier series), the theorem, although important, is fairly weak. It does give the following corollary, though.

Corollary 1.1. *Let $f : \mathbb{T} \to \mathbb{C}$ be twice continuously differentiable. Then there is a constant C (not depending on n) for which $|\hat{f}(n)| \leq C(1/|n|^2)$ whenever $n \neq 0$. Thus, by Theorem 1.1, the partial sums S_N converge absolutely and uniformly to f.*

The next theorem pertains to the norm convergence of partial sums. Recall from above that we are working in the space of integrable functions on $\mathbb{T}$, with inner product $\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)}dx$. The associated norm is the L^2 norm: $||f||_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx$. Thus, the distance between two functions f and g is given by

$$||f - g||_2 = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x) - g(x)|^2 dx}.$$

Theorem 1.2. *If $f : \mathbb{T} \to \mathbb{C}$ is integrable, then*

$$\lim_{N \to \infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x) - S_N(x)|^2 dx = 0.$$

In other words, S_N converges to f in the L^2-norm.

The assumptions of this theorem are, of course, just about as weak as possible. However, convergence in the L^2-norm is certainly not a strong notion of convergence, since it does not imply pointwise convergence anywhere (for example, there are functions that converge to zero in the L^2-norm but converge nowhere pointwise). The following theorem is a bit stronger.

Theorem 1.3. *If $f : \mathbb{T} \to \mathbb{C}$ is integrable, then for any x at which f is differentiable, the partial sums $S_N(x)$ converge to $f(x)$ as $N \to \infty$.*

This theorem, although stronger than its predecessors, still is not a satisfactory answer to our questions of convergence. In 1964, however, Carleson proved the strongest possible result regarding pointwise convergence of Fourier series.

Theorem 1.4 (Carleson)**.** *If $f : \mathbb{T} \to \mathbb{C}$ is integrable, then the set of points x for which $S_N(x)$ does not converge to $f(x)$ has Lebesgue measure zero. In other words, the partial sums of the Fourier series of f converge pointwise almost-everywhere on $\mathbb{T}$.*

We say that this theorem is the strongest possible pointwise convergence result because of the following theorem due to Kahane and Katznelson. In essence, their theorem says that for any set of measure zero in $\mathbb{T}$, there is a continuous function whose Fourier series behaves terribly on that set.

Theorem 1.5 (Kahane-Katznelson). *Let $E \subset \mathbb{T}$ have Lebesgue measure zero. Then there exists a function $f : \mathbb{T} \to \mathbb{C}$ such that f is continuous on $\mathbb{T}$ but*

$$\limsup_{N\to\infty} |S_N(x)| = \infty$$

for all $x \in E$.

For example, we can take $E = \mathbb{Q} \cap [-\pi, \pi]$. Despite E being dense in $\mathbb{T}$, there is a continuous function whose Fourier series does not even come close to converging at any point of E. Together, the results of Carleson and Kahane-Katznelson completely solve the pointwise convergence question about Fourier series. They also vividly show us why questions of convergence on all of $\mathbb{T}$ are so difficult.

There are several other results regarding the convergence of Fourier series, and readers who wish to learn more may consult any text in Fourier analysis. For our purposes, these theorems give us a flavor of classical Fourier analysis and will help us to appreciate discrete Fourier analysis, where we will not have to worry about any questions of convergence. We now discuss some non-convergence results in classical Fourier analysis. Later, we will see that these results have direct analogues in discrete Fourier analysis.

One of the most important tools in classical Fourier analysis is a type of smoothing, or averaging, operation between two functions, called the convolution.

Definition 4. If $f, g : \mathbb{T} \to \mathbb{C}$ are integrable, then we define their convolution as a function $f * g : \mathbb{T} \to \mathbb{C}$ where

$$(f * g)(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)g(x-y)dy$$

for $x \in [-\pi, \pi]$.

In general, it is helpful to think of the convolution as a weighted average. Observe that if we let g be identically one on $\mathbb{T}$, then $(f * g)(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)dy$, which is simply the average value of f over $\mathbb{T}$. Thus, even when g is not a constant function, we can consider the convolution to be an average of f, weighted by g. Additionally, we can think of the convolution as a smoothing operation because $f * g$ is continuous on $\mathbb{T}$ for any integrable f and g.

The convolution also has several desirable algebraic properties. First, it is linear in that $f * (g + h) = (f * g) + (f * h)$ and $(cf) * g = c(f * g) = f * (cg)$ for all integrable functions f, g, h on $\mathbb{T}$ and any $c \in \mathbb{C}$. This is a simple consequence of the linearity of the integral. Second, the convolution is associative and commutative: $(f * g) * h = f * (g * h)$ and $f * g = g * f$ for all integrable functions f, g, h on $\mathbb{T}$. This tells us that not only can we think of $f * g$ as an average of f weighted by g but also as an average of g weighted by f. Third, and perhaps most important, the convolution gives the Fourier coefficients a certain multiplicative property:

$$\widehat{f * g}(n) = \hat{f}(n)\hat{g}(n)$$

for all integrable functions f, g on $\mathbb{T}$ and any $n \in \mathbb{N}$.

We will not give a justification of these properties of the convolution here. However, we shall see that the convolution defined in discrete Fourier analysis has many of the same properties, and we will provide proofs there.

We now present an important identity in classical Fourier analysis that will have a direct analogue in the finite version. Diverging from our desired brevity in this section, we give a proof of this result.

Theorem 1.6 (Parseval's Formula). *Let $f : \mathbb{T} \to \mathbb{C}$ be integrable. Then*

$$\sum_{n\in\mathbb{Z}} |\hat{f}(n)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2 dx.$$

Proof. First, recall from before that the collection $\{e_n : n \in \mathbb{Z}\}$, where $e_n(x) = e^{inx}$, is an orthonormal set with respect to the inner product

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}dx.$$

Also, observe that

$$\hat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}dx = \langle f, e_n\rangle$$

for each $n \in \mathbb{Z}$. Therefore, given $N \in \mathbb{N}$, we have $S_N = \sum_{|n|\leq N} \langle f, e_n\rangle e_n$. We claim that $f - S_N$ is orthogonal to each e_n for which $|n| \leq N$. Indeed, for such an n, we have

$$\begin{aligned}\langle f - S_N, e_n\rangle &= \langle f, e_n\rangle - \langle S_N, e_n\rangle = \langle f, e_n\rangle - \left\langle \sum_{|m|\leq N} \langle f, e_m\rangle e_m, e_n \right\rangle \\ &= \langle f, e_n\rangle - \sum_{|m|\leq N} \langle f, e_m\rangle\langle e_m, e_n\rangle = \langle f, e_n\rangle - \langle f, e_n\rangle = 0,\end{aligned}$$

where we have used the orthonormality of the set $\{e_n\}$. Thus, the claim is true. By linearity, this implies that $f - S_N$ is orthogonal to any combination $\sum_{|n|\leq N} b_n e_n$ where $b_n \in \mathbb{C}$. In particular, we can let $b_n = \hat{f}(n)$ so that $f - S_N$ is orthogonal to S_N. Now, write $f = (f - S_N) + S_N$, and observe that the Pythagorean theorem implies

$$||f||_2^2 = ||f - S_N||_2^2 + ||S_N||_2^2.$$

Also observe that

$$\begin{aligned}||S_N||_2^2 = \langle S_N, S_N\rangle &= \left\langle \sum_{|n|\leq N} \hat{f}(n)e_n, \sum_{|m|\leq N} \hat{f}(m)e_m \right\rangle \\ &= \sum_{|n|\leq N}\sum_{|m|\leq N} \left\langle \hat{f}(n)e_n, \hat{f}(m)e_m \right\rangle = \sum_{|n|\leq N}\sum_{|m|\leq N} \hat{f}(n)\overline{\hat{f}(m)}\langle e_n, e_m\rangle \\ &= \sum_{|n|\leq N} \hat{f}(n)\overline{\hat{f}(n)} = \sum_{|n|\leq N} |\hat{f}(n)|^2,\end{aligned}$$

where we have again used the orthonormality of $\{e_n\}$. Therefore, we have

$$||f||_2^2 = ||f - S_N||_2^2 + \sum_{|n|\leq N} |\hat{f}(n)|^2$$

for each $N \in \mathbb{N}$. Now, let N approach ∞. By Theorem 1.2, we see that $||f - S_N||_2^2$ approaches 0. Hence, we have

$$||f||_2^2 = \lim_{N\to\infty} \sum_{|n|\leq N} |\hat{f}(n)|^2 = \sum_{n\in\mathbb{Z}} |\hat{f}(n)|^2,$$

which is precisely

$$\sum_{n\in\mathbb{Z}} |\hat{f}(n)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x)|^2 dx.$$

□

There is certainly more to classical Fourier analysis than we have covered here. In fact, there is a nice extension of the periodic theory that we have discussed to non-harmonic functions defined on all of $\mathbb{R}$. A result of this extension is a beautiful relationship between functions and their Fourier transforms, called the Fourier inversion formula. These and similar topics, though interesting, will not give us much additional insight into the foundations of discrete Fourier analysis. Thus, we now turn exclusively to the finite version.

1.2 Discrete Fourier Analysis

We saw in the previous section that given an integrable function f on the unit circle $\mathbb{T}$ in the complex plane, we can define Fourier coefficients for this function that allow us to write f as an infinite linear combination of functions from an orthonormal set, using the Fourier coefficients as the coefficients in the linear combination. We also saw that f and this linear combination may not agree at some points, but they will agree on almost all of $\mathbb{T}$. If we want to develop a discrete version of this theory, our first task is to determine on what domain we should define our functions f. In particular, is there a discrete version of $\mathbb{T}$?

There is, of course, such a set: the group of N-th roots of unity. We denote this group by $\mathbb{Z}(N)$. Recall that a complex number z is an N-th root of unity if and only if $z^N = 1$. It is well known that $\mathbb{Z}(N) = \left\{1, e^{2\pi i/N}, \ldots, e^{2(N-1)\pi i/N}\right\}$ where the group operation is complex multiplication. Since complex multiplication is commutative, $\mathbb{Z}(N)$ is an abelian group for all N.

For our purposes, however, it will be easier to use $\mathbb{Z}(N)$ in a different form, namely in additive form. Let $\mathbb{Z}_N$ denote the quotient group $\mathbb{Z}/\mathbb{Z}N$ (that is, the group of congruence classes modulo N). There is a group isomorphism between $\mathbb{Z}(N)$ and $\mathbb{Z}_N$ given by $e^{2\pi ki/N} \leftrightarrow k + \mathbb{Z}N$. Thus, $\mathbb{Z}(N)$ is isomorphic to $\mathbb{Z}_N$, where the group operation in $\mathbb{Z}_N$ is addition of congruence classes. In fact, we can think of $\mathbb{Z}_N$ as the set of integers $\{0, 1, \ldots, N-1\}$ under addition modulo N. We will always consider $\mathbb{Z}_N$ in this form.

We therefore wish to study functions defined on $\mathbb{Z}_N$ that take values in $\mathbb{C}$. It is important first to observe that such functions form a vector space over $\mathbb{C}$ with inner product

$$\langle f, g\rangle = \sum_{k=0}^{N-1} f(k)\overline{g(k)}.$$

This inner product induces the l^2 norm

$$||f||_2^2 = \sum_{k=0}^{N-1} |f(k)|^2.$$

It is clear that this vector space has dimension N. Indeed, the collection of characteristic functions $\chi_n : \mathbb{Z}_N \to \mathbb{C}$ defined by

$$\chi_n(k) = \begin{cases} 0 & \text{if } k \neq n \\ 1 & \text{if } k = n \end{cases}$$

for $n \in \mathbb{Z}_N$ spans the vector space since

$$f = \sum_{n=0}^{N-1} f(n)\chi_n$$

for each $f : \mathbb{Z}_N \to \mathbb{C}$. It is also easy to see that this collection is linearly independent.

A different (and more important) collection of functions in this vector space are the exponential functions. For $n \in \mathbb{Z}_N$, define $e_n : \mathbb{Z}_N \to \mathbb{C}$ by $e_n(k) = e^{2\pi kni/N}$. We claim that this collection is an orthogonal set. Indeed, for any n and m in $\mathbb{Z}_N$, we have

$$\langle e_n, e_m \rangle = \sum_{k=0}^{N-1} e^{2\pi kni/N} e^{-2\pi kmi/N} = \sum_{k=0}^{N-1} e^{2\pi(n-m)ki/N} = \begin{cases} 0 & \text{if } n = m \\ N & \text{if } n \neq m \end{cases}$$

where the last equality follows from the identity $1 + x + x^2 + \ldots + x^{N-1} = (1 - x^N)/(1 - x)$ for $x \neq 1$ and letting $x = e^{2\pi(n-m)i/N}$ when $n \neq m$. When $n = m$, the equality is trivial. Thus, the collection of functions e_n for $n \in \mathbb{N}$ is orthogonal; and more is true, namely that each e_n has norm squared equal to N. The orthogonality of this collection implies immediately that the elements are linearly independent. But this, in turn, means that they form a basis for the vector space since there are N of them. Thus, given an arbitrary function $f : \mathbb{Z}_N \to \mathbb{C}$, we can express it as a linear combination

$$f = \sum_{n=0}^{N-1} a_n e_n.$$

Notice that we do not have to worry about convergence as we did in the classical analogue. But we can ask the same question as we did before: what are the coefficients a_n?

Suppose $f = \sum_{n=0}^{N-1} a_n e_n$. Then for each $n \in Z_N$, we have the following:

$$\begin{aligned} a_n &= \sum_{m=0}^{N-1} a_m \left(\frac{1}{N} \langle e_m, e_n \rangle \right) = \sum_{m=0}^{N-1} a_m \left(\frac{1}{N} \sum_{k=0}^{N-1} e_m(k) \overline{e_n(k)} \right) \\ &= \frac{1}{N} \sum_{m=0}^{N-1} \sum_{k=0}^{N-1} a_m e_m(k) \overline{e_n(k)} = \frac{1}{N} \sum_{k=0}^{N-1} \left(\sum_{m=0}^{N-1} a_m e_m(k) \right) \overline{e_n(k)} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} f(k) e^{-2\pi kni/N}. \end{aligned}$$

We therefore introduce the following definition.

Definition 5. Let $f : \mathbb{Z}_N \to \mathbb{C}$. Then for $n \in \mathbb{Z}_N$, we say that

$$\hat{f}(n) = \frac{1}{N} \sum_{k=0}^{N-1} f(k) e^{-2\pi kni/N}$$

is the n-th Fourier coefficient of f.

Using this definition, we have already proven the following theorem, which is the finite analogue of the Fourier inversion from the classical theory.

Theorem 1.7 (Inversion formula)**.** *Let $f : \mathbb{Z}_N \to \mathbb{C}$. Then*

$$f(k) = \sum_{n=0}^{N-1} \hat{f}(n) e_n(k)$$

for all $k \in \mathbb{Z}_N$.

As in the classical theory, we can define the convolution of two complex valued functions on $\mathbb{Z}_N$.

Definition 6. If $f, g : \mathbb{Z}_N \to \mathbb{C}$, then we define their convolution $f \star g : \mathbb{Z}_N \to \mathbb{C}$ as

$$(f \star g)(n) = \frac{1}{N} \sum_{k=0}^{N-1} f(k) g(n-k)$$

for $n \in \mathbb{Z}_N$.

We claim that like its classical analogue, the discrete convolution has several nice algebraic and multiplicative properties. This is the content of the following proposition.

Proposition 1.1. *For $f, g, h : \mathbb{Z}_N \to \mathbb{C}$, the following are true:*

i. $f \star (g + h) = (f \star g) + (f \star h)$

ii. $(cf) \star g = c(f \star g)$ *if* $c \in \mathbb{C}$

iii. $(f \star g) \star h = f \star (g \star h)$

iv. $f \star g = g \star f$

v. $\widehat{f \star g} = \hat{f} \cdot \hat{g}$

Proof. Parts (i) and (ii) follow directly from the definition of convolution. The proofs of (iii) and (iv) primarily use interchanging sums and re-indexing.
(iii): For each $n \in \mathbb{Z}_N$, we have

$$\begin{aligned}[(f \star g) \star h](n) &= \frac{1}{N} \sum_{k=0}^{N-1} (f \star g)(k) h(n-k) = \frac{1}{N} \sum_{k=0}^{N-1} \left(\frac{1}{N} \sum_{s=0}^{N-1} f(s) g(k-s) \right) h(n-k) \\ &= \frac{1}{N^2} \sum_{s=0}^{N-1} f(s) \sum_{k=0}^{N-1} g(k-s) h(n-k).\end{aligned}$$

Now, for s fixed, let $y = k - s$. Note that as k runs through $\mathbb{Z}_N$, y also runs through $\mathbb{Z}_N$. And of course, with $y = k - s$, we have $n - k = n - s - y$. Thus,

$$\sum_{k=0}^{N-1} g(k-s)h(n-k) = \sum_{y=0}^{N-1} g(y)h(n-s-y) = N(g \star h)(n-s).$$

Therefore, we see that

$$[(f \star g) \star h]\,(n) = \frac{1}{N}\sum_{s=0}^{N-1} f(s)(g \star h)(n-s) = [f \star (g \star h)]\,(n)$$

as desired.
(iv): For each $n \in \mathbb{Z}_N$, we have

$$(f \star g)(n) = \frac{1}{N}\sum_{k=0}^{N-1} f(k)g(n-k)$$

by definition. Let $y = n - k$, and again, as k runs through $\mathbb{Z}_N$, y also runs through $\mathbb{Z}_N$. Therefore,

$$\frac{1}{N}\sum_{k=0}^{N-1} f(k)g(n-k) = \frac{1}{N}\sum_{y=0}^{N-1} f(n-y)g(y) = (g \star f)(n),$$

so we indeed have $(f \star g)(n) = (g \star f)(n)$ as desired.
(v): For each $n \in \mathbb{Z}_N$, we have

$$\begin{aligned}\widehat{(f \star g)}(n) &= \frac{1}{N}\sum_{k=0}^{N-1} (f \star g)(k)e^{-2\pi kni/N} = \frac{1}{N}\sum_{k=0}^{N-1}\left(\frac{1}{N}\sum_{s=0}^{N-1} f(s)g(k-s)\right)e^{-2\pi kni/N} \\ &= \frac{1}{N^2}\sum_{s=0}^{N-1} f(s)\sum_{k=0}^{N-1} g(k-s)e^{-2\pi(s+k-s)ni/N} \\ &= \frac{1}{N^2}\sum_{s=0}^{N-1} f(s)e^{-2\pi sni/N}\sum_{k=0}^{N-1} g(k-s)e^{-2\pi(k-s)ni/N}.\end{aligned}$$

For s fixed, let $y = k - s$. Then as k runs through $\mathbb{Z}_N$, y runs through $\mathbb{Z}_N$ as well. Thus,

$$\sum_{k=0}^{N-1} g(k-s)e^{-2\pi(k-s)ni/N} = \sum_{y=0}^{N-1} g(y)e^{-2\pi yni/N} = N\hat{g}(n).$$

This gives us

$$\widehat{(f \star g)}(n) = \frac{1}{N^2}\sum_{s=0}^{N-1} f(s)e^{-2\pi sni/N}N\hat{g}(n) = \hat{g}(n)\frac{1}{N}\sum_{s=0}^{N-1} f(s)e^{-2\pi sni/N} = \hat{f}(n)\hat{g}(n)$$

as desired.

□

At this point, we observe that indexing our sums by $k = 0, 1, \ldots, N-1$ is not really accurate. Indeed, when we re-indexed the sums above, we did not actually use the integers $0, 1, \ldots, N-1$ (if we had, it would have been necessary to consider negative integers or integers greater than N in the new index). Instead, we simply used the fact that the sum was over all of the elements of a group. Thus, from here on, we will use the simple notation $\sum_k$ to denote the sum over all elements of $\mathbb{Z}_N$ (what we have, until now, been denoting by $\sum_{k=0}^{N-1}$). It will always be clear which N we are referring to when we use $\sum_k$.

The final result in discrete Fourier analysis that is a direct analogue of a result in the classical theory is Parseval's formula. In fact, Parseval's formula here is an easy consequence of the fact that the complex valued functions on $\mathbb{Z}_N$ form a finite dimensional vector space.

Theorem 1.8 (Parseval's formula)**.** *If $f : \mathbb{Z}_N \to \mathbb{C}$, then*

$$\sum_n |\hat{f}(n)|^2 = \frac{1}{N} \sum_k |f(k)|^2.$$

Proof. Given a function f, recall that the l^2 norm of f is defined by

$$||f||_2^2 = \langle f, f \rangle = \sum_k f(k)\overline{f(k)} = \sum_k |f(k)|^2.$$

We also know by the inversion formula that

$$f = \sum_n \hat{f}(n) e_n,$$

and consequently, we have

$$\begin{aligned} \langle f, f \rangle &= \left\langle \sum_n \hat{f}(n) e_n, \sum_k \hat{f}(k) e_k \right\rangle = \sum_n \sum_k \hat{f}(n)\overline{\hat{f}(k)} \langle e_n, e_k \rangle \\ &= \sum_n \hat{f}(n)\overline{\hat{f}(n)} \langle e_n, e_n \rangle = N \sum_n |\hat{f}(n)|^2, \end{aligned}$$

where we have used the fact that $\langle e_n, e_k \rangle$ is 0 if $n \neq k$ but is N if $n = k$. Therefore, we see that

$$\sum_k |f(k)|^2 = \langle f, f \rangle = N \sum_n |\hat{f}(n)|^2,$$

which is what we wanted to show.

□

It is interesting to note that the theory developed here is not specific to the group $\mathbb{Z}_N$. In fact, discrete Fourier analysis can be used on any finite abelian group G. In essence, the Fourier coefficients of a complex valued function on G are defined in a way analogous to what we did above. The primary difference is that instead of using the basis $\{e_n : n \in \mathbb{Z}_N\}$, we use the characters of G (a function $e : G \to \mathbb{C}$ is a character of G if $e(a \cdot b) = e(a)e(b)$ for all $a, b \in G$). It turns out that for any such group, the characters of G form a basis for the vector space of complex valued functions on G, just as $\{e_n : n \in \mathbb{Z}_N\}$ forms a basis for the vector space of complex valued functions on $\mathbb{Z}_N$. In fact, $\{e_n : n \in \mathbb{Z}_N\}$ is precisely the set of characters of $\mathbb{Z}_N$. A more complete explanation of this theory can be found in [23], but since it is not necessary for our purposes, we will not go into more detail. Instead, we will shift our attention to Szemerédi's Theorem and some concepts that Gowers uses in his proof of this important result.

Chapter 2

The Work of Roth and Gowers on Szemerédi's Theorem

Szemerédi's Theorem was initially conjectured by Erdös, Szekeres, and Turán in 1936 [6]. The fundamental question behind the conjecture is how dense an unstructured subset of $[1, N]$ can be. (Here, we use the notation $[x, y] = \{n \in \mathbb{Z} : x \leq n \leq y\}$. Whenever we use this interval notation, it will denote a set of integers, not an interval in $\mathbb{R}$.) Alternatively, for a fixed density, can we find $[1, N]$ long enough to guarantee that every subset with that density is structured? In order to answer these questions, we first need a notion of structure. Naturally, we consider a set of integers A to be structured if it contains long arithmetic progressions: sets of the form $\{a + qr : q = 0, 1, \ldots, n\}$ where $a \in \mathbb{Z}$ and $r, n \in \mathbb{N}$. For example, the set $\{2, 5, 8, 11, 14, 17, 20\}$ is an arithmetic progression because it can be written in the form $\{2 + 3q : q = 0, 1, \ldots, 6\}$. As a result, the set $A = \{2, 5, 6, 8, 11, 14, 15, 16, 17, 20, 26\}$ is fairly structured because it contains the arithmetic progressions $\{2, 5, 8, 11, 14, 17, 20\}$ and $\{2, 8, 14, 20, 26\}$ (among others), which are long compared to the size of A itself.

If $A = \{a + qr : q = 0, 1, \ldots, n\}$ is an arithmetic progression, then we say that the length of the progression is $|A| = n + 1$ and the common difference is r. Note that any one or two element set is trivially an arithmetic progression, so we disregard these types of progressions when considering the amount of structure in a given set. We will now introduce Szemerédi's Theorem in two (equivalent) ways. The first will be easier to state, but the second will give us more insight into the difficulty of the problem.

Suppose that k is a positive integer and $\delta > 0$. Is it possible to find N large enough so that any subset A of $[1, N]$ with size at least δN contains an arithmetic progression of length k? Here, we consider δ to be the density of A in $[1, N]$, and we are asking whether or not there is a number N large enough for which any subset of $[1, N]$ with density at least δ must contain a k-term progression. Szemerédi's Theorem answers this question in the affirmative.

Theorem 2.1 (Szemerédi's Theorem). *Let $k \in \mathbb{N}$ and $\delta > 0$. There is an $N_0 \in \mathbb{N}$ for which any subset $A \subset [1, N]$ with size at least δN contains an arithmetic progression of length k whenever $N \geq N_0$.*

Unfortunately, this form of Szemerédi's Theorem does not immediately tell us much about how difficult the problem is. To gain some insight into its level of difficulty, we consider the following similar problem. Given $N \in \mathbb{N}$, we wish to find a $\delta > 0$ for which every set $A \subset [1, N]$ of cardinality at least δN contains a long arithmetic progression (that is, every subset with high enough density has additive structure). In particular, we want to find a minimal such δ. This will then allow us to determine the highest density that an unstructured set may have.

Note first that if $\delta > \frac{N-1}{N}$, then the only subset of $[1, N]$ with cardinality at least δN is the entire set itself, so the problem is trivial. We can therefore suppose that $\delta \leq \frac{N-1}{N}$. Now consider the set $A = [1, N] \backslash \{\lceil \frac{N}{2} \rceil\}$. Observe that if $k \geq \lceil \frac{N}{2} \rceil$, then A cannot contain any arithmetic progression of length k and common difference at least 2 (such an arithmetic progression would have either an element greater than N or the element $\lceil \frac{N}{2} \rceil$). Also, the longest arithmetic progression of common difference 1 that A contains is precisely $\{1, \ldots, \lceil \frac{N}{2} \rceil - 1\}$. Hence, A does not contain any progression of length k if $k \geq \lceil \frac{N}{2} \rceil$. Therefore, when we are given a set $A \subset [1, N]$ and want to determine its structure, we are interested only in arithmetic progressions of length $k \leq \lfloor \frac{N}{2} \rfloor$.

Now, if N is given and $k \leq \lfloor \frac{N}{2} \rfloor$, it is easy to find a $0 < \delta < 1$ for which any subset $A \subset [1, N]$ with cardinality $|A| \geq \delta N$ contains an arithmetic progression of length k. For example, we can do the following. Let l the largest positive integer such that $k \leq \lfloor \frac{N}{l} \rfloor = m$, and since $k \leq \lfloor \frac{N}{2} \rfloor$, we know that $l \geq 2$. Partition $[1, N]$ into the disjoint intervals

$$[1, m], [m + 1, 2m], \ldots, [(l - 1)m + 1, lm], [lm + 1, N]$$

where the last interval in this list might be empty. In this list, there are exactly l intervals with length m. Let $\delta = \frac{N-l+1}{N}$, so if A has size at least δN, then the complement of A in $[1, N]$ contains at most $l - 1$ points. The pigeonhole principle then tells us that there is an interval $[jm + 1, (j + 1)m]$, where $0 \leq j \leq l - 1$, for which the complement of A is disjoint from $[jm + 1, (j + 1)m]$. Hence, $[jm + 1, (j + 1)m] \subset A$, so A contains an arithmetic progression of length $m = \lfloor \frac{N}{l} \rfloor \geq k$. And since we had $l \geq 2$, we know that $0 < \delta < 1$, as claimed. Note also that δ depends only on N and k.

Of course, our method for finding such a δ is incredibly crude. Since our δ in fact guarantees that A will contain an interval of length k, we should be able to find a much smaller δ that still guarantees an arithmetic progression of length k in A. Let us say that $\delta > 0$ is (N, k)-good if every $A \subset [1, N]$ of cardinality at least δN contains a progression of length k, and define

$$r_{N,k} = \inf(\delta > 0 : \delta \text{ is } (N, k)\text{-good}).$$

It is clear that if δ is $(N; k)$-good, then ϵ is also $(N; k)$-good whenever $\epsilon \geq \delta$. Thus, we see that $N \cdot r_{N,k}$ is the size of the largest subset of $[1, N]$ that does not contain an arithmetic progression of length k. In other words, $r_{N,k}$ is the highest density that a set in $[1, N]$ may have if it contains no progressions of length k. It is natural, now, to ask whether we can find good upper bounds on $r_{N,k}$. This is the substance of the second form of Szemerédi's Theorem.

Theorem 2.2 (Szemerédi's Theorem). *Let k be a positive integer. Then*

$$\lim_{N \to \infty} r_{N,k} = 0.$$

Before we justify the equivalence of the two statements, observe the following. If we let $\delta_{N,k}$ denote the $\delta > 0$ we found by our crude use of the pigeonhole principle above, then for each $k \geq 1$, we have $\lim_{N \to \infty} \delta_{N,k} = (k - 1)/k$. Thus, the conclusion of Szemerédi's

Theorem is much stronger than what we found. We now prove that the two forms of Szemerédi's Theorem are equivalent.

Proof. Assume that Theorem 2.2 is true, so for each positive integer k, we have $r_{N,k} \to 0$ as $N \to \infty$. Let $k \in \mathbb{N}$ and $\delta > 0$. Then there exists an N_0 large enough for which $r_{N,k} < \delta$ whenever $N \geq N_0$. Suppose that A is a subset of $[1, N]$ where $N \geq N_0$ and $|A| \geq \delta N$. If A does not contain an arithmetic progression of length k, then δ is not (N, k)-good. But this implies that $\delta \leq r_{N,k}$, which is a contradiction. Hence, A must contain a k-term progression, so Theorem 2.1 holds.

Assume now that Theorem 2.1 is true. Let k be a fixed positive integer, and let $\delta > 0$. Choose $N_0 \in \mathbb{N}$ such that any subset $A \subset [1, N]$ with size at least δN contains an arithmetic progression of length k whenever $N \geq N_0$. Then δ is (N, k)-good for each $N \geq N_0$, so we must have $r_{N,k} \leq \delta$ whenever $N \geq N_0$. Therefore, $\lim_{N\to\infty} r_{N,k} = 0$, so Theorem 2.2 holds. □

From here on, we will primarily consider Szemerédi's Theorem to be of the form in Theorem 2.1, but it is useful to keep its other form in mind.

Szemerédi first proved this result in 1975 using combinatorial arguments [25]. Other proofs followed, including a proof by Furstenberg in 1982 that primarily used ergodic theory [7], and a proof by Gowers in 2001 that used discrete Fourier analysis [9]. It is this latter proof that interests us. As mentioned at the beginning of this chapter, we will not present Gowers's entire proof of Szemerédi's Theorem. Instead, we will present his proof of the case where $k = 4$ [10]. Thus, we want to show that there is an $N_0 \in \mathbb{N}$ for which any subset $A \subset [1, N]$ with size at least δN contains an arithmetic progression of length four whenever $N \geq N_0$, or equivalently, that $\lim_{N\to\infty} r_{N,4} = 0$. We are content with presenting the proof of this case only because the methods used in it are very similar to the general proof but are easier to work with.

Gowers, however, was not the first person to approach Szemerédi's Theorem with analytic methods. In 1953, Roth proved the theorem for progressions of length three using discrete Fourier analysis [20]. In fact, the upper bound that Roth obtained on $r_{N,3}$ were much better than those obtained by Szemerédi and Furstenberg. As a result, it was natural to ask whether some of Roth's ideas could be extended to a general proof of Szemerédi's Theorem using discrete Fourier analysis. If so, it would be reasonable to expect that one could obtain much better bounds for general k than Szemerédi or Furstenberg had found. Gowers did precisely this in his 2001 proof. Before beginning our exposition on Gowers's proof, we give a sketch of Roth's proof for three-term progressions. This will familiarize readers with some of the basic ideas that Gowers uses in his work.

2.1 Roth's Work on Progressions of Length Three

First recall that for $f : \mathbb{Z}_N \to \mathbb{C}$, we define the discrete Fourier coefficients as

$$\hat{f}(s) = \frac{1}{N} \sum_x f(x) e^{-2\pi i x s / N}$$

for $s \in \mathbb{Z}_N$. A natural question to consider is how large these coefficients can be. We give a few examples that will motivate Roth's approach – an approach that relies heavily on the sizes of Fourier coefficients for certain functions.

