On Arithmetic Structures in Dense Sets of Integers[1]

Ben Green[2]

Abstract

We prove that if $A \subseteq \{1, \dots, N\}$ has density at least $(\log\log N)^{-c}$, where c is an absolute constant, then A contains a triple $(a, a+d, a+2d)$ with $d = x^2 + y^2$ for some integers x, y, not both zero. We combine methods of Gowers and Sárkőzy with an application of Selberg's Sieve. The result may be regarded as a step towards establishing a fully quantitative version of the polynomial Szemerédi Theorem of Bergelson and Leibman.

1. Introduction. If one believes that mathematics is the study of patterns then it is of no surprise that the following result of Szemerédi [16] is often regarded as one of the highlights of all combinatorics.

Theorem 1 (Szemerédi) *Let $\alpha > 0$ be a real number and let k be a positive integer. Then there is $N_0 = N_0(k, \alpha)$ such that any subset $A \subseteq \{1, \dots, N\}$ of size at least αN contains an arithmetic progression of length k, provided that $N \geq N_0(k, \alpha)$.*

Szemerédi's proof was long and combinatorial but just two years later Furstenburg provided a completely different proof of Theorem 1 using ergodic theory. Furstenburg's methods have proved extremely amenable to generalisation, and Furstenburg himself proved the following result.

Theorem 2 *Fix $\alpha > 0$ and let $A \subseteq \{1, \dots, N\}$ have size αN. Then provided $N > N_1(\alpha)$ is sufficiently large one can find two distinct elements $x, x' \in A$ whose difference $x - x'$ is a perfect square.*

As with all such applications of ergodic theory Furstenburg's approach gave no bound on $N_1(\alpha)$. At about the same time Sárkőzy [12] proved the same result in a completely different manner. Sárkőzy's argument took inspiration from a much earlier paper of Roth [10] in which Szemerédi's Theorem was proved for progressions of length 3. The method used is analytic in spirit and does lead to an effective bound on $N_1(\alpha)$, albeit one which is a great distance from the conjectured truth.

The paper [12] was the first in a series of three, and in the final paper [13] of this series an analytic proof of a generalisation of Theorem 2 was outlined. This generalisation says that the squares may, in the formulation of that theorem, be replaced by the set $\{p(d) : d \in \mathbb{N}\}$ where p is any polynomial which maps $\mathbb{N}$ to itself and has an integer root. To see that some

[1]Mathematics Subject Classification (MSC): 11B25 (05D10, 11N36).

[2]The author is supported by an EPSRC research grant, and has also enjoyed the hospitality of Princeton University for some of the period during which this work was carried out.

restriction on the polynomial is necessary for such a result to hold, we invite the reader to construct a set with density $\frac{1}{3}$ containing no difference of the form x^2+1.

Since the late 1970s there have been several significant advances in our understanding of these and related questions. Firstly we mention the result of Bergelson and Leibman from 1996 [1], which is an example of how far Furstenburg's ergodic-theoretic methods have been able to take us.

Theorem 3 (Bergelson-Leibman) *Fix $\alpha > 0$ and let $A \subseteq \{1,\dots,N\}$ have size αN. Let $p_1,\dots,p_r$ be polynomials with $p(\mathbb{N}) \subseteq \mathbb{N}$ and $p(0)=0$. Then provided $N > N_2(\alpha)$ is large enough (exactly how large will depend on the polynomials p_i as well as on α) we can find $a,d \in \mathbb{N}$ for which all r of the numbers $a+p_i(d)$ lie in A.*

This extends both Theorem 1 and Theorem 2 and implies, amongst other things, that dense subsets of the integers contain arbitrarily long arithmetic progressions whose common difference is a non-zero square. Secondly there is a result of Gowers [4], which gives the first bounds for Szemerédi's Theorem.

Theorem 4 (Gowers) *Let $k > 0$ be an integer. Then there is an effectively computable constant $c(k)$ such that any subset of $\{1,\dots,N\}$ of density at least $(\log\log N)^{-c(k)}$ contains a k-term arithmetic progression.*

Gowers' argument takes inspiration from Roth's paper [10]. As we have already remarked, Sárkőzy's methods also bear some resemblance to those of Roth. It is therefore natural to ask whether Gowers' techniques can be adapted to give bounds for questions related to the polynomial Szemerédi Theorem.

In §3 we give a new variant on Sárkőzy's proof of Theorem 2 which we believe to be substantially easier to understand than the original (though it gives a worse bound for $N_1(\alpha)$). Perhaps more importantly we demonstrate that this argument can be made to fit almost entirely into the general methodology of [4], which we shall outline below. It should be pointed out that in 1985 Srinivasan [15] gave a still different argument which seems to be rather simpler than that of Sárkőzy, but more complex than the one we shall give here. The rest of the paper is devoted to a proof of the following result.

Theorem 5 *There is a constant c such that any subset of $\{1,\dots,N\}$ of density at least $(\log\log N)^{-c}$ contains a 3-term arithmetic progression whose common difference is positive and of the form x^2+y^2.*

The proof of this result involves finding a mutual generalisation of the methods of Gowers and Sárkőzy, together with a slightly surprising use of the Selberg Sieve.

The reader will find it hard to understand this section unless she has a working knowledge of the methods of Gowers, such as can be obtained by reading [3] or, even better, parts of [4].

It would be a sizeable undertaking to summarise those papers in any detail here, and indeed there seems little point in doing so.

What we will do is offer a crude outline of the top-level structure of Gowers' proof of Szemerédi's Theorem. All our arguments in this paper will have this broad structure at their heart. Suppose then that we have a set $A \subseteq \{1, \dots, N\}$ of size δN which contains no arithmetic progression of length k. Set $A_0 = A$, $\delta_0 = \delta$ and $N_0 = N$.

- At the ith stage of our argument we will have a set $A_i \subseteq \{1, \dots, N_i\}$ with density δ_i which contains no arithmetic progression of length k.
- The fact that A_i contains no progressions of length k implies that A_i is *non-random* in a certain rather precise sense involving certain Fourier coefficients being large. This means that A_i does not satisfy a property which is possessed by almost all sets of density δ_i. In Gowers' proof this property is known as $(k-2)$-uniformity.
- If a set is not $(k-2)$-uniform then, by a long and complicated argument, it is possible to show that A_i has density at least $\delta_i + \eta(\delta_i)$ on a fairly long progression P, where η is an increasing function of δ_i.
- Define A_{i+1} to be $A_i \cap P$ and rescale so that P has common difference 1. Set $\delta_{i+1} = \delta_i + \eta(\delta_i)$ and $N_{i+1} = |P|$.

Iterating this argument leads to an effective version of Szemerédi's Theorem since after a finite number of steps the density δ_i will exceed 1, a contradiction.

2. Notation and Basic Concepts. We shall make substantial use of Fourier analysis on finite cyclic groups, so we would like to take this opportunity to give the reader a swift introduction. If nothing else this will serve to clarify notation.