Let f_0 be the function that is identically δ on $\mathbb{Z}_N$, where δ is a positive real number. Then, we have

$$\hat{f_0}(s) = \frac{\delta}{N}\sum_x e^{-2\pi ixs/N} = \begin{cases} \delta & \text{if } s = 0 \\ 0 & \text{if } s \neq 0 \end{cases}.$$

We see here that the Fourier coefficients are all small, with the exception of $\hat{f_0}(0)$. Often, we will not want to consider the 0-th coefficient, so we can say that all of the non-trivial Fourier coefficients are small. (The trivial coefficient is always the 0-th coefficient.)

Now let f be a "random" function on $\mathbb{Z}_N$. Namely, for each $x \in \mathbb{Z}_N$, let $f(x)$ be an independent random variable where

$$f(x) = \begin{cases} 1 & \text{with probability } \delta \\ 0 & \text{with probability } 1-\delta \end{cases}$$

where $0 < \delta < 1$ is fixed. We can think of f, then, as the characteristic function of a set of cardinality approximately δN. It is then true that with high probability, $\hat{f}(s)$ is close to $\hat{f_0}(s)$ for all $s \in \mathbb{Z}_N$, as long as N is large enough. More precisely, we can show that for any $\lambda > 0$,

$$\text{Prob}(|\hat{f}(s) - \hat{f_0}(s)| > \lambda \text{ for some } s \in \mathbb{Z}_N) \leq e^{-CN}$$

where C is a constant depending only on δ and λ. Thus, by taking N large enough, the probability that $\hat{f}(s)$ and $\hat{f_0}(s)$ are far apart for some s becomes arbitrarily small. We see then that for a random function defined in this way, all of the non-trivial Fourier coefficients are small (as long as N is large enough, of course).

We now concern ourselves with functions on $\mathbb{Z}_N$, where N is prime, that are defined using a subset of $\mathbb{Z}_N$. The easiest such functions are characteristic functions: if $A \subset \mathbb{Z}_N$, we define the characteristic function as

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

Note that we have identified the set A with its characteristic function. We will continue to do this throughout the next two chapters. Now, observe that for any $s \in \mathbb{Z}_N$,

$$\hat{A}(s) = \frac{1}{N}\sum_x A(x)e^{-2\pi isx/N} = \frac{1}{N}\sum_{x\in A} e^{-2\pi isx/N}.$$

By the triangle inequality, we then have

$$|\hat{A}(s)| \leq \frac{1}{N}|A| = \delta.$$

We therefore see that the Fourier coefficients of A can be no larger than δ. It is easy to see that when $s = 0$, we actually have $\hat{A}(0) = \delta$. A natural question then is whether or not there exists a nonzero $s \in A$ for which $|\hat{A}(s)|$ is close to δ. The answer to this question depends on the structure of A. In particular, if A is a "random" set in that its characteristic function behaves like a random function, then the non-trivial Fourier coefficients will all be small. But if A is not a random set, then there will be a non-trivial Fourier coefficient that is large compared to δ. One more example will serve to highlight this distinction.

Let $A = \{a, a+r, \ldots, a+(m-1)r\}$ be an arithmetic progression in $\mathbb{Z}_N$ of length m and common difference $r \neq 0$. Suppose also that N is large compared to m. Then A certainly does not resemble a random set since it is highly unlikely that m points, chosen from $\mathbb{Z}_N$ independently, would form an arithmetic progression. We claim that this lack of randomness forces the characteristic function of A to have at least one non-trivial Fourier coefficient that is large compared to $\delta = m/N$.

Since we are working under the assumption that N is prime, we can find an $s \in \mathbb{Z}_N$ such that $rs = 1$. It then turns out that $e^{-2\pi ixs/N}$ is almost constant on the progression A. Indeed, if $x \in A$, then $x = a + jr$ for some $0 \leq j \leq m$. We then have

$$\left|e^{-2\pi ixs/N} - e^{-2\pi ias/N}\right| = \left|e^{-2\pi i(a+jr)s/N} - e^{-2\pi ias/N}\right|$$

$$= \left|e^{-2\pi ijrs/N} - 1\right| = \left|e^{-2\pi ij/N} - 1\right| = \sqrt{2 - 2\cos(2\pi j/N)}$$

$$= 2\sin(\pi j/N) \leq \frac{2\pi j}{N} \leq \frac{2\pi m}{N} = 2\pi\delta$$

and recall that δ is small since N is much larger than m. Looking at the Fourier coefficient $\hat{A}(s)$, we then have

$$\hat{A}(s) = \frac{1}{N}\sum_{x\in A} e^{-2\pi ixs/N} = \frac{1}{N}\sum_{x\in A}\left(e^{-2\pi ia/N} + e^{-2\pi ixs/N} - e^{-2\pi ia/N}\right)$$

$$= \delta e^{-2\pi ia/N} + \frac{1}{N}\sum_{x\in A}\left(e^{-2\pi ixs/N} - e^{-2\pi ia/N}\right)$$

We can consider $\frac{1}{N}\sum_{x\in A}\left(e^{-2\pi ixs/N} - e^{-2\pi ia/N}\right)$ to be an error term because of what we found above:

$$\left|\frac{1}{N}\sum_{x\in A}\left(e^{-2\pi ixs/N} - e^{-2\pi ia/N}\right)\right| \leq \frac{1}{N}\sum_{x\in A}\left|e^{-2\pi ixs/N} - e^{-2\pi ia/N}\right|$$

$$\leq \frac{1}{N}\sum_{x\in A} 2\pi\delta = 2\pi\delta^2$$

Thus, the error term is bounded in absolute value by a constant times δ^2, which is quite small. As a result, $\hat{A}(s)$ is close to $\delta e^{-2\pi ia/N}$, so the size of this Fourier coefficient, $|\hat{A}(s)|$, is close to δ.

We now ask an inverse question to what we have just found. If a subset A in $\mathbb{Z}_N$ of size δN has a Fourier coefficient whose size is approximately δ, can we say something about the structure of A? It turns out that we can; the precise answer is given in the following lemma.

Lemma 2.1. *Let $A \subset \mathbb{Z}_N$ have cardinality $|A| = \delta N$, and suppose that there is a non-zero $s \in \mathbb{Z}_N$ such that $|\hat{A}(s)| > \epsilon$, where $\epsilon > 0$. Then there exists an arithmetic progression $P = \{a, a+r, \ldots, a+(m-1)r\}$ where m is approximately $\epsilon\sqrt{N}$ and $|A \cap P| \geq (\delta + \frac{\epsilon}{2})|P|$.*

Proof Sketch. Let s be the element in $\mathbb{Z}_N$ for which $|\hat{A}(s)| > \epsilon$. From what we did earlier, we might expect that if A has a large intersection with an arithmetic progression of common difference r, then $rs = 1$ modulo N. Indeed, that would explain why $|\hat{A}(s)|$ is large.

Unfortunately, since we are no longer working under the assumption that N is prime, such an r might not exist. To remedy this, we do the following. Choose $r \in \mathbb{Z}_N$ such that $1 \leq r \leq \sqrt{N}$ and $\left\{\frac{rs}{N}\right\} \leq \frac{1}{\sqrt{N}}$. Here, $\left\{\frac{rs}{N}\right\}$ denotes the fractional part of $\frac{rs}{N}$, which is equal to $\frac{rs}{N} - \lfloor\frac{rs}{N}\rfloor$. Note that such an r must exist by a simple pigeon-hole argument. It is entirely possible that rs is not equal to 1 modulo N, but we can consider rs to be much closer to 1 than it is to N.

We now want to cover $\mathbb{Z}_N$ by arithmetic progressions $P_j = \{a_j, a_j + r, \ldots, a_j + (m_j - 1)r\}$, each of common difference r and length approximately $\epsilon\sqrt{N}$. We also want these progressions to be pairwise disjoint. We now observe that the function $e^{-2\pi isx/N}$ is almost constant on P_j for each j. Although this is not a direct consequence of what we found earlier (precisely because we do not necessarily have $rs = 1$ modulo N), we can justify the observation in the same way as before, using the fact that $\left\{\frac{rs}{N}\right\} \leq \frac{1}{\sqrt{N}}$. It turns out that each $e^{-2\pi isx/N}$ is approximately equal to $e^{-2\pi isa_j/N}$ for $x \in P_j$.

We now consider the large Fourier coefficient $\hat{A}(s)$. By splitting sums, it can be shown that

$$\hat{A}(s) = \frac{1}{N}\sum_x (A(x) - \delta)\, e^{-2\pi ixs/N}.$$

(This representation of the non-trivial Fourier coefficients of a characteristic function will be important in Gowers's work.) We then have

$$\begin{aligned}
\epsilon < |\hat{A}(s)| &= \frac{1}{N}\left|\sum_x (A(x) - \delta)\, e^{-2\pi ixs/N}\right| \leq \frac{1}{N}\sum_j \left|\sum_{x\in P_j} (A(x) - \delta)\, e^{-2\pi ixs/N}\right| \\
&\approx \frac{1}{N}\sum_j \left|\sum_{x\in P_j} (A(x) - \delta)\, e^{-2\pi isa_j/N}\right| = \frac{1}{N}\sum_j |e^{-2\pi isa_j/N}| \left|\sum_{x\in P_j} (A(x) - \delta)\right| \\
&= \frac{1}{N}\sum_j \left|\sum_{x\in P_j} A(x) - \delta\right|
\end{aligned}$$

where the $\approx$ symbol represents an approximation. We therefore know that there exists a j for which

$$\left|\sum_{x\in P_j} A(x) - \delta\right| > \epsilon|P_j|.$$

But the sum on the left side is equal to $|A \cap P_j| - \delta|P_j|$, so we have

$$|A \cap P_j| > (\epsilon + \delta)|P_j|.$$

Note that this seems to be a stronger result than the lemma actually states, since the statement of the lemma has $\epsilon/2$ instead of ϵ. The need for $\epsilon/2$ is a consequence of approximating $e^{-2\pi ixs/N}$ by $e^{-2\pi isa_j/N}$.

□

We have, of course, omitted many details from this proof sketch. When we present Gowers's proof of Szemerédi's Theorem for progressions of length four, we will use many of the same ideas that we used here. There, we will provide all of the details that are necessary to justify each step. For now, though, Lemma 2.1 will allow us to give Roth's proof of Szemerédi's Theorem for progressions of length three.

Theorem 2.3 (Roth)**.** *Let $A \subset [1, N]$ have cardinality $|A| \geq \delta N$, where $\delta > \frac{c}{\log\log N}$. Then A contains an arithmetic progression of length three. Here, c is an absolute constant that does not depend on N.*

The idea of the proof of Roth's Theorem is to consider two cases: first, if A is a random set in that all of its Fourier coefficients are small; and second, if A has at least one large Fourier coefficient. In the first case, we will be able to show that A contains many arithmetic progressions of length three, and in the second case, we will be able to apply Lemma 2.1 to obtain an arithmetic progression P on which A has increased density. We can then iterate the argument, replacing A and $\mathbb{Z}_N$ with $A \cap P$ and P, respectively. The details are as follows.

Proof. We first work in the setting of $\mathbb{Z}_N$ rather than $\mathbb{Z}$, so for now, let $A \subset \mathbb{Z}_N$ with cardinality δN, and identify A with its characteristic function. Define

$$\Lambda(A) = \frac{1}{N^2} \sum_{x,r} A(x)A(x+r)A(x+2r).$$

Note that for each pair x, r we have

$$A(x)A(x+r)A(x+2r) = \begin{cases} 0 & \text{if } \{x, x+r, x+2r\} \not\subset A \\ 1 & \text{if } \{x, x+r, x+2r\} \subset A \end{cases}$$

Therefore, $N^2\Lambda(A)$ counts the number of three-term, mod-N arithmetic progressions in A. Included in this count, though, are trivial progressions – those with $r = 0$. Since there are $|A| = \delta N$ trivial progressions, we see that the number of non-trivial three-term progressions in A is $N^2\Lambda(A) - \delta N$.

Using Fourier arguments, it is possible to show that

$$\Lambda(A) = \sum_{s} \hat{A}(s)^2 \hat{A}(-2s).$$

In the following chapter, we will show and justify a similar identity, so we will not justify this identity here. We then have

$$\Lambda(A) = \hat{A}(0)^3 + \sum_{s \neq 0} \hat{A}(s)^2 \hat{A}(-2s).$$

We know from earlier that $\hat{A}(0) = \delta$, but we now wish to bound the term $\sum_{s\neq 0} \hat{A}(s)^2\hat{A}(-2s)$, which we consider to be an error term.

Recall that by Parseval's formula, $\sum_s |\hat{A}(s)|^2 = \frac{1}{N}\sum_x |A(x)|^2$. Thus,

$$\sum_{s} |\hat{A}(s)|^2 = \frac{1}{N}\delta N = \delta.$$

As a result, we see that

$$\left| \sum_{s \neq 0} \hat{A}(s)^2 \hat{A}(-2s) \right| \leq \max_{s \neq 0} |\hat{A}(s)| \cdot \sum_{s} |\hat{A}(s)|^2 = \delta \max_{s \neq 0} |\hat{A}(s)|.$$

We now consider two different cases. First assume that all non-trivial Fourier coefficients are small in that $|\hat{A}(s)| \leq \frac{\delta^2}{2}$ for each $s \neq 0$. This is essentially the case where A resembles a random set. We then have

$$\left| \sum_{s \neq 0} \hat{A}(s)^2 \hat{A}(-2s) \right| \leq \frac{\delta^3}{2},$$

and consequently,

$$|\Lambda(A)| \geq |\hat{A}(0)^3| - \left| \sum_{s \neq 0} \hat{A}(s)^2 \hat{A}(-2s) \right| \geq \delta^3 - \frac{\delta^3}{2} = \frac{\delta^3}{2}.$$

The number of non-trivial three-term progressions in A is therefore at least $\frac{\delta^3 N^2}{2} - \delta N$. But this quantity is strictly positive by our assumption that $\delta > \frac{c}{\log \log N}$, as long as $c \geq \sqrt{2}$. Thus, A contains at least one mod-N arithmetic progression of length three (of course, this might not be a valid progression in $\mathbb{Z}$ due to wrap-around in $\mathbb{Z}_N$).

For the second case (the case corresponding to A being a more structured set), suppose that $|\hat{A}(s)| > \frac{\delta^2}{2}$ for some $s \neq 0$. Then by Lemma 2.1, there exists an arithmetic progression $P = \{a, a+r, \ldots, a+(M-1)r\}$ of length approximately $\frac{\delta^2}{2}\sqrt{N}$ where $|A \cap P| \geq (\delta + \frac{\delta^2}{4})|P|$. Now, we iterate the argument, replacing A and $\mathbb{Z}_N$ by $A \cap P$ and P. (Even though P itself may not be of the form $[1, |P|]$, we can identify it with this interval and perform the iteration validly.) At each step in the iteration, the density of A inside an arithmetic progression increases. In fact, it increases in such a way that the iteration can be repeated only finitely-many times before this density becomes greater than one. Of course, the density of a set in an interval can never be greater than one, so at some step, the set that we have replaced A with must not have any large non-trivial Fourier coefficients. This set will then contain a three-term progression, and this gives a three-term progression in A. The maximum number of times that the iteration must be done to reach density one gives the bound $\delta > \frac{c}{\log \log N}$. We will not show the calculations needed to justify this here because Gowers's proof concludes with a similar statement that we will justify there.

We have therefore shown that if $A \subset \mathbb{Z}_N$ has cardinality $|A| \geq \delta N$, where $\delta > \frac{c}{\log \log N}$, then A contains a mod-N arithmetic progression of length three. Unfortunately, this does not immediately imply that the result extends when A is a subset of $[1, N]$. Indeed, the progression we found when we considered $A \subset \mathbb{Z}_N$ might wrap around the group and therefore may not be a progression in the interval $[1, N]$. Remedying this problem requires some technical work that we will have to do in Gowers's proof. But here, we will simply state that it can be done to prove Roth's Theorem.

□

It is easy to see that Roth's Theorem proves Szemerédi's Theorem for progressions of length three. Indeed, given $\delta > 0$, choose $N_0 > \exp\exp(c/\delta)$. Then, for each $N \geq N_0$, we have $\delta > c/(\log \log N)$, so by Roth's Theorem, any subset $A \subset [1, N]$ of size at least δN contains an arithmetic progression of length three. Alternatively, Roth's Theorem tells us that for each N, the density $c/\log \log N$ is $(N, 3)$-good. Thus, $r_{N,3} \leq c/\log \log N$, so as N approaches infinity, $r_{N,3}$ must approach zero.

2.2 Preliminaries for Gowers's Work on Progressions of Length Four

Gowers's work on Szemerédi's Theorem for progressions of length four follows the same basic ideas that Roth used in his work on progressions of length three. The fundamental concept is to use Fourier coefficients of the characteristic function of a set to determine whether the set is random or structured. If it is random, prove that it must contain a progression of length four; if it is structured, find a long progression on which the set has increased density so that we may iterate the argument. The details of Gowers's proof, however, are much more involved than those in Roth's proof. To deal with technicalities, Gowers introduces some new notation and redefines some traditional concepts in discrete Fourier analysis. Most of these new definitions are essentially the same as their traditional counterparts; Gowers simply modifies them so that they work better for his purposes. Before beginning his proof, though, it is necessary to introduce these changes.

In order to simplify notation, Gowers defines $\omega = e^{2\pi i/N}$ when N is clear from context. We will also use this notation, and each time it is used, the N we are using will be clear (in fact, we primarily use ω in sums that are taken over $\mathbb{Z}_N$). Gowers then defines the Fourier coefficients as follows.

Definition 7. If $f : \mathbb{Z}_N \to \mathbb{C}$, then the n-th Fourier coefficient of f is defined as

$$\tilde{f}(n) = \sum_k f(k)\omega^{-kn}.$$

Note that this new definition relates to the traditional definition by $\tilde{f}(n) = N\hat{f}(n)$ for each $n \in \mathbb{Z}_N$. As a result of this change, we have the following facts:

$$f(k) = \sum_n \hat{f}(n)e_n(k) = \sum_n \frac{1}{N}\tilde{f}(n)\omega^{-nk} = \frac{1}{N}\sum_n \tilde{f}(n)\omega^{-nk}$$

for each $k \in \mathbb{Z}_N$, and

$$\sum_k |f(k)|^2 = N\sum_n |\hat{f}(n)|^2 = N\sum_n |\frac{1}{N}\tilde{f}(n)|^2 = \frac{1}{N}\sum_n |\tilde{f}(n)|^2.$$

Acknowledging an abuse of terminology, we will call the former identity the inversion formula and the latter we will call Parseval's formula. Whenever we refer to these formulae, we will be referring to Gowers's version, not to the traditional version. The other primary abuse of terminology that we use following Gowers is the definition of convolution.

Definition 8. If $f, g : \mathbb{Z}_N \to \mathbb{C}$, then we define their convolution $f * g : \mathbb{Z}_N \to \mathbb{C}$ as

$$(f * g)(n) = \sum_k f(k)\overline{g(k-n)}$$

for $n \in \mathbb{Z}_N$.

Despite some important differences between Gowers's definition of convolution and the traditional definition, all of the algebraic properties of the traditional convolution have analogues for Gowers's convolution. These are given in the following proposition.

Proposition 2.1. *For $f, g, h : \mathbb{Z}_N \to \mathbb{C}$, the following are true:*

i. $f * (g + h) = (f * g) + (f * h)$

ii. $(cf) * g = c(f * g)$ *if* $c \in \mathbb{C}$

iii. $(f * g) * h = Nf * (g \star h)$

iv. $(f * g)(n) = (\overline{g} * \overline{f})(-n)$ *for each* $n \in \mathbb{Z}_N$

v. $\widetilde{f * g} = \tilde{f} \cdot \overline{\tilde{g}}$

Proof. As in Proposition 1.1, parts (i) and (ii) follow directly from the definition of the convolution and the fact that $\overline{g + h} = \overline{g} + \overline{h}$. We prove the other three parts primarily by interchanging and re-indexing sums as we did before.
(iii): For $n \in \mathbb{Z}_N$, we have

$$\begin{aligned}
[(f * g) * h]\,(n) &= \sum_k (f * g)(k)\overline{h(k-n)} = \sum_k \left(\sum_t f(t)\overline{g(t-k)} \right) \overline{h(k-n)} \\
&= \sum_t f(t) \left(\sum_k \overline{g(t-k)h(k-n)} \right) = \sum_t f(t) \left(\sum_y \overline{g(y)h(t-n-y)} \right) \\
&= \sum_t f(t) N\overline{(g \star h)(t-n)} = N\,[f * (g \star h)]\,(n),
\end{aligned}$$

where we have used the substitution $y = t - k$.
(iv): For $n \in \mathbb{Z}_N$, we have

$$\begin{aligned}
(f * g)(n) &= \sum_k f(k)\overline{g(k-n)} = \sum_y f(y+n)\overline{g(y)} \\
&= \sum_y \overline{g}(y)\overline{\overline{f}}(y - (-n)) = (\overline{g} * \overline{f})(-n),
\end{aligned}$$

where this time, we have used the substitution $y = k - n$.
(v): For $n \in \mathbb{Z}_N$, we have

$$\begin{aligned}
\widetilde{f * g}(n) &= \sum_k (f * g)(k)\omega^{-kn} = \sum_k \left(\sum_t f(t)\overline{g(t-k)} \right) \omega^{-kn} \\
&= \sum_t \sum_k f(t)\overline{g(t-k)}\omega^{(-t+t-k)n} = \sum_t f(t)\omega^{-tn} \left(\sum_k \overline{g(t-k)}\omega^{(t-k)n} \right) \\
&= \sum_t f(t)\omega^{-tn} \left(\sum_y \overline{g(y)}\omega^{yn} \right) = \sum_t f(t)\omega^{-tn} \left(\sum_y \overline{g(y)\omega^{-yn}} \right) \\
&= \sum_t f(t)\omega^{-tn} \left(\overline{\tilde{g}(n)} \right) = \tilde{f}(n)\overline{\tilde{g}(n)},
\end{aligned}$$

where we have again used the substitution $y = t - k$.

□

The only property of the convolution that we will use directly in the proof of Szemerédi's Theorem is (v); but it is useful to know the other properties to understand the convolution operation better.

It is also important to introduce a characteristic-type function that is associated with a given subset of $\mathbb{Z}_N$. Namely, if $A \subset \mathbb{Z}_N$ has cardinality δN, we define the balanced function of A to be

$$f_A(n) = A(n) - \delta = \begin{cases} -\delta & \text{if } n \notin A \\ 1-\delta & \text{if } n \in A. \end{cases}$$

where, as before, we have identified the set A with its characteristic function. Observe now that for any subset A of $\mathbb{Z}_N$ with $|A| = \delta N$, we have

$$\begin{aligned} \tilde{f_A}(0) &= \sum_k f_A(k)\omega^{-k\cdot 0} = \sum_k f_A(k) = \sum_{k\notin A} f_A(k) + \sum_{k\in A} f_A(k) \\ &= (N-\delta N)(-\delta) + (\delta N)(1-\delta) = 0. \end{aligned}$$

Also, we see that for any $n \neq 0$,

$$\begin{aligned} \tilde{f_A}(n) &= \sum_k f_A(k)\omega^{-kn} = \sum_{k\notin A} -\delta\omega^{-kn} + \sum_{k\in A}(1-\delta)\omega^{-kn} \\ &= -\delta\sum_{k\notin A}\omega^{-kn} - \delta\sum_{k\in A}\omega^{-kn} + \sum_{k\in A}\omega^{-kn} = -\delta\sum_k \omega^{-kn} + \sum_{k\in A}\omega^{-kn} \\ &= \sum_{k\in A}\omega^{-kn} = \sum_k A(k)\omega^{-kn} = \tilde{A}(n). \end{aligned}$$

Thus, the Fourier coefficients of A and f_A coincide, except for $n = 0$, where the coefficient of f_A vanishes. Because of this, it will be more useful for us to analyze the balanced function than the characteristic function. Indeed, to determine the additive structure of a set A, we want to know how large the non-trivial Fourier coefficients of A are, and it is generally easier to consider $\max_{n\in\mathbb{Z}_N}|\tilde{f_A}(n)|$ than $\max_{n\neq 0}|\tilde{A}(n)|$. Certainly, both quantities are equal. We therefore use balanced functions, rather than characteristic functions, to study the structure of sets.

In order to estimate the size of certain quantities (usually sums of Fourier coefficients), we will use two classical inequalities often: Hölder's inequality (and its consequence, the Cauchy-Schwarz inequality) and Minkowski's inequality.

Theorem 2.4 (Hölder). *If a_n and b_n are real numbers for all $1 \leq n \leq N$ and $p, q > 1$ satisfy $1/p + 1/q = 1$, then*

$$\left|\sum_{n=1}^N a_n b_n\right| \leq \left(\sum_{n=1}^N |a_n|^p\right)^{1/p}\left(\sum_{n=1}^N |b_n|^q\right)^{1/q}.$$

The Cauchy-Schwarz inequality is the inequality obtained by setting $p = q = 2$.

Theorem 2.5 (Minkowski). *If a_n and b_n are real numbers for all $1 \leq n \leq N$ and $p \geq 1$, then*

$$\left(\sum_{n=1}^N |a_n + b_n|^p\right)^{1/p} \leq \left(\sum_{n=1}^N |a_n|^p\right)^{1/p} + \left(\sum_{n=1}^N |b_n|^p\right)^{1/p}.$$

We make one more note before proceeding to Gowers's proof. In the section on discrete Fourier analysis earlier, we worked in the inner product space of complex valued functions on $\mathbb{Z}_N$, where the inner product was defined by

$$\langle f, g\rangle = \sum_k f(k)\overline{g(k)}.$$

This induces the l^2 norm

$$||f||_2 = \sqrt{\sum_k |f(k)|^2}.$$

As we will see, this norm is important in Gowers's proof of Szemerédi's Theorem, but two other norms are important as well – the l^1 norm and the uniform (or l_∞) norm. The l^1 norm is defined by

$$||f||_1 = \sum_k |f(k)|$$

for $f : \mathbb{Z}_N \to \mathbb{C}$, and the l_∞ norm is defined by

$$||f||_\infty = \max_{k\in\mathbb{Z}_N} |f(k)|.$$

These will be the only three norms needed here, and they will be distinguished by their subscripts.

Chapter 3

Gowers's Proof of Szemerédi's Theorem for Progressions of Length Four

In the previous section, we gave a short outline of the strategy in Gowers's proof. In particular, we will use the dichotomy between structure and randomness in subsets of $[1, N]$ that naturally arises by considering sizes of Fourier coefficients. We now give a more detailed outline of Gowers's methods that will serve as a guide for where we are going. First, it is necessary to define the following concept.

Definition 9. Let $f : \mathbb{Z}_N \to \mathbb{D}$ be a function to the closed unit disk in $\mathbb{C}$. Then f is α-uniform if

$$\sum_r |\tilde{f}(r)|^4 \leq \alpha N^4.$$

Our first task in the proof of Szemerédi's Theorem will be to give equivalent definitions of α-uniformity. As we shall see, a function f is α-uniform if and only if there is an upper bound (a constant depending on α, multiplied by N) on the sizes of the Fourier coefficients of f. Therefore, we can say that f is α-uniform for a small α if and only if f behaves like a random function.

We now want to use α-uniformity of functions to say something about subsets of $\mathbb{Z}_N$. Recall from earlier that we consider a set $A \subset \mathbb{Z}_N$ to be random if all of the non-trivial Fourier coefficients of its characteristic function are small. We also remarked, however, that it will be more convenient to use the balanced function of A rather than the characteristic function. Thus, if all Fourier coefficients of the balanced function of A are small, then A is, in some sense, random. We therefore say that A is α-uniform if and only if its balanced function is α-uniform. As a result, we see that A is random if it is α-uniform for small enough α.

From the proof of Roth's Theorem, we know that if $A \subset \mathbb{Z}_N$ is α-uniform for a sufficiently small α, then A must contain a progression of length three. Unfortunately, α-uniformity does not guarantee a progression of length four (see [8], Section 3, for a short discussion about why uniformity is not strong enough). As a result, we will need a stronger notion of randomness. The following definition gives us just that.

Definition 10. Let $f : \mathbb{Z}_N \to \mathbb{D}$ be a function to the unit disk in $\mathbb{C}$. Then f is quadratically α-uniform if

$$\sum_u \sum_v \left| \sum_s f(s)\overline{f(s-u)f(s-v)}f(s-u-v) \right|^2 \leq \alpha N^4.$$

Note here that if we define $\Delta(f;k)(s) = f(s)\overline{f(s-k)}$, then the sum on the left-hand side in Definition 10 is equivalent to

$$\sum_u \sum_v \left| \left[\Delta(f;u) * \Delta(f;u)\right](v) \right|^2 .$$

The collection of functions $\Delta(f;k)$ for $k \in \mathbb{Z}_N$ will play an important role in Gowers's proof.

As with uniformity, it will be necessary for us to find equivalent definitions of quadratic α-uniformity so that we can relate this concept to the sizes of Fourier coefficients. We shall see that a function f is quadratically α-uniform if and only if most of the Fourier coefficients of $\Delta(f;k)$, as k ranges over $\mathbb{Z}_N$, are small. Of course, the notion of "smallness" and "most of" will depend on α. We can therefore consider f to be quadratically α-uniform if and only if the collection of functions $\Delta(f;k)$ are somewhat random.

We now say that a set $A \subset \mathbb{Z}_N$ is quadratically α-uniform if its balanced function is quadratically α-uniform. For small enough α, we can then think of quadratically α-uniform sets as pseudo-random sets. We shall see that unlike α-uniformity, quadratic α-uniformity will guarantee a progression of length four, as long as α is sufficiently small. Gowers's general argument goes as follows.

Let A be a subset of $\mathbb{Z}_N$. First, suppose that A is quadratically α-uniform for a sufficiently small α. Our main goal here is first to prove that there are many "possible" arithmetic progressions, and as a result, at least one of them must be an actual arithmetic progression. Namley, suppose that we can guarantee many elements of the form $(a, a+d, a+2d, a+3d)$ in the set $A \cap [2N/5, 3N/5) \times A \cap [2N/5, 3N/5) \times A \times A$ where $d \in \mathbb{Z}_N$. These are the "possible" arithmetic progressions because if $d \neq 0$, then the quadruple corresponds to an arithmetic progression in $\mathbb{Z}_N$. Observe, though, that since we restrict $a \in A \cap [2N/5, 3N/5)$, there can be at most δN quadruples where $d = 0$. We can therefore conclude that if there are strictly more than δN quadruples $(a, a+d, a+2d, a+3d)$, then there is at least one with $d \neq 0$. This gives us a mod-N progression in A. But also notice that since a and $a+d$ are in $[2N/5, 3N/5)$, we have $a+2d$ and $a+3d$ in $[0, N)$. Hence, the arithmetic progression in $\mathbb{Z}_N$ is actually an arithmetic progression in $\mathbb{Z}$. The main difficulty in this case, of course, is showing that there are many elements of the form $(a, a+d, a+2d, a+3d)$ in $A \cap [2N/5, 3N/5) \times A \cap [2N/5, 3N/5) \times A \times A$. Most of the ideas we use to do this will parallel the ideas developed in Roth's proof.

Now suppose the second case; that is, A is not quadratically α-uniform for a sufficiently small α. This is certainly the more difficult case, and it requires some heavy machinery such as Freiman's Theorem and a discrete version of Weyl's Equidistribution Theorem. A helpful concept in this part of the proof is that of additive quadruples. Namely, if $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$, then we call a quadruple $(a, b, c, d) \in \mathbb{Z}_N^4$ additive if $a + b = c + d$ and $\phi(a) + \phi(b) = \phi(c) + \phi(d)$.

An overview of the approach is as follows. As we shall see, the failure of A to be quadratically α-uniform implies that there is some set $B \subset \mathbb{Z}_N$ with $|B| \geq \alpha N/2$ and a function $\phi : B \to \mathbb{Z}_N$ such that ϕ has many additive quadruples $(a,b,c,d) \in B^4$. We then use this fact to show that there is an arithmetic progression P in $\mathbb{Z}_N$ such that for any $s \in \mathbb{Z}_N$, the progression $P+s$ can be partitioned into smaller progressions $P_{s1}, \ldots, P_{sm}$, where each of these smaller progressions is actually an arithmetic progression in $\mathbb{Z}$. We then want to show that for some s, at least one of the P_{sj} has a large intersection with A (that is, the density of A in P_{sj} is greater than the density of A in $[1,N]$) just as we did in the proof of Roth's Theorem.

Once we have obtained this P_{sj}, we iterate the argument. Namely, we replace A and $[1,N]$ with $A \cap P_{sj}$ and P_{sj}, respectively, and we use the fact that the density of $A \cap P_{sj}$ in P_{sj} is higher than the original density of A in $[1,N]$. In fact, the increase in density that we find will allow us to conclude that this argument can be iterated only a finite number of times before the density exceeds one. As a result, at some step in the iteration, the smaller set $A \cap P_{sj}$ will be quadratically α-uniform for a sufficiently small α. By the first case, then, $A \cap P_{sj}$ (and hence, A itself) must contain an arithmetic progression of length four.

We now begin Gowers's proof of Szemerédi's Theorem for progressions of length four. We will break the proof into sections for ease.

3.1 Quadratic Uniformity

We first want to prove some helpful results regarding α-uniform and quadratically α-uniform sets. The properties we find will give us a straightforward proof that if A is quadratically α-uniform for a sufficiently small α, then A must contain a four-term arithmetic progression. We begin with a lemma that gives equivalent definitions for α-uniformity. In this lemma, when we say that one bound depending on c_i implies another bound depending on c_j, we mean that c_j is an absolute constant multiplied by some power of c_i.

Lemma 3.1. *Let $f : \mathbb{Z}_N \to \mathbb{D}$ be a function to the closed unit disk in $\mathbb{C}$. Then the following are equivalent.*

1. $\sum_r \left|\tilde{f}(r)\right|^4 \leq c_1 N^4$
2. $\max_r \left|\tilde{f}(r)\right| \leq c_2 N$
3. $\sum_k \left|\sum_s f(s)\overline{f(s-k)}\right|^2 \leq c_3 N^3$
4. $\sum_k \left|\sum_s f(s)\overline{g(s-k)}\right|^2 \leq c_4 N^2 ||g||_2^2$ *for each function $g : \mathbb{Z}_n \to \mathbb{C}$*

Proof. First, note that for any $f, g : \mathbb{Z}_N \to \mathbb{D}$, we have

$$\sum_k \left|\sum_s f(s)\overline{g(s-k)}\right|^2 = \sum_k |(f*g)(k)|^2 = \frac{1}{N}\sum_k \left|\widetilde{(f*g)}(k)\right|^2$$
$$= \frac{1}{N}\sum_k \left|\tilde{f}(k)\overline{\tilde{g}(k)}\right|^2 = \frac{1}{N}\sum_k \left|\tilde{f}(k)\right|^2 |\tilde{g}(k)|^2.$$

(1) ⇔ (3): Since $\sum_k \left|\sum_s f(s)\overline{f(s-k)}\right|^2 = \frac{1}{N}\sum_k \left|\tilde{f}(k)\right|^4$ by what we just found, it is immediate that $\sum_r \left|\tilde{f}(r)\right|^4 \leq c_1 N^4$ if and only if $\sum_k \left|\sum_s f(s)\overline{f(s-k)}\right|^2 \leq c_3 N^3$ where $c_1 = c_3$.

(4) ⇒ (3): Observe that since $|f(s)| \leq 1$ for all $s \in \mathbb{Z}_N$, we have $||f||_2^2 = \sum_r |f(r)|^2 \leq N$. Hence, we see that

$$\sum_k \left| \sum_s f(s)\overline{f(s-k)} \right|^2 \leq c_4 N^2 ||f||_2^2 \leq c_4 N^3 \leq c_3 N^3$$

if $c_4 \leq c_3$. Note that we have used the fact that $||f||_2^2 \leq N$.

(1) ⇒ (4) : Using the Cauchy-Schwarz inequality and Parseval's formula, we have

$$\begin{aligned}\sum_k \left| \sum_s f(s)\overline{g(s-k)} \right|^2 &= \frac{1}{N} \sum_k \left|\tilde{f}(k)\right|^2 |\tilde{g}(k)|^2 \leq \frac{1}{N} \left(\sum_k \left|\tilde{f}(k)\right|^4 \right)^{1/2} \left(\sum_k |\tilde{g}(k)|^4 \right)^{1/2} \\ &\leq \frac{1}{N} \sqrt{c_1} N^2 \left(\sum_k |\tilde{g}(k)|^4 \right)^{1/2} \leq N\sqrt{c_1} \left(\sum_k |\tilde{g}(k)|^2 \right) \\ &= N\sqrt{c_1} \left(N \sum_k |g(k)|^2 \right) = \sqrt{c_1} N^2 ||g||_2^2 \leq c_4 N^2 ||g||_2^2\end{aligned}$$

as long as $c_1 \leq c_4^2$.

We now know that (1), (3), and (4) are equivalent since (3) ⇒ (1) ⇒ (4) ⇒ (3). Thus, it is sufficient to show that (1) and (2) are equivalent.

(1) ⇒ (2): Trivially, we have $\left(\max_r \left|\tilde{f}(r)\right|\right)^4 \leq \sum_r \left|\tilde{f}(r)\right|^4$, so

$$\max_r \left|\tilde{f}(r)\right| \leq \left(\sum_r \left|\tilde{f}(r)\right|^4 \right)^{1/4} \leq \left(c_1 N^4\right)^{1/4} \leq c_2 N$$

as long as $c_1 \leq c_2^4$.

(2) ⇒ (1): Using Parseval's formula and the fact that $||f||_2^2 \leq N$, we have

$$\begin{aligned}\sum_r \left|\tilde{f}(r)\right|^4 &\leq \max_r \left|\tilde{f}(r)\right|^2 \sum_r \left|\tilde{f}(r)\right|^2 \leq (c_2 N)^2 \sum_r \left|\tilde{f}(r)\right|^2 \\ &= c_2^2 N^2 \left(N \sum_r |f(r)|^2 \right) = c_2^2 N^3 ||f||_2^2 \leq c_2^2 N^4 \leq c_1 N^4\end{aligned}$$

as long as $c_1 \geq c_2^2$.

Therefore, (1) is equivalent to (2), so the proof is complete.

□

Recall that we say f is α-uniform if condition (1) holds for $\alpha = c_1$, and a set $A \subset \mathbb{Z}_N$ is α-uniform if its balanced function is. The following is an easy fact about sums of uniform functions.

Lemma 3.2. *Suppose that for $1 \leq i \leq k$, $f_i : \mathbb{Z}_N \to \mathbb{D}$ is α_i-uniform. Then the sum $f_1 + \ldots + f_k$ is $(\alpha_1^{1/4} + \ldots + \alpha_k^{1/4})^4$-uniform.*

Proof. First note that by definition,

$$(f_1 \widetilde{+ \ldots +} f_k)(r) = \sum_s (f_1 + \ldots + f_k)(s)\omega^{-rs} = \sum_s f_1(s)\omega^{-rs} + \ldots + \sum_s f_k(s)\omega^{-rs}$$
$$= \tilde{f}_1(r) + \ldots + \tilde{f}_k(r)$$

for each $r \in \mathbb{Z}_N$. Thus, we have

$$\sum_r \left|(f_1 \widetilde{+ \ldots +} f_k)(r)\right|^4 \leq \sum_r \left(|\tilde{f}_1(r)| + \ldots + |\tilde{f}_k(r)|\right)^4.$$

As a result, using Minkowski's inequality and the fact that each f_i is α_i-uniform, we see that

$$\left[\sum_r \left|(f_1 \widetilde{+ \ldots +} f_k)(r)\right|^4\right]^{1/4} \leq \left[\sum_r \left(|\tilde{f}_1(r)| + \ldots + |\tilde{f}_k(r)|\right)^4\right]^{1/4}$$
$$\leq \left(\sum_r |\tilde{f}_1(r)|^4\right)^{1/4} + \ldots + \left(\sum_r |\tilde{f}_k(r)|^4\right)^{1/4}$$
$$\leq (\alpha_1 N^4)^{1/4} + \ldots + (\alpha_k N^4)^{1/4} = N(\alpha_1^{1/4} + \ldots + \alpha_k^{1/4}).$$

Hence, we have

$$\sum_r \left|(f_1 \widetilde{+ \ldots +} f_k)(r)\right|^4 \leq (\alpha_1^{1/4} + \ldots + \alpha_k^{1/4})^4 N^4,$$

so $f_1 + \ldots + f_k$ is $(\alpha_1^{1/4} + \ldots + \alpha_k^{1/4})^4$-uniform. □

In the proof of Roth's Theorem, we saw that α-uniformity for small enough α was enough to guarantee that a set had many three-term progressions in $\mathbb{Z}_N$. As we discussed earlier, it is necessary to strengthen our notion of randomness to deal with progressions of length four. We now present the concept of quadratic uniformity more formally.

Recall that given a function $f : \mathbb{Z}_N \to \mathbb{C}$ and $k \in \mathbb{Z}_N$, we define $\Delta(f;k) : \mathbb{Z}_N \to \mathbb{C}$ by $\Delta(f;k)(s) = f(s)\overline{f(s-k)}$. We have the following result, where equivalence has the same definition as it did in Lemma 3.1.

Lemma 3.3. *Let $f : \mathbb{Z}_N \to \mathbb{D}$ be a function to the closed unit disk in $\mathbb{C}$. Then the following are equivalent:*

1. $\sum_u \sum_v \left|\sum_s f(s)\overline{f(s-u)f(s-v)}f(s-u-v)\right|^2 \leq c_1 N^4$
2. $\sum_k \sum_r \left|\widetilde{\Delta(f;k)}(r)\right|^4 \leq c_2 N^5$
3. $\left|\widetilde{\Delta(f;k)}(r)\right| \geq c_3 N$ *for at most* $c_3^2 N$ *pairs* (k, r)
4. *For all but $c_4 N$ values of k, the function $\Delta(f;k)$ is c_4-uniform*

Proof. (1) ⇔ (2): Using Parseval's formula and the multiplicative property of the convolution, we have the following:

$$\begin{aligned}\sum_u\sum_v\left|\sum_s f(s)\overline{f(s-u)f(s-v)}f(s-u-v)\right|^2 &= \sum_u\sum_v\left|\sum_s \Delta(f;u)(s)\overline{\Delta(f;u)(s-v)}\right|^2\\ = \sum_u\sum_v\left|\,[\Delta(f;u)*\Delta(f;u)]\,(v)\right|^2 &= \sum_u\frac{1}{N}\sum_v\left|[\widetilde{\Delta(f;u)*\Delta(f;u)}](v)\right|^2\\ = \sum_u\frac{1}{N}\sum_v\left|\widetilde{\Delta(f;u)}(v)\overline{\widetilde{\Delta(f;u)}(v)}\right|^2 &= \frac{1}{N}\sum_u\sum_v\left|\widetilde{\Delta(f;u)}(v)\right|^4,\end{aligned}$$

so the equivalence between (1) and (2) follows immediately if $c_1 = c_2$.

(2) ⇒ (3): Suppose that $\left|\widetilde{\Delta(f;k)}(r)\right| \geq c_3N$ for more than c_3^2N pairs (k,r), so $\left|\widetilde{\Delta(f;k)}(r)\right|^4 \geq c_3^4N^4$ for more than c_3^2N pairs (k,r). Then

$$\sum_k\sum_r\left|\widetilde{\Delta(f;k)}(r)\right|^4 > (c_3^4N^4)(c_3^2N) = c_3^6N^5 \geq c_2N^5$$

if $c_2 \leq c_3^6$. By taking the contrapositive of this, we see that (2) implies (3) if $c_2 \leq c_3^6$.

(3) ⇒ (2): Again, we prove this implication by proving the contrapositive. Observe first that for any r and k in $\mathbb{Z}_N$, we have

$$\sum_r\left|\widetilde{\Delta(f;k)}(r)\right|^4 = N\sum_v\left|\sum_s f(s)\overline{f(s-u)f(s-v)}f(s-u-v)\right|^2$$

by the proof of (1) ⇔ (2). Using the triangle inequality and the fact that f takes values in the unit disk, we then have

$$\begin{aligned}\sum_r\left|\widetilde{\Delta(f;k)}(r)\right|^4 &\leq N\sum_v\left(\sum_s |f(s)||\overline{f(s-u)}||\overline{f(s-v)}||f(s-u-v)|\right)^2\\ &\leq N\sum_v\left(\sum_s 1\right)^2 = N^4.\end{aligned}$$

We now claim that if (2) does not hold, then there are more than $c_2N/2$ values of k such that $\sum_r\left|\widetilde{\Delta(f;k)}(r)\right|^4 > c_2N^4/2$. Indeed, if there were no more than $c_2N/2$ such k, then we would have at least $N - c_2N/2$ values of k for which $\sum_r\left|\widetilde{\Delta(f;k)}(r)\right|^4 \leq c_2N^4/2$. Hence,

$$\begin{aligned}\sum_k\sum_r\left|\widetilde{\Delta(f;k)}(r)\right|^4 &\leq (N - c_2N/2)(c_2N^4/2) + (c_2N/2)(N^4)\\ &= (c_2 - c_2^2/4)N^5 \leq c_2N^5.\end{aligned}$$

so (2) would hold. Therefore, we have more than $c_2N/2$ values of k such that

$$\sum_r\left|\widetilde{\Delta(f;k)}(r)\right|^4 > c_2N^4/2.$$

For each such k, Lemma 3.1 tells us that since $\sum_r \left|\widetilde{\Delta(f;k)}(r)\right|^4 > c_2 N^4/2$, we must have $\max_r \left|\widetilde{\Delta(f;k)}(r)\right| > (c_2/2)^{1/2} N$. Thus, if $c_2 \geq 2c_3^2$, then there are more than $c_3^2 N$ values of k for which there is an r such that $\left|\widetilde{\Delta(f;k)}(r)\right| > c^3 N$. Hence, (3) does not hold, as desired. We conclude that (3) implies (2) if $c_2 \geq 2c_3^2$.

We now have equivalence among (1), (2), and (3). It suffices to show that (2) and (4) are equivalent.

(4) ⇒ (2): If $\Delta(f;k)$ is c_4-uniform for all but $c_4 N$ values of k, then there are at least $N - c_4 N$ values of k for which $\sum_r \left|\widetilde{\Delta(f;k)}(r)\right|^4 \leq c_4 N^4$. Recall also that for the other $c_4 N$ values of k, we still have the bound $\sum_r \left|\widetilde{\Delta(f;k)}(r)\right|^4 \leq N^4$. Hence,

$$\sum_k \sum_r \left|\widetilde{\Delta(f;k)}(r)\right|^4 \leq (N - c_4 N)(c_4 N^4) + (c_4 N)(N^4) = (2c_4 - c_4^2)N^5 \leq c_2 N^5$$

as long as $c_2 \geq 2c_4$.

(2) ⇒ (4): We prove the contrapositive again. If (4) does not hold, then there are more than $c_4 N$ values of k such that $\sum_r \left|\widetilde{\Delta(f;k)}(r)\right|^4 > c_4 N^4$. Thus,

$$\sum_k \sum_r \left|\widetilde{\Delta(f;k)}(r)\right|^4 > (c_4 N)(c_4 N^4) = c_4^2 N^5 \geq c_2 N^5$$

as long as $c_2 \leq c_4^2$. Therefore, (2) implies (4) if $c_2 \leq c_4^2$.

Hence, we see that (2) and (4) are equivalent, so the proof is complete.

□

Recall that we say f is quadratically α-uniform if condition (1) in Lemma 3.3 holds for $\alpha = c_1$, and a set $A \subset \mathbb{Z}_N$ is quadratically α-uniform if its balanced function is. Above, we claimed that quadratic uniformity is somehow a stronger condition than uniformity. The next two lemmas formalize this claim.

Lemma 3.4. *If f is quadratically α-uniform and real-valued, then f is $\alpha^{1/2}$-uniform.*

Proof. If f is quadratically α-uniform, then we have the following:

$$\sum_k \left|\sum_s f(s)f(s-k)\right|^2 = \sum_k \left|\left(\sum_s f(s)f(s-k)\right)^2\right|$$

$$= \sum_k \left|\sum_{u,s} f(s)f(s-k)f(s-u)f(s-u-k)\right|$$

by expansion and the fact that $s-u$ runs through $\mathbb{Z}_N$ as u varies

$$\leq \sum_k \sum_u \left|\sum_s f(s)f(s-k)f(s-u)f(s-u-k)\right|$$

by the triangle inequality

$$\leq \sum_k \left[\left(\sum_u \left| \sum_s f(s)f(s-k)f(s-u)f(s-u-k) \right|^2 \right)^{1/2} (N)^{1/2} \right]$$

by the Cauchy-Schwarz inequality

$$= N^{1/2} \sum_k \left(\sum_u \left| \sum_s f(s)f(s-k)f(s-u)f(s-u-k) \right|^2 \right)^{1/2}$$

$$\leq N^{1/2} \left(\sum_k \sum_u \left| \sum_s f(s)f(s-k)f(s-u)f(s-u-k) \right|^2 \right)^{1/2} (N)^{1/2}$$

again by the Cauchy-Schwarz inequality

$$= N \left(\sum_k \sum_u \left| \sum_s f(s)f(s-k)f(s-u)f(s-u-k) \right|^2 \right)^{1/2}$$

$$\leq N(\alpha N^4)^{1/2}$$

since f is real-valued and quadratically α-uniform

$$= \alpha^{1/2} N^3.$$

Hence, by (3) in Lemma 3.1, f is $\alpha^{1/2}$-uniform as desired. □

Lemma 3.5. *Let $A \subset N$ have cardinality δN, and suppose that A is quadratically α-uniform. Then for all but at most $\alpha^{1/2}N$ values of k, $A \cap (A+k)$ is $81\alpha^{1/2}$-uniform. Furthermore, for all but at most $\alpha^{1/4}N$ values of k, $\big||A \cap (A+k)| - \delta^2 N\big| \leq \alpha^{1/8}N$.*

Proof. Let f be the balanced function of A, and identify $A \cap (A+k)$ with its characteristic function. Then for each $s \in \mathbb{Z}_N$, we have

$$A \cap (A+k)(s) = \begin{cases} 1 \text{ if } s \in A \cap (A+k) \\ 0 \text{ if } s \notin A \cap (A+k) \end{cases} = \begin{cases} \delta + (1-\delta) \text{ if } s \in A \cap (A+k) \\ \delta + (-\delta) \text{ if } s \notin A \cap (A+k) \end{cases}$$
$$= (\delta + f(s))(\delta + f(s-k)) = \delta^2 + \delta f(s) + \delta f(s-k) + f(s)f(s-k).$$

By assumption, f is quadratically α-uniform, so by Lemma 3.3 (specifically the cases of (1) implies (2) with $c_2 = \alpha$ and (2) implies (4) with $c_2 = c_4^2$), we know that for all but at most $\alpha^{1/2}N$ values of k, the function $\Delta(f;k)$ is $\alpha^{1/2}$-uniform. Also, because f is real-valued, Lemma 3.4 implies that f itself is $\alpha^{1/2}$-uniform.

For a given k where $\Delta(f;k)$ is $\alpha^{1/2}$-uniform, define f_1, f_2, f_3 as $f_1(s) = \delta f(s)$, $f_2(s) = \delta f(s-k)$, and $f_3(s) = f(s)f(s-k) = \Delta(f;k)(s)$. Observe that

$$\sum_r \left|\tilde{f}_1(r)\right|^4 = \sum_r \delta^4 \left|\tilde{f}(r)\right|^4 \leq \delta^4 \alpha^{1/2} N^4 \leq \alpha^{1/2} N^4$$

since $\delta \leq 1$ and f is $\alpha^{1/2}$-uniform. Thus, f_1 is $\alpha^{1/2}$-uniform. Similarly, we know that f_2 is $\alpha^{1/2}$-uniform since $s-k$ varies over $\mathbb{Z}_N$ as s does. Lastly, f_3 is also $\alpha^{1/2}$-uniform since $f_4 = \Delta(f;k)$.

By Lemma 3.2, we therefore know that that the function $A \cap (A+k) - \delta^2$ is $81\alpha^{1/2}$-uniform. Note, however, that any non-trivial Fourier coefficient of $A \cap (A+k) - \delta^2$ equals the corresponding coefficient of $A \cap (A+k)$. Thus, if g_k is the balanced function of the set $A \cap (A+k)$, then

$$\sum_r |\tilde{g_k}(r)|^4 = \sum_{r\neq 0} \left| A \widetilde{\cap (A+} k)(r) \right|^4 = \sum_{r \neq 0} \left| [A \cap \widetilde{(A+k)} - \delta^2](r) \right|^4 \leq 81\alpha^{1/2}N^4,$$

so g_k is $81\alpha^{1/2}$-uniform. By definition, then, the set $A \cap (A+k)$ is $81\alpha^{1/2}$-uniform. Since this holds for each k such that $\Delta(f;k)$ is $\alpha^{1/2}$-uniform, we know that $A \cap (A+k)$ is $81\alpha^{1/2}$-uniform for all but at most $\alpha^{1/2}N$ values of k, as desired.

Now consider $|A \cap (A+k)|$. For any k, note that

$$\begin{aligned}
|A \cap (A+k)| &= \sum_s A \cap (A+k)(s) \\
&= \sum_s \left[\delta^2 + \delta f(s) + \delta f(s-k) + f(s)f(s-k)\right] \\
&= \sum_s \delta^2 + \sum_s \delta f(s) + \sum_s \delta f(s-k) + \sum_s f(s)f(s-k) \\
&= \delta^2 N + \sum_{s\in A} \delta f(s) + \sum_{s\notin A} \delta f(s) + \sum_{s \in A+k} \delta f(s-k) \\
&\quad + \sum_{s \notin A+k} \delta f(s-k) + \sum_s f(s)f(s-k) \\
&= \delta^2 N + |A|\delta(1-\delta) + (N-|A|)\delta(-\delta) + |A+k|\delta(1-\delta) \\
&\quad + (N - |A+k|)\delta(-\delta) + \sum_s f(s)f(s-k) \\
&= \delta^2 N + \sum_s f(s)f(s-k)
\end{aligned}$$

Thus, $|A \cap (A+k)| - \delta^2 N = \sum_s f(s)f(s-k)$, so $\big||A \cap (A+k)| - \delta^2 N\big| = |\sum_s f(s)f(s-k)|$. As a result, if we had $\big||A \cap (A+k)| - \delta^2 N\big| > \alpha^{1/8}N$ for more than $\alpha^{1/4}N$ values of k, then we would have

$$\sum_k \left|\sum_s f(s)f(s-k)\right|^2 > (\alpha^{1/4}N)(\alpha^{1/8}N)^2 = \alpha^{1/2}N^3,$$

which by Lemma 3.1 would contradict the fact that f is $\alpha^{1/2}$-uniform. Thus, for all but at most $\alpha^{1/4}N$ values of k, $\big||A \cap (A+k)| - \delta^2 N\big| \leq \alpha^{1/8}N$, as desired.

□

We now recall the standard norms for functions $f : \mathbb{Z}_N \to \mathbb{C}$. The l^1 norm is given by $||f||_1 = \sum_s |f(s)|$, and the l^2 norm is given by $||f||_2 = \left(\sum_s |f(s)|^2\right)^{1/2}$. From the Cauchy-Schwarz inequality, we see that for any f,

$$||f||_1 = \sum_s |f(s)| \leq \left(\sum_s |f(s)|^2\right)^{1/2} \left(\sum_s 1\right)^{1/2} = N^{1/2}||f||_2.$$

Hence, if $||f||_1 = wN$, then we have $||f||_2^2 \geq w^2 N$. We now wish to show that if equality almost holds in this inequality, then f is almost constant over $\mathbb{Z}_N$. The following lemma makes this clear.

Lemma 3.6. *Let $f : \mathbb{Z}_N \to \mathbb{R}^+$ map $\mathbb{Z}_N$ to the non-negative reals with $||f||_1 = wN$. Suppose that $||f||_2^2 \leq (1+\epsilon)w^2N$ for some $\epsilon > 0$. Then for each $A \subset \mathbb{Z}_N$, we have*

$$\left| \sum_{s \in A} f(s) - w|A| \right| \leq \epsilon^{1/2} w N^{1/2} |A|^{1/2}.$$

Proof. We first claim that the the variance of f is no more than ϵw^2. Observe that by definition of the mean and using the fact that f is non-negative, we have

$$\mathbb{E}[f] = \frac{1}{N}\sum_s f(s) = \frac{1}{N}\sum_s |f(s)| = \frac{1}{N}||f||_1 = \frac{1}{N}wN = w.$$

Similarly, observe that

$$\mathbb{E}[f^2] = \frac{1}{N}\sum_s f(s)^2 = \frac{1}{N}\sum_s |f(s)|^2 = \frac{1}{N}||f||_2^2 \leq (1+\epsilon)w^2.$$

Now, by the definition of variance, we have

$$\mathrm{var}(f) = \mathbb{E}[(f-w)^2] = \mathbb{E}[f^2] - 2w\mathbb{E}[f] + \mathbb{E}[w^2] \leq (1+\epsilon)w^2 - 2w^2 + w^2 = \epsilon w^2$$

where we have used the linearity of the functional $\mathbb{E}$. We can now prove the lemma easily. Indeed, by the Cauchy-Schwarz inequality, we have

$$\left| \sum_{s \in A} f(s) - w|A| \right| \leq \sum_{s \in A} |f(s) - w| \leq \left(\sum_{s \in A} (f(s)-w)^2 \right)^{1/2} \left(\sum_{s \in A} 1 \right)^{1/2}$$

$$\leq |A|^{1/2} \left(\sum_s (f(s)-w)^2 \right)^{1/2} = |A|^{1/2} \left(N \cdot \mathrm{var}(f)\right)^{1/2} \leq \epsilon^{1/2} w N^{1/2} |A|^{1/2}$$

which is what we wanted to show. □

In the next lemma, we prove a result regarding the number of arithmetic progressions in uniform sets. However, this result only tells us about progressions of length three because we are assuming uniformity (rather than quadratic uniformity). When we eventually add an assumption about quadratic uniformity, we will be able to say something about the number of progressions of length four. The following lemma will be useful when we add in the stronger assumption.

Before stating the lemma, though, it is important to point out a method that is common in Fourier analytic approaches to additive number theory. Notice that if A is a subset of $\mathbb{Z}_N$, then any progression of length k in A can be identified with a k-tuple in A^k that has the form $(a, a-r, a-2r, \ldots, a-(k-1)r)$ for some $a \in A$ and $r \geq 1$. If this is the case, then we see that $a \in A \cap (A+r) \cap \ldots \cap (A+(k-1)r)$. Therefore, if we are interested in the number of progressions of length k in A, we consider the sum

$$\sum_r |A \cap (A+r) \cap \ldots \cap (A+(k-1)r)|.$$

However, for each r, we can express $|A \cap (A+r) \cap \ldots \cap (A+(k-1)r)|$ in terms of this set's characteristic function, namely

$$\sum_s \left[A \cap (A+r) \cap \ldots \cap (A+(k-1)r)\right](s).$$

But this can be written as

$$\sum_{s} A(s)A(s-r)\cdots A(s-(k-1)r).$$

Therefore, the number of progressions of length k in A (including those with common difference $r=0$) is given by

$$\sum_{r}\sum_{s} A(s)A(s-r)\cdots A(s-(k-1)r),$$

which lends itself nicely to Fourier analytic arguments.

Lemma 3.7. *Let A, B, and C be subsets of $\mathbb{Z}_N$ with cardinalities αN, βN, and γN respectively. If C is η-uniform, then*

$$\left|\sum_{r} |A\cap(B+r)\cap(C+2r)| - \alpha\beta\gamma N^2\right| \leq \eta^{1/4}N^2.$$

Proof. As usual, identify A, B, and C with their characteristic functions. Then for each pair $s,r\in\mathbb{Z}_N$, we have

$$A(s)B(s-r)C(s-2r) = \begin{cases} 1 & \text{if } s\in A\cap(B+r)\cap(C+2r) \\ 0 & \text{if not.}\end{cases}$$

Therefore, we see that

$$\sum_{r} |A\cap(B+r)\cap(C+2r)| = \sum_{r}\sum_{s} A(s)B(s-r)C(s-2r).$$

But observe the following. For any pair s,r, if we let $x=s$ and $y=s-r$, then $A(s)B(s-r)C(s-2r) = A(x)B(y)C(-x+2y) = A(x)B(y)C(z)$ where $x-2y+z=0$. Conversely, for any triple x,y,z such that $x-2y+z=0$, we can let $s=x$ and $r=x-y$. Then $y=s-r$ and $z=2y-x=s-2r$, so $A(x)B(y)C(z)=A(s)B(s-r)C(s-2r)$. Hence, we have

$$\sum_{r}\sum_{s} A(s)B(s-r)C(s-2r) = \sum_{x-2y+z=0} A(x)B(y)C(z).$$

Now recall that for fixed x,y,z, we have

$$\sum_{p} \omega^{-p(x-2y+z)} = \begin{cases} N & \text{if } x-2y+z=0 \\ 0 & \text{if not.}\end{cases}$$

Thus, we can write the sum $\sum_{x-2y+z=0} A(x)B(y)C(z)$ as

$$\sum_{x,y,z} A(x)B(y)C(z)\left(\frac{1}{N}\sum_{p}\omega^{-p(x-2y+z)}\right).$$

This, in turn, equals

$$\frac{1}{N}\sum_{p}\sum_{x,y,z} A(x)B(y)C(z)\omega^{-p(x-2y+z)},$$

which we can then write as

$$\frac{1}{N}\sum_p\left(\sum_x A(x)\omega^{-px}\right)\left(\sum_y B(y)\omega^{2py}\right)\left(\sum_z C(z)\omega^{-pz}\right).$$

By definition of the Fourier coefficients, this is simply $\frac{1}{N}\sum_p \tilde{A}(p)\tilde{B}(-2p)\tilde{C}(p)$. Therefore, we have found that

$$\sum_r |A\cap(B+r)\cap(C+2r)| = \frac{1}{N}\sum_p \tilde{A}(p)\tilde{B}(-2p)\tilde{C}(p).$$

Note that for $p=0$, we have

$$\frac{1}{N}\tilde{A}(0)\tilde{B}(0)\tilde{C}(0) = \frac{1}{N}\left(\sum_s A(s)\right)\left(\sum_s B(s)\right)\left(\sum_s C(s)\right) = \alpha\beta\gamma N^2.$$

As a result,

$$\left|\sum_r |A\cap(B+r)\cap(C+2r)| - \alpha\beta\gamma N^2\right| = \left|\frac{1}{N}\sum_{p\neq 0}\tilde{A}(p)\tilde{B}(-2p)\tilde{C}(p)\right|,$$

so it suffices to show that $\left|\sum_{p\neq 0}\tilde{A}(p)\tilde{B}(-2p)\tilde{C}(p)\right| \leq \eta^{1/4}N^3$. We have the following:

$$\begin{aligned}
&\left|\sum_{p\neq 0}\tilde{A}(p)\tilde{B}(-2p)\tilde{C}(p)\right| \leq \sum_{p\neq 0}\left|\tilde{A}(p)\right|\left|\tilde{B}(-2p)\right|\left|\tilde{C}(p)\right| \\
&= \sum_{p\neq 0}\left|\tilde{A}(p)\right|\left|\tilde{B}(-2p)\right|\left|\tilde{f_C}(p)\right| \\
&\qquad \text{since } \tilde{f_C}(p) = \tilde{C}(p) \text{ for } p \neq 0 \\
&\leq \eta^{1/4}N\sum_{p\neq 0}\left|\tilde{A}(p)\right|\left|\tilde{B}(-2p)\right| \\
&\qquad \text{since } C \text{ is } \eta\text{-uniform implies } \max_p\left|\tilde{f_C}(p)\right| \leq \eta^{1/4}N \\
&\leq \eta^{1/4}N\left(\sum_{p\neq 0}\left|\tilde{A}(p)\right|^2\right)^{1/2}\left(\sum_{p\neq 0}\left|\tilde{B}(-2p)\right|^2\right)^{1/2} \\
&\qquad \text{by the Cauchy-Schwarz inequality} \\
&\leq \eta^{1/4}N\left(\sum_p\left|\tilde{A}(p)\right|^2\right)^{1/2}\left(\sum_p\left|\tilde{B}(p)\right|^2\right)^{1/2} \\
&\qquad \text{since } -2p \text{ runs through elements in } \mathbb{Z}_N \text{ as } p \text{ runs through all of } \mathbb{Z}_N \\
&= \eta^{1/4}N\left(N\sum_p|A(p)|^2\right)^{1/2}\left(N\sum_p|B(p)|^2\right)^{1/2} \\
&\qquad \text{by Parseval's formula} \\
&= \eta^{1/4}N\left(N|A|\right)^{1/2}\left(N|B|\right)^{1/2} \\
&\leq \eta^{1/4}N^3,
\end{aligned}$$

which is what we wanted to show. □

We now wish to prove a similar lemma, except we want to extend the result to four sets. In doing this, we will need to make some stronger assumptions, including quadratic uniformity. We will also use Lemma 3.7 a few times, though the general argument will be more complicated.