Let N be a fixed positive integer, and write $\mathbb{Z}_N$ for the cyclic group with N elements. Let ω denote the complex number $e^{2\pi i/N}$. Although ω clearly depends on N, we shall not indicate this dependence in the rest of the paper, trusting that the value of N is clear from context. Let $f : \mathbb{Z}_N \to \mathbb{C}$ be any function. Then for $r \in \mathbb{Z}_N$ we define the Fourier transform

$$\hat{f}(r) \;=\; \sum_{x \in \mathbb{Z}_N} f(x)\omega^{rx}.$$

We shall repeatedly use two important properties of the Fourier transform. The first is Parseval's identity, which states that if $f : \mathbb{Z}_N \to \mathbb{C}$ and $g : \mathbb{Z}_N \to \mathbb{C}$ are two functions then

$$N \sum_{x \in \mathbb{Z}_N} f(x)\overline{g(x)} \;=\; \sum_{r \in \mathbb{Z}_N} \hat{f}(r)\overline{\hat{g}(r)}.$$

The second is the interaction of *convolutions* with the Fourier transform. If $f, g : G \to \mathbb{C}$ are two functions on an abelian group G we define the convolution

$$(f * g)(x) \;=\; \sum_{y \in G} f(y)g(x-y).$$

It is easy to check that

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

One more piece of notation: we will often use the same letter to denote both a set and its characteristic function.

It is now possible to formally introduce the concepts of uniformity and quadratic uniformity, the only types of uniformity that will feature in this paper. These correspond to what we called 1-uniformity and 2-uniformity in our earlier outline.

Let $A \subseteq \mathbb{Z}_N$ be a set of size αN and let $f = A - \alpha$ be its *balanced function*. We say that A is η-uniform if $\|\hat{f}\|_\infty \leq \eta N$. To define quadratic uniformity we need some extra notation. If $f : G \to \mathbb{C}$ is any function on an abelian group we write

$$\Delta(f; h)(x) \;=\; f(x)f(x-h).$$

Let f be the balanced function of A. Then A is said to be *quadratically η-uniform* if

$$\sum_h \|\Delta(f; h)\hat{}\,\|_\infty^2 \;\leq\; \eta N^3. \tag{1}$$

These definitions are very similar, though not completely identical, to those in [4]. For a detailed discussion of the basic properties of uniformity and quadratic uniformity the reader should consult [4]. It turns out that being quadratically uniform is a stronger requirement than being uniform (and so knowing a set fails to be quadratically uniform gives less information than knowing it fails to be uniform). We will need the following quantitative statement of this fact later on.

Proposition 6 *If A is quadratically η-uniform then it is $\eta^{1/4}$-uniform.*

Proof By (1) we have

$$\sum_h |\Delta(f; h)\hat{}(0)|^2 \;\leq\; \eta N^3.$$

Writing this statement out in full gives

$$\sum_{a+b=c+d} f(a)f(b)f(c)f(d) \;\leq\; \eta N^3$$

which implies, by Parseval's theorem, that

$$\sum_r |\hat{f}(r)|^4 \ \leq \ \eta N^4.$$

It follows immediately that $\|\hat{f}\|_\infty^4 \leq \eta N^4$, which implies the proposition. □

3. A Short Proof of Sárkőzy's Theorem for Squares. In this section we prove Theorem 2 using an iterative argument of the type outlined above. For obvious reasons we say that A has a *square difference* if there are distinct elements x and x' in A with $x - x'$ a square. Let us recall Theorem 2.

Theorem 2 *Fix $\alpha > 0$ and let $A \subseteq \{1, \dots, N\}$ have size αN. Then provided $N > N_1(\alpha)$ is sufficiently large A has a square difference.*

As the first part of our argument we shall show that a set $A \subseteq \{1, \dots, N\}$ with size αN which contains no square difference fails to be η-uniform for some reasonably large η. In proving this statement we shall use just one fact about the squares, an elementary proof of which may be found in [8].

Lemma 7 *Let $r_6(n)$ denote the number of ordered sextuples $(a_1, \dots, a_6)$ with $a_1^2 + \cdots + a_6^2 = n$. Then*

$$n^2 \ \leq \ r_6(n) \ \leq \ 40n^2.$$

In fact we shall have no use for the lower bound in the lemma, but have included it to emphasise that r_6 is comparable to n^2. □

Now let S be the set of non-zero squares less than $N/2$, and let $B = A \cap [0, N/2]$. Regard S, A and B as subsets of $\mathbb{Z}_N$. Assume without loss of generality that $|B| \geq \alpha N/2$. If $A - A$ contains no square then certainly

$$\sum_{x,d} B(x)A(x+d)S(d) \ = \ 0,$$

where the sums are over $\mathbb{Z}_N$. It follows easily from Parseval's Theorem and the triangle inequality that

$$\sum_{r \neq 0} |\hat{S}(r)||\hat{B}(r)||\hat{A}(r)| \ \geq \ |\hat{S}(0)||\hat{B}(0)||\hat{A}(0)| \ \geq \ \tfrac{1}{4}\alpha^2 N^{5/2}. \qquad (2)$$

The left-hand side here is at most

$$\sup_{r \neq 0} |\hat{A}(r)|^{1/6} \sum_r |\hat{S}(r)||\hat{B}(r)||\hat{A}(r)|^{5/6}$$

which, by Hölder's inequality, is at most

$$\sup_{r\neq 0}|\hat{A}(r)|^{1/6}\left(\sum_r|\hat{S}(r)|^{12}\right)^{1/12}\left(\sum_r|\hat{B}(r)|^2\right)^{1/2}\left(\sum_r|\hat{A}(r)|^2\right)^{5/12}. \quad (3)$$

Now (by Parseval again) $\sum_r|\hat{S}(r)|^{12}$ is equal to $N\sum_x R_6(x)^2$, where $R_6(x)$ is the number of solutions to $a_1^2+\cdots+a_6^2\equiv x(\mathrm{mod}N)$ with $a_i^2\in(0,N/2]$. An easy exercise using Lemma 7 shows that $R_6(x)\leq 600N^2$ for all x, which gives the bound

$$\sum_r|\hat{S}(r)|^{12}\ \leq\ 2^{19}N^6. \quad (4)$$

Parseval also gives $\sum_r|\hat{A}(r)|^2\leq N^2$ and $\sum_r|\hat{B}(r)|^2\leq N^2$. Substituting these and (4) into (3) gives that there is $r\neq 0$ for which

$$|\hat{A}(r)|\ \geq\ 2^{-30}\alpha^{11/2}|A|. \quad (5)$$

In other words, A is non-uniform.

To complete our proof of Theorem 2 by the iteration method we are going to use (5) to show that A has increased density on a *square-difference* AP. If we pass to such a subprogression and then rescale this subprogression to have common difference 1, the resulting set A' will still not contain a square difference. To find a subprogression of the desired type requires a further standard fact about the squares, which is a quantitative version of the fact that the squares are a Heilbronn Set (see [7]). As noted in [4] it is rather difficult to find a precise statement of this result in the literature. Fortunately however we can use Lemma 5.5 of [4] to simply write down the next lemma.