Lemma 3.8. *Let A, B, C, and D be subsets of $\mathbb{Z}_N$ with cardinalities αN, βN, γN, and δN respectively. If C and D are quadratically η^4-uniform for some $\eta \leq 2^{-20}$, where $\eta < \beta^2\gamma^2\delta^2$, then*

$$\left|\sum_r |A \cap (B+r) \cap (C+2r) \cap (D+3r)| - \alpha\beta\gamma\delta N^2\right| \leq \frac{3\eta^{1/16}N^2}{\beta\gamma\delta}.$$

Proof. As in the previous lemma, we identify sets with their characteristic functions. Now, define a non-negative, real-valued function f by $f(s) = \sum_r B(s-r)C(s-2r)D(s-3r)$. Note that if we sum $f(s)$ over all $s \in \mathbb{Z}_N$, we obtain an expression that resembles expressions we dealt with in Lemma 3.7. Recalling that this sum is precisely the l^1 norm of f, we have

$$\begin{aligned} ||f||_1 &= \sum_s f(s) = \sum_s \sum_r B(s-r)C(s-2r)D(s-3r) \\ &= \sum_r \sum_s B(s-r)C(s-2r)D(s-3r) = \sum_r |(B+r) \cap (C+2r) \cap (D+3r)|. \end{aligned}$$

Note that if we fix $r \in \mathbb{Z}_N$, the set $B \cap (C+r) \cap (D+2r)$ is simply a shift of $(B+r) \cap (C+2r) \cap (D+3r)$ by $-r$. As a result,

$$|(B+r) \cap (C+2r) \cap (D+3r)| = |B \cap (C+r) \cap (D+2r)|$$

for each $r \in \mathbb{Z}_N$, so

$$\sum_r |(B+r) \cap (C+2r) \cap (D+3r)| = \sum_r |B \cap (C+r) \cap (D+2r)|.$$

By assumption, D is quadratically η^4-uniform, so by Lemma 3.4, we know that D is η^2-uniform. We can now apply Lemma 3.7 to obtain

$$\left| ||f||_1 - \beta\gamma\delta N^2 \right| = \left|\sum_r |B \cap (C+r) \cap (D+2r)| - \beta\gamma\delta N^2\right| \leq \eta^{1/2}N^2.$$

Hence, we know that

$$(\beta\gamma\delta - \eta^{1/2})N^2 \leq ||f||_1 \leq (\beta\gamma\delta + \eta^{1/2})N^2.$$

We now wish to estimate the l^2 norm of f. By simple expansion, we have

$$\begin{aligned} ||f||_2^2 &= \sum_s \left(\sum_r B(s-r)C(s-2r)D(s-3r)\right)\left(\sum_q B(s-q)C(s-2q)D(s-3q)\right) \\ &= \sum_s \sum_{r,q} B(s-r)B(s-q)C(s-2r)C(s-2q)D(s-3r)D(s-3q) \end{aligned}$$

If we let $p = q - r$, then this becomes

$$\sum_s \sum_{r,p} B(s-r)B(s-r-p)C(s-2r)C(s-2r-2p)D(s-3r)D(s-3r-3p)$$

since each pair r, q is uniquely associated with a pair r, p with the property that $p = q - r$. We can then write this as

$$\begin{aligned}
&\sum_{r,p}\Big|(B+r)\cap(B+r+p)\cap(C+2r)\cap(C+2r+2p)\cap(D+3r)\cap(D+3r+3p)\Big| \\
&\quad = \sum_{r,p}\Big|B\cap(B+p)\cap(C+r)\cap(C+r+2p)\cap(D+2r)\cap(D+2r+3p)\Big| \\
&\quad = \sum_{r,p}\Big|(B\cap(B+p))\cap\Big((C\cap(C+2p))+r\Big)\cap\Big((D\cap(D+3p))+2r\Big)\Big|.
\end{aligned}$$

Note that for each p, we can work with the three sets $B\cap(B+p)$, $C\cap(C+2p)$, and $D\cap(D+3p)$. Then the sum

$$\sum_{r}\Big|(B\cap(B+p))\cap\Big((C\cap(C+2p))+r\Big)\cap\Big((D\cap(D+3p))+2r\Big)\Big|$$

resembles the sum dealt with in Lemma 3.7. To apply the results from that lemma, though, we need to know something about the uniformity of $D\cap(D+3p)$.

For this, we recall Lemma 3.5. Since D is quadratically η^4-uniform, there are at most $\eta^2 N$ values of k for which the set $D\cap(D+k)$ is not $81\eta^2$-uniform. Thus, there are at most $\eta^2 N$ values of p for which the set $D\cap(D+3p)$ is not $81\eta^2$-uniform (if N is coprime to 3, then the number of such values of k will be the same as the number of such values of p, while if N is not coprime to 3, the number of such values of k may be greater than the number of such values of p). Let U denote the set of such p, so $|U| \le \eta^2 N$. Then for each $p \notin U$ (that is, when $D\cap(D+3p)$ is $81\eta^2$-uniform), we can apply Lemma 3.7 to obtain

$$\begin{aligned}
&\sum_{r}\Big|(B\cap(B+p))\cap\Big((C\cap(C+2p))+r\Big)\cap\Big((D\cap(D+3p))+2r\Big)\Big| \\
&\quad \le \frac{|B\cap(B+p)|}{N}\frac{|C\cap(C+2p)|}{N}\frac{|D\cap(D+3p)|}{N}N^2 + 81^{1/4}\eta^{1/2}N^2 \\
&\quad = \frac{1}{N}|B\cap(B+p)||C\cap(C+2p)||D\cap(D+3p)| + 3\eta^{1/2}N^2
\end{aligned}$$

We therefore can say the following:

$$\begin{aligned}
||f||_2^2 &= \sum_{r,p}\Big|(B\cap(B+p))\cap\Big((C\cap(C+2p))+r\Big)\cap\Big((D\cap(D+3p))+2r\Big)\Big| \\
&= \sum_{p\in U}\sum_{r}\Big|(B\cap(B+p))\cap\Big((C\cap(C+2p))+r\Big)\cap\Big((D\cap(D+3p))+2r\Big)\Big| \\
&\quad + \sum_{p\notin U}\sum_{r}\Big|(B\cap(B+p))\cap\Big((C\cap(C+2p))+r\Big)\cap\Big((D\cap(D+3p))+2r\Big)\Big| \\
&\le \sum_{p\in U}\sum_{r}N + \sum_{p\notin U}\left(\frac{1}{N}|B\cap(B+p)||C\cap(C+2p)||D\cap(D+3p)| + 3\eta^{1/2}N^2\right) \\
&= |U|N^2 + 3\eta^{1/2}N^3 + \frac{1}{N}\sum_{p}|B\cap(B+p)||C\cap(C+2p)||D\cap(D+3p)|
\end{aligned}$$

$$\leq \eta^2 N^3 + 3\eta^{1/2}N^3 + \frac{1}{N}\sum_p |B\cap(B+p)||C\cap(C+2p)||D\cap(D+3p)|$$
$$\leq 4\eta^{1/2}N^3 + \frac{1}{N}\sum_p |B\cap(B+p)||C\cap(C+2p)||D\cap(D+3p)|.$$

The last inequality follows from $\eta < 1$, and we use it to make computations more simple.

We now wish to use the second result from Lemma 3.5. Since C and D are quadratically η^4-uniform, we know that for all but at most ηN values of k, $\big||C\cap(C+k)| - \gamma^2 N\big| \leq \eta^{1/2}N$, and for all but at most ηN values of m, $\big||D\cap(D+m)| - \delta^2 N\big| \leq \eta^{1/2}N$. Hence, we know that for all but at most ηN values of k, $|C\cap(C+2k)| \leq \gamma^2 N + \eta^{1/2}N$, and for all but at most ηN values of m, $|D\cap(D+3m)| \leq \delta^2 N + \eta^{1/2}N$. Therefore, both $|C\cap(C+2p)| \leq \gamma^2 N + \eta^{1/2}N$ and $|D\cap(D+3p)| \leq \delta^2 N + \eta^{1/2}N$ hold for all but at most $2\eta N$ values of p. Let J denote the set of p such that these inequalities hold, so the cardinality of the complement of J satisfies $|J^c| \leq 2\eta N$.

Continuing with the chain of inequalities from before, we have

$$\begin{aligned}
&\frac{1}{N}\sum_p |B\cap(B+p)||C\cap(C+2p)||D\cap(D+3p)| + 4\eta^{1/2}N^3\\
&\quad\leq \frac{1}{N}\sum_{p\in J}|B\cap(B+p)|(\gamma^2 N + \eta^{1/2}N)(\delta^2 N + \eta^{1/2}N)\\
&\qquad + \frac{1}{N}\sum_{p\notin J}(\beta N)(\gamma N)(\delta N) + 4\eta^{1/2}N^3\\
&\quad\leq \frac{1}{N}(\gamma^2\delta^2 + \gamma^2\eta^{1/2} + \delta^2\eta^{1/2} + \eta)N^2\sum_{p\in J}|B\cap(B+p)| + \frac{1}{N}|J^c|N^3 + 4\eta^{1/2}N^3\\
&\quad\leq N(\gamma^2\delta^2 + 2\eta^{1/2} + \eta)\sum_p |B\cap(B+p)| + 2\eta N^3 + 4\eta^{1/2}N^3.
\end{aligned}$$

Now note that

$$\sum_p |B\cap(B+p)| = \sum_p\left(\sum_s B(s)B(s-p)\right) = \sum_s B(s)\sum_p B(s-p) = \beta^2 N^2.$$

Therefore, if we continue from above, we obtain

$$\begin{aligned}
N(\gamma^2\delta^2 + 2\eta^{1/2} + \eta)&\sum_p |B\cap(B+p)| + 2\eta N^3 + 4\eta^{1/2}N^3\\
&= \beta^2 N^3(\gamma^2\delta^2 + 2\eta^{1/2} + \eta) + 2\eta N^3 + 4\eta^{1/2}N^3\\
&\leq N^3(\beta^2\gamma^2\delta^2 + 2\eta^{1/2} + 3\eta^{1/2}\eta^{1/2} + 4\eta^{1/2})\\
&\leq N^3(\beta^2\gamma^2\delta^2 + 7\eta^{1/2})\\
&\leq N^3(\beta^2\gamma^2\delta^2 + 3\eta^{1/8})
\end{aligned}$$

where the last two inequalities hold since $\eta \leq 2^{-20}$. We therefore have shown that

$$||f||_2^2 \leq N^3(\beta^2\gamma^2\delta^2 + 3\eta^{1/8}).$$

Our wish now is to find an $\epsilon > 0$ such that

$$||f||_2^2 \leq \frac{1}{N}||f||_1^2(1+\epsilon).$$

This will allow us to use Lemma 5, and as a result, we will obtain the desired bound on

$$\sum_r |A \cap (B+r) \cap (C+2r) \cap (D+3r)|.$$

First, recall that by assumption, $\eta \neq \beta^2\gamma^2\delta^2$, so we can write

$$N^3(\beta^2\gamma^2\delta^2 + 3\eta^{1/8}) = N^3 \frac{(\beta^2\gamma^2\delta^2 + 3\eta^{1/8})}{\beta^2\gamma^2\delta^2} \frac{\beta^2\gamma^2\delta^2}{(\beta\gamma\delta - \eta^{1/2})^2} \left(\beta\gamma\delta - \eta^{1/2}\right)^2.$$

Continuing, we have

$$\begin{aligned}
&= N^3 \left(\beta\gamma\delta - \eta^{1/2}\right)^2 \left(1 + \frac{3\eta^{1/8}}{\beta^2\gamma^2\delta^2}\right) \left(\frac{\beta\gamma\delta - \eta^{1/2}}{\beta\gamma\delta}\right)^{-2} \\
&= \frac{1}{N}\left[N^2\left(\beta\gamma\delta - \eta^{1/2}\right)\right]^2 \left(1 + \frac{3\eta^{1/8}}{\beta^2\gamma^2\delta^2}\right) \left(1 - \frac{\eta^{1/2}}{\beta\gamma\delta}\right)^{-2} \\
&\leq \frac{1}{N}||f||_1^2 \left(1 + \frac{3\eta^{1/8}}{\beta^2\gamma^2\delta^2}\right) \left(1 - \frac{\eta^{1/2}}{\beta\gamma\delta}\right)^{-2}
\end{aligned}$$

where the last inequality comes from the bound $N^2(\beta\gamma\delta - \eta^{1/2}) \leq ||f||_1$ which we found at the beginning of the proof, along with the fact that $\beta\gamma\delta - \eta^{1/2} > 0$ (so squaring does not switch the inequality).

We consider now the product

$$\left(1 + \frac{3\eta^{1/8}}{\beta^2\gamma^2\delta^2}\right) \left(1 - \frac{\eta^{1/2}}{\beta\gamma\delta}\right)^{-2}.$$

Using the fact that that $(1-x)^{-2} = \sum_{n=1}^{\infty} nx^{n-1}$ for any $|x| < 1$ and our assumption that $\eta \leq 2^{-20}$, we know that this is bounded above by $1 + 4\eta^{1/8}/(\beta^2\gamma^2\delta^2)$. Therefore, we see that

$$||f||_2^2 \leq \frac{1}{N}||f||_1^2 \left(1 + 4\frac{\eta^{1/8}}{\beta^2\gamma^2\delta^2}\right).$$

Let $\epsilon = 4\eta^{1/8}/(\beta^2\gamma^2\delta^2)$ and $w = ||f||_1/N$ so that $||f||_2^2 \leq (1+\epsilon)w^2N$. By Lemma 3.6, we then know that

$$\left|\sum_{s\in A} f(s) - w|A|\right| \leq \epsilon^{1/2} w N^{1/2} |A|^{1/2}.$$

Recalling that $|w - \beta\gamma\delta N| = (1/N)\left|||f||_1 - \beta\gamma\delta N^2\right| \leq \eta^{1/2}N$, we then have

$$\begin{aligned}
&\left|\sum_{s\in A} f(s) - \alpha\beta\gamma\delta N^2\right| \leq \left|\sum_{s\in A} f(s) - w|A|\right| + \left|w|A| - \alpha\beta\gamma\delta N^2\right| \\
&\leq \epsilon^{1/2} w N^{1/2}|A|^{1/2} + |A|\eta^{1/2}N \leq \epsilon^{1/2}||f||_1 + \eta^{1/2}N^2 \\
&\leq \frac{2\eta^{1/16}}{\beta\gamma\delta}\left(\beta\gamma\delta + \eta^{1/2}\right)N^2 + \eta^{1/2}N^2 \leq \frac{3\eta^{1/16}N^2}{\beta\gamma\delta}
\end{aligned}$$

where the last inequality follows from our assumption that $\eta \leq 2^{-20}$. But by definition of f,

$$\sum_{s\in A} f(s) = \sum_r |A \cap (B+r) \cap (C+2r) \cap (D+3r)|,$$

so we have our desired result. □

We are almost ready to show that if a subset of $\mathbb{Z}_N$ is quadratically α-uniform for a sufficiently small α, then the set must contain a progression of length four that is a genuine progression when considered as a subset of $\mathbb{Z}$. We first need to prove one small lemma, though.

Lemma 3.9. *Let $A \subset \mathbb{Z}_N$ be α-uniform with cardinality δN, and let P be an interval of the form $[a, a+M]$ where $M = \beta N$. Then $\big||A \cap P| - \beta\delta N\big| \leq 3\alpha^{1/4}N$.*

Proof. Departing from our standard notation, we consider the representatives of the congruence classes of $\mathbb{Z}_N$ to lie in the interval $[-\lfloor N/2\rfloor, \lfloor N/2\rfloor]$. Identifying the set P with its characteristic function, we have

$$\begin{aligned}\left|\tilde{P}(r)\right| &= \left|\sum_s P(s)\omega^{-rs}\right| = \left|\sum_{s=1}^{M} \omega^{-r(a+s)}\right| = \left|\omega^{-ra}\right| \cdot \left|\sum_{s=1}^{M} \omega^{-rs}\right| \\ &= \left|\omega^{-r}\right| \cdot \left|\frac{1-\omega^{-rM}}{1-\omega^{-r}}\right| \leq \frac{2}{|1-\omega^{-r}|}\end{aligned}$$

for each non-zero $r \in \mathbb{Z}_N$. But note that

$$\begin{aligned}\left|1-\omega^{-r}\right| &= \left|1-e^{-2\pi i r/N}\right| = \left|1-\cos\left(-2\pi r/N\right) - i\sin\left(-2\pi r/N\right)\right| \\ &= \sqrt{2-2\cos\left(-2\pi r/N\right)} = 2\left|\sin\left(\pi r/N\right)\right|.\end{aligned}$$

Since we have assumed $-N/2 \leq r \leq N/2$ and we know that $|\sin(x)| \geq (2/\pi)|x|$ for $-\pi/2 \leq x \leq \pi/2$, we have $|\sin(\pi r/N)| \geq 2|r|/N$. Therefore $|1-\omega^{-r}| \geq 4|r|/N$, and as a result,

$$\left|\tilde{P}(r)\right| \leq \frac{2}{4|r|/N} = \frac{N}{2|r|}.$$

Observe now that

$$\begin{aligned}&\sum_{r\neq 0}\left|\tilde{P}(r)\right|^{4/3} \leq \sum_{r\neq 0}\left(\frac{N}{2|r|}\right)^{4/3} = \left(\frac{N}{2}\right)^{4/3} \sum_{r\in[-N/2,N/2]\setminus\{0\}} \frac{1}{|r|^{4/3}} \\ &= 2\left(\frac{N}{2}\right)^{4/3}\sum_{r=1}^{\lfloor N/2\rfloor}\frac{1}{r^{4/3}} \leq 2\left(\frac{N}{2}\right)^{4/3}\sum_{r=1}^{\infty}\frac{1}{r^{4/3}} \\ &\leq 2\left(\frac{N}{2}\right)^{4/3}\left(1+\sum_{r=1}^{\infty}\int_r^{r+1} x^{-4/3}dx\right) = 2\left(\frac{N}{2}\right)^{4/3}\left(1+\int_1^{\infty} x^{-4/3}dx\right) \\ &= 8\left(\frac{N}{2}\right)^{4/3} \leq 4N^{4/3}.\end{aligned}$$

Also note that if we identify A with its characteristic function, the size of $A \cap P$ is given by

$$|A \cap P| = \sum_s A(s)P(s) = \frac{1}{N}\sum_r \sum_{s,t} A(s)P(t)\omega^{-r(s-t)} = \frac{1}{N}\sum_r \tilde{A}(r)\tilde{P}(-r).$$

Using Hölder's inequality, we then have

$$\begin{aligned}\big||A \cap P| - \beta\delta N\big| &= \frac{1}{N}\left|\sum_r \tilde{A}(r)\tilde{P}(-r) - \beta\delta N^2\right| = \left|\sum_{r\neq 0} \tilde{A}(r)\tilde{P}(-r)\right| \\ &\leq \frac{1}{N}\left(\sum_{r\neq 0}|\tilde{A}(r)|^4\right)^{1/4}\left(\sum_{r\neq 0}|\tilde{P}(r)|^{4/3}\right)^{3/4}.\end{aligned}$$

Since A is α-uniform, we know that $\sum_{r\neq 0}|\tilde{A}(r)|^4 = \sum_r |\tilde{f_A}(r)|^4 \leq \alpha N^4$, where f_A denotes, as usual, the balanced function of A. Therefore,

$$\big||A \cap P| - \beta\delta N\big| \leq \frac{1}{N}\left(\alpha N^4\right)^{1/4}\left(4N^{4/3}\right)^{3/4} \leq 3\alpha^{1/4}N$$

as desired.

□

We are now fully equipped to prove the main result of this section.

Corollary 3.1. *Let $A_0 \subset \mathbb{Z}_N$ be quadratically α-uniform with size $|A_0| = \delta N$, where $\alpha \leq 2^{-832}\delta^{448}$ and $N > 200\delta^{-3}$. Then A_0 contains an arithmetic progression of length four.*

Proof. Let $A = A_0 \cap [2N/5, 3N/5) = B$ and $C = A_0 = D$. Note that since A_0 is quadratically, α-uniform, it is $\alpha^{1/2}$-uniform by Lemma 3.4. First we claim that A and B have size at least $\delta N/10$. To show this, take $P = [2N/5, 3N/5)$ in Lemma 3.9, so $|P| = \beta N$, where $\beta = \lfloor N/5 \rfloor / N \geq 1/5 - 1/N$. We then have $\big||A_0 \cap P| - \beta\delta N\big| \leq 3\alpha^{1/8}N$, and as a result,

$$\begin{aligned}|A| = |B| = &|A_0 \cap P| \geq \beta\delta N - 3\alpha^{1/8}N \geq \left(\frac{\delta}{5} - \frac{\delta}{N} - 3\cdot 2^{-104}\delta^{56}\right)N \\ &\geq \left(\frac{\delta}{5} - \frac{\delta}{10}\right)N = \frac{\delta N}{10}.\end{aligned}$$

We now wish to apply Lemma 3.8 to the sets A, B, C, and D. For convenience, let $|A| = \alpha' N$, $|B| = \beta' N$, $|C| = \gamma' N$, and $|D| = \delta' N$. Note that C and D are quadratically α-uniform for some $\alpha \leq 2^{-832}\delta^{448}$, so they are quadratically η^4-uniform for some $\eta \leq 2^{-208}\delta^{112} \leq 2^{-20}$. Also, we have the bound

$$\beta'^2\gamma'^2\delta'^2 \geq \left(\frac{\delta}{10}\right)^2\delta^4 = \frac{\delta^6}{100} > 2^{-208}\delta^{112} \geq \eta,$$

so the hypotheses of Lemma 3.8 are satisfied here. As a result, we obtain

$$\left|\sum_r |A \cap (B+r) \cap (C+2r) \cap (D+3r)| - \alpha'\beta'\gamma'\delta' N^2\right| \leq \frac{3\eta^{1/16}N^2}{\beta'\gamma'\delta'},$$

so the number of elements $(a, a+r, a+2r, a+3r) \in A \times B \times C \times D$ is at least

$$\alpha'\beta'\gamma'\delta' N^2 - \frac{3\eta^{1/16}N^2}{\beta'\gamma'\delta'}.$$

From the sizes of A, B, C, and D and our bounds on η, we know that this is at least

$$\frac{\delta}{10}\frac{\delta}{10}\delta\delta N^2 - \frac{3\eta^{1/16}N^2}{(\delta/10)\delta\delta} \geq \frac{\delta^4 N^2}{100} - \frac{30\cdot 2^{-13}\delta^7 N^2}{\delta^3} \geq \frac{\delta^4 N^2}{100} - \frac{\delta^4 N^2}{200} = \frac{\delta^4 N^2}{200}.$$

Thus, there are at least $\delta^4 N^2/200$ such elements in $A \times B \times C \times D$. Of course, not every such element is actually an arithmetic progression; it could be the case that $r = 0$. But there are at most $|C| = \delta N$ quadruples $(a, a+r, a+2r, a+3r) \in A \times B \times C \times D$ with $r = 0$, so there are at least $\delta^4 N^2/200 - \delta N$ quadruples with $r \neq 0$. By our strict lower bound on N, we have

$$\frac{\delta^4 N^2}{200} - \delta N > \frac{\delta^4 N\left(200\delta^{-3}\right)}{200} - \delta N = 0,$$

so in particular, there is at least one quadruple with $r \neq 0$. If $(a, a+r, a+2r, a+3r)$ is this quadruple, then we immediately see that $\{a, a+r, a+2r, a+3r\}$ is a progression in $\mathbb{Z}$, not just in $\mathbb{Z}_N$, because of our restriction of A and B to the interval $[2N/5, 3N/5)$. Indeed, with this restriction we must have $a \in [2N/5, 3N/5)$ and $r \in (0, N/5)$, so $a + 2r \in [2N/5, 4N/5)$ and $a + 3r \in [2N/5, N)$. Thus, $\{a, a+r, a+2r, a+3r\}$ is a genuine arithmetic progression in A_0.

□

3.2 Sets that Fail to be Quadratically Uniform

In the previous section, we showed that if a subset of $\{1, \ldots, N\}$ is quadratically α-uniform for a small enough value α, then the set must contain an arithmetic progression of length four. We now must deal with the more difficult case for when the set is not quadratically α-uniform.

If $A \subset \mathbb{Z}_N$ is not quadratically α-uniform, then by definition, its balanced function f fails to be quadratically α-uniform. From Lemma 3.3, this means that

$$\sum_k \sum_r \left|\widetilde{\Delta(f;k)}(r)\right|^4 \geq \alpha N^5.$$

Recall from the proof of (3) ⇒ (2) in that lemma, we showed that if (2) does not hold, then there are at least $\alpha N/2$ values of k for which there exists an r such that $\left|\widetilde{\Delta(f;k)}(r)\right| \geq (\alpha/2)^{1/2}N$. In other words, there are many values of k for which $\Delta(f;k)$ has a large Fourier coefficient. Let B be the set of such k, so $|B| \geq (\alpha/2)N$, and define a function $\phi : B \to \mathbb{Z}_N$ by $\phi(k) = r$, where r is one of the elements of $\mathbb{Z}_N$ for which $\left|\widetilde{\Delta(f;k)}(r)\right| \geq (\alpha/2)^{1/2}N$ (there might be more than one such r, but in this case, just choose one of them). We then have

$$\sum_{k\in B} \left|\widetilde{\Delta(f;k)}(\phi(k))\right|^2 \geq \sum_{k\in B} (\alpha/2)N^2 \geq (\alpha/2)^2 N^3.$$

It is therefore reasonable to begin our study of sets that fail to be quadratically uniform with the following proposition. As we shall see, the consequences of this result will eventually allow us to find an arithmetic progression $P \subset \mathbb{Z}$ such that A has increased density on P, just as we did in the proof of Roth's Theorem.

Proposition 3.1. *Let $\alpha > 0$, $f : \mathbb{Z}_N \to \mathbb{D}$, $B \subset \mathbb{Z}_N$, and $\phi : B \to \mathbb{Z}_N$ such that*

$$\sum_{k \in B} \left| \widetilde{\Delta(f;k)}(\phi(k)) \right|^2 \geq \alpha N^3.$$

Then there are at least $\alpha^4 N^3$ quadruples $(a,b,c,d) \in B^4$ such that $a+b = c+d$ and $\phi(a) + \phi(b) = \phi(c) + \phi(d)$.

Proof. For each $k \in B$, we have

$$\left| \widetilde{\Delta(f;k)}(\phi(k)) \right|^2 = \left(\sum_s \Delta(f;k)(s)\omega^{-s\phi(k)} \right) \left(\overline{\sum_t \Delta(f;k)(t)\omega^{-t\phi(k)}} \right),$$

so by definition of $\Delta(f;k)$,

$$\begin{aligned} \left| \widetilde{\Delta(f;k)}(\phi(k)) \right|^2 &= \left(\sum_s f(s)\overline{f(s-k)}\omega^{-s\phi(k)} \right) \left(\sum_t \overline{f(t)}f(t-k)\omega^{t\phi(k)} \right) \\ &= \sum_{s,t} f(s)\overline{f(s-k)f(t)}f(t-k)\omega^{-\phi(k)(s-t)} \end{aligned}$$

Hence, the inequality in the proposition is equivalent to

$$\sum_{k \in B} \sum_{s,t} f(s)\overline{f(s-k)f(t)}f(t-k)\omega^{-\phi(k)(s-t)} \geq \alpha N^3.$$

Now, let $u = s - t$, so we can rewrite this as

$$\sum_{k \in B} \sum_{s,u} f(s)\overline{f(s-k)f(s-u)}f(s-k-u)\omega^{-\phi(k)u} \geq \alpha N^3.$$

Since $\alpha > 0$, we have

$$\begin{aligned} \alpha N^3 &\leq \left| \sum_{k \in B} \sum_{s,u} f(s)\overline{f(s-k)f(s-u)}f(s-k-u)\omega^{-\phi(k)u} \right| \\ &= \left| \sum_{s,u} f(s)\overline{f(s-u)} \sum_{k \in B} \overline{f(s-k)}f(s-k-u)\omega^{-\phi(k)u} \right| \\ &\leq \sum_{s,u} |f(s)|\,|f(s-u)| \left| \sum_{k \in B} \overline{f(s-k)}f(s-k-u)\omega^{-\phi(k)u} \right| \\ &\leq \sum_{s,u} \left| \sum_{k \in B} \overline{f(s-k)}f(s-k-u)\omega^{-\phi(k)u} \right| \\ &\quad \text{since } |f(x)| \leq 1 \end{aligned}$$

$$\leq \left(\sum_{s,u} 1^2\right)^{1/2} \left(\sum_{s,u} \left|\sum_{k\in B} \overline{f(s-k)} f(s-k-u)\omega^{-\phi(k)u}\right|^2\right)^{1/2}$$

by applying the Cauchy-Schwarz inequality twice

$$= N\left(\sum_{s,u} \left|\sum_{k\in B} \overline{f(s-k)} f(s-k-u)\omega^{-\phi(k)u}\right|^2\right)^{1/2}.$$

Consequently, we see that

$$\sum_{s,u} \left|\sum_{k\in B} \overline{f(s-k)} f(s-k-u)\omega^{-\phi(k)u}\right|^2 \geq \alpha^2 N^4.$$

For each u, define $\gamma(u)$ as

$$\gamma(u) = \frac{1}{N^3}\sum_{s} \left|\sum_{k\in B} \overline{f(s-k)} f(s-k-u)\omega^{-\phi(k)u}\right|^2,$$

so that

$$\sum_u \gamma(u) \geq \alpha^2 N.$$

For each u, we therefore have

$$\begin{aligned}\gamma(u)N^3 &= \sum_{s} \left|\sum_{k\in B} \overline{f(s-k)} f(s-k-u)\omega^{-\phi(k)u}\right|^2 \\ &= \sum_{s} \left|\sum_{k} B(k)\omega^{-\phi(k)u}\overline{\Delta(f;u)(s-k)}\right|^2 \\ &= \sum_{s} \left|\sum_{k} B(-k)\omega^{-\phi(-k)u}\overline{\Delta(f;u)(k-s)}\right|^2.\end{aligned}$$

Consider the function $\Delta(f;u)$. Using the fact that $|f(x)| \leq 1$, we have

$$||\Delta(f;u)||_2^2 = \sum_t |\Delta(f;u)(t)|^2 = \sum_t \left|f(t)\overline{f(t-u)}\right|^2 \leq N.$$

Hence, we see that

$$\sum_{s} \left|\sum_{k} B(-k)\omega^{-\phi(-k)u}\overline{\Delta(f;u)(k-s)}\right|^2 \geq \gamma(u)N^2||\Delta(f;u)||_2^2,$$

so by the implication of (4) from (1) in Lemma 3.1, we have

$$\sum_r |\tilde{g}(r)|^4 \geq \gamma(u)^2 N^4$$

where $g(k) = B(-k)\omega^{-\phi(-k)u}$. But for each r, we know that

$$\begin{aligned}\tilde{g}(r) &= \sum_k g(k)\omega^{-kr} = \sum_k B(-k)\omega^{-\phi(-k)u}\omega^{-kr} \\ &= \sum_k B(k)\omega^{-\phi(k)u}\omega^{kr} = \sum_{k\in B} \omega^{-\phi(k)u+kr} \\ &= \overline{\sum_{k\in B} \omega^{\phi(k)u-kr}},\end{aligned}$$

so we see that

$$\sum_r |\tilde{g}(r)|^4 = \sum_r \left|\sum_{k\in B} \omega^{\phi(k)u-kr}\right|^4,$$

and therefore

$$\sum_r \left|\sum_{k\in B} \omega^{\phi(k)u-kr}\right|^4 \geq \gamma(u)^2 N^4.$$

We know, though, that $\sum_u \gamma(u) \geq \alpha^2 N$, so using the Cauchy-Schwarz inequality, we have $\sum_u \gamma(u)^2 \geq \alpha^4 N$. Therefore, summing the above inequality over $u \in \mathbb{Z}_N$, we obtain

$$\sum_u \sum_r \left|\sum_{k\in B} \omega^{\phi(k)u-kr}\right|^4 \geq \alpha^4 N^5.$$

We now expand the left hand side of this inequality to conclude the proof of the proposition. Observe that for each u, r we have

$$\begin{aligned}&\left|\sum_{k\in B} \omega^{\phi(k)u-kr}\right|^4 \\ &= \left(\sum_{a\in B} \omega^{\phi(a)u-ar}\right)\left(\sum_{b\in B} \omega^{\phi(b)u-br}\right)\left(\overline{\sum_{c\in B} \omega^{\phi(c)u-cr}}\right)\left(\overline{\sum_{d\in B} \omega^{\phi(d)u-dr}}\right) \\ &= \left(\sum_{a\in B} \omega^{\phi(a)u-ar}\right)\left(\sum_{b\in B} \omega^{\phi(b)u-br}\right)\left(\sum_{c\in B} \omega^{-\phi(c)u+cr}\right)\left(\sum_{d\in B} \omega^{-\phi(d)u+dr}\right) \\ &= \sum_{a,b,c,d\in B} \omega^{\phi(a)u-ar+\phi(b)u-br-\phi(c)u+cr-\phi(d)u+dr} \\ &= \sum_{a,b,c,d\in B} \omega^{u(\phi(a)+\phi(b)-\phi(c)-\phi(d))}\omega^{-r(a+b-c-d)}.\end{aligned}$$

Therefore, we see that

$$\sum_{u,r} \sum_{a,b,c,d\in B} \omega^{u(\phi(a)+\phi(b)-\phi(c)-\phi(d))}\omega^{-r(a+b-c-d)} \geq \alpha^4 N^5.$$

But also, we have

$$\begin{aligned}&\sum_{u,r} \sum_{a,b,c,d\in B} \omega^{u(\phi(a)+\phi(b)-\phi(c)-\phi(d))}\omega^{-r(a+b-c-d)} \\ &\quad= \sum_{a,b,c,d\in B} \left(\sum_u \omega^{u(\phi(a)+\phi(b)-\phi(c)-\phi(d))}\right)\left(\sum_r \omega^{-r(a+b-c-d)}\right) \\ &\quad= \sum_{a,b,c,d\in B} \chi(a,b,c,d)\psi(a,b,c,d)\end{aligned}$$

where $\chi(a,b,c,d) = \sum_u \omega^{u(\phi(a)+\phi(b)-\phi(c)-\phi(d))}$ and $\psi(a,b,c,d) = \sum_r \omega^{-r(a+b-c-d)}$. Note that

$$\chi(a,b,c,d) = \begin{cases} 0 & \text{if } \phi(a)+\phi(b) \neq \phi(c)+\phi(d) \\ N & \text{if } \phi(a)+\phi(b) = \phi(c)+\phi(d) \end{cases}$$

and

$$\psi(a,b,c,d) = \begin{cases} 0 & \text{if } a+b \neq c+d \\ N & \text{if } a+b = c+d \end{cases}$$

Therefore,

$$\sum_{a,b,c,d \in B} \chi(a,b,c,d)\psi(a,b,c,d) = N^2 Q,$$

where Q is the number of quadruples $(a,b,c,d) \in B$ such that $a+b = c+d$ and $\phi(a)+\phi(b) = \phi(c)+\phi(d)$. This gives

$$\sum_{u,r} \sum_{a,b,c,d \in B} \omega^{u(\phi(a)+\phi(b)-\phi(c)-\phi(d))} \omega^{-r(a+b-c-d)} = N^2 Q,$$

and since

$$\sum_{u,r} \sum_{a,b,c,d \in B} \omega^{u(\phi(a)+\phi(b)-\phi(c)-\phi(d))} \omega^{-r(a+b-c-d)} \geq \alpha^4 N^5,$$

we obtain our desired result that

$$Q \geq \alpha^4 N^3.$$

□

These quadruples will be important to us throughout the remainder of the proof, so we give them a name. For $\phi : B \to \mathbb{Z}_N$, we say that $(a,b,c,d) \in B^4$ is ϕ-additive if $a+b = c+d$ and $\phi(a)+\phi(b) = \phi(c)+\phi(d)$. In the cases where the function ϕ is understood, we may say that such a quadruple is additive. We now want to show that for a given ϕ, the ϕ-additive quadruples have structure. This will require the invocation of two strong theorems, one of which we will not prove (Freiman's Theorem), and the other of which we will prove (the Balog-Szemerédi Theorem). In fact, Gowers gives a new proof of the Balog-Szemerédi Theorem, in which he obtains better bounds than previous proofs did. These new bounds allow us to find significantly better bounds for Szemerédi's Theorem itself.