Lemma 8 *Let $a\in\mathbb{Z}_N$, let $t\leq N$, and suppose that $t\geq 2^{2^{128}}$. Then there is $p\leq t$ such that $|p^2a|\leq t^{-1/16}N$.*

Now let r be such that $|\hat{A}(r)|$ is large. Put $t=N^{1/4}$ in Lemma 8 and let $p\leq N^{1/4}$ be such that $|p^2r|\leq N^{127/128}$. Let B be the arithmetic progression $p^2,2p^2,\ldots,Lp^2$ where $L=\frac{1}{20}N^{1/128}$. One calculates

$$\begin{aligned}|\hat{B}(r)|\ &\geq\ L\sup_{x=1,\ldots,L}\left(1-\left|1-\omega^{rp^2x}\right|\right)\\ &\geq\ L\left(1-\frac{2\pi|rp^2|L}{N}\right)\\ &\geq\ L/2.\end{aligned}$$

From this and the fact that $|\hat{A}(r)|\geq 2^{-30}\alpha^{11/2}|A|$ we get that

$$\begin{aligned}N\sum_x|A\cap(B+x)|^2\ &=\ \sum_r|\hat{A}(r)|^2|\hat{B}(r)|^2\\ &\geq\ \left(1+2^{-62}\alpha^{11}\right)|A|^2L^2. \quad (6)\end{aligned}$$

Now certain of the translates $B+x$ have the rather unfortunate property of splitting into two smaller progressions when we "unravel" $\mathbb{Z}_N$ to recover $\{1,\dots,N\}$. We shall call these values of x bad, and denote the set of good values by $\mathcal{G}$. Now $p \le N^{1/4}$ and so B has diameter less than $N^{2/3}$. It follows that there are at most $N^{2/3}$ bad values of x, and so their total contribution to $N\sum_x |A\cap(B+x)|^2$ does not exceed $L^2N^{5/3}$. Assuming that $\alpha \ge 32N^{-1/39}$ (which it certainly will be) one sees from (6) that

$$N\sum_{x\in\mathcal{G}} |A\cap(B+x)|^2 \ \ge\ \left(1+2^{-63}\alpha^{11}\right)|A|^2L^2.$$

The left hand side here is at most

$$N\sup_{x\in\mathcal{G}} |A\cap(B+x)|\cdot|A|L,$$

from which we deduce that there is $x\in\mathcal{G}$ for which

$$|A\cap(B+x)| \ \ge\ \left(\alpha+2^{-63}\alpha^{12}\right)|B|.$$

We have deduced, from the assumption that $A-A$ does not contain a square and that $N\ge 2^{2^{130}}$, that A has density at least $\alpha+2^{-63}\alpha^{12}$ on a subprogression with length at least $\frac{1}{20}N^{1/128}$ and square common difference. Iterating this argument leads to the following result.

Proposition 9 *There is a constant C such that, if A is a subset of $\{1,\dots,N\}$ with density at least $C(\log\log N)^{-1/11}$, then A contains two elements a,a' with $a-a'$ a non-zero square.*

In order to prove Sárkőzy's result in the simplest possible manner we have not worried too much about sacrificing the quality of the bound obtained. There is a variation of the above argument in which one shows that A is what I call *arithmetically non-uniform*, which means that some $|\hat{A}(r)|$ is large with r approximately equal to a rational with small denominator. This allows one to perform the iteration more efficiently, and doing this gives a bound of form $(\log N)^{-c}$ in Proposition 9. I intend to discuss this and related matters in a future paper. The current best known bound of $(\log N)^{-c\log\log\log\log N}$ for this problem is due to Pintz, Steiger and Szemerédi [9] and makes use of some rather involved Fourier arguments. There is still a massive gap in our knowledge, and I cannot resist closing this section with the following open problem.

Problem 10 *Let $\epsilon>0$ and let $N>N_0(\epsilon)$ be sufficiently large. Does there exist a set $A\subseteq\{1,\dots,N\}$ with $|A|\ge N^{1-\epsilon}$, such that A does not contain two elements that differ by a square?*

The best that is known is that one can take $\epsilon=0.267$, a result due to Ruzsa [11].

4. APs with Common Difference x^2+y^2. In this section we turn our attentions to the

main business of this paper, a proof of Theorem 5.

Theorem 5 *There is a constant c such that any subset of $\{1, \ldots, N\}$ of density at least $(\log\log N)^{-c}$ contains a 3-term arithmetic progression whose common difference is non-zero and of the form $x^2 + y^2$.*

Our proof of Theorem 5 will be by the iteration method, and as such will fall into two parts. In the first part, containing most of our original work on the problem, we show that a quadratically uniform set contains roughly the expected number of progressions $(a, a + d, a + 2d)$ with $d = x^2 + y^2$. The second part of the proof follows [4] extremely closely. Our objective there is to show that a set which fails to be quadratically uniform has increased density on a long subprogression with square common difference.

We shall start our treatment in quite a general setting. Let D be a subset of $\mathbb{N}$, and for $N \in \mathbb{N}$ regard $D_N = D \cap \{1, \ldots, N\}$ as a subset of $\mathbb{Z}_N$ in the natural way. Suppose that for some $k \in \mathbb{N}$ we have

$$\sum_r \left|\hat{D}_N(r)\right|^{2k} \ \leq \ C|D_N|^{2k}$$

and

$$|D_{2N}| \ \leq \ C|D_N|,$$

where C is independent of N. Then we shall say that D is *uniformly k-dense*. The obvious non-trivial example in view of our earlier discussions is the set S of squares, which is uniformly 6-dense by (4). Our nomenclature is non-standard, but quite convenient.

Much of our later work depends on another instance of this phenomenon.

Proposition 11 *The set E of primes of form $4k + 1$ is uniformly 2-dense.*

We will prove this proposition in a number of stages. We begin with a brief resumé of the results from Sieve Theory we will need both here and later on. The standard reference for this subject is [6] but in our unbiased opinion a good way to understand the necessary background is to read [5], which has been specially updated for this purpose. We will only be concerned with sieving polynomial sequences, which is the simplest situation covered by the Selberg Sieve. Let h be a polynomial with integer coefficients, and let $\mathcal{A}$ denote the sequence $\{h(1), \ldots, h(N)\}$. Let $\mathcal{P}$ denote the set of primes. Then the Selberg Sieve gives upper bounds for $S(\mathcal{A}, \mathcal{P}, z)$, which is defined to be the number of $x \in \mathcal{A}$ which are not divisible by any prime $p \leq z$. This upper bound is given in terms of a function ω defined at primes p to be the number of elements in $\{h(1), \ldots, h(p)\}$ which are divisible by p. To put it another way, the proportion of elements of $\mathcal{A}$ which are divisible by p is roughly $\omega(p)/p$. The key result we shall require is the following, which is Theorem 11 in [5] (we should also note that it can be read out of [6]).

Theorem 12 *Suppose that $\omega(p) \leq C$ for all primes p. Then*

$$S(\mathcal{A}, \mathcal{P}, N^{1/16C}) \ll N \prod_{p \leq N^{1/16C}} \left(1 - \frac{\omega(p)}{p}\right),$$

where the implied constant depends only on C.

Proposition 13 *Let $n \in \mathbb{N}$. Then the number of representations of n as the sum of two elements of E, $r_2(E, n)$ satisfies*

$$r_2(E, n) \ll \frac{n}{(\log n)^2} \prod_{p|n} \left(1 + \frac{1}{p}\right).$$

Proof Clearly $r_2(E, n) \leq r_2(\mathcal{P}, n)$. Consider the polynomial $h(x) = x(n - x)$ and let $\mathcal{A} = \{h(1), \ldots, h(n)\}$. We wish to count the number of $x \leq n$ for which x and $n - x$ are both prime. For such x we either have $x \leq n^{1/2}$, $x \geq n - n^{1/2}$ or else $h(x)$ has no prime factor less than $n^{1/2}$. It follows that $r_2(\mathcal{P}, n)$ is bounded above by $2n^{1/2} + S(\mathcal{A}, \mathcal{P}, n^{1/2})$.

It is easy to see that, in the notation of our potted introduction to sieve theory, one has $\omega(p) = 2$ for all p except when $p|n$, in which case $\omega(p) = 1$. Thus by Theorem 12 one has

$$r_2(E, n) \ll n^{1/2} + n \prod_{p \leq n^{1/32}} \left(1 - \frac{2}{p}\right) \prod_{\substack{p \leq n^{1/32} \\ p|n}} \left(1 - \frac{2}{p}\right)^{-1} \left(1 - \frac{1}{p}\right). \tag{7}$$

Now recall that $\prod_{p \leq m} \left(1 - \frac{1}{p}\right)^{-1} \asymp \log m$ and that $\prod_p \left(1 - \frac{\lambda}{p^2}\right)$ converges for any real λ. Armed with these two facts one sees from (7) that

$$r_2(E, n) \ll \frac{n}{(\log n)^2} \prod_{p|n} \left(1 + \frac{1}{p}\right)$$

as claimed. □

This is of course a very standard deduction from the Selberg Sieve, but we have included it to ensure that the reader is happy with our notation. It turns out to be extremely convenient to write $\xi(n)$ for the quantity $\prod_{p|n} \left(1 + \frac{1}{p}\right)$ appearing here. In a short while we will use Proposition 13 to show that E is uniformly 2-dense. Before doing this however it is necessary to give a crude estimate for the moments $\sum_{u \leq N} \xi(u)^s$ of ξ.

Lemma 14 *Let $s \geq 1$ be real. Then*

$$\sum_{u \leq N} \xi(u)^s \leq 2^{2s2^{2s}} N.$$

Proof First observe that for any $x \leq 1$ one has

$$(1+x)^s \ \leq \ 1+2^s x.$$

Secondly, if $p_1, \dots, p_r$ are distinct primes then for any C one has

$$\prod_{i=1}^r \frac{C}{p_i} \ = \ \prod_{p_i \leq C^2} \frac{C}{p_i} \prod_{p_i > C^2} \frac{C}{p_i} \ \leq \ C^{C^2} \prod_{p_i > C^2} \frac{1}{\sqrt{p_i}} \ \leq \ C^{C^2} \prod_{i=1}^r \frac{1}{\sqrt{p_i}}\,.$$

Thus

$$\sum_{u \leq N} \xi(u)^s \ = \ \sum_{u \leq N} \prod_{p|u} \left(1+\frac{1}{p}\right)^s \ \leq \ \sum_{u \leq N} \prod_{p|u} \left(1+\frac{2^s}{p}\right) \ \leq \ 2^{s2^{2s}} \sum_{u \leq N} \sum_{d|u} \frac{1}{\sqrt{d}}\,,$$

which is at most $2^{s2^{2s}} N \sum_{d \leq N} d^{-3/2}$. This implies the result. □

Proof of Proposition 11 Regard $E_N = E \cap \{1, \dots, N\}$ as a subset of $\mathbb{Z}_N$, as we must do to even make sense of what it means to be uniformly 2-dense. It is easy to see that

$$\sum_r |\hat{E}_N(r)|^4 \ = \ N \sum_{a+b=c+d} E(a)E(b)E(c)E(d),$$

where the equation $a+b=c+d$ is taken in $\mathbb{Z}_N$. This is clearly at most

$$N \sum_{n \leq N} \left(r_2(E,n) + r_2(E, N+n)\right)^2,$$

which by Proposition 13 is at most a constant times

$$\frac{N^3}{(\log N)^4} \sum_{n \leq N} \left(\xi(n) + \xi(N+n)\right)^2.$$

The sum, by the Cauchy-Schwarz inequality, is at most $4 \sum_{n \leq 2N} \xi(n)^2$, a quantity which we know to be $O(N)$ by Lemma 14. Thus

$$\sum_r |\hat{E}_N(r)|^4 \ \ll \ \left(\frac{N}{\log N}\right)^4.$$

Since $|E_N| \sim N/2\log N$ it follows that E is indeed uniformly 2-dense. □

If $A \subseteq \{1, \dots, N\}$ is a set of density α then we write $f = A - \alpha$ for its balanced function. Write $B = A \cap \{1, \dots, N/3\}$, let β be the density of B and let $g = B - \beta$ be its balanced function. Finally if D is any set then we say that an arithmetic progression $(x, x+d, x+2d)$ with $d \in D$ is a *D-progression*.

Proposition 15 *Suppose that A does not contain a D-progression, where D is uniformly k-dense. Then either*

$$\sup_{r\neq 0}|\hat{A}(r)| \gg \alpha^{2k}N \tag{8}$$

or else we have

$$\sum_x \sum_{d\in U} g(x)f(x+d)f(x+2d) \gg \alpha^3 N|U|. \tag{9}$$

where the sum is over $\mathbb{Z}_N$ and $U = D \cap \{1,\ldots,N/3\}$.

Proof To begin with we show that the conclusion holds with room to spare if β is much smaller than expected. Suppose that $\beta \leq \alpha/12$, and let I be the characteristic function of $\{-N/12,\ldots,N/12\} \subseteq \mathbb{Z}_N$. Then

$$\frac{6}{N}\sum_x A(x)(I*I)(x+N/6) \leq \beta N \leq \alpha N/12.$$

Taking Fourier transforms gives

$$\sum_r \omega^{-rN/6}\hat{A}(r)|\hat{I}(r)|^2 \leq \alpha N^3/72,$$

which implies by the triangle inequality that

$$\sum_{r\neq 0}|\hat{A}(r)||\hat{I}(r)|^2 \geq \alpha N^3/72.$$

However Parseval's identity gives $\sum|\hat{I}(r)|^2 = N^2/6$, from which it follows immediately that

$$\sup_{r\neq 0}|\hat{A}(r)| \geq \alpha N/12.$$

Assume now that $\beta \geq \alpha/12$. It is easy to see that there is no *modular* progression of form $(x, x+d, x+2d)$, $d \in U$, in $B \times A \times A$, and so we have

$$\sum_{x\in\mathbb{Z}_N}\sum_{d\in U} B(x)A(x+d)A(x+2d) = 0.$$

Writing $A = f + \alpha$ and $B = g + \beta$ we may expand this as a sum of eight terms. Of these we have a term $T = \sum_x \sum_{d\in U} g(x)f(x+d)f(x+2d)$, a term $\alpha^2\beta N|U|$ and three terms which are identically zero. If $T \geq \alpha^2\beta N|U|/4$ then we are done. Failing this the triangle inequality implies that one of the other three terms must be at least $\alpha^2\beta N|U|/4$, which in turn is not less than $\alpha^3 N|U|/48$. These other three terms are very similar to one another, and brushing a little work under the carpet we suppose without loss of generality that

$$\alpha\sum_x\sum_{d\in U} g(x)f(x+d) \geq \alpha^3 N|U|/48. \tag{10}$$

The LHS may be written in terms of Fourier coefficients as $\alpha N^{-1}\sum_r \hat{g}(r)\hat{f}(-r)\hat{U}(r)$. This sum may be estimated using Hölder's Inequality exactly as in §3. It is at most

$$\alpha N^{-1}\sup_r \left|\hat{f}(r)\right|^{1/k}\left(\sum_r \left|\hat{U}(r)\right|^{2k}\right)^{1/2k}\left(\sum_r |\hat{g}(r)|^2\right)^{1/2}\left(\sum_r \left|\hat{f}(r)\right|^2\right)^{(k-1)/2k} \quad (11)$$