3.3 Some Combinatorial Considerations

We briefly depart from the setting of $\mathbb{Z}_N$ and discuss sets in $\mathbb{Z}$. Given finite subsets $A, B \subset \mathbb{Z}$, we define the sumset $A+B$ to be

$$A + B = \{a + b : a \in A, b \in B\}.$$

It is natural to ask how large $A+B$ is, and the answer certainly depends on the structure of A and B. The following result gives upper and lower bounds on sumsets.

Theorem 3.1. *Let A and B be finite subsets of $\mathbb{Z}$. Then $|A| + |B| - 1 \leq |A+B| \leq |A| \cdot |B|$.*

It is also natural to ask which sets A and B have a small sumset. In particular, can we characterize the sets for which $|A+B| = |A| + |B| - 1$? The following result answers this question in the affirmative.

Theorem 3.2. *Let A and B be finite subsets of $\mathbb{Z}$ with $|A| \geq 2$ and $|B| \geq 2$. If $|A+B| = |A| + |B| - 1$, then A and B are arithmetic progressions of the same common difference.*

A much more interesting result in the theory of sumsets is due to Freiman. After seeing Theorem 3.2, an obvious question is what happens when $A + B$ is close to, but not exactly, the minimal size. Freiman answers this question for sumsets of the form $A + A$. Essentially, Freiman's Theorem says that if $|A + A|$ is on the order of $|A|$, then A is contained in a reasonably small generalized arithmetic progression, where a generalized arithmetic progression is simply a sumset of arithmetic progressions. If $P_1, \ldots, P_d$ are arithmetic progressions in $\mathbb{Z}$, then we say that the generalized arithmetic progression $Q = P_1 + \ldots + P_d$ has dimension d. The formal statement of Freiman's Theorem is the following.

Theorem 3.3. *Let C be a positive constant. Then there exist constants d and K which depend only on C such that for each finite subset $A \subset \mathbb{Z}$ with $|A + A| \leq C|A|$, there exists a generalized arithmetic progression Q of dimension at most d for which $|Q| \leq K|A|$ and $A \subset Q$.*

Note that the conclusion that Q is small (that is, the dimension of Q is no greater than d and the size of Q is on the order of $|A|$) is vital in this theorem. Indeed, any finite set $A \subset \mathbb{Z}$ is trivially contained in the generalized arithmetic progression $\{-m, \ldots, 0, \ldots, m\}$, where $m = \max_{a \in A}\{|a|\}$. Thus, ensuring that Q can be small compared to A is an important part of Freiman's Theorem.

It will be necessary for us to apply Freiman's Theorem to subsets of $\mathbb{Z}^D$ rather than to subsets of $\mathbb{Z}$. Fortunately, there is a method of embedding finite subsets of $\mathbb{Z}^D$ into $\mathbb{Z}$ isomorphically, namely, using Freiman isomorphisms.

Definition 11. Let A and B be subsets of some additive groups G and H. We say that a function $\phi : A \to B$ is a k-homomorphism if $\phi(x_1) + \ldots + \phi(x_k) = \phi(y_1) + \ldots + \phi(y_k)$ whenever $x_1 + \ldots + x_k = y_1 + \ldots + y_k$ for $x_i, y_i \in A$. If there is also an inverse map $\phi^{-1} : B \to A$ that is a k-homomorphism, then we call ϕ a k-isomorphism.

It is easy to see that Freiman homomorphisms are a weaker type of map than group homomorphisms (where we require that $\phi(x + y) = \phi(x) + \phi(y)$ for all elements x and y in the group). The important aspect of Freiman homomorphisms is that they preserve the additive structure of their domains. The following fact tells us precisely this, which we state without proof.

Fact 3.1 (Proposition 5.24 in [30]). *Let $\phi : A \to B$ be a k-homomorphism where $k \geq 2$, and let $Q = P_1 + \ldots + P_d = \{a_0 + a_1x_1 + \ldots a_dx_d : 0 \leq x_i \leq n_i\}$ be a generalized arithmetic progression in A. Then $\phi(Q)$ is a generalized arithmetic progression in B with the same dimension as Q. Furthermore, if ϕ is a k-isomorphism, then Q is a proper progression if and only if $\phi(Q)$ is a proper progression.*

Here, we say that a generalized arithmetic progression $Q = P_1 + \ldots + P_d$ is proper if each $q \in Q$ has a unique representation $q = x_1 + \ldots x_d$ where $x_i \in P_i$. Thus, Q is a proper progression if and only if $|Q| = |P_1| \cdots |P_d|$.

In order to apply Freiman's Theorem to subsets of $\mathbb{Z}^D$, it will be convenient for us to consider the following k-homomorphism from $\mathbb{Z}^D$ into $\mathbb{Z}$. Given a finite set $A \subset \mathbb{Z}^D$ such that all coordinates of points in A are positive, define $\phi_b : A \to \mathbb{Z}$ by

$$\phi_b(a_1, a_2, \ldots, a_D) = a_1 + a_2 b + \ldots + a_D b^{D-1}$$

for $b \in \mathbb{Z}$. In other words, we are simply sending the D-tuple $(a_1, a_2, \ldots, a_D)$ to the integer whose base b expansion corresponds to $(a_1, a_2, \ldots, a_D)$. Then ϕ_b is certainly a k-homomorphism for all $k \in \mathbb{N}$. More importantly, if b is chosen large enough (k times the absolute value of the largest coordinate of an element in A), then ϕ_b is a k-isomorphism between A and $\phi_b(A)$. Indeed, this follows from the uniqueness of base b expansions.

Now, let $C > 0$ be given and let A be a subset of $\mathbb{Z}^D$ for which $|A + A| \leq C|A|$. Assume that all coordinates of points in A are positive. We can then apply the Freiman isomorphism ϕ_b to A, so that $B = \phi_b(A)$ is a subset of $\mathbb{Z}$ with the same cardinality as A. Since ϕ_b preserves the additive structure of A, it is easy to see that $|B + B| = |A + A|$. Therefore, we have $|B + B| \leq C|A| = C|B|$, where B is a finite set of integers. By Freiman's Theorem, there exist constants d and K, depending only on C, and a generalized arithmetic progression Q in $\mathbb{Z}$ of dimension at most d for which $|Q| \leq K|B|$ and $B \subset Q$. Now consider the inverse map $\phi_b^{-1} : Q \to \mathbb{Z}^D$ which takes an integer in Q to the D-tuple corresponding to its base b expansion. Of course, it is possible for the base b-expansion of some $q \in Q$ to exceed D in length. If this happens, then we simply allow the coefficient of b^{D-1} to exceed b. This ensures that the image of Q is indeed a subset of $\mathbb{Z}^D$.

It is not difficult to show that the map ϕ_b^{-1} is a k-isomorphism between Q and the image of Q. And of course, $R = \phi_b^{-1}(Q)$ contains the set A since ϕ_b^{-1} sends B back up to A. Then R is a generalized arithmetic progression in $\mathbb{Z}^D$ that contains A. Furthermore, since ϕ^{-1} preserves cardinalities, we know that R has dimension at most d and $|R| \leq K|A|$. Here, d and K are the same constants as in the previous paragraph.

Now, if $A \subset \mathbb{Z}^D$ has some elements with negative coordinates, we can still obtain the same result. Indeed, simply shift A by a k-isomorphism (certainly, any shift is an isomorphism) to obtain a set $A' \subset \mathbb{Z}^D$ where all coordinates of points in A' are positive. Then find a generalized arithmetic progression R' in $\mathbb{Z}^D$ as we did above that covers A'. Finally, shift R' back towards A using the inverse shift to obtain a progression R that covers A. It is easy to see that R has the desired properties. Thus, Freiman's Theorem holds for subsets of $\mathbb{Z}^D$.

As it turns out, we will have to consider subsets A of $\mathbb{Z}^D$ where $|A - A| \leq C|A|$ for some constant C. Of course, Freiman's Theorem directly tells us nothing about A since we are assuming a bound on the difference set rather than on the sumset. Fortunately, Ruzsa proved that the same result holds when we consider difference sets [21]. In fact, his theorem is more general.

Theorem 3.4 (Ruzsa). *Let C be a positive constant. Then there exist constants $d \leq 2^{18}C^{32}$ and $K \leq C^5 2^d$ which depend only on C such that for each pair of finite subsets $A, B \subset \mathbb{Z}$ with $|A| = |B|$ and $|A + B| \leq C|A|$, there exists a proper generalized arithmetic progression Q with dimension at most d for which $|Q| \leq K|A|$ and $A \subset Q$.*

Note that by taking $B = -A$ in this theorem we obtain the desired result for difference sets. Our primary goal for the remainder of this section will now be to prove the Balog-Szemerédi Theorem:

Theorem 3.5. *Let A be a subset of $\mathbb{Z}^D$ with cardinality m such that $||A * A||_2^2 \geq c_0 m^3$. Then there are constants c and C depending only on c_0, and a subset $A'' \subset A$ of cardinality at least cm for which $|A'' - A''| \leq Cm$.*

Note here that the norm $||A * A||_2^2$ has an important combinatorial interpretation. Indeed, for each $x \in \mathbb{Z}^D$, we have

$$A * A(x) = \sum_{y \in \mathbb{Z}^D} A(y)A(y-x) = \sum_{y-z=x} A(y)A(z),$$

which equals the number of pairs $(y,z) \in A^2$ such that $y - z = x$. As a result,

$$\begin{aligned} ||A * A||_2^2 &= \sum_{x \in \mathbb{Z}^D} [A * A(x)]^2 = \sum_{x \in \mathbb{Z}^D} \left(\sum_{y-z=x} A(y)A(z) \right) \left(\sum_{u-w=x} A(u)A(w) \right) \\ &= \sum_{x \in \mathbb{Z}^D} \sum_{y-z=x=u-w} A(y)A(z)A(u)A(w) = \sum_{y-z=u-w} A(y)A(z)A(u)A(w), \end{aligned}$$

which equals the number the quadruples $(u,w,y,z) \in A^4$ such that $u - w = y - z$. Thus, the Balog-Szemerédi Theorem tells us that if A is a set in $\mathbb{Z}^D$ containing many quadruples (u,w,y,z) for which $u - w = y - z$, then there is a large subset A'' of A whose difference set is small. Of course, we can then apply Freiman's Theorem to A'' (we will use Ruzsa's form of it) to obtain the following beautiful corollary.

Corollary 3.2. *Let $A \subset \mathbb{Z}^D$ have cardinality m and suppose that there are at least $c_0 m^3$ quadruples $(x,y,z,w) \in A^4$ such that $x - y = z - w$. Then there is a proper generalized arithmetic progression Q of cardinality at most Cm and dimension at most d such that $|A \cap Q| \geq cm$, where the constants C, d, and c depend only on c_0.*

We now give the proof of Theorem 3.5, which will first require a technical lemma.

Lemma 3.10. *Let X be a set of cardinality m, let $\delta > 0$, and let $A_1, \ldots, A_n$ be subsets of X such that $\sum_{x=1}^n \sum_{y=1}^n |A_x \cap A_y| \geq \delta^2 mn^2$. Then there is a subset $K \in [1,n]$ of cardinality at least $2^{-1/2}\delta^5 n$ such that for at least 90% of the pairs $(x,y) \in K^2$, the intersection $A_x \cap A_y$ has size at least $\delta^2 m/2$. In particular, this result holds if $|A_x| \geq \delta m$ for each $x \in [1,n]$.*

Proof. For each $j \in X$, define $B_j = \{i : j \in A_i\}$ to be the set of indices i for which A_i contains j. Then let $E_j = B_j^2$, so $E_j \subset [1,n]^2$. Now, randomly choose five elements $j_1, \ldots, j_5$ from X uniformly and independently, so that the probability of choosing a given j is $1/m$.

Let $Y = E_{j_1} \cap \cdots \cap E_{j_5} \subset [1,n]^2$. Then $(x,y) \in Y$ if and only if A_x and A_y both contain all of $j_1, \ldots, j_5$. We now consider probabilities associated with this set Y. Note that for any $(x,y) \in [1,n]^2$ and $r \in \{1, \ldots, 5\}$, we have

$$\Pr\left((x,y) \in E_{j_r}\right) = \Pr\left(x \in B_{j_r} \text{ and } y \in B_{j_r}\right) = \Pr\left(j_r \in A_x \cap A_y\right) = \frac{|A_x \cap A_y|}{m}$$

so let $p_{xy} = |A_x \cap A_y|/m$ be this probability. Since we have chosen the five j_r uniformly and independently, we see that

$$\Pr\left((x,y) \in Y\right) = [\Pr\left((x,y) \in E_{j_r}\right)]^5 = p_{xy}^5.$$

By assumption, we have $\sum_{x,y=1}^{n} |A_x \cap A_y| \geq \delta^2 mn^2$, so

$$\sum_{x,y=1}^{n} p_{xy} = \sum_{x,y=1}^{n} \frac{|A_x \cap A_y|}{m} \geq \delta^2 n^2.$$

Using Hölder's inequality twice, we have

$$\begin{aligned}
\delta^2 n^2 \leq \sum_{x,y=1}^{n} p_{xy} \leq \sum_{x=1}^{n} \left[\left(\sum_{y=1}^{n} p_{xy}^5 \right)^{1/5} \left(n^{4/5} \right) \right] &= n^{4/5} \sum_{x=1}^{n} \left(\sum_{y=1}^{n} p_{xy}^5 \right)^{1/5} \\
\leq n^{4/5} \left[\left(\sum_{x=1}^{n} \sum_{y=1}^{n} p_{xy}^5 \right)^{1/5} \left(n^{4/5} \right) \right] &= n^{8/5} \left(\sum_{x,y=1}^{n} p_{xy}^5 \right)^{1/5}
\end{aligned}$$

and therefore,

$$\sum_{x,y=1}^{n} p_{xy}^5 \geq \delta^{10} n^2.$$

Thus, the expected size of Y is

$$\mathbb{E}[|Y|] := \sum_{x,y=1}^{n} \Pr\left((x,y) \in Y\right) \geq \delta^{10} n^2.$$

Now, let $Z = \{(x,y) \in Y : |A_x \cap A_y| < \delta^2 m/2\}$, which is simply the set of pairs (x,y) in Y such that $p_{xy} < \delta^2/2$. For any $(x,y) \in [1,n]^2$, we then have

$$\begin{aligned}
\Pr\left((x,y) \in Z\right) &= \Pr\left((x,y) \in Y \text{ and } p_{xy} < \frac{\delta^2}{2}\right) \\
&= \Pr\left(p_{xy} < \frac{\delta^2}{2}\right) \cdot \Pr\left((x,y) \in Y \middle| p_{xy} < \frac{\delta^2}{2}\right) \\
&\leq \Pr\left((x,y) \in Y \middle| p_{xy} < \frac{\delta^2}{2}\right) \leq \left(\frac{\delta^2}{2}\right)^5
\end{aligned}$$

since $\Pr\left((x,y) \in Y\right) = p_{xy}^5$. As a result, the expected size of Z is

$$\mathbb{E}[|Z|] := \sum_{x,y=1}^{n} \Pr\left((x,y) \in Z\right) \leq \frac{\delta^{10} n^2}{32}.$$

We therefore see that

$$\mathbb{E}[|Y| - 16|Z|] = \mathbb{E}[|Y|] - 16\mathbb{E}[|Z|] \geq \frac{\delta^{10} n^2}{2},$$

so in particular, there exist elements $j_1, \ldots, j_5$ in X such that the corresponding sets Y and Z satisfy $|Y| \geq 16|Z|$ and $|Y| \geq (\delta^{10} n^2)/2$.

Let $K = B_{j_1} \cap \ldots \cap B_{j_5}$, so $K^2 = E_{j_1} \cap \ldots \cap E_{j_5} = Y$. We then have $|K| = \sqrt{|Y|} \geq \sqrt{(\delta^{10} n^2)/2} = 2^{-1/2} \delta^5 n$, which is the desired lower bound on the size of K. We also see that since Z is a subset of Y and $|Y| \geq 16|Z|$, the size of Z is at most a sixteenth the size of

Y. But the elements (x, y) of Z are precisely those in $Y = K^2$ for which $|A_x \cap A_y| < \delta^2 m/2$. Therefore, at least fifteen sixteenths of the elements (x, y) in K^2 have $|A_x \cap A_y| \geq \delta^2 m/2$, as desired.

Now, suppose that $|A_x| \geq \delta m$ for each $x \in [1, n]$. We wish to show that

$$\sum_{x,y=1}^{n} |A_x \cap A_y| \geq \delta^2 m n^2.$$

For each $j \in X$, let $R(j) = |B_j|$, where, as before, $B_j = \{i : j \in A_i\}$. Thus, $R(j)$ counts the number of sets A_i to which j belongs. If we put a uniform distribution on X, then we can view R as a random variable on X with expected value $\mathbb{E}[R] = (1/m)\sum_{j \in X} R(j)$. We then know that the variance of R satisfies

$$0 \leq \mathrm{Var}(R) = \mathbb{E}[(R - \mathbb{E}[R])^2] = \mathbb{E}[R^2] - \mathbb{E}[R]^2,$$

so in particular, $\mathbb{E}[R^2] \geq \mathbb{E}[R]^2$. Note, though, that

$$m\mathbb{E}[R] = \sum_{j \in X} R(j) = \sum_{x=1}^{n} |A_x| \geq \delta m n$$

by the principle of inclusion/exclusion and our assumption that $|A_x| \geq \delta m$ for each x. Thus, $\mathbb{E}[R]^2 \geq \delta^2 n^2$. Additionally, observe that

$$m\mathbb{E}[R^2] = \sum_{j \in X} R(j)^2 = \sum_{x,y=1}^{n} |A_x \cap A_y|.$$

Indeed, if $R(j) = k$, then there exist $i_1, \ldots, i_k$ such that $j \in A_{i_1} \cap \cdots \cap A_{i_k}$, and as a result, j contributes k^2 to the sum $\sum_{x,y=1}^{n} |A_x \cap A_y|$. As a result, we have

$$\sum_{x,y=1}^{n} |A_x \cap A_y| = m\mathbb{E}[R^2] \geq m\mathbb{E}[R]^2 \geq \delta^2 m n^2,$$

which is what we wanted to show.

□

Lemma 3.10 will not only help us to give a new proof of Theorem 3.5, but it also will allow us to obtain much better bounds on the constants C, d, and c than previous proofs have done. Interestingly, the new proof is essentially a graph theoretic proof where we will apply Lemma 3.10 to a set of vertices and their neighborhoods in a conveniently defined graph.

Proof of Theorem 3.5. Let $f : \mathbb{Z}^D \to \mathbb{Z}$ be defined by

$$f(x) = A * A(x) = \sum_{y \in \mathbb{Z}^D} A(y)A(y - x).$$

We can interpret $f(x)$ as the number of distinct representations x has in the set $A - A$. Thus, f is certainly non-negative, and by the assumption that $||A * A||_2^2 \geq c_0 m^3$, we have

$||f||_2^2 \geq c_0 m^3$. It is also clear that $f(x) \leq m$ for each $x \in \mathbb{Z}^D$ since $\sum_{y \in \mathbb{Z}^D} A(y)A(y-x) \leq \sum_{y \in \mathbb{Z}^D} A(y) = m$. Furthermore, we have

$$\begin{aligned}\sum_{x \in \mathbb{Z}^D} f(x) &= \sum_{x \in \mathbb{Z}^D} \sum_{y \in \mathbb{Z}^D} A(y)A(y-x) = \sum_{y \in \mathbb{Z}^D} A(y) \sum_{x \in \mathbb{Z}^D} A(y-x) \\ &= \sum_{y \in \mathbb{Z}^D} A(y)|y-A| = m \sum_{y \in \mathbb{Z}^D} A(y) = m^2,\end{aligned}$$

so $||f||_1 = m^2$. This leads us to the additional observation that $||f||_2^2 = \sum_{x \in \mathbb{Z}^D} f(x)^2 \leq m \sum_{x \in \mathbb{Z}^D} f(x) = m^3$, so it is true that $c_0 \leq 1$.

Now, observe that the support of f is precisely the set $A - A$. It is entirely possible for $A - A$ to be large, though. Indeed, consider the set

$$A = \{(0, ar) : a \in [1, m/2]\} \cup \{(ar, 0) : a \in [1, m/2]\}$$

for some $r \in \mathbb{N}$. Then A contains many quadruples $(u, w, y, z) \in A^4$ for which $u - w = y - z$, so $||A * A||_2^2$ is large (proportional to m^3). But if r is large enough, $A - A$ will be proportional to m^2. Thus, we can only hope to find a subset A'' of A such that $|A'' - A''|$ is proportional to $|A''|$. Fortunately, this can be done.

To find such an A'', first observe that $f(x) \geq c_0 m/2$ for at least $c_0 m/2$ values of x. Indeed, if this was not true, then we would have

$$\begin{aligned}||f||_2^2 &= \sum_{x \in \mathbb{Z}^D} f(x)^2 = \sum_{\{x : f(x) \geq c_0 m/2\}} f(x)^2 + \sum_{\{x : f(x) < c_0 m/2\}} f(x)^2 \\ &< m^2 \left(\frac{c_0 m}{2}\right) + \frac{c_0 m}{2} \left(\sum_{\{x : f(x) < c_0 m/2\}} f(x) \right) \leq \frac{c_0 m^3}{2} + \frac{c_0 m}{2} ||f||_1 = c_0 m^3,\end{aligned}$$

which contradicts our bound on $||f||_2^2$ from earlier. Let us call x a popular difference if $f(x) \geq c_0 m/2$, so there are at least $c_0 m/2$ popular differences. Observe that for any $x \in \mathbb{Z}^D$,

$$f(x) = \sum_{y \in \mathbb{Z}^D} A(y)A(y-x) = \sum_{z \in \mathbb{Z}^D} A(z)A(z+x) = f(-x),$$

by setting $z = y - x$. Hence, x is a popular difference if and only if $-x$ is also. Now, define a graph G with vertex set A by placing an edge between a and b if and only if $a - b$ is a popular difference (or equivalently, if and only if $b - a$ is a popular difference).

We now wish to know how many edges there are in G. Recall that $f(x)$ is the number of representations x has in the set $A - A$, so in particular, $f(0) = m$, which is the number of loops in G. And since $f(x) = f(-x)$, we know that the number of non-loop edges in G is given by

$$E_0 = \frac{1}{2} \sum_{\{x \neq 0 : f(x) \geq c_0 m/2\}} f(x).$$

As a result, the total number of edges in G is

$$E = \frac{1}{2} f(0) + \frac{1}{2} \sum_{\{x : f(x) \geq c_0 m/2\}} f(x) \geq \frac{m}{2} + \frac{1}{2} \left(\frac{c_0 m}{2}\right)^2.$$

Now, let $d(a)$ be the degree of vertex a in G, so we have

$$\sum_{a\in A} d(a) = 2E_0 + \#(\text{loops}) = 2E - m \geq \left(\frac{c_0 m}{2}\right)^2.$$

We now claim that there must be at least $c_0^2 m/8$ vertices a such that $d(a) \geq c_0^2 m/8$. Indeed, if not, we would have

$$\begin{aligned}\sum_{a\in A} d(a) &= \sum_{\{a:d(a)\geq c_0^2 m/8\}} d(a) + \sum_{\{a:d(a)< c_0^2 m/8\}} d(a)\\ &< m\left(\frac{c_0^2 m}{8}\right) + \frac{c_0^2 m}{8}\left(\sum_{\{a:d(a)<c_0^2 m/8\}} 1\right)\\ &\leq \frac{c_0^2 m^2}{8} + \frac{c_0^2 m^2}{8} = \left(\frac{c_0 m}{2}\right)^2,\end{aligned}$$

which contradicts what we just found. Let $\delta = c_0^2/8$ and let $X = \{a \in A : d(a) \geq \delta m\}$, so $n = |X| \geq \delta m$. Let $a_1, \ldots, a_n$ be the vertices in X and let $A_1, \ldots, A_n$ be their respective neighborhoods (that is, let A_i be the set of vertices in G that are joined to a_i by an edge). Since $|A_i| \geq \delta m$, we can apply Lemma 3.10 to find a set $A' \subset X$ with $|A'| \geq \delta^5 n/\sqrt{2} \geq \delta^6 m/\sqrt{2}$ such that for at least 90% of the pairs $a_i, a_j \in A'$, the intersection $A_i \cap A_j$ has size at least $\delta^2 m/2$. For simplicity, let $\alpha = \delta^6/\sqrt{2}$, so $|A'| \geq \alpha m$.

We now define a second graph, H, with vertex set A'. Place an edge between a_i and a_j if and only if $|A_i \cap A_j| \geq \delta^2 m/2$. We again ask how many edges are in this graph. Similar to before, the number of non-loop edges is given by

$$E_0 = \frac{1}{2}\#\left((a_i, a_j) \in A' \times A' : |A_i \cap A_j| \geq \frac{\delta^2 m}{2} \text{ and } i \neq j\right).$$

If we let L denote the number of loops in H, then the total number of edges is given by

$$\begin{aligned}E &= L + \frac{1}{2}\#\left((a_i, a_j) \in A' \times A' : |A_i \cap A_j| \geq \frac{\delta^2 m}{2} \text{ and } i \neq j\right)\\ &= \frac{1}{2}L + \frac{1}{2}\#\left((a_i, a_j) \in A' \times A' : |A_i \cap A_j| \geq \frac{\delta^2 m}{2}\right)\\ &\geq \frac{1}{2}L + \frac{1}{2}\frac{9}{10}|A'|^2\end{aligned}$$

since at least 90% of the pairs $a_i, a_j \in A'$ have $|A_i \cap A_j| \geq \delta^2 m/2$. Letting $d(a)$ denote the degree of vertex a in H, we have

$$\sum_{a\in A'} d(a) = 2E_0 + L = 2E - L \geq \frac{9}{10}|A'|^2.$$

We now claim that there are at least $(1/2)|A'|$ vertices with degree at least $(4/5)|A'|$. Indeed, if not, then we would have

$$\begin{aligned}\sum_{a\in A'} d(a) &= \sum_{\{a:d(a)\geq(4/5)|A'|\}} d(a) + \sum_{\{a:d(a)<(4/5)|A'|\}} d(a)\\ &< |A'|\left(\frac{|A'|}{2}\right) + \frac{4|A'|}{5}\left(\sum_{\{a:d(a)<(4/5)|A'|\}} 1\right)\\ &\leq \frac{|A'|^2}{2} + \frac{4|A'|}{5}\left(|A'| - \frac{1}{2}|A'|\right) = \frac{9}{10}|A'|^2,\end{aligned}$$

which contradicts what we just found. Thus, let A'' be the set of $a \in A'$ such that $d(a) \geq (4/5)|A'|$. Then we have $|A''| \geq (1/2)|A'| \geq \alpha m/2$, where we recall that $\alpha = \delta^5/\sqrt{2}$. We now wish to show that $A'' - A''$ has small cardinality (that is, on the order of $|A''|$).

Let a_i and a_j be elements of A''. Since the degrees of a_i and a_j in H are each at least $(4/5)|A'|$, we know that there are at least $(4/5)|A'| + (4/5)|A'| - |A'| = (3/5)|A'|$ elements $a_k \in A'$ such that a_k is joined to both a_i and a_j. For each such a_k, this implies that

$$|A_i \cap A_k| \geq \frac{\delta^2 m}{2} \text{ and } |A_j \cap A_k| \geq \frac{\delta^2 m}{2}.$$

Now, let $b \in A_i \cap A_k$. By definition of A_i and A_k, we then know that b is joined to both a_i and a_k in the graph G, so $a_i - b$ and $a_k - b$ are popular differences. Thus, $f(a_i - b) \geq c_0 m/2$ and $f(a_k - b) \geq c_0 m/2$, which implies that $a_i - b$ and $a_k - b$ each have at least $c_0 m/2$ representations in $A - A$ (recall that $f(x)$ is the number of distinct representations of x in the set $A - A$). If $a_i - b = p - q$ and $a_k - b = r - s$ are such representations (so $p, q, r, s \in A$), then we can write

$$a_i - a_k = a_i - b - (a_k - b) = p - q - (r - s).$$

Hence, there are at least $(c_0 m/2)^2$ ways of writing $a_i - a_k$ as $p - q - (r - s)$ where $(p, q, r, s) \in A^4$, $p - q = a_i - b$ and $r - s = a_k - b$. Note also that if b_1 and b_2 are elements of $A_i \cap A_k$, then the set of quadruples $(p, q, r, s) \in A^4$ where $p - q = a_i - b_1$ and $r - s = a_k - b_1$ is disjoint from the set of quadruples $(p, q, r, s) \in A^4$ where $p - q = a_i - b_2$ and $r - s = a_k - b_2$. Thus, if we sum over all $b \in A_i \cap A_k$, we see that there are at least

$$|A_i \cap A_k| \left(\frac{c_0 m}{2}\right)^2 \geq \frac{\delta^2 m}{2} = \frac{\delta^2 c_0^2 m^3}{8}$$

quadruples $(p, q, r, s) \in A^4$ such that $a_i - a_k = p - q - (r - s)$. Of course, the same is true for $a_j - a_k$.