To deal with this observe that since D is uniformly k-dense we have

$$\begin{aligned}\sum_r |\hat{U}(r)|^{2k} &= N\sum_{\substack{a_1+\cdots+a_k\\=b_1+\cdots+b_k}} U(a_1)\dots U(a_k)U(b_1)\dots U(b_k)\\ &\leq N\sum_{\substack{a_1+\cdots+a_k\\=b_1+\cdots+b_k}} D_N(a_1)\dots D_N(a_k)D_N(b_1)\dots D_N(b_k)\\ &= \sum_r |\hat{D}_N(r)|^{2k}\\ &\ll |D_N|^{2k}\\ &\ll |U|^{2k}.\end{aligned}$$

Using Parseval's identity on the other two factors in (11) we can bound that expression above by a constant multiple of

$$\sup_r |\hat{f}(r)|^{1/k}\cdot N^{1-1/k}|U|.$$

It follows from (10) that $\sup_r|\hat{f}(r)| \gg \alpha^{2k}N$, and we are done. □

Suppose now that we have a set $A \subseteq \{1,\dots,N\}$ with density α which contains no D-progressions, where D is the set of sums of two squares. A classical number-theoretic fact allows us to make the key observation of this paper, namely that A does not contain any E progressions (where E, as before, is the set of primes of the form $4k+1$). Using Propositions 11 and 15, we have the following.

Proposition 16 *Let $A \subseteq \{1,\dots,N\}$ have density α, and suppose that A does not contain a triple $(a, a+d, a+2d)$ with $d = x^2+y^2$. Then either*

$$\sup_{r\neq 0}|\hat{A}(r)| \gg \alpha^4 N \quad (12)$$

or

$$\sum_x\sum_{d\in V} g(x)f(x+d)f(x+2d) \gg \frac{\alpha^3N^2}{\log N}. \quad (13)$$

where $V = E\cap\{1,\dots,N/3\}$.

If (12) holds then further progress is comparitively easy, so we assume for the moment that (13) is satisfied. The Cauchy-Schwarz inequality together with the fact that $\|g\|_\infty \leq 1$ gives that

$$\sum_x \left| \sum_d f(x+d)f(x+2d)V(d) \right|^2 \gg \frac{\alpha^6 N^3}{(\log N)^2}.$$

Multiplying out, rearranging and changing the summation variables, this implies that

$$\sum_h \sum_x \sum_d \Delta(f;h)(x)\Delta(f;2h)(x+d)\Delta(V;h)(d) \gg \frac{\alpha^6 N^3}{(\log N)^2} \tag{14}$$

We are going to use this to show that, for many h, $\Delta(f;h)$ has a large Fourier coefficient. We will do this by applying the Selberg Sieve again to show that $\Delta(V;h)$ is uniformly 2-dense *on average*, a concept we shall not define precisely.

Proposition 17 *Let $r_h(n)$ denote the number of ways of expressing n as a difference of 2 elements of $\Delta(V;h)$. Then*

$$r_h(n) \ll \frac{N}{(\log N)^4} \frac{\xi(h)^2\xi(n)^2\xi(n+h)\xi(n-h)}{\xi((n,h))^3}.$$

Proof We bring the Selberg Sieve to bear on this by observing that $r_h(n)$ is at most the number of $x \leq N$ for which x, $x-n$, $x-h$ and $x-n-h$ are all prime. With the exception of at most $8N^{1/2}$ of these x the polynomial

$$h(x) = x(x-n)(x-h)(x-n-h)$$

has no prime factor less than $N^{1/2}$. Writing $\mathcal{A} = \{h(1), \ldots, h(N)\}$ this implies that

$$r_h(n) \ll 8N^{1/2} + S(\mathcal{A}, \mathcal{P}, N^{1/2}).$$

We would clearly like to apply Theorem 12, but first we must think about $\omega(p)$. Suppose that $p \geq 3$. Then it is reasonably easy to see that $\omega(p) = 4$ unless one of the following four possibilities occurs: (i) $p|n$; (ii) $p|h$; (iii) $p|(n+h)$ and (iv) $p|(n-h)$. Furthermore these possibilities are mutually exclusive unless p divides both n and h, in which case they all occur. If they do all occur then $\omega(p) = 1$. If (i) or (ii) occurs then $\omega(p) = 2$, and if (iii) or (iv) occurs then $\omega(p) = 3$. If $p = 2$ the behaviour is more subtle, but we will not concern ourselves with this as each individual prime only contributes a bounded multiplicative factor to the sieve estimate of Theorem 12.

Theorem 12 (with $C = 4$) certainly applies to this situation, then, and with a modicum of effort one can verify that the key quantity

$$N \prod_{p \leq N^{1/64}} \left(1 - \frac{\omega(p)}{p}\right)$$

is equal, up to a product of terms of the form $\prod_p(1-\lambda p^{-2})$, to the delightful expression

$$\prod_{p\leq N^{1/64}}\Big(1-\frac{4}{p}\Big)\prod_{p|n}\Big(1-\frac{2}{p}\Big)^{-1}\prod_{p|h}\Big(1-\frac{2}{p}\Big)^{-1}$$
$$\times\prod_{p|(n+h)}\Big(1-\frac{1}{p}\Big)^{-1}\prod_{p|(n-h)}\Big(1-\frac{1}{p}\Big)^{-1}\prod_{p|(n,h)}\Big(1-\frac{1}{p}\Big)^{3}$$

where the final five products are also constrained to be over $p\leq N^{1/64}$. Since an integer less than N cannot have more than 64 prime factors $p\geq N^{1/64}$, this restriction can be removed at the expense of introducing another bounded multiplicative constant. The resulting expression is easily seen to be bounded above by a constant multiple of the one appearing in the statement of the proposition. □

Proposition 18 *For any h we have*

$$\sum_{n\leq N}r_h(n)^2\ \ll\ \frac{N^3}{(\log N)^8}\xi(h)^4,$$

where the implied constant is independent of h.

Proof By Proposition 17 we have

$$\sum_{n\leq N}r_h(n)^2\ \ll\ \frac{N^2}{(\log N)^8}\xi(h)^4\sum_{n\leq N}\xi(n)^4\xi(n+h)^2\xi(n-h)^2.$$

By Hölder's Inequality the sum over n is at most $\sum_n\xi(n)^8$. We are therefore done by Lemma 14. □

The next result clarifies the sense in which $\Delta(V;h)$ is, on average, uniformly 2-dense. In the following proposition $\Delta(V;h)$ is regarded as a subset of $\mathbb{Z}_N$.

Proposition 19

$$\sum_r|\Delta(V;h)\hat{}(r)|^4\ \ll\ \frac{N^4}{(\log N)^8}\xi(h)^4. \tag{15}$$

Proof Immediate from Proposition 18 and the fact that $\Delta(V,h)$ is supported in an interval of size $N/3$ (so that there is no problem with modular addition not being the same as "ordinary" addition). □

Now we come to use (14). We shall use it to treat each h separately, which we do by noting that from (14) follows

$$\sum_x\sum_d\Delta(f;h)(x)\Delta(f;2h)(x+d)\Delta(V;h)(d)\ \gg\ \gamma(h)\frac{\alpha^6N^2}{(\log N)^2},$$

where $\sum_h \gamma(h) = N$. Taking Fourier Coefficients, this implies that

$$\sum_r \Delta(f;h)\hat{}(r)\Delta(f;2h)\hat{}(-r)\Delta(V;h)\hat{}(r) \gg \gamma(h)\frac{\alpha^6 N^3}{(\log N)^2}. \tag{16}$$

Applying Hölder gives

$$\sup_r |\Delta(f;h)\hat{}(r)|^{1/2}\left(\sum_r |\Delta(f;h)\hat{}(r)|^2\right)^{1/4}\left(\sum_r |\Delta(f;2h)\hat{}(r)|^2\right)^{1/2}\left(\sum_r |\Delta(V;h)\hat{}(r)|^4\right)^{1/4} \gg \gamma(h)\frac{\alpha^6 N^3}{(\log N)^2}.$$

The first two bracketed expressions may be bounded above using Parseval, and the third is subject to the upper bound (15). One gets

$$\sup_r |\Delta(f;h)\hat{}(r)| \gg \frac{\gamma(h)^2\alpha^{12}N}{\xi(h)^2}. \tag{17}$$

Thus there is a function $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ such that

$$\sum_h |\Delta(f;h)\hat{}(\phi(h))|^2 \gg \alpha^{24}N^2\sum_h \frac{\gamma(h)^4}{\xi(h)^4}.$$

This would imply that A fails to be quadratically uniform if we could show that the sum here is $\gg N$. To do this, we recall that $\sum_h \gamma(h) = N$ and use Hölder, getting

$$\sum_h \frac{\gamma(h)^4}{\xi(h)^4} \gg N^4\left(\sum_h \xi(h)^{4/3}\right)^{-3}.$$

This is indeed $\gg N$ by Lemma 14.

We have now shown that if (14) holds then A is not quadratically $C\alpha^{24}$-uniform for some C. We also know that if A does not contain any E-progressions then either (12) or (14) must hold. However (12) is just the statement that A is not $C\alpha^4$-uniform for some C. Thus by Proposition 6 we may incorporate everything we have done so far into the following.

Proposition 20 *Suppose that $A \subseteq \{1,\dots,N\}$ has density α yet does not contain a progression $(x, x+d, x+2d)$ with d a positive sum of two squares. Then A is not quadratically $C\alpha^{24}$-uniform for some C.*

To spell it out, the conclusion of this proposition implies that

$$\sum_u |\Delta(f;u)\hat{}(\phi(u))|^2 \gg \alpha^{24}N^3$$

for some function $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$.

5. Increasing the Density on a Special Subprogression. Sets that fail to be quadratically uniform have an interesting structure, as was established by Gowers [4] in the course of proving Szemerédi's Theorem for progressions of length 4 (a slightly weaker result was established in [3]). In that paper a result along the following lines is proved.

Theorem 21 (Gowers' Inverse Theorem) *Suppose that $A \subseteq \{1, \dots, N\}$ has density α and that A fails to be quadratically $C_1\alpha^{m_1}$-uniform. Then there is an arithmetic progression L of length $|L| \gg N^{C_2\alpha^{m_2}}$ on which A has density at least $\alpha + C_3\alpha^{m_3}$, where C_2, C_3, m_2 and m_3 depend only on m_1 and C_1.*

The main result of this section is a version of Theorem 21 in which L has square common difference. Unfortunately we have found it necessary to go a fair way into the detailed workings of Gowers' argument in order to obtain this modification. Therefore in order to fully understand this section the reader will need to be conversant with Chapters 6, 7 and 8 of [4]. We shall require one additional ingredient, which is a version of the following simultaneous approximation result of Schmidt [14].

Theorem 22 (Schmidt) *Let $r_1, \dots, r_h \in \mathbb{Z}_N$ and let $t \leq N$. Suppose that $t \geq N_0(h)$ is sufficiently large. Then there is $p \leq t$ such that $|p^2 r_i| \leq t^{-1/3h^2} N$ for all i, $1 \leq i \leq h$.*

Unfortunately there does not seem to be any place in the literature where an explicit value of $N_0(h)$ is derived. It would be possible to work through the proof in [14] and derive such an explicit value (which would probably not be too large) but we will prove our own, substantially weaker, version of Theorem 22 with explicit constants.

Proposition 23 *Let $r_1, \dots, r_h \in \mathbb{Z}_N$ and let $t \leq N$. Suppose that $t \geq 2^{2^{7h+129}}$. Then there is $p \leq t$ such that $|p^2 r_i| \leq t^{-2^{-(7h+6)}} N$ for all i, $1 \leq i \leq h$.*

Proof We make extensive use of Lemma 8. Choose $p_1, p_2, \dots$ inductively so that

$$p_i \;\leq\; t^{2^{-(7i+1)}} \tag{18}$$

and

$$\left|p_i^2 p_{i-1}^2 \dots p_1^2 r_i\right| \;\leq\; t^{-2^{-(7i+5)}} N \tag{19}$$

for each $i = 1, \dots, h$. Let $p = p_1 p_2 \dots p_h$. It is easy to check that $p \leq t$. Furthermore

$$\begin{aligned} |p^2 r_i| &\leq |p_h|^2 |p_{h-1}^2| \dots |p_{i+1}^2| \left|p_i^2 \dots p_1^2 r_i\right| \\ &\leq t^{2^{-(7i+6)}} \cdot t^{-2^{-(7i+5)}} N \\ &\leq t^{-2^{-(7h+6)}} N \end{aligned}$$

as required. The repeated applications of Lemma 8 require certain things to be sufficiently large. The most difficult condition to satisfy is that arising from the last step in our inductive construction, where we require

$$t^{2^{-(7h+1)}} \geq 2^{2^{128}}.$$

This is where the restriction in the proposition comes from. □

Let us assume then that $f : \mathbb{Z}_N \to \mathbb{C}$ has $\|f\|_\infty \leq 1$ and that

$$\sum_k |\Delta(f;k)\hat{}\,(\phi(k))|^2 \geq 2\beta N^3 \tag{20}$$

for some function $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$. We start with a trivial deduction from this, which is proved by a simple averaging argument: there is a set $B \subseteq \mathbb{Z}_N$ with $|B| \geq \beta N$ and

$$|\Delta(f;r)\hat{}(\phi(r))| \geq \beta^{1/2} N \tag{21}$$

for all $k \in B$. We have then

$$\sum_{k\in B} |\Delta(f;k)\hat{}\,(\phi(k))|^2 \geq \beta^2 N^3. \tag{22}$$

If $K \subseteq \mathbb{Z}_N$ and $\eta \in (0,1)$ write $\mathcal{B}(K,\eta)$ for the set of all $n \in \mathbb{Z}_N$ such that $|nk| \leq \eta N$ for all $k \in K$. We call such a set a *Bohr Neighbourhood*. The following may be deduced from Proposition 6.1, Corollary 7.6, Lemma 7.8 and Corollary 7.9 of [4]. This deduction is logically extremely straightforward, but just a touch messy when it comes to actually working out values.