But now note that we can write $a_i - a_j = a_i - a_k - (a_j - a_k)$, so for each pair of quadruples $(p, q, r, s) \in A^4$ and $(t, u, v, w) \in A^4$ such that $a_i - a_k = p - q - (r - s)$ and $a_j - a_k = t - u - (v - w)$, we have

$$a_i - a_j = p - q - (r - s) - [t - u - (v - w)].$$

Thus, for each k, there are at least $(\delta^2 c_0^2 m^3/8)^2$ ways of writing $a_i - a_j$ as $p - q - (r - s) - [t - u - (v - w)]$ where $(p, q, r, s, t, u, v, w) \in A^8$, $p - q - (r - s) = a_i - a_k$, and $t - u - (v - w) = a_j - a_k$. Note again that if k_1 and k_2 are such that the vertices a_{k_1} and a_{k_2} in H are both joined to a_i and a_j, then the set of such octuples corresponding to k_1 is disjoint from the set of such octuples corresponding to k_2. Hence, if we sum over all k such that a_k is joined to a_i and a_j in H, then we find that there are at least

$$\left(\frac{3}{5}|A'|\right)\frac{\delta^4 c_0^4 m^6}{64} \geq \frac{3}{320}\alpha m \delta^4 c_0^4 m^6 > \frac{1}{120}\alpha m^7 \delta^4 c_0^4$$

octuples $(p,q,r,s,t,u,v,w) \in A^8$ such that $a_i - a_j = p - q - (r-s) - [t - u - (v-w)]$.

This result holds for all differences $a_i - a_j$ in $A'' - A''$. Of course, there are m^8 octuples in A^8 and each such octuple (p,q,r,s,t,u,v,w) can satisfy $a_i - a_j = p - q - (r-s) - [t-u-(v-w)]$ for at most one difference $a_i - a_j$. As a result, we see that there are at most

$$\frac{m^8}{\frac{1}{120}\alpha m^7 \delta^4 c_0^4} = \frac{2^{30}120\sqrt{2}m}{c_0^{24}} < \frac{2^{38}m}{c_0^{24}}$$

differences in $A'' - A''$. Hence,

$$|A'' - A''| \leq \frac{2^{38}}{c_0^{24}}m$$

as desired. In addition, recall that we have

$$|A''| \geq \alpha m/2 = \frac{c_0^{12}m}{2\cdot 8^6\sqrt{2}} > \frac{c_0^{12}}{2^{20}}m,$$

so the theorem is proved with $c = 2^{-20}c_0^{12}$ and $C = 2^{38}c_0^{-24}$.

□

We now conclude this combinatorial section with the important corollary that links our results from Lemma 3.1 and Theorem 3.5.

Corollary 3.3. *Let $B \subset \mathbb{Z}_N$ have cardinality βN, and let $\phi : B \to \mathbb{Z}_N$ be a function with at least $c_0 N^3$ additive quadruples. Then there are constants γ and η depending only on β and c_0, a mod-N arithmetic progression $P \subset \mathbb{Z}_N$ with $|P| \geq N^\gamma$, and a linear function $\psi : P \to \mathbb{Z}_N$ such that $\phi(s)$ is defined and equal to $\psi(s)$ for at least $\eta|P|$ values of $s \in P$.*

Proof. Let Γ be the graph of ϕ, so $\Gamma = \{(b,\phi(b)) : b \in B\}$. For now, we consider Γ to be a subset of $\mathbb{Z}^2$ rather than of $(\mathbb{Z}_N)^2$. We then have $|\Gamma| = |B| = \beta N$, and since $\Gamma \subset \mathbb{Z}^2$, we know from before that $||\Gamma * \Gamma||_2^2$ is the number of quadruples $(x,y,z,w) \in \Gamma^4$ for which $x - y = z - w$. But this is simply the number of quadruples

$$\Big((b_1,\phi(b_1)),(b_2,\phi(b_2)),(b_3,\phi(b_3)),(b_4,\phi(b_4))\Big) \in \Gamma^4$$

such that $(b_1,\phi(b_1)) - (b_2,\phi(b_2)) = (b_3,\phi(b_3)) - (b_4,\phi(b_4))$, which is equal to the number of quadruples $(b_1,b_2,b_3,b_4) \in B^4$ such that $b_1 - b_2 = b_3 - b_4$ and $\phi(b_1) - \phi(b_2) = \phi(b_3) - \phi(b_4)$. By rearranging these equalities, it is clear that this equals the number of quadruples $(b_1,b_4,b_3,b_2) \in B^4$ such that $b_1 + b_4 = b_3 + b_2$ and $\phi(b_1) + \phi(b_4) = \phi(b_3) + \phi(b_2)$, which is simply the number of ϕ-additive quadruples. By assumption, ϕ has at least c_0N^3 additive quadruples, so

$$||\Gamma * \Gamma||_2^2 \geq c_0 N^3 = \frac{c_0}{\beta^3}|\Gamma|^3.$$

By Corollary 3.2, we then know that there exists a proper generalized arithmetic progression $Q = P_1 + \ldots + P_d$ with $|Q| \leq CN$ and $|\Gamma \cap Q| \geq cN$, where the constants d, C, and c depend only on c_0 and β.

Without loss of generality, we can assume that $|P_i| \leq |P_{i+1}|$ for each $i \in [1, d-1]$ (rearranging the order of the sums in a sumset does not alter the sumset itself). Since Q is a proper progression (that is, each element in Q has a unique representation as a sum of elements in $P_1, \ldots, P_d$), we know that $|Q| = |P_1| \cdots |P_d|$, and therefore,

$$|P_d|^d \geq |Q| \geq |\Gamma \cap Q| \geq cN,$$

so $|P_d| \geq (cN)^{1/d}$. Now, for each $x \in P_1 + \ldots + P_{d-1}$, let $R_x = x + P_d$. Then we have $|R_x| = |P_d| \geq (cN)^{1/d}$ for each x, and $Q = \cup R_x$, where the union is taken over all $x \in P_1 + \ldots + P_{d-1}$. Furthermore, this is a disjoint union because of the fact that Q is proper. We claim that there is an x such that $|\Gamma \cap R_x| \geq cC^{-1}|R_x|$. If not, we would have

$$\begin{aligned} |\Gamma \cap Q| &= \left| \bigcup_x (\Gamma \cap R_x) \right| = \sum_x |\Gamma \cap R_x| < \sum_x cC^{-1}|R_x| \\ &= cC^{-1}|Q| \leq cC^{-1}(CN) = cN \end{aligned}$$

which contradicts our assumption that $|\Gamma \cap Q| \geq cN$. Therefore, there is a one-dimensional progression R such that $|R| \geq (cN)^{1/d}$ and $|\Gamma \cap R| \geq cC^{-1}|R|$.

Assume now that $|\Gamma \cap R| \geq 2$. Let P be the projection of R onto the horizontal axis, and let A be the projection of R onto the vertical axis. Since R is a one-dimensional progression, we know that P and A are progressions also. Furthermore, since R and Γ share at least two points, and Γ is the graph of a function, we know that R is not vertical. As a result, we must have $|P| = |R|$. Define a function $\psi : P \to \mathbb{Z}$ by letting $\psi(s)$ be the integer in A such that $(s, \psi(s)) \in R$. Note that this is well-defined because R is not vertical, and it is linear since R is a one-dimensional progression. We then know that Γ contains $|R \cap \Gamma|$ pairs $(s, \psi(s))$ where $s \in P$. Since $|R \cap \Gamma| \geq cC^{-1}|R| = cC^{-1}|P|$, we see that there are at least $cC^{-1}|P|$ values of $s \in P$ for which $\psi(s) = \phi(s)$. Also, $|P| \geq (cN)^{1/d}$ again by the fact that $|P| = |R|$.

Now suppose that $\Gamma \cap R = \{(s, r)\}$. Let $P = \{s + q : q = 0, \ldots, |R| - 1\}$ so that P is an arithmetic progression in $\mathbb{Z}$ and $|P| = |R| \geq (cN)^{1/d}$. Define a function $\psi : P \to \mathbb{Z}$ by $\psi(s + q) = r + q$, so ψ is certainly linear. Since $(s, r) \in \Gamma$, we know that $\phi(s) = r = \psi(s)$, so ϕ and ψ agree on $\{s\}$, which is a set of size $1 \geq cC^{-1}|R| = cC^{-1}|P|$. Thus, the same result still holds.

We have therefore found a progression P in $\mathbb{Z}$ that almost satisfies the conclusions of the corollary; recall that we want P to be a mod-N progression. Note that for each pair $(s, \psi(s)) \in \Gamma \cap R$, we have $s \in [0, N-1]$ and $\phi(s) \in [0, N-1]$ by the fact that Γ is the graph of a function (embedded into $\mathbb{Z}^2$) from a subset of $\mathbb{Z}_N$ into $\mathbb{Z}_N$. Thus, if we reduce the elements of P modulo N, the values $s \in P$ for which $(s, \psi(s)) \in \Gamma \cap R$ do not change, and they form a progression in $\mathbb{Z}_N$. Let P' be the set of such s, considered as a subset of $\mathbb{Z}_N$ rather than of $\mathbb{Z}$, and since P' is a progression, let $P' = \{a + qr : q = 0, \ldots, m-1\}$. It is clear that $m \leq |P|$, so extend P' to the mod-N progression $P'' = \{a + qr : q = 0, \ldots, |P| - 1\}$. Also, extend ψ to a linear function $\psi'' : P'' \to \mathbb{Z}$ so that $\psi''(s) = \psi(s)$ for each $s \in P'$. We then know that $|P''| = |P| \geq (cN)^{1/d}$ and there are at least $cC^{-1}|P''|$ values of $s \in P''$ such that $\psi''(s) = \phi(s)$ (namely, those s for which $(s, \psi(s)) \in \Gamma \cap R$).

As a result, we can set η to be cC^{-1}, but our lower bound on $|P''|$ is not of the form N^γ. In order to find such a γ, we do the following. Let N_0 be the smallest integer such that

$cN_0 > 1$ (note that N_0 depends only on c, which in turn depends only on c_0 and β). Then, let $\gamma = \log(cN_0)/(\log(N_0^d))$. For all $N \geq N_0$, we have $\gamma \leq \log(cN)/(d\log(N))$, so $N^\gamma \leq (cN)^{1/d}$. Also, for each $N < N_0$, we know that $cN \leq 1$, so the statement that $|P''| \geq (cN)^{1/d}$ is equivalent to the statement that $|P''| \geq 1$. This, in turn, is equivalent to $|P''| \geq N^\gamma$ because $N^\gamma \leq 1$ if and only if $\log(cN)/d \leq 0$. Therefore, $\gamma = \log(cN_0)/(d\log(N_0))$ and $\eta = cC^{-1}$ are our desired constants, and these depend only on c_0 and β.

□

Gowers mentions that the bounds in Theorem 3.4 (Ruzsa's version of Freiman's Theorem) imply the existence of an absolute constant K that allow us to take $\gamma = c_0^K$ and $\eta = \exp\left(-(1/c_0)^K\right)$ in Corollary 3.3. These are the bounds that we will use at the end of Gowers's proof.

3.4 Quadratic Fourier Sums

Recall from the beginning of section 3.2 that if A fails to be quadratically α-uniform, then there is a subset $B \subset \mathbb{Z}_N$ with $|B| \geq (\alpha/2)N$ and a function $\phi : B \to \mathbb{Z}_N$ for which $\left|\widetilde{\Delta(f;k)}(\phi(k))\right| \geq (\alpha/2)^{1/2}N$ whenever $k \in B$. In particular,

$$\sum_{k\in B}\left|\widetilde{\Delta(f;k)}(\phi(k))\right|^2 \geq (\alpha/2)^2N^3.$$

From Proposition 3.1, we then know that B contains at least $(\alpha/2)^8N^3$ ϕ-additive quadruples. Then by Corollary 3.3, there are constants γ and η, depending only on α, for which we can find a mod-N arithmetic progression P with $|P| \geq N^\gamma$ and a linear function $\psi : P \to \mathbb{Z}_N$ such that $\phi(s)$ agrees with $\psi(s)$ for at least $\eta|P|$ values of $s \in P$. Our goal in this section is to find an arithmetic progression on which the density of A actually increases. To do this, we must use a type of quadratic Fourier analysis.

Proposition 3.2. *Let $A \subset \mathbb{Z}_N$ and let f be the balanced function of A. Let P be an arithmetic progression in $\mathbb{Z}_N$ of cardinality T, and suppose there exist λ, μ such that $\sum_{k\in P}\left|\widetilde{\Delta(f;k)}(\lambda k+\mu)\right|^2 \geq \beta N^2T$. Then there exist quadratic polynomials $\psi_0, \psi_1, \ldots, \psi_{N-1} : \mathbb{Z}_N \to \mathbb{Z}_N$ for which*

$$\sum_s\left|\sum_{z\in P+s} f(z)\omega^{-\psi_s(z)}\right| \geq \frac{\beta NT}{\sqrt{2}}.$$

Proof. For technical reasons, let $Q = -P$ and $\eta = -\lambda$. Then we see that

$$\sum_{k\in Q}\left|\widetilde{\Delta(f;k)}(\eta k+\mu)\right|^2 \geq \beta N^2T.$$

By expanding the left-hand side, we have the following:

$$\beta N^2T \leq \sum_{k\in Q}\left|\sum_s \Delta(f;k)(s)\omega^{-(\eta k+\mu)s}\right|^2$$

$$\begin{aligned}
&= \sum_{k\in Q}\left(\sum_{s}\Delta(f;k)(s)\omega^{-(\eta k+\mu)s}\right)\overline{\left(\sum_{t}\Delta(f;k)(t)\omega^{-(\eta k+\mu)t}\right)}\\
&= \sum_{k\in Q}\sum_{s,t}\Delta(f;k)(s)\omega^{-(\eta k+\mu)s}\overline{\Delta(f;k)(t)}\omega^{-(\eta k+\mu)t}\\
&= \sum_{k\in Q}\sum_{s,t} f(s)f(s-k)f(t)f(t-k)\omega^{-(\eta k+\mu)(s-t)}.
\end{aligned}$$

Let $u = s - t$ so that if s is fixed, u varies over $\mathbb{Z}_N$ as t does. Thus, this becomes

$$= \sum_{k\in Q}\sum_{s,u} f(s)f(s-k)f(s-u)f(s-k-u)\omega^{-(\eta k+\mu)u}.$$

Since Q is an arithmetic progression modulo N, we can write it in the form $Q = \{x+d, x+2d, \ldots, x+Td\}$. This allows us to express the above sum as

$$= \sum_{i=1}^{T}\sum_{s,u} f(s)f(s-x-id)f(s-u)f(s-x-id-u)\omega^{-(\eta x+\eta id+\mu)u}.$$

Observe that for each $u \in \mathbb{Z}_N$, there are T ways of writing $u = y + jd$ where $y \in \mathbb{Z}_N$ and $j = 1, \ldots, T$. We can therefore write the sum as

$$\begin{aligned}
&= \sum_{i=1}^{T}\sum_{s} f(s)f(s-x-id)\left(\frac{1}{T}\sum_{y}\sum_{j=1}^{T} f(s-y-jd)\right.\\
&\qquad \left.\times f(s-x-id-y-jd)\omega^{-(\eta x+\eta id+\mu)(y+jd)}\right)\\
&= \frac{1}{T}\sum_{s,y}\sum_{i,j=1}^{T} f(s)f(s-x-id)f(s-y-jd)\\
&\qquad \times f(s-x-id-y-jd)\omega^{-(\eta x+\eta id+\mu)(y+jd)}.
\end{aligned}$$

Define a function $\gamma : \mathbb{Z}_N \times \mathbb{Z}_N \to \mathbb{R}$ by

$$\gamma(s,y) = \frac{1}{T^2}\left|\sum_{i,j=1}^{T} f(s-x-id)f(s-y-jd)\right.$$
$$\left.\times f(s-x-id-y-jd)\omega^{-(\eta x+\eta id+\mu)(y+jd)}\right|.$$

We then have

$$\begin{aligned}
\beta N^2 T &\le \frac{1}{T}\sum_{s,y}\left|\sum_{i,j=1}^{T} f(s)f(s-x-id)f(s-y-jd)\right.\\
&\qquad \left.\times f(s-x-id-y-jd)\omega^{-(\eta x+\eta id+\mu)(y+jd)}\right|\\
&= \frac{1}{T}\sum_{s,y}|f(s)|\gamma(s,y)T^2 \le T\sum_{s,y}\gamma(s,y)
\end{aligned}$$

since $|f(s)| \leq 1$ for each s. Therefore, we see that $(1/N^2)\sum_{s,y}\gamma(s,y) \geq \beta$, or in other words, the average value of γ as it ranges over $\mathbb{Z}_N \times \mathbb{Z}_N$ is at least β.

We now briefly move to a more general setting. Suppose that f_1, f_2, and f_3 are real-valued functions on $\mathbb{Z}_N$ such that $|f_i(s)| \leq 1$ for all s and

$$\left|\sum_{i,j=1}^{T} f_1(i)f_2(j)f_3(i+j)\omega^{-(ai+bj-2cij)}\right| \geq \delta T^2$$

for some real constants a, b, c, δ. Because $2ij = (i+j)^2 - i^2 - j^2$, we can write this as

$$\begin{aligned}\delta T^2 &\leq \left|\sum_{i,j=1}^{T} f_1(i)f_2(j)f_3(i+j)\omega^{-\left(ai+bj-c\left[(i+j)^2-i^2-j^2\right]\right)}\right| \\ &= \left|\sum_{i,j=1}^{T} f_1(i)\omega^{-(ai+ci^2)}f_2(j)\omega^{-(bj+cj^2)}f_3(i+j)\omega^{c(i+j)^2}\right| \\ &= \left|\sum_{i,j=1}^{T} f_1(i)\omega^{-(ai+ci^2)}f_2(j)\omega^{-(bj+cj^2)}\left(\sum_{k=1}^{2T} f_3(k)\omega^{ck^2}\left[\frac{1}{N}\sum_r \omega^{-r(i+j-k)}\right]\right)\right|,\end{aligned}$$

where we have substituted k for $i+j$. This is valid because we multiplied by the sum $\frac{1}{N}\sum_r \omega^{-r(i+j-k)}$, which vanishes when $k \neq i+j$ and is 1 when $k = i+j$. We can then write this as

$$= \frac{1}{N}\left|\sum_r \sum_{i,j=1}^{T}\sum_{k=1}^{2T} f_1(i)\omega^{-(ai+ci^2)}f_2(j)\omega^{-(bj+cj^2)}f_3(k)\omega^{ck^2}\omega^{-r(i+j-k)}\right|.$$

Now, let $g_1, g_2, g_3 : \mathbb{Z}_N \to \mathbb{C}$ be defined by $g_1(r) = \sum_{i=1}^{T} f_1(i)\omega^{-(ai+ci^2)}\omega^{-ri}$, $g_2(r) = \sum_{j=1}^{T} f_2(j)\omega^{-(bj+cj^2)}\omega^{-rj}$, and $g_3(r) = \sum_{k=1}^{2T} f_3(k)\omega^{-ck^2}\omega^{-rk}$ so that we have

$$\begin{aligned}&\frac{1}{N}\left|\sum_r \sum_{i,j=1}^{T}\sum_{k=1}^{2T} f_1(i)\omega^{-(ai+ci^2)}f_2(j)\omega^{-(bj+cj^2)}f_3(k)\omega^{ck^2}\omega^{-r(i+j-k)}\right| \\ &\quad= \frac{1}{N}\left|\sum_r g_1(r)g_2(r)\overline{g_3(r)}\right| \leq \frac{1}{N}||g_1||_\infty \sum_r |g_2(r)| \cdot |g_3(r)| \\ &\quad\leq \frac{1}{N}||g_1||_\infty \left(\sum_r |g_2(r)|^2\right)^{1/2}\left(\sum_r |g_3(r)|^2\right)^{1/2} = \frac{1}{N}||g_1||_\infty ||g_2||_2 ||g_3||_2.\end{aligned}$$

Therefore, $||g_1||_\infty ||g_2||_2 ||g_3||_2 \geq \delta T^2 N$. We now wish to find a lower bound for $||g||_\infty$, so we will estimate $||g_2||_2$ and $||g_3||_2$. To do this, note first that for each $r \in \mathbb{Z}_N$, we have $g_2(r) = \tilde{h_2}(r)$ and $g_3(r) = \tilde{h_3}(r)$ where h_2 and h_3 are given by

$$h_2(r) = \begin{cases} f_2(x)\omega^{-(bx+cx^2)} & \text{if } x \in [1,T] \\ 0 & \text{otherwise}\end{cases}$$

and

$$h_3(x) = \begin{cases} f_3(x)\omega^{-cx^2} & \text{if } x \in [1,2T] \\ 0 & \text{otherwise}\end{cases}$$

As a result, we have the estimate

$$||g_2||_2^2 = \sum_r |g_2(r)|^2 = \sum_r |\tilde{h_2}(r)|^2 = N \sum_x |h_2(x)|^2 = N \sum_{x=1}^{T} |h_2(x)|^2 \leq NT,$$

where we have used Parseval's formula and the assumption that $|f_2(x)| \leq 1$ for all $x \in \mathbb{Z}_N$. Similarly, we have an estimate on $||g_3||_2$:

$$||g_3||_2^2 = \sum_r |g_3(r)|^2 = \sum_r |\tilde{h_3}(r)|^2 = N \sum_x |h_3(x)|^2 = N \sum_{x=1}^{2T} |h_3(x)|^2 \leq 2NT.$$

Therefore, we see that $\delta T^2 \leq \frac{1}{N} ||g_1||_\infty (NT)^{1/2} (2NT)^{1/2} = ||g_1||_\infty \sqrt{2} T$, so in particular, there exists some $r \in \mathbb{Z}_N$ for which $|g_1(r)| \geq \delta T / \sqrt{2}$. By definition of g_1, we then have

$$\left| \sum_{i=1}^{T} f_1(i) \omega^{-(ai+ci^2+ri)} \right| \geq \frac{\delta T}{\sqrt{2}}.$$

If we let $\phi(i) = (a+r)i + ci^2$, then we obtain a quadratic polynomial for which

$$\left| \sum_{i=1}^{T} f_1(i) \omega^{-\phi(i)} \right| \geq \frac{\delta T}{\sqrt{2}}.$$

We now wish to apply this fact to our original function f, the balanced function of A. For each fixed pair s, y, let $f_1(i) = f(s - x - id)$, $f_2(j) = f(s - y - jd)$, and $f_3(k) = f(s - x - y - kd)$. We then have

$$\begin{aligned} &\left| \sum_{i,j=1}^{T} f_1(i) f_2(j) f_3(i+j) \omega^{-(ai+bj-2cij)} \right| \\ &= \left| \sum_{i,j=1}^{T} f(s-x-id) f(s-y-jd) f(s-x-y-id-jd) \omega^{-(ai+bj-2cij)} \right| \cdot \left| \omega^{-\eta xy - \mu y} \right| \\ &= \left| \sum_{i,j=1}^{T} f(s-x-id) f(s-y-jd) f(s-x-id-y-jd) \omega^{-(\eta x + \eta id + \mu)(y+jd)} \right| \\ &= \gamma(s,y) T^2 \end{aligned}$$

for $a = \eta yd$, $b = \eta xd + \mu d$, and $c = -(1/2)\eta d^2$. We can therefore apply the general fact that we found, with $\delta = \gamma(s,y)$. This fact tells us that for each pair (s,y) in $\mathbb{Z}_N \times \mathbb{Z}_N$, there is a quadratic polynomial $\phi_{s,y}$ such that

$$\left| \sum_{i=1}^{T} f(s-x-id) \omega^{-\phi_{s,y}(i)} \right| \geq \frac{\gamma(s,y) T}{\sqrt{2}}.$$

Let $\gamma(s) = (1/N) \sum_y \gamma(s,y)$ be the average of $\gamma(s,y)$ over y. We then know that given s, there is a y for which

$$\left| \sum_{i=1}^{T} f(s-x-id) \omega^{-\phi_{s,y}(i)} \right| \geq \frac{\gamma(s) T}{\sqrt{2}}.$$

Call this quadratic polynomial ϕ_s so that

$$\left|\sum_{i=1}^{T} f(s-x-id)\omega^{-\phi_s(i)}\right| \geq \frac{\gamma(s)T}{\sqrt{2}}.$$

For each s, define the linear function $h_s(t) = -(1/d)(t+x-s)$, and let ψ_s be the quadratic polynomial defined by $\psi_s(t) = \phi_s(h_s(t))$. We then have

$$\begin{aligned}\left|\sum_{i=1}^{T} f(s-x-id)\omega^{-\psi_s(s-x-id)}\right| &= \left|\sum_{i=1}^{T} f(s-x-id)\omega^{-\phi_s(h_s(s-x-id))}\right| \\ &= \left|\sum_{i=1}^{T} f(s-x-id)\omega^{-\phi_s(i)}\right| \geq \frac{\gamma(s)T}{\sqrt{2}}.\end{aligned}$$

But notice that

$$\left|\sum_{i=1}^{T} f(s-x-id)\omega^{-\psi_s(s-x-id)}\right| = \left|\sum_{z\in Q} f(s-z)\omega^{-\psi_s(s-z)}\right| = \left|\sum_{z\in P+s} f(z)\omega^{-\psi_s(z)}\right|$$

since $Q = -P$. We therefore have

$$\left|\sum_{z\in P+s} f(z)\omega^{-\psi_s(z)}\right| \geq \frac{\gamma(s)T}{\sqrt{2}},$$

so summing over s, we find that

$$\sum_s \left|\sum_{z\in P+s} f(z)\omega^{-\psi_s(z)}\right| \geq \frac{T}{\sqrt{2}}\sum_s \gamma(s).$$

Note, however, that $\sum_s \gamma(s)$ is simply the product of N and the average value of $\gamma(s,y)$ over all pairs $(s,y) \in \mathbb{Z}_N \times \mathbb{Z}_N$. Earlier, we showed that this average value is at least β, so we know that $\sum_s \gamma(s)$ is at least βN. This gives the desired result.

□

It should be noted here that the above proof is not entirely complete. Observe that we allowed the coefficients of the quadratic functions ψ_s to be non-integer rationals, even though we claimed that $\psi_s : \mathbb{Z}_N \to \mathbb{Z}_N$. We can, however, modify the ψ_s in order to ensure that they take values in $\mathbb{Z}_N$, but we will not concern ourselves with the details here.

As we shall see later, the existence of an arithmetic progression P and its corresponding quadratic polynomials $\psi_0, \ldots, \psi_{N-1}$ will be important if A is not quadratically uniform. It does not necessarily follow, however, that A has increased density on this progression P, or even on any of its translates $P+s$. Instead, we will have to partition each progression $P+s$ into sub-progressions, and it will be one of these sub-progressions on which A has increased density. Of course, we cannot partition $P+s$ arbitrarily; our task now is to find what properties these sub-progressions must have in order to guarantee that A will have increased density on at least one of them. To do this, we begin with a discrete analogue of a famous result due to Weyl.

Theorem 3.6. *Let N be sufficiently large, and let $a \in \mathbb{Z}_N$. Then for any $t \leq N$, there exists a $p \leq t$ such that $\left|p^2 a\right| \leq Ct^{-1/8}N$, where C is an absolute constant.*

Gowers omits the proof of this result, and we shall do the same here. He mentions, though, that the bounds in this statement of the inequality are not the best known. His larger paper [9] includes a proof of the result along with better bounds. This inequality will help us divide progressions into sub-progressions, but we first need a lemma that is essentially a combinatorial result.

Lemma 3.11. *Let $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ be linear of the form $\phi(x) = ax + b$, and let $r, s \leq N$. Then for some $m \leq (8rN/s)^{1/2}$, the set $\{0, 1, \ldots, r-1\}$ can be partitioned into arithmetic progressions $P_1, \ldots, P_m$ (in $\mathbb{Z}$, not just in $\mathbb{Z}_N$) for which the diameter of $\phi(P_j)$ is at most s for every j. Furthermore, the sizes of the P_j differ by at most one.*

Proof. Let $t = \left\lceil (rN/2s)^{1/2} \right\rceil$. Consider the elements $\phi(0), \ldots, \phi(t)$, but instead of considering them as elements of $\mathbb{Z}_N$, consider them as integers in the natural way. Since there are $t+1$ such integers, we know by the pigeon-hole principle that there exist $i \neq j$ in $[0, t]$ such that $|\phi(i) - \phi(j)| \leq N/t$. But by the linearity of ϕ, we then have

$$|\phi(i-j) - \phi(0)| = |\phi(j-i) - \phi(0)| = |\phi(i) - \phi(j)| \leq N/t.$$

Thus, let $u = |i - j|$ so that $|\phi(u) - \phi(0)| \leq N/t$.

Now take the interval $\{0, 1, \ldots, r-1\}$, and split it into u congruence classes modulo u. Then each congruence class has size either $\lceil r/u \rceil$ or $\lfloor r/u \rfloor$, and each class is an arithmetic progression in $\mathbb{Z}$ with common difference u. Suppose now that P is a set consisting of consecutive elements of one of these congruence classes with $|P| \leq \lceil st/N \rceil$. We can then write P in the form $P = \{x + qu : q = 0, \ldots, |P| - 1\}$ where x is an integer in $\{0, 1, \ldots, r-1\}$. Then, for any $q_1, q_2 \in [0, |P| - 1]$, we have

$$\begin{aligned}
&|\phi(x + q_1 u) - \phi(x + q_2 u)| = |a(x + q_1 u) + b - a(x + q_2 u) - b| = |ua(q_1 - q_2)| \\
&\leq (|P| - 1)\,|ua| = (|P| - 1)\,|\phi(u) - \phi(0)| \leq \frac{st}{N} \cdot \frac{N}{t} = s,
\end{aligned}$$

so in particular, the diameter of $\phi(P)$ is at most s. We now wish to partition each congruence class into progressions with size at most $\lceil st/N \rceil$ so that the total number of progressions is at most $(2rN/s)^{1/2}$ and the progressions differ in size by at most 1.

To do this, divide $\lfloor 2rN/(ust) \rfloor$ into $\lfloor r/u \rfloor$, thereby obtaining integers m_1 and m_2 such that

$$\left\lfloor \frac{r}{u} \right\rfloor = m_1 \left\lfloor \frac{2rN}{ust} \right\rfloor + m_2$$

and $0 \leq m_2 < \lfloor 2rN/(ust) \rfloor$. Let $C = \{x + qu : q = 0, \ldots, |C| - 1\}$ be a given congruence class, so that $|C| = \lfloor r/u \rfloor$ or $|C| = \lceil r/u \rceil$. Now lay down arithmetic progressions

$$Q_j = \{x + qu : q = jm_1, \ldots, (j+1)m_1 - 1\}$$

for $j = 0, 1, \ldots, \lfloor 2rN/(ust) \rfloor - 1$. We have therefore covered the first $m_1 \lfloor 2rN/(ust) \rfloor$ elements of C with $\lfloor 2rN/(ust) \rfloor$ progressions, each of size m_1. As a result, there are either m_2 or $m_2 + 1$ elements of C that have not been covered (depending on the size of C). Recall that $m_2 < \lfloor 2rN/(ust) \rfloor$, so the number of elements in C that have not been covered is at most the number of progressions that we have already laid down. As a result, we can shift the progressions Q_j to the right, and add at most one element to each progression so that the new progressions P_j cover C completely. We still have $\lfloor 2rN/(ust) \rfloor$ progressions, but now the size of $|P_j|$ is either m_1 or $m_1 + 1$.