Proposition 24 *Let $\delta = 2^{-1849}\beta^{5821}$. There is a set $B' \subseteq B$ with $|B'| \geq \delta N$, and a set $K \subseteq \mathbb{Z}_N$ with $|K| \leq 16\delta^{-2}$, such that the following is true. If m is any positive integer and $d \in \mathcal{B}(K, \delta/100m)$ then there is c such that $\phi(x) - \phi(y) = c(x-y)$ whenever $x, y \in B'$ and $x - y$ belongs to the set $\{jd : -m \leq j \leq m\}$.*

Write P_0 for the arithmetic progression $\{d, 2d, \ldots, md\}$. Choose a translate $P = P_0 + z$ for which $|P \cap B'| \geq \delta m$, and let $H = P \cap B'$. If x and y lie in H then $x - y$ is in $\{jd : -m \leq j \leq m\}$ and so $\phi|_H$ is the restriction of a linear function from $\mathbb{Z}_N$ to itself, by Proposition 24.

We will soon find ourselves dealing with some reasonably large numbers, and it is convenient to have a shorthand notation for them. If $n \gg 1$ is a parameter then we write $C_0(n)$ for any polynomial in n. $C_1(n)$ will mean a function of type $2^{p(n)}$, where p is a polynomial, and finally $C_2(n)$ will be a function of type $2^{2^{p(n)}}$. We will feel at liberty to use these symbols several times, sometimes in the same formula, to denote different functions.

Proposition 23 allows us, provided $|K|$ and $\delta/100m$ satisfy a certain inequality, to conclude

that d may be taken to be a small square number. Take $t = N^{1/4}$ in Proposition 23 and recall that $|K| \leq 16\delta^{-2}$. Then there is $p \leq N^{1/4}$ such that $p^2 \in B(K, \delta/100m)$ provided that

$$\frac{\delta}{100m} \;\geq\; N^{-1/C_1(\delta^{-1})}$$

and that $N \geq C_2(\delta^{-1})$. For N greater than some $C_2(\delta^{-1})$ we can pick $m = N^{1/C_1(\delta^{-1})}$ so that this is satisfied by a colossal margin.

Putting everything together, and recalling that δ is $2^{-1849}\beta^{5821}$, we have the following.

Proposition 25 *Suppose that* (22) *holds and that* $N \geq C_2(\beta^{-1})$. *Then there is a progression* $P \subseteq \mathbb{Z}_N$ *with common difference* p^2, *where* $p \leq N^{1/4}$, *and length at least* $N^{1/C_1(\beta^{-1})}$, *with the following property. There is* $H \subseteq P \cap B$ *with* $|H| \geq 2^{-1849}\beta^{5821}|P|$ *and* $\lambda, \mu \in \mathbb{Z}_N$ *such that* $\phi(s) = \lambda s + \mu$ *for all* $s \in H$.

Let us now make a deduction from Proposition 8.1 of [4]. This is done using the previous proposition and (21).

Proposition 26 *Suppose that* (20) *holds and that* $N \geq C_2(\beta^{-1})$. *Then there is an arithmetic progression* P *with common difference* p^2, *where* $p \leq N^{1/4}$, *and length at least* $N^{1/C_1(\beta^{-1})}$ *and quadratic polynomials* $\psi_0, \psi_1, \dots, \psi_{N-1}$ *such that*

$$\sum_s \left| \sum_{z \in P+s} f(z)\omega^{-\psi_s(z)} \right| \;\geq\; 2^{-1850}\beta^{5822}N|P|.$$

Before stating and proving the next Lemma we need a version of Lemma 8 for fourth powers. Once again we can simply read it from [4].

Lemma 27 *Let* $a \in \mathbb{Z}_N$, *let* $t \leq N$ *and suppose that* $t \geq 2^{2^{512}}$. *Then there is* $p \leq t$ *such that* $|p^4 a| \leq t^{-1/128}N$.

Proposition 28 *Let* $\psi(x) = ax^2 + bx + c$ *be a quadratic polynomial with coefficients in* $\mathbb{Z}_N$. *Let* $L \leq N$, *and let* $P \subseteq \mathbb{Z}_N$ *be an arithmetic progression of length* L *with square common difference. Let* $W \leq L^{2^{-22}}$. *Then there is a partition of* P *into subprogressions* $R_1, \dots, R_m$, *with* R_i *having length between* W *and* $2W$ *and square common difference, such that* $\mathrm{Diam}(\psi(R_i)) \leq L^{-2^{-20}}N$ *for all* i.

Proof We may rescale and assume that $P = \{1, \dots, L\}$ with no loss of generality. For any x_0, λ, d we have

$$\psi(x_0 + \lambda d^2) - \psi(x_0) \;=\; a\lambda^2 d^4 + (2ax_0 + b)\lambda d^2.$$

By Proposition 27 we may choose $d \leq L^{1/8}$ such that $|ad^4| \leq L^{-2^{-10}}N$. Partition $\{1, \ldots, L\}$ into progressions $P_1, \ldots, P_l$ with common difference d^2 and lengths lying between $L^{2^{-13}}$ and $L^{2^{-12}}$. The diameter of ψ on P_i then satisfies

$$\mathrm{Diam}(\psi(P_i)) \;\leq\; L^{-2^{-11}}N + \mathrm{Diam}(\theta_i(P_i)), \tag{23}$$

where θ_i is a linear polynomial depending on i. Fix i, and for ease of notation rescale P_i to $\{1, \ldots, K\}$ (by the square scaling factor d^{-2}) where $K \geq L^{2^{-13}}$. Suppose that $\theta_i(x) = rx + s$ under this rescaling. Clearly

$$\theta_i(x_1 + \mu e^2) - \theta_i(x_1) \;=\; r\mu e^2.$$

By Lemma 8 we may choose $e \leq K^{1/4}$ so that $|re^2| \leq K^{-1/64}N$. Divide P_i into further subprogressions E_j with common difference e^2 and lengths lying between $K^{1/256}$ and $K^{1/128}$. On these subprogressions we will have $\mathrm{Diam}(\theta_i(E_j)) \leq K^{-1/128}N$. Do this for each i, rescale the resultant progressions by the factor d^2, and then perform a further subdivision to satisfy the technical condition on the lengths of the R_i in the statement. Recalling (23) we get the result. □

Call an arithmetic progression $Q \subseteq \mathbb{Z}_N$ *nice* if it does not wrap in $\mathbb{Z}_N$, by which we mean that the length $L(Q)$ and the common difference $d(Q)$ satisfy $L(Q)d(Q) \leq N$. Observe that the progression P found in Proposition 25 is nice, because we constructed it to have small common difference. We observe that any translate of a nice progression is nice, as is any subprogression.

Let us now combine Propositions 26 and 28 in the obvious way.

Proposition 29 *Suppose that* (20) *holds and that* $N \geq C_2(\beta^{-1})$. *Then there is* $W = N^{1/C_1(\beta^{-1})}$, *nice arithmetic progressions* $R_{i,s}$, $1 \leq i \leq m, 1 \leq s \leq N$, *each having length between* W *and* $2W$ *and square common difference, and quadratic polynomials* $\psi_0, \ldots, \psi_{N-1}$ *such that*

$$\sum_i \sum_s \left| \sum_{z \in R_{i,s}} f(z)\omega^{\psi_s(z)} \right| \;\geq\; 2^{-1850}\beta^{5822} \sum_{i,s} |R_{i,s}|. \tag{24}$$

Furthermore every point of $\mathbb{Z}_N$ *lies in the same number of* $R_{i,s}$ *and*

$$\mathrm{Diam}(\psi_s(R_{i,s})) \;\leq\; W^{-2}N. \tag{25}$$

Let us use this proposition immediately. For each i, s choose $z_0 \in R_{i,s}$. Then using (25) and

the fact that $\|f\|_\infty \leq 1$ we have

$$\begin{aligned} & \left| \sum_{z \in R_{i,s}} f(z) \left(\omega^{\psi_s(z)} - \omega^{\psi_s(z_0)} \right) \right| \\ \leq \; & 2\pi W^{-2} \sum_{z \in R_{i,s}} |f(z)| \\ \leq \; & 4\pi W^{-1} \\ \leq \; & 2^{-1851} \beta^{5822} |R_{i,s}| \end{aligned}$$

provided that $W \geq C_0(\beta^{-1})$. This will be the case if N is at least some suitably large $C_2(\beta^{-1})$ (though N might need to be larger than it had to be before). It now follows from (24) that

$$\sum_i \sum_s \left| \sum_{z \in R_{i,s}} f(z) \right| \; \geq \; 2^{-1851} \beta^{5822} \sum_{i,s} |R_{i,s}|.$$

We are now almost home. The fact that every point lies in the same number of $R_{i,s}$ implies that

$$\sum_i \sum_s \sum_{z \in R_{i,s}} f(z) \; = \; 0,$$

and so

$$\sum_i \sum_s \left(\left| \sum_{z \in R_{i,s}} f(z) \right| + \sum_{z \in R_{i,s}} f(z) \right) \; \geq \; 2^{-1851} \beta^{5822} \sum_{i,s} |R_{i,s}|.$$

From this it follows immediately that

$$\sum_{z \in R} f(z) \; \geq \; 2^{-1852} \beta^{5822} |R|$$

for some $R = R_{i,s}$.

The one remaining obstacle is the fact that whilst R is nice, it might still straddle 0 in $\mathbb{Z}_N$. Thus when $\mathbb{Z}_N$ is "unwrapped", R might become two arithmetic progressions R_1 and R_2. Both of these will still have square common difference, so we can be hopeful of extricating ourselves. The following lemma, proved by a simple averaging argument, covers the situation.

Lemma 30 *Suppose that $\|f\|_\infty \leq 1$ and that $\sum_{z \in R} f(z) \geq \eta |R|$, where R is a nice arithmetic progression in $\mathbb{Z}_N$. Suppose that $R = R_1 \cup R_2$, where neither R_1 nor R_2 straddles 0. Then, for some $i \in \{1, 2\}$, R_i has length at least $\eta |R|/3$ and $\sum_{z \in R_i} f(z) \geq \eta |R_i|/3$.*

Finally, we can deduce the following variant of Gowers' Inverse Theorem.

Theorem 31 *Let $f : \mathbb{Z}_N \to \mathbb{C}$ have $\|f\|_\infty \leq 1$, and suppose that there is a function $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ such that*

$$\sum_r |\Delta(f;r)\hat{}(\phi(r))|^2 \geq \beta N^3.$$

Suppose that $N \geq C_2(\beta^{-1})$. Then there is an arithmetic progression $S \subseteq \mathbb{Z}$ with square common difference and length at least $N^{1/C_1(\beta)}$ such that

$$\sum_{x \in S} f(x) \geq 2^{-7677}\beta^{5822}|S|.$$

The sudden change in the powers of two here comes from the fact that we replaced 2β in (20) with β.

By the main result of §4 we have shown that if A contains no triples $(a, a+d, a+2d)$ with $d = x^2 + y^2$ then A has density $\alpha + \Omega(\alpha^{139728})$ on a progression of size $N^{1/C_1(\alpha^{-1})}$ with square common difference. The new set A' must have the same property - it cannot contain an arithmetic progression with sum-of-two-squares common difference. How often can this argument be iterated?

After $O(\alpha^{-139728})$ iterations we will have reached density 1, by which time the length of our subprogression will still be of the form $N' = N^{1/C_1(\alpha^{-1})}$. In order for the iteration step to work we require that $N' \geq C_2(\alpha^{-1})$ at all times. We therefore have a contradiction provided that $N^{1/C_1(\alpha^{-1})} \geq C_2(\alpha^{-1})$, and it is easy to see that this is implied by some bound $\alpha \gg (\log\log N)^{-c}$. This concludes the proof of Theorem 5. □

According to my calculations, $c = 10^{-6}$ is admissible.

References

[1] Bergelson, V. and Leibman, A. *Polynomial Extensions of Van der Waerden's and Szemerédi's Theorems*, J. Amer. Math. Soc. **9** (1996), no 3, 725 – 753.

[2] Furstenburg, H. *Ergodic behaviour of diagonal measures and a theorem of Szemeredi on arithmetic progressions*, J. Analyse. Math. **31** (1977) 204 – 256.

[3] Gowers, W.T. *A New Proof of Szemerédi's Theorem for Progressions of Length 4*, Geom. Funct. Anal. **8** (1998), no. 3, 529–551.

[4] Gowers, W.T. *A New Proof of Szemerédi's Theorem*, to appear in Geom. Funct. Analysis. Available at `http://www.dpmms.cam.ac.uk/~wtg10`.

[5] Green, B. J. *Notes on Sieve Theory: The Selberg Sieve*, available at `http://www.dpmms.cam.ac.uk/~bjg23/expos.html`.

[6] Halberstam, H. and Richert, H. -E. *Sieve Methods*, Academic Press 1974.

[7] Montgomery, H.L. *Ten Lectures on the Interface between Analytic Number Theory and Harmonic Analysis*, CBMS Regional Conference Series in Mathematics 84, AMS 1994.

[8] Nathanson, M.B. *Elementary Methods in Number Theory*, Springer 2000.

[9] Pintz, J., Steiger, W.L. and Szemerédi, E. *On Sets of Natural Numbers whose Difference Set Contains No Squares*, J. London Math. Soc. (2) **37** (1988) 219 – 231.

[10] Roth, K.F. *On Certain Sets of Integers*, J. London Math Soc **28** (1953) 104 – 109.

[11] Ruzsa, I.Z. *Difference Sets Without Squares*, Period. Math. Hungar. **15** (1984), no. 3, 205 – 209.

[12] Sárkőzy, A. *On Difference Sets of Sequences of Integers I*, Acta. Math. Acad. Sci. Hungar **31** (1978), nos. 1 – 2, 125–149.

[13] Sárkőzy, A. *On Difference Sets of Sequences of Integers III*, Acta Math. Acad. Sci. Hungar **31** (1978), nos. 3 – 4 355 – 386.

[14] Schmidt, W.M. *Small Fractional Parts of Polynomials*, Regional Conference Series in Mathematics **32**, AMS 1977.

[15] Srinivasan, S. *On a Result of Sarkozy and Furstenburg*, Nieuw. Arch. Wisk (4) **3** (1985), no. 3, 275 – 280.

[16] Szemerédi, E. *On Sets of Integers Containing No k Elements In Arithmetic Progression*, Acta. Arith **27** (1975), 199 – 245.