Since C was an arbitrary congruence class, it is clear that we can do this with each of the u classes. We then obtain a partition of $\{0, 1, \ldots, r-1\}$ into at most $u \cdot \lfloor 2rN/(ust) \rfloor$ arithmetic progressions. But by our choice of t, we have

$$u \left\lfloor \frac{2rN}{ust} \right\rfloor \leq \frac{2rN}{st} \leq \left(\frac{8rN}{s} \right)^{1/2}.$$

It therefore remains to be shown that the diameter of each $\phi(P_j)$ is at most s. To do this, we need only to show that $|P_j| \leq \lceil st/N \rceil$. This will hold, though, if we show that $m_1 + 1 \leq \lceil st/N \rceil$.

Suppose for a contradiction that $m_1 + 1 > \lceil st/N \rceil$, so that $m_1 \geq \lceil st/N \rceil$. We then have

$$m_1 \left\lfloor \frac{2rN}{ust} \right\rfloor \geq \left\lceil \frac{st}{N} \right\rceil \cdot \left\lfloor \frac{2rN}{ust} \right\rfloor > \frac{st}{N} \left(\frac{2rN}{ust} - 1 \right) = \frac{r}{u} + \left(\frac{r}{u} - \frac{st}{N} \right).$$

But by our choice of t and the fact that $u \leq t$, we have $ut \leq rN/2s$. Hence, $st/N \leq r/2u < r/u$, so we see that $m_1 \lfloor 2rN/ust \rfloor > r/u$, which is a contradiction since $m_2 \geq 0$. Thus, the desired result holds. □

Recall that by Proposition 3.2, certain conditions on a balanced function imply the existence of quadratic polynomials $\psi_0, \ldots, \psi_{N-1}$. We eventually will want to partition an arithmetic progression into smaller progressions $P_1, \ldots, P_m$ so that the diameter of $\psi_s(P_j)$ is not very large for any s and j. It is clear, then, that we want to extend Lemma 3.11 to quadratic polynomials. To do this, we will need to use the discrete version of Weyl's inequality, Theorem 3.6.

Proposition 3.3. *There is an absolute constant C with the following property. Let $\psi : \mathbb{Z}_N \to \mathbb{Z}_N$ be a quadratic polynomial and let $r \in [1, N]$. Then for some $m \leq Cr^{1-1/256}$, the set $\{0, 1, \ldots, r-1\}$ can be partitioned into arithmetic progressions $P_1, \ldots, P_m$ (in $\mathbb{Z}$, not just in $\mathbb{Z}_N$) for which the diameter of $\psi(P_j)$ is at most $Cr^{-1/128}$ for every j. Furthermore, the sizes of the P_j differ by at most one.*

Proof. Let ψ be given by $\psi(x) = ax^2 + bx + c$. Then by Theorem 3.6, there exists a $p \leq r^{1/2}$ for which $|ap^2| \leq C_1 r^{-1/16} N$, where C_1 is some absolute constant. We now wish to split the quadratic function into a "constant" part and a "linear" part. We certainly cannot do this on all of $\mathbb{Z}_N$, but we can do it on progressions of common difference p. Observe that for any x and s we have

$$\psi(x + sp) = a(x + sp)^2 + b(x + sp) + c = as^2p^2 + ax^2 + 2axsp + bx + bsp + c.$$

Let $\theta(x, s) = ax^2 + 2axsp + bx + bsp + c$, so θ is linear in s, and $\psi(x + sp) = s^2(ap^2) + \theta(x, s)$. Of course, the term $s^2(ap^2)$ is not constant on progressions of common difference p, but it is possible to bound this term. Indeed, note that for any u, the set $\{s^2(ap^2) : 0 \leq s < u\}$ has diameter given by

$$\max_{0 \leq s,t < u} \left| s^2(ap^2) - t^2(ap^2) \right| = \max_{0 \leq s,t < u} \left| ap^2 \right| \cdot \left| s^2 - t^2 \right| \leq C_1 u^2 r^{-1/16} N.$$

Now, suppose that $2 \leq u \leq r^{1/4}$. We claim that we can then partition the set $\{0, \ldots, r-1\}$ into arithmetic progressions of the form

$$Q_j = \{x_j, x_j + p, \ldots, x_j + (|Q_j| - 1)p\}$$

so that the size of Q_j satisfies $u - 1 \leq |Q_j| \leq u$. We do this by a process similar to that in the previous lemma.

First, reduce $\{0, \ldots, r-1\}$ modulo p and look at the resulting p congruence classes. Then, divide $u - 1$ into $\lfloor r/p \rfloor$ to obtain integers m_1 and m_2 such that

$$\left\lfloor \frac{r}{p} \right\rfloor = m_1(u-1) + m_2$$

and $0 \leq m_2 < u - 1$. Now, let $C = \{x + qp : q = 0, \ldots, |C| - 1\}$ be a given congruence class, so we know that $|C| = \lfloor r/p \rfloor$ or $|C| = \lceil r/p \rceil$. Lay down arithmetic progressions

$$A_j = \{x + qp : q = j(u-1), \ldots, (j+1)(u-1) - 1\}$$

for $j = 0, 1, \ldots, m_1 - 1$. These m_1 progressions, each of size $u - 1$, then cover the first $m_1(u-1)$ elements of C. As a result, there are m_2 or $m_2 + 1$ elements of C that remain to be covered. We claim that the number of elements remaining, though, is no greater than the number of progressions that we have already laid down. Indeed, if this was not true, then we would have $m_1 < m_2 + 1$. This would then give,

$$\left\lfloor \frac{r}{p} \right\rfloor < (m_2 + 1)(u-1) + m_2 \leq (u-1)^2 + u - 2 = u^2 - u - 1 < r^{1/2} - 1.$$

But note that since $p \leq r^{1/2}$, we have

$$\left\lfloor \frac{r}{p} \right\rfloor \geq \left\lfloor r^{1/2} \right\rfloor > r^{1/2} - 1,$$

so we obtain $r^{1/2} - 1 < \lfloor r/p \rfloor < r^{1/2} - 1$, which is a contradiction. Therefore, it is possible to shift the progressions A_j to the right, and add at most one element to each progression, so that the resulting progressions cover all of C. Since C was an arbitrary congruence class, it is clear that we can perform this process for all p classes. This gives us the desired collection of progressions Q_j with size $u - 1 \leq |Q_j| \leq u$. Of course, each of these progressions is a progression in the integers, not just in $\mathbb{Z}_N$.

Now, for each $Q_j = \{x_j, x_j + p, \ldots, x_j + (|Q_j| - 1)p\}$, consider the function $\theta(x_j, s)$ which is linear in s on the progression Q_j. We can then extend $\theta(x_j, s)$ to a linear function ϕ_j that is defined on all of $\mathbb{Z}_N$ such that $\theta(x_j, s) = \phi_j(x_j + sp)$ for all $s = 0, \ldots, |Q_j| - 1$. Therefore, using the fact that the set $\{s^2(ap^2) : 0 \leq s < u\}$ has diameter at most $C_1 u^2 r^{-1/16} N$, we see that for any subset $P \subset Q_j$,

$$\operatorname{diam}(\psi(P)) \leq C_1 u^2 r^{-1/16} N + \operatorname{diam}(\phi_j(P)).$$

Let $u = \lceil r^{1/64} \rceil$. For r large enough, we know that this implies $2 \leq u \leq r^{1/4}$ (for smaller r, we can obtain the desired result simply by changing the constant C so that it swallows up these cases). Now, apply Lemma 3.11 to each Q_j. More specifically, divide each Q_j into sub-progressions $P_{j1}, \ldots, P_{jq_j}$ such that $\mathrm{diam}(\phi_j(P_{jt})) \leq 2u^{1/2}N$ for each t and the sizes of the P_{jt} differ by at most one as t varies. Additionally, we can guarantee that q_j, the number of sub-progressions P_{jt} that we use, is at most $(8u_jN/(2u^{-1/2}N))^{1/2}$ for each j. Since $u_j \leq u$, this gives $q_j \leq 2u^{3/4}$ for each j. Furthermore, since the larger progressions Q_j differ in size by at most one, we can also guarantee that the sizes of the P_{jt} differ by at most one as j and t vary.

We then have, for each j and t,

$$\begin{aligned}\mathrm{diam}(\psi(P_{jt})) &\leq C_1u^2r^{-1/16}N + 2u^{1/2}N \leq N\left(C_1r^{-1/32}r^{-1/16} + 2r^{-1/128}\right)\\ &\leq (C_1+2)r^{-1/128}N.\end{aligned}$$

It now remains to be shown that the total number of progressions P_{jt} that we have used is bounded above by $Cr^{1-1/128}$. Observe first that the number of sets Q_j from before is pm_1 since there were p congruence classes and we partitioned each congruence class into m_1 progressions to obtain the Q_j. Then, since $q_j \leq 2u^{3/4}$ for each j, we see that the total number of progressions, m, satisfies

$$\begin{aligned}m &\leq 2pm_1u^{3/4} \leq 2p\frac{\lfloor r/p \rfloor}{u}u^{3/4} \leq 2ru^{-1/4} = 2r\left\lceil r^{1/64}\right\rceil^{-1/4}\\ &\leq 2r\left(r^{1/64}\right)^{-1/4} = 2r^{1-1/256}.\end{aligned}$$

Note that we have used the inequality $m_1 \leq \lfloor r/p \rfloor / u$, which follows from the fact that $m_2 \leq m_1$. Therefore, if we let $C = 2 + C_1$, we find that $m \leq Cr^{1-1/256}$ and $\mathrm{diam}(\psi(P_{jt})) \leq Cr^{-1/128}N$, as desired.

□

In essence, Proposition 3.3 tells us that we can partition the interval $[0, r-1]$ into arithmetic progressions on which the quadratic function ψ is almost constant. Therefore, we can view this result as a quadratic analogue to what we did in the proof sketch of Lemma 2.1. Continuing the analogy with the proof of that lemma, the next step for us is not surprising.

Corollary 3.4. *Let $\psi : \mathbb{Z}_N \to \mathbb{Z}_N$ be a quadratic polynomial and let $r \in [1, N]$. Then for some $m \leq Cr^{1-1/256}$, the set $\{0, 1, \ldots, r-1\}$ can be partitioned into arithmetic progressions $P_1, \ldots, P_m$ for which the sizes of the P_j differ by at most one, and if $f : \mathbb{Z}_N \to \mathbb{D}$ is any function with*

$$\left|\sum_{x=0}^{r-1} f(x)\omega^{-\psi(x)}\right| \geq \alpha r,$$

then

$$\sum_{j=1}^{m}\left|\sum_{x\in P_j} f(x)\right| \geq \frac{\alpha r}{2}.$$

Proof. By Proposition 3.3, we can find progressions $P_1, \ldots, P_m$ that partition the interval $[0, r-1]$ so that the sizes of the P_j differ by at most one and $m \leq Cr^{1-1/256}$. In addition, the diameter of $\psi(P_j)$ is at most $Cr^{-1/128}$ for each j. Let r be large enough so that $Cr^{-1/128} \leq \alpha N/4\pi$. Since there are only finitely many r that do not satisfy this inequality, we can change the constant C to absorb these cases. Observe now that by the triangle inequality and the fact that the collection $P_1, \dots, P_m$ partitions $[0, r-1]$, we have

$$\alpha r \leq \left| \sum_{x=0}^{r-1} f(x)\omega^{-\psi(x)} \right| \leq \sum_{j=1}^{m} \left| \sum_{x \in P_j} f(x)\omega^{-\psi(x)} \right|.$$

For each j, choose a fixed $x_j \in P_j$. Then for any $x \in P_j$, we have the estimate

$$\begin{aligned} \left|\omega^{-\psi(x)} - \omega^{-\psi(x_j)}\right| &= \left|e^{-2\pi i\psi(x)/N} - e^{-2\pi i\psi(x_j)/N}\right| = \left|e^{-2\pi i(\psi(x)-\psi(x_j))/N} - 1\right| \\ &= \sqrt{2 - 2\cos\left(\frac{2\pi(\psi(x)-\psi(x_j))}{N}\right)} = 2\left|\sin\left(\frac{\pi(\psi(x)-\psi(x_j))}{N}\right)\right| \\ &\leq 2\left|\frac{\pi(\psi(x)-\psi(x_j))}{N}\right| \leq \frac{2\pi}{N}\left(\frac{\alpha N}{4\pi}\right) = \frac{\alpha}{2} \end{aligned}$$

where the last inequality comes from $\left|\psi(x) - \psi(x_j)\right| \leq \operatorname{diam}(\psi(P_j)) \leq Cr^{-1/128} \leq \alpha N/4\pi$. Since the function f takes values in the unit disk, we then have $|f(x)| \cdot \left|\omega^{-\psi(x)} - \omega^{-\psi(x_j)}\right| \leq \alpha/2$ if $x \in P_j$. As a result, we see that for each j,

$$\begin{aligned} \frac{\alpha}{2}\left|P_j\right| &\geq \sum_{x \in P_j} \left|f(x)\omega^{-\psi(x)} - f(x)\omega^{-\psi(x_j)}\right| \geq \left|\sum_{x \in P_j} \left(f(x)\omega^{-\psi(x)} - f(x)\omega^{-\psi(x_j)}\right)\right| \\ &= \left|\sum_{x \in P_j} f(x)\omega^{-\psi(x)} - \sum_{x \in P_j} f(x)\omega^{-\psi(x_j)}\right| \geq \left|\sum_{x \in P_j} f(x)\omega^{-\psi(x)}\right| - \left|\sum_{x \in P_j} f(x)\omega^{-\psi(x_j)}\right|. \end{aligned}$$

Therefore, we have

$$\left|\sum_{x \in P_j} f(x)\omega^{-\psi(x_j)}\right| \geq \left|\sum_{x \in P_j} f(x)\omega^{-\psi(x)}\right| - \frac{\alpha}{2}\left|P_j\right|$$

for each j. This then gives the following:

$$\begin{aligned} \sum_{j=1}^{m}\left|\sum_{x \in P_j} f(x)\right| &= \sum_{j=1}^{m}\left|\omega^{-\psi(x_j)}\right| \cdot \left|\sum_{x \in P_j} f(x)\right| = \sum_{j=1}^{m}\left|\sum_{x \in P_j} f(x)\omega^{-\psi(x_j)}\right| \\ &\geq \sum_{j=1}^{m}\left|\sum_{x \in P_j} f(x)\omega^{-\psi(x)}\right| - \sum_{j=1}^{m}\frac{\alpha}{2}\left|P_j\right| \geq \alpha r - \frac{\alpha r}{2} = \frac{\alpha r}{2} \end{aligned}$$

which is what we wanted to show.

□

3.5 Putting Everything Together

We are now ready to prove the main result of the paper.

Theorem 3.7. *There is an absolute constant C with the following property. If $A \subset \mathbb{Z}_N$ has cardinality δN where $N \geq \exp\exp\exp\left(1/\delta^C\right)$, then A contains an arithmetic progression of length four.*

Proof. First observe that If δ is close to one, then the theorem holds trivially. Also note that that as long as $C_0 \geq 1$, we will have $\exp\exp\exp\left(1/\delta^{C_0}\right) \geq 200\delta^{-3}$. Therefore, assume that $A \subset \mathbb{Z}_N$ has size $|A| = \delta N$ where $N \geq \exp\exp\exp\left(1/\delta^{C_0}\right)$ and $C_0 \geq 1$. Then by Corollary 3.1, we know that if A is quadratically $2^{-832}\delta^{448}$-uniform, A must contain a progression of length four.

Let f be the balanced function of A. We can therefore assume that f is not quadratically α-uniform, where $\alpha = 2^{-832}\delta^{448}$. By the discussion at the beginning of section 3.4, we can find a subset $B \subset \mathbb{Z}_N$ with $|B| \geq (\alpha/2)N$ and a function $\phi : B \to \mathbb{Z}_N$ for which $\left|\widetilde{\Delta(f;k)}(\phi(k))\right| \geq (\alpha/2)^{1/2}N$ whenever $k \in B$. In particular,

$$\sum_{k\in B}\left|\widetilde{\Delta(f;k)}(\phi(k))\right|^2 \geq (\alpha/2)^2N^3.$$

Additionally, recall that B must contain at least $(\alpha/2)^8N^3$ ϕ-additive quadruples, and as a result, there are constants γ and η, depending only on α, for which we can find a mod-N arithmetic progression P with $|P| \geq N^\gamma$ and a linear function $\psi : P \to \mathbb{Z}_N$ such that $\phi(s)$ agrees with $\psi(s)$ for at least $\eta|P|$ values of $s \in P$. Let $|P'|$ be the set of such s and let ψ have the form $\psi(x) = \lambda k + \mu$. We then see that

$$\begin{aligned}\sum_{k\in P}\left|\widetilde{\Delta(f;k)}(\lambda k+\mu)\right|^2 &\geq \sum_{k\in P'}\left|\widetilde{\Delta(f;k)}(\lambda k+\mu)\right|^2 = \sum_{k\in P'}\left|\widetilde{\Delta(f;k)}(\psi(k))\right|^2\\ &\geq \frac{\alpha}{2}N^2|P'| \geq \frac{\alpha}{2}\eta N^2|P|\end{aligned}$$

where the second-to-last inequality follows from the fact that $P' \subset B$. Also recall from the discussion directly following the proof of Corollary 3.3, there exists an absolute constant K_0 for which we can take $\gamma = (\alpha/2)^{8K_0}$ and $\eta = \exp\left(-(2/\alpha)^{8K_0}\right)$. Using the fact that $\alpha = 2^{-832}\delta^{448}$, we can express these as $\gamma = c_1^{-1}\delta^{K_1}$ and $\eta = \exp\left(-c_1/\delta^{K_1}\right)$ where $c_1 = 2^{843\cdot 8K_0}$ and $K_1 = 448 \cdot 8K_0$. Now, let K_2 be another constant (not depending on δ) that is large enough to ensure that $\gamma \geq \delta^{K_2}$ and $(\alpha/2)\eta \geq \exp\left(-(1/\delta)^{K_2}\right)$. Setting $\beta = (\alpha/2)\eta$ and $T = |P|$, we see that

$$\sum_{k\in P}\left|\widetilde{\Delta(f;k)}(\lambda k+\mu)\right|^2 \geq \beta N^2T.$$

Thus, by Proposition 3.2, we can find quadratic polynomials $\psi_0, \psi_1, \ldots, \psi_{N-1} : \mathbb{Z}_N \to \mathbb{Z}_N$ for which

$$\sum_s\left|\sum_{x\in P+s} f(x)\omega^{-\psi_s(x)}\right| \geq \frac{\beta NT}{\sqrt{2}}.$$

For each s, we want to apply Corollary 3.4 to the progression $P+s$. Note that even though we proved this corollary for intervals of the form $\{0,\ldots,r-1\}$, it applies just as well to arithmetic progressions. If c_2 denotes the absolute constant in Corollary 3.4, then we know that for each s, there is a partition of the set $P+s$ into progressions $P_{s1},\ldots,P_{sm_s}$, where

$m_s \leq c_2 T^{1-1/256}$, for which the sizes of the P_{sj} differ by at most one and if $g : \mathbb{Z}_N \to \mathbb{D}$ satisfies

$$\left| \sum_{x \in P+s} g(x) \omega^{-\psi_s(x)} \right| \geq \alpha_s T,$$

then

$$\sum_{j=1}^{m_s} \left| \sum_{x \in P_{sj}} g(x) \right| \geq \frac{\alpha_s T}{2}.$$

Observe that as we vary s, the sets $P + s$ simply shift. As a result, we may assume that m_s is the same for all s. Let this number be m. Also, note that for each s, the average length of the progressions P_{sj} is given by T/m, which is at least $T/(c_2 T^{1-1/256}) = c_3 T^{1/256}$, where $c_3 = 1/c_2$ is an absolute constant. Now, by the fact that the progressions P_{sj} differ in size by at most one, we know that all of them must have size at least $c_3 T^{1/256}$ and none may be longer than twice this length. Furthermore, Corollary 3.4 ensures that each P_{sj} is a genuine progression in $\mathbb{Z}$, not just in $\mathbb{Z}_N$. We now wish to show that it is on one of these progressions P_{sj} that A has increased density.

For each s, let

$$\alpha_s = \frac{1}{T} \left| \sum_{x \in P+s} f(x) \omega^{-\psi_s(x)} \right|.$$

We then have

$$\sum_{j=1}^{m} \left| \sum_{x \in P_{sj}} f(x) \right| \geq \frac{\alpha_s T}{2}.$$

Summing over s, we then find that

$$\sum_s \sum_{j=1}^{m} \left| \sum_{x \in P_{sj}} f(x) \right| \geq \sum_s \frac{\alpha_s T}{2} = \frac{1}{2} \sum_s \left| \sum_{x \in P+s} f(x) \omega^{-\psi_s(x)} \right| \geq \frac{\beta N T}{2\sqrt{2}}.$$

Now, for each s and j, let $p_{sj} = \sum_{x \in P_{sj}} f(x)$. Observe that for any $x \in \mathbb{Z}_N$, there are exactly $T = |P|$ elements s for which $x \in P + s$. Indeed, the set of such s is precisely $x - P$. Since the progressions $P_{s1}, \ldots, P_{sm}$ partition $P + s$ for each s, we then know that for each $s \in x - P$, there is a unique j for which $x \in P_{sj}$. As a result, if we consider the elements covered by P_{sj} as we let s vary over $\mathbb{Z}_N$ and j vary between 1 and m, we see that x is counted exactly T times. This then implies that

$$\sum_s \sum_{j=1}^{m} p_{sj} = \sum_s \sum_{j=1}^{m} \sum_{x \in P_{sj}} f(x) = T \sum_x f(x).$$

But we know that $\sum_x f(x) = \tilde{f}(0) = 0$ since f is the balanced function of A. Therefore, $\sum_s \sum_{j=1}^{m} p_{sj} = 0$. Now, define $q_{sj} = \max\left(p_{sj}, 0\right)$. We then have

$$\sum_s \sum_{j=1}^{m} q_{sj} = \frac{1}{2} \sum_s \sum_{j=1}^{m} |p_{sj}| = \frac{1}{2} \sum_s \sum_{j=1}^{m} \left| \sum_{x \in P_{sj}} f(x) \right| \geq \frac{\beta N T}{4\sqrt{2}}.$$

By averaging, we then know that there must exist an s and j for which

$$\sum_{x \in P_{sj}} f(x) = q_{sj} \geq \frac{\beta T}{4m\sqrt{2}} \geq \frac{\beta c_3 T^{1/256}}{4\sqrt{2}}$$

where the last inequality follows from the fact that $T/m \geq c_3 T^{1/256}$. We therefore have

$$\frac{\beta c_3 T^{1/256}}{4\sqrt{2}} \leq \sum_{x \in P_{sj}} f(x) = \sum_{x \in A \cap P_{sj}} (1-\delta) + \sum_{x \in P_{sj} \setminus A} (-\delta) = \left|A \cap P_{sj}\right| - \delta \left|P_{sj}\right|.$$

Recall, though, that the size of P_{sj} must satisfy $c_3 T^{1/256} \leq \left|P_{sj}\right| \leq 2c_3 T^{1/256}$. As a result,

$$\left|A \cap P_{sj}\right| \geq \frac{\beta c_3 T^{1/256}}{4\sqrt{2}} + \delta \left|P_{sj}\right| \geq \frac{\beta \left|P_{sj}\right|}{8\sqrt{2}} + \delta \left|P_{sj}\right| = (\beta c_4 + \delta) \left|P_{sj}\right| \geq (\epsilon c_4 + \delta) \left|P_{sj}\right|$$

where $c_4 = 2^{-7/2}$ and $\epsilon = \exp\left(-(1/\delta)^{K_2}\right)$. We have therefore found an arithmetic progression (recall that P_{sj} is a genuine progression in $\mathbb{Z}$, not just in $\mathbb{Z}_N$) on which the density of A is increased.

We now wish to iterate the argument, as we did in the proof of Roth's Theorem. Here we give details. Replace the sets A and $[1, N]$ with $A \cap P_{sj}$ and P_{sj}, respectively. Even though P_{sj} is not an interval itself, we can identify it with the interval $[1, |P_{sj}|]$. We then know that the density of $A \cap P_{sj}$ in P_{sj} is at least $\delta + c_4\epsilon$ and the size of P_{sj} is at least $N^\gamma \geq N^{\delta^{K_2}}$.

Let $\delta_0 = \delta$ be the density of A in $[1, N]$ that we begin with (that is, the density at the 0-th step of the iteration) and let $A_0 = A$ and $N_0 = N$. For each $r \in \mathbb{N}$, let A_r be the set that we have replaced A with at the r-th step, and let N_r be the size of the progression Q_r that we have replaced $[1, N]$ with. For example, we have $A_1 = A \cap P_{sj}$, $Q_1 = P_{sj}$, and $N_1 = |P_{sj}|$. Then let δ_r be the density of A_r in Q_r, and let ϵ_r be the ϵ obtained at the r-th step of the iteration. By definition, then, $\epsilon_r = \exp\left(-(1/\delta_r)^{K_2}\right)$.

In order for the r-th step of the iteration to hold, we need $N_r \geq 200\delta^{-3}$. Indeed, this was the only hypothesis we assumed regarding the relationship between N and δ in the 0-th step. We now wish to show that as long as $r \leq \exp\left(1/\delta^{K_3}\right)$, where $K_3 = 2K_2$, the desired inequality will hold.

Observe first that $N_1 = |P_{sj}| \geq N_0^{\delta_0^{K_2}}$. Continuing inductively, we see that $N_r \geq N_{r-1}^{\delta_{r-1}^{K_2}}$ for each $r \geq 1$. We claim now that

$$N_r \geq N^{\left(\delta^{K_2}\right)^r}$$

for all r. The proof is by induction. When $r = 0$, the claim holds trivially. Now suppose that it holds for some r. We then have

$$N_{r+1} \geq N_r^{\delta_r^{K_2}} \geq \left(N^{\left(\delta^{K_2}\right)^r}\right)^{\delta_r^{K_2}} \geq \left(N^{\left(\delta^{K_2}\right)^r}\right)^{\delta^{K_2}} = N^{\left(\delta^{K_2}\right)^{r+1}}$$

where the last inequality follows from $\delta_r \geq \delta$. Thus, the claim is true.

We now show that $N_r \geq 200\delta^{-3}$ whenever $r \leq \exp\left(1/\delta^{K_3}\right)$, as long as we begin with N large enough. Observe first that since δ is not close to one and K_3 is large, we have

$$\exp\left(1/\delta^{4K_3}\right) \geq 2\exp\left(1/\delta^{2K_3}\right) \geq \exp\left(2/\delta^{K_3}\right) + \exp\left(1/\delta^{K_3}\right).$$

Therefore,

$$\exp\left(1/\delta^{K_3}\right)\exp\left(1/\delta^{K_3}\right) \leq \exp\left(1/\delta^{4K_3}\right) - \exp\left(1/\delta^{K_3}\right),$$

so if $r \leq \exp\left(1/\delta^{K_3}\right)$, then

$$r \leq \frac{\exp\left(1/\delta^{4K_3}\right) - \exp\left(1/\delta^{K_3}\right)}{\exp\left(1/\delta^{K_3}\right)} \leq \frac{\exp\left(1/\delta^{4K_3}\right) - \log\log\left(1/\delta^{K_3}\right)}{\log\left(1/\delta^{K_3}\right)}.$$

As a result, we have

$$K_3 r \log(1/\delta) \leq \exp\left(1/\delta^{4K_3}\right) - \log\log\left(1/\delta^{K_3}\right).$$

If we assume now that $N \geq \exp\exp\exp\left(1/\delta^{4K_3}\right)$, then we see that $\log\log N \geq \exp\left(1/\delta^{4K_3}\right)$. Therefore, we obtain

$$K_3 r \log(1/\delta) \leq \log\log N - \log\log\left(1/\delta^{K_3}\right).$$

This then gives us

$$\frac{1}{\delta^{K_3 r}} \leq \frac{\log N}{\log\left(1/\delta^{K_3}\right)},$$

so by taking reciprocals, we have

$$\delta^{K_3 r} \geq \frac{K_3 \log(1/\delta)}{\log N} \geq \frac{\log\left(1/\delta^3\right)}{\log N} + \frac{\log 200}{\log N}$$

since K_3 is assumed to be large. We then see that

$$\left(\delta^{K_3}\right)^r \log N \geq \log 200 + \log\left(1/\delta^3\right),$$

and as a consequence,

$$N^{\left(\delta^{K_3}\right)^r} \geq 200\delta^{-3}.$$

We have therefore shown that if $N \geq \exp\exp\exp\left(1/\delta^{4K_3}\right)$, then for each $r \leq \exp\left(1/\delta^{K_3}\right)$, we have

$$N_r \geq N^{\left(\delta^{K_2}\right)^r} \geq N^{\left(\delta^{K_3}\right)^r} \geq 200\delta^{-3}.$$

Hence, the r-th step of the iteration holds for these values of r.

We now wish to show that the iteration can be repeated at most $\exp\left(1/\delta^{K_3}\right)$ times. If we are able to verify this, then the theorem will follow because of the fact that we are allowed to perform the iteration this many times.

Recall that δ_r is the density of A_r in $[1, N_r]$ and $\epsilon_r = \exp\left(-(1/\delta_r)^{K_2}\right)$. By the increase in density that resulted from the 0-th step of the argument, we have $\delta_1 = \delta_0 + c_4\epsilon_0$. Continuing inductively, we see that for each $r \geq 1$, the density δ_r is precisely $\delta_r = \delta_{r-1} + c_4\epsilon_{r-1}$. We now claim that $\delta_r \geq \delta\left(1 + c_4\epsilon\right)^r$ for each n. Again, we prove this inductively.

Of course, the claim holds when $r = 0$. Now, suppose that it holds for some r. Then we have

$$\delta_{r+1} = \delta_r + c_4\epsilon_r \geq \delta\left(1 + c_4\epsilon\right)^r + c_4\epsilon_r.$$

Observe, though, that since $\delta_r \geq \delta$, we know that

$$\epsilon_r = \exp\left(-(1/\delta_r)^{K_2}\right) \geq \exp\left(-(1/\delta)^{K_2}\right) = \epsilon,$$

so

$$\delta_{r+1} \geq \delta\left(1 + c_4\epsilon\right)^r + c_4\epsilon.$$

Since ϵ is small and δ is not close to one, we also have

$$\delta\left[\left(1 + c_4\epsilon\right)^{r+1} - \left(1 + c_4\epsilon\right)^r\right] \leq c_4\epsilon.$$

Thus, we have the desired result $\delta_{r+1} \geq \delta\left(1 + c_4\epsilon\right)^{r+1}$. This tells us that at each step of the iteration, the density increases by at least a factor of $1 + c_4\epsilon$. We now claim that repeating the iteration $\exp\left(1/\delta^{K_3}\right)$ times would give us a density greater than one.

Let $r \geq \exp\left(1/\delta^{K_3}\right) = \exp\left(1/\delta^{2K_2}\right)$. Using the fact that $c_4 = 2^{-7/2}$ is a small constant and K_2 is large, we have

$$r > \frac{1}{c_4}\exp\left(1/\delta^{K_2+2}\right) \geq \frac{1}{c_4}\exp\left(1/\delta^{K_2+1} + 1/\delta\right) \geq \frac{1}{c_4}\log\left(1/\delta\right)\exp\left(1/\delta^{K_2+1}\right).$$

As a result, we see that

$$rc_4\exp\left(-(1/\delta)^{K_2+1}\right) > \log\left(1/\delta\right).$$

Recall now, by properties of logarithms, that for x close to zero, $\log(1+x) \approx x$. Indeed, the Taylor expansion for this function is $\log(1+x) = x - x^2/2 + x^3/3 - x^4/4 + \ldots$, so if x is small, the higher-order terms are negligible. Since ϵ is small, we therefore have the estimate

$$\log(1 + c_4\epsilon) = \log\left(1 + c_4\exp\left(-(1/\delta)^{K_2}\right)\right) \geq c_4\exp\left(-(1/\delta)^{K_2+1}\right),$$

where we have increased the power of $1/\delta$ to take care of the higher-order terms. This then gives us

$$\log\left(1/\delta\right) < r\log\left(1 + c_4\epsilon\right),$$

which implies that $\delta\left(1 + c_4\epsilon\right)^r > 1$. Thus, the density δ_r is greater than one, which certainly cannot happen when we iterate the argument.

Therefore, if we begin the iteration process with N large enough, as we continue to iterate the argument, it must happen that some set A_r is quadratically $2^{-832}\delta_r^{448}$-uniform in the progression Q_r. Otherwise, we would continue the iteration through $\exp\left(1/\delta^{K_3}\right)$ steps, thereby obtaining a density greater than one. Taking $C = 4K_3$ in the statement of the theorem, we obtain the stated result.

□

Although the bounds that this proof gives are better than those in most other proofs (e.g. the combinatorial proof by Szemerédi and the ergodic theoretic proof by Furstenberg), they are not the best that Gowers has obtained. In his longer paper detailing the complete proof of Szemerédi's Theorem, he obtains the following bounds for progressions of length four. Essentially, he is able to get rid of one exponential.

Theorem 3.8 (Bounds in [9] for Four-Term Progressions). *Let $\delta > 0$ and let $A \subset [1, N]$ have cardinality at least δN. If $N \geq \exp\exp\left(1/\delta^C\right)$, where C is some absolute constant, then A contains an arithmetic progression of length four.*

Chapter 4

Extensions of Gowers's Work

The most natural extension of Gowers's work on progressions of length four is to progressions of arbitrary length. As we have mentioned earlier, Gowers was able to do this in [9], and surprisingly, the methods used in his argument for arbitrary k are generally analogous (albeit, significantly more complicated) to the methods in the previous chapter.

The first necessity in dealing with progressions of arbitrary length is to find more general types of uniformity. Recall that in Roth's work on progressions of length three, uniformity was a strong enough notion of randomness. For progressions of length four, however, uniformity was not strong enough, so we had to introduce quadratic uniformity. As one might expect, progressions of longer length require even stronger notions of randomness. To develop such a notion, Gowers generalizes the Δ operator we used earlier. Namely, for a function $f : \mathbb{Z}_N \to \mathbb{C}$, define

$$\Delta(f; a_1, \ldots, a_d) = \Delta(\Delta(f; a_1, \ldots, a_{d-1}); a_d)$$

inductively. Note that the Δ operator we used was simply the case of $d = 1$. Working out the induction, it is possible to show that we can define $\Delta(f; a_1, \ldots, a_d)$ equivalently as

$$\Delta(f; a_1, \ldots, a_d)(s) = \prod_{\epsilon_1, \ldots, \epsilon_d} \left(C^{\epsilon_1 + \ldots + \epsilon_d} f\right) \left(s - \sum_{i=1}^{d} a_i \epsilon_i\right)$$

for $s \in \mathbb{Z}_N$, where the product is over all sequences $\epsilon_1, \ldots, \epsilon_d$ with $\epsilon_i \in \{0, 1\}$ and C is the operator that takes a complex-valued function to its pointwise complex conjugate. When $d = 1$, we see that the only such sequences are $\epsilon_1 = 0$ and $\epsilon_1 = 1$. Thus, we have

$$\Delta(f; a)(s) = \left(C^0 f(s - a \cdot 0)\right) \left(C^1 f(s - a \cdot 1)\right) = f(s)\overline{f(s - a)}$$

which is precisely how we defined Δ in the previous chapter. It is also easy to check that when $d = 2$, the Δ operator becomes

$$\Delta(f; a, b)(s) = f(s)\overline{f(s - a) f(s - b)} f(s - a - b)$$

and when $d = 3$, it is

$$\begin{aligned}\Delta(f; a, b, c)(s) = f(s)&\overline{f(s - a) f(s - b) f(s - c)} \\ &\cdot f(s - a - b) f(s - a - c) f(s - b - c) \overline{f(s - a - b - c)}.\end{aligned}$$

Note here that our definition in the previous chapter for a function f being quadratically α-uniform is equivalent to

$$\sum_{a,b}\left|\sum_{s}\Delta(f;a,b)(s)\right|^2 \leq \alpha N^4.$$

More generally, we say that $f : \mathbb{Z}_N \to \mathbb{D}$ is α-uniform of degree d if

$$\sum_{a_1,\ldots,a_d}\left|\sum_{s}\Delta(f;a_1,\ldots,a_d)(s)\right|^2 \leq \alpha N^{d+2}.$$

We then say that a set $A \subset \mathbb{Z}_N$ is α-uniform of degree d if its balanced function is. It turns out that this notion of randomness is precisely what we need to extend Gowers's arguments from progressions of length four to progressions of arbitrary length.

As he did in his work on progressions of length four, Gowers then shows that if A is α-uniform of degree $d-2$ for a sufficiently small α, then A contains many mod-N arithmetic progressions of length d. And of course, with α small enough, at least one of these will be a progression in $\mathbb{Z}$. As one might expect from the inductive definition of higher-degree uniformity, most of the results obtained regarding uniform sets are direct analogues of those found in the previous chapter. Thus, proving that a set must contain a progression of length d if it is α-uniform of degree $d-2$ for small α is not the difficult part of generalizing Gowers's methods. Instead, it is when A fails to be α-uniform of degree $d-2$ that is the more difficult situation to deal with. Gowers does this by generalizing the notion of "quadratic bias" that we used in the previous chapter to polynomial bias. Namely, if A fails to be α-uniform of degree $d-2$, then we use polynomials of degree $d-2$ to find a progression on which A has increased density. After an impressive and intricate argument, Gowers is able to establish the following result.

Theorem 4.1 (Szemerédi's Theorem with Gowers's Bounds)**.** *Let $0 < \delta \leq 1/2$, let k be a positive integer, and let $A \subset [1,N]$ have cardinality at least δN. If $N \geq 2 \uparrow 2 \uparrow \delta^{-1} \uparrow 2 \uparrow 2 \uparrow (k+9)$, then A contains an arithmetic progression of length k.*

Here, the notation $a \uparrow b$ simply means a^b. Of course, this is a gigantic bound on N, and the most obvious question is whether or not this bound can be improved. In other words, given δ and k, what is the smallest N_0 for which any set $A \subset [1,N]$ of size at least δN contains a progression of length k if $N \geq N_0$? Note, of course, that the bound given in Theorem 3.8 for progressions of length four is significantly better than the bound given here for $k=4$. But we can just as well ask whether or not this better bound can be further improved.

According to Laba, as of January 2008, Gowers's bounds were the best known for $k \geq 5$ [15]. For $k=4$, however, Green and Tao were able to improve Gowers's bounds slightly [11]. In this same paper, Green and Tao claimed to have a further improvement; namely, if $A \subset [1,N]$ has size at least δN and $N \geq \exp(1/d^C)$, then A contains a progression of length four. Here, C is an absolute constant. As of now, these bounds seem to be the best known.

We now briefly depart from discussing progressions of arbitrary length so that we can focus on bounds for progressions of length three. We will present these in terms of bounds on δ rather than on N for simplicity, but it is clear that we can go from one to the other easily. Recall Roth's Theorem from a previous chapter.

Theorem 4.2 (Roth)**.** *Let $N \in \mathbb{N}$ and $A \subset [1, N]$ have cardinality $|A| \geq \delta N$, where $\delta > c/(\log \log N)$. Then A contains an arithmetic progression of length three. Here, c is an absolute constant.*

This bound on δ is not the best known, though. In 1999, Bourgain proved that the result holds if $\delta > (\log \log N)^{1/2}/(\log N)^{1/2}$ using much deeper Fourier analytic methods than Roth originally did [3]. According to Tao, Bourgain's bound is currently the best known [27]. It is generally believed that these bounds can be significantly improved. This belief is primarily a result of attempts to construct sets in $\mathbb{Z}$ with no three-term arithmetic progression. Over sixty years ago, Behrend proved the following classic result [2].

Theorem 4.3 (Behrend)**.** *Let N be a sufficiently large integer. Then there are constants c, C and a subset $A \subset [1, N]$ of cardinality $|A| \geq cN \exp(-C\sqrt{\log N})$ that contains no arithmetic progressions of length three.*

An immediate consequence of this construction is that the density δ in Roth's Theorem (and more generally in Szemerédi's Theorem) must be greater than $\exp(-C\sqrt{\log N})$. Surprisingly, after sixty years of work on this problem, no one has been able to improve Behrend's construction significantly. In other words, no one can find a subset of $[1, N]$ with size much larger than $\exp(-C\sqrt{\log N})$ that does not contain any three-term progressions. For this reason, many mathematicians believe that the bounds on δ in Roth's Theorem can be significantly improved, perhaps down to somewhere near $\exp(-C\sqrt{\log N})$. What may be more surprising, though, is that Behrend's construction is quite elementary. We are therefore able to give the proof here.

Proof of Behrend's Theorem. Observe first that no sphere in $\mathbb{R}^d$ can contain an arithmetic progression of length three. Indeed, suppose that there was a sphere

$$S_r = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_1^2 + \ldots + x_d^2 = r\}$$

(without loss of generality, we can assume that it is centered at the origin) that contained the progression

$$P = \{(x_1, \ldots, x_d), (x_1 + r_1, \ldots, x_d + r_d), (x_1 + 2r_1, \ldots, x_d + 2r_d)\}.$$

Since $x_1^2 + \ldots + x_d^2 = r = (x_1 + r_1)^2 + \ldots + (x_d + r_d)^2$, we know that

$$r_1^2 + \ldots + r_d^2 = -2(x_1 r_1 + \ldots + x_2 r_2).$$

But also, $(x_1 + 2r_1)^2 + \ldots + (x_d + 2r_d)^2 = r$, so

$$r_1^2 + \ldots + r_d^2 = -(x_1 r_1 + \ldots + x_d r_d).$$

This implies, however, that $r_1^2 + \ldots + r_d^2 = 0$, so P is a trivial progression. Thus, S_r contains no non-trivial three-term arithmetic progressions.

We now wish to consider the set of integers

$$S_{n,d,r} = \{(x_1, \ldots, x_d) \in [1, n]^d : x_1^2 + \ldots + x_d^2 = r\}.$$

Note here that we are still using the notation $[1, n] = \{1, 2, \ldots, n\}$. Then $S_{n,d,r}$ is certainly a subset of a sphere in $\mathbb{R}^d$, so it contains no progressions of length three. We now want to map this set back to an interval $[1, N]$ in $\mathbb{Z}$; of course, we will want to use Freiman isomorphisms to do this.

First, though, observe that if we let r range from n to dn^2, the sets $S_{n,d,r}$ cover the entire cube $[1, n]^d$. Since this cube has size n^d, the pigeon-hole principle tells us that there must be an R with $|S_{n,d,R}| \geq n^d/(dn^2)$. Consider the function $\phi : [1, n]^d \to \mathbb{Z}$ defined by

$$\phi(x_1, \ldots, x_d) = \sum_{i=1}^{d} x_i (2n)^{i-1}.$$

Recall from the previous chapter (section 3.3) that this is a Freiman homomorphism of order 2, and if we restrict the codomain to be the image of ϕ, then it becomes a Freiman isomorphism. As a result, $\phi(S_{n,d,R})$ has cardinality at least $n^d/(dn^2)$, and it contains no three-term progressions. Furthermore, if we let N be large enough, then $\phi(S_{n,d,R})$ is contained in $[1, N]$. It turns out that $N \geq C_1^d n^d$ will work, for an absolute constant C_1. Therefore, if we set $n = c_1 N^{1/d}$, where $c_1 = 1/C_1 < 1$, then we see that $A = \phi(S_{n,d,R})$ is a subset of $[1, N]$ that contains no progressions of length three and has size

$$|A| \geq c_1^{d-2} \frac{N}{dN^{2/d}}.$$

It is not difficult to show, then, that there are constants c_2, C_2 such that

$$|A| \geq c_2 N \exp\left(-C_2 d - \frac{2}{d} \log N - \log d\right).$$

If we then choose d approximately $\sqrt{\log N}$, we have

$$|A| \geq c_2 N \exp\left(-C_2\sqrt{\log N} - 2\sqrt{\log N} - \log\sqrt{\log N}\right).$$

Again, it is possible to find constants c, C such that

$$|A| \geq cN \exp(-C\sqrt{\log N})$$

so the theorem is proved. □

It is natural to ask whether this method can be extended to construct large subsets of the integers that contain no arithmetic progressions of length k. It can, and according to Tao, the largest subsets of $[1, N]$ that are known to contain no k-term progressions have size approximately $N \exp\left(-C(\log N)^{1/(k-1)}\right)$. It seems likely, then, that Gowers's bounds for Szemerédi's Theorem can be significantly improved for all k.

We transition now to a brief discussion about sets in the integers that contain arbitrarily long arithmetic progressions. We say that a subset $A \subset \mathbb{Z}$ contains arbitrarily long arithmetic progressions if A contains a k-term progression for all $k \in \mathbb{N}$. If A has this property, then we can think of A as a fairly structured set. A classic question in this field of study is whether or not the prime numbers contain arbitrarily long progressions. In 2006, Green and Tao answered this in the affirmative [12]. Interestingly, their proof used Szemerédi's Theorem (though certainly not in a straightforward way).

Theorem 4.4 (Green-Tao). *The set of prime numbers contains arbitrarily long arithmetic progressions.*

This result has been, arguably, the most important of the decade. It also gives strong evidence that the primes are far from being randomly distributed; in fact, they contain a large amount of structure. It is necessary to note that the Green-Tao Theorem is a special case of a very general conjecture by Erdös and Turán in [6] that seems far from being resolved.

Conjecture 4.1 (Erdös-Turán)**.** *Let A be a subset of the positive integers such that $\sum_{n \in A} 1/n = \infty$. Then A contains arbitrarily long arithmetic progressions.*

Of course, if P denotes the set of prime numbers, then $\sum_{p \in P} 1/p = \infty$, so one may consider the Green-Tao Theorem to be evidence that the Erdös-Turán Conjecture is true. It is easy to see that the converse of this conjecture is not true though. Indeed, for each $k \geq 3$, let

$$P_k = \{k^4 - k + 1, k^4 - k + 2, \ldots, k^4\}.$$

Then P_k is an arithmetic progression of length k, so let

$$A = \bigcup_{k \geq 3} P_k.$$

But note that for each k, we have

$$\sum_{n \in P_k} \frac{1}{n} \leq \frac{k}{k^3(k-1)} = \frac{1}{k^2(k-1)} \leq \frac{1}{k^2}.$$

Therefore,

$$\sum_{n \in A} \frac{1}{n} \leq \sum_{k} \frac{1}{k^2} < \infty$$

so A contains arbitrarily long arithmetic progressions but it's reciprocals do not diverge.

4.1 Sumsets in Cyclic Groups of Prime Order

We now conclude with a discussion of sumsets in groups $\mathbb{Z}_p$, where p is prime. Recall from section 3.3 in the previous chapter the following basic theorems.

Theorem 4.5. *Let A and B be finite subsets of $\mathbb{Z}$. Then $|A| + |B| - 1 \leq |A + B| \leq |A| \cdot |B|$.*

Theorem 4.6. *Let A and B be finite subsets of $\mathbb{Z}$ with $|A| \geq 2$ and $|B| \geq 2$. If $|A + B| = |A| + |B| - 1$, then A and B are arithmetic progressions of the same common difference.*

It turns out that neither of these results is difficult to prove, primarily because nothing strange happens when we sum two integers. For example, if we sum the sets A and B, we know that the smallest element of $A + B$ will be precisely the sum of the smallest element of A and the smallest element of B. Similarly, the largest element of $A + B$ will be the sum of the largest element of A and the largest element of B. And of course, elements in A and B that are between these two extremes will sum to an element that is between the extremes in the sumset. Essentially, we are only saying that the sum is linear in regard to order.

Consider instead sumsets in the group $\mathbb{Z}_N$. As an example, take $N = 10$ and look at $A = \{5, 6, 9\}$ and $B = \{2, 3\}$. We then have $A + B = \{1, 2, 7, 8, 9\}$. In particular, the largest element in A and the largest element in B sums to the smallest element in $A + B$, while the smallest element in A and the smallest element in B sums to the median of $A + B$. It is clear that we have lost order-linearity of the sum due to "wrap-around." Namely, if a sum becomes too large (exceeds N), then it reduces to a small number modulo N. As one should expect, this simple phenomenon makes questions about sumsets in $\mathbb{Z}_N$ much more difficult than those about sumsets in $\mathbb{Z}$. Despite this difficulty, though, there do happen to be direct analogues to Theorems 4.5 and 4.6 in $\mathbb{Z}_p$, as long as p is prime. The first is due (independently) to Cauchy and Davenport [18].

Theorem 4.7. *Let A and B be non-empty subsets of $\mathbb{Z}_p$, where p is prime. Then $|A| + |B| - 1 \leq |A + B| \leq |A| \cdot |B|$.*

The upper bound on $|A + B|$ is, of course, just as trivial as it was in Theorem 4.5. The lower bound, however, is significantly more difficult to prove. Still, it is not as hard as the following result, which is due to Vosper.

Theorem 4.8 (Vosper)**.** *Let A and B be non-empty subsets of $\mathbb{Z}_p$, where p is prime, and suppose that $A + B \neq \mathbb{Z}_p$. Then $|A + B| = |A| + |B| - 1$ if and only if one of the following holds.*

1. $\min(|A|, |B|) = 1$,
2. $|A + B| = p - 1$ *and* $A = \mathbb{Z}_p \setminus (c - B)$ *where* $\{c\} = \mathbb{Z}_p \setminus (A + B)$,
3. *A and B are both arithmetic progressions with the same common difference.*

We can consider conditions 1 and 2 to be somewhat degenerate cases, so in general, it is safe to think of arithmetic progressions with the same common difference as the "only" sets whose sumset is of minimal size. The traditional proofs of both the Cauchy-Davenport Theorem and Vosper's Theorem are entirely combinatorial. Recently, though, some work in discrete Fourier analysis – namely, the finite uncertainty principle – has had a surprising impact on these topics. In a 2005 paper, Tao established the following theorem [29].

Theorem 4.9 (Tao)**.** *Let $f : \mathbb{Z}_p \to \mathbb{C}$ be a non-zero function, where p is prime. Then*

$$|supp(f)| + \left|supp(\hat{f})\right| \geq p + 1.$$

Conversely, if A and B are non-empty subsets of $\mathbb{Z}_p$ with $|A| + |B| \geq p + 1$, then there exists a function $f : \mathbb{Z}_p \to \mathbb{C}$ for which $supp(f) = A$ and $supp(\hat{f}) = B$.

Here, $\text{supp}(f)$ denotes the support of f – the set of elements in $\mathbb{Z}_p$ on which f is non-zero. Thus, the first part of this result is essentially saying that there is no non-zero function from $\mathbb{Z}_p$ to $\mathbb{C}$ such that both it and its Fourier transform vanish at many points. It is important to note now that we have returned to the standard definition for the Fourier transform $\hat{f}$, as given in definition 5 from chapter 1. As a surprising corollary to Tao's theorem, we can prove the Cauchy-Davenport Theorem easily.

Proof of Cauchy-Davenport via Tao's Theorem. Let A and B be non-empty subsets of $\mathbb{Z}_p$. If $|A| + |B| \leq p + 1$, let $X = [0, p - |A|]$ and $Y = [|B| - 1, p - 1]$. Then we see that $|X| = p + 1 - |A|$ and $|Y| = p + 1 - |B|$. Furthermore, $X \cap Y = [|B| - 1, p - |A|]$, so $|X \cap Y| = p + 2 - |A| - |B| = |X| + |Y| - p \geq 1$. Now suppose that $|A| + |B| > p + 1$. In this case, let $X = [1, p+1-|A|]$ and $Y = [p+1-|A|, 2p+1-|A|-|B|]$. We then have $|X| = p+1-|A|$ and $|Y| = p+1-|B|$ as in the previous case, but here, $X \cap Y = \{p+1-|A|\}$, so $|X \cap Y| = 1$. Note also, however, that $|X|+|Y|-p = p+2-|A|-|B| \leq 0$. Therefore, we can combine the two cases into the following statement. There exist sets X and Y satisfying $|X| = p+1-|A|$, $|Y| = p+1-|B|$, and $|X \cap Y| = \max(|X|+|Y|-p, 1)$.

Since $|A|+|X| = p+1$ and $|B|+|Y| = p+1$, Theorem 4.9 tells us that there exist functions $f, g : \mathbb{Z}_p \to \mathbb{C}$ for which $\operatorname{supp}(f) = A$, $\operatorname{supp}(\hat{f}) = X$, $\operatorname{supp}(g) = B$, and $\operatorname{supp}(\hat{g}) = Y$. Consider the convolution $f \star g$, defined by

$$f \star g(s) = \sum_t f(t)g(s-t).$$

Recall that by the properties of the convolution, we know that $\widehat{f \star g}(s) = \hat{f}(s) \cdot \hat{g}(s)$. As a result, $\operatorname{supp}(\widehat{f \star g}) = X \cap Y$, and since $|X \cap Y| \geq 1$, we know that $f \star g$ is non-zero. In addition, observe that if $s \in \operatorname{supp}(f \star g)$, then $f(t)g(s-t) \neq 0$ for some t. This implies that $f(t) \neq 0 \neq g(s-t)$, so in particular, $t \in A$ and $s-t \in B$. Thus, $s \in A+B$. We therefore have $\operatorname{supp}(f \star g) \subset A+B$.

Using the first part of Theorem 4.9, we then see that

$$|A+B| + |X \cap Y| \geq |\operatorname{supp}(f \star g)| + \left|\operatorname{supp}(\widehat{f \star g})\right| \geq p+1$$

since $f \star g$ is non-zero. Consequently, we have

$$\begin{aligned} |A+B| &\geq p+1-|X \cap Y| = p+1-\max(p+2-|A|-|B|, 1) \\ &= p+1+\min(|A|+|B|-p-2, -1) \\ &= \min(|A|+|B|-1, p) \end{aligned}$$

which is the desired inequality.

□

It is reasonable now to ask whether or not one may use similar ideas to obtain an analytic proof of Vosper's Theorem. It seems like this should be possible, but we have been unable to do so. This remains an open problem.

Starting with these classical problems regarding sumsets in $\mathbb{Z}_p$, it is possible to go in several different directions. We will discuss two in particular; the first branches off of the Cauchy-Davenport Theorem, and the second branches off of Vosper's Theorem.

Given sets A and B in $\mathbb{Z}_p$, we may wish to control the types of sums that we allow to be in the sumset. For example, define the restricted sumset $\widehat{A+B}$ to be

$$\widehat{A+B} = \{a+b : a \in A, b \in B, \text{ and } a \neq b\}.$$

Of course, if A and B are disjoint, there is no difference between $A+B$ and $\widehat{A+B}$. If A and B have a non-empty intersection, though, the restriction $a \neq b$ may severely change the size of the sumset. The following result, conjectured by Erdös and Heilbronn in 1964 [5] and proved by Dias da Silva and Hamidoune in 1994 [4], is therefore slightly surprising. In essence, it says that the minimal size of a restricted sumset is not much smaller than the minimal size of a traditional sumset.

Theorem 4.10 (Dias da Silva–Hamidoune). *Let A be a non-empty subset of $\mathbb{Z}_p$, where p is prime. Then* $\left|\widehat{A+A}\right| \geq \min\left(p, 2|A|-3\right)$.

In 2006, Alon, Nathanson, and Ruzsa gave a simpler proof of this theorem using a new "polynomial method" [1]. They were also able to solve several other types of modified sumset problems. Their version of Theorem 4.10 is the following.

Theorem 4.11 (Alon–Nathanson–Ruzsa). *Let A and B be non-empty subsets of $\mathbb{Z}_p$, where p is prime and $|A| \neq |B|$. Then* $\left|\widehat{A+B}\right| \geq \min\left(p, |A|+|B|-2\right)$.

It is clear that if we take $B = A\backslash\{a\}$, where a is any element in A, then we obtain the result in Theorem 4.10. We should note here that Theorems 4.10 and 4.11 both have generalizations to restricted sumsets of the form

$$A_1 \widehat{+\ldots+} A_n = \{a_1 + \ldots + a_n : a_i \in A_i \text{ and } a_i \neq a_j \text{ for } i \neq j\}.$$

A different way to generalize restricted sumset problems is to alter the restriction we place on the sums. Let $f : \mathbb{Z}_p \times \mathbb{Z}_p \to \mathbb{Z}_p$ be a polynomial in $\mathbb{Z}_p[x,y]$, and for subsets A and B in $\mathbb{Z}_p$, define

$$\widehat{A+B}_f = \{a+b : a \in A, b \in B, \text{ and } f(a,b) \neq 0\}.$$

According to Lev, it is possible to use the polynomial method of Alon, Nathanson, and Ruzsa to prove a lower bound on the cardinality of $\widehat{A+B}_f$. He then conjectures the following [17].

Conjecture 4.2 (Lev). *Let A and B be non-empty subsets of $\mathbb{Z}_p$, where p is prime, and let f be an arbitrary injective map from A to B. Then*

$$|\{a+b : a \in A, b \in B, \textit{ and } f(a) \neq b\}| \geq \min\left(p-2, |A|+|B|-3\right).$$

We now ask whether or not these restricted sumset problems have analytic proofs. Considering that Tao's finite uncertainty principle, Theorem 4.9, gave an analytic proof of the Cauchy-Davenport Theorem, it seems possible. It is not too surprising, then, that Guo and Sun were recently able to obtain the following result by extending Tao's methods [13].

Theorem 4.12 (Guo-Sun). *Let A, B, and S be subsets of $\mathbb{Z}_p$, where p is prime, and $A \neq \emptyset \neq B$. Then*

$$|\{a+b : a \in A, b \in B, \textit{ and } a-b \notin S\}| \geq \min\left(p, |A|+|B|-2|S|-1\right).$$

It is clear that if we take $S = \emptyset$, then we obtain the Cauchy-Davenport Theorem, and if we take $S = \{0\}$, then we get the result of Theorem 4.10. Unfortunately, Theorem 4.12 does not immediately give us information about restricted sumsets $\widehat{A+B}_f$ or about those in Lev's conjecture. The primary problem is that in Guo and Sun's result, the forbidden set is, in some sense, fixed. Indeed, for a given $b \in B$, the only elements of A that we are allowed to consider are $a \notin S+b$. Thus, as we vary $a \in A$, the forbidden set does not change. This is not the case, however, for $\widehat{A+B}_f$ and for the sumsets in Lev's conjecture. Even if we fix $b \in B$, the forbidden set for a changes as a varies over A. Nevertheless, Theorem 4.12 is a strong result, and it suggests that we should be able to modify Tao's method further to deal with sumsets such as $\widehat{A+B}_f$ and those in Lev's conjecture.

We now return to problems regarding the traditional sumset $A+B$ in $\mathbb{Z}_p$. From Vosper's Theorem, we know that if $A+B$ has minimal size, then (apart from two degenerate cases) A and B must be arithmetic progressions with the same common difference. What if we relax the assumption that $A+B$ is strictly minimal? In other words, what if $A+B$ has size $|A|+|B|$, or $|A|+|B|+1$? Then A and B are not progressions with the same difference, but is it true that they are "almost" arithmetic progressions. Essentially, we are asking for a result in $\mathbb{Z}_p$ that parallels Freiman's Theorem for sumsets in $\mathbb{Z}$. The following theorem gives a partial affirmative answer to this question, and it seems appropriate that it bears the names of both Freiman and Vosper [18].

Theorem 4.13 (Freiman-Vosper). *Let A be a non-empty subset of $\mathbb{Z}_p$, where p is prime. Assume that $|A| \leq p/35$ and $|A+A| = 2|A|-1+r \leq (12|A|/5)-3$. Then A is contained in a mod-p arithmetic progression of length $|A|+r$.*

Several mathematicians believe that it should be possible to find a more general result (that is, to get rid of the upper bound assumption on $|A|$ and to improve the upper bound on $|A+A|$) [17], [19]. In particular, the following conjecture would strengthen the Freiman-Vosper Theorem.

Conjecture 4.3. *Let A be a non-empty subset of $\mathbb{Z}_p$, where p is prime and is sufficiently large. Assume that $|A+A| \leq \min(p-1, 3|A|-4)$. Then A is contained in a mod-p arithmetic progression of length at most $|2A|-|A|+1$.*

The only known proof of the Freiman-Vosper Theorem relies heavily on discrete Fourier analysis. Unfortunately, these methods do not seem to be well-suited for Conjecture 4.3. Instead, it is likely that purely combinatorial arguments are needed to generalize the Freiman-Vosper Theorem. We therefore see that although discrete Fourier analysis plays a central role in additive number theory, there are instances where other types of arguments may be necessary.

Bibliography

[1] Alon, N., M.B. Nathanson, and I. Ruzsa. *The polynomial method and restricted sums of congruence classes*. J. Number Theory, **56** (1996), 404–417.

[2] Behrend, F.A. *On sets of integers which contain no three terms in arithmetic progression*. Proc. Nat. Acad. Sci., **32** (1946), 331–332.

[3] Bourgain, J. *On triples in arithmetic progression*. GAFA, Geom. Funct. Anal., **9** (1999), 968–984.

[4] Dias da Silva, J.A. and Y.O. Hamidoune. *Cyclic spaces for Grassmann derivatives and additive theory*. Bull. London Math. Soc., **26** (1994), 140–146.

[5] Erdös, P. and H. Heilbronn. *On the addition of residue classes mod p*. Acta Arith., **9** (1964), 149–159.

[6] Erdös, P. and P. Turán. *On some sequences of integers*. J. London Math. Soc., **11** (1936), 261–264.

[7] Furstenberg, H. *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*. J. Analyse Mat., **31** (1977), 204–256.

[8] Gowers, W.T. *Fourier analysis and Szemerédi's Theorem*. Doc. Math., Extra Vol. ICM, **1** (1998), 617–629.

[9] Gowers, W.T. *A new proof of Szemerédi's Theorem*. GAFA, Geom. Funct. Anal., **11** (2001), 465–588.

[10] Gowers, W.T. *A new proof of Szemerédi's Theorem for arithmetic progressions of length four*. GAFA, Geom. Funct. Anal., **8** (1998), 529–551.

[11] Green, B. and T. Tao. *New bounds for Szemerédi's Theorem, II: A new bound for $r_4(N)$*. arXiv:math/0610604v1, 2006.

[12] Green, B. and T. Tao. *The primes contain arbitrarily long arithmetic progressions*. Ann. of Math., **167** (2008), 481–547.

[13] Guo, S. and Z.W. Sun. *A variant of Taos method with application to restricted sumsets*. J. Number Theory, **129** (2009), 434–438.

[14] Körner, T.W. Fourier Analysis. Cambridge Univ. Press, 1988.

[15] Laba, I. *From Fourier analysis to arithmetic combinatorics*. Bull. of the Amer. Math. Soc., **45** (2008), 77–115.

[16] Laba, I. ”Harmonic Analysis and Additive Combinatorics.” Lecture 12880, Math. Sci. Res. Inst., Aug. 2008, http://www.msri.org/communications/vmath/VMathVideos.

[17] Lev, V. ”Generalized Erdös-Heilbronn and Bipartite Graphs.” Personal website, http://math.haifa.ac.il/∼seva/problems.html.

[18] Nathanson, M.B. Additive Number Theory: Inverse Problems and the Geometry of Sumsets. Graduate Texts in Math. 165, Springer, 1996.

[19] Rødseth, O.J. Personal communication with M. O’Neill, 2007.

[20] Roth, K.F. *On certain sets of integers*. J. London Math. Soc., **28** (1953), 245–252.

[21] Ruzsa, I. *Generalized arithmetic progressions and sumsets*. Acta Math. Hung., **65** (1994), 379–388.

[22] Soundararajan, K. ”Additive Combinatorics: Winter 2007.” Stanford Univ., http://math.stanford.edu/∼ksound/Notes.pdf.

[23] Stein, E.M. and R. Shakarchi. Fourier Analysis: An Introduction. Princeton Univ. Press, 2003.

[24] Szemerédi, E. *Integer Sets Containing No Arithmetic Progressions*. Acta Math. Hung., **56** (1990), 155–158.

[25] Szemerédi, E. *On sets of integers containing no k elements in arithmetic progression*. Acta Arith., **27** (1975), 299–345.

[26] Tao, T. *The dichotomy between structure and randomness, arithmetic progressions, and the primes*. arXiv:math/0512114v2, 2005.

[27] Tao, T. ”Lecture Notes 5 for 254A.” Univ. Calif. Los Angeles, http://www.math.ucla.edu/∼tao/254a.1.03w.

[28] Tao, T. *Structure and randomness in combinatorics*. arXiv:0707.4269v2, 2007.

[29] Tao, T. *An uncertainty principle for cyclic groups of prime order*. Math. Res. Lett., **12** (2005), 121–128.

[30] Tao, T. and V.H. Vu. Additive Combinatorics. Cambridge Studies in Advanced Math. 105, 2006.