GAFA, Geom. funct. anal.
Vol. 11 (2001) 465 – 588
1016-443X/01/030465-124 $ 1.50+0.20/0

© Birkhäuser Verlag, Basel 2001

GAFA Geometric And Functional Analysis

A NEW PROOF OF SZEMERÉDI'S THEOREM

W.T. Gowers

Contents

The research for this article was carried out in part for the Clay Mathematics Institute.

1 Introduction

In 1927 van der Waerden published a celebrated theorem, which states that if the positive integers are partitioned into finitely many classes, then at least one of these classes contains arbitrarily long arithmetic progressions. This is one of the fundamental results of Ramsey theory, and it has been strengthened in many different directions. A more precise statement of the theorem is as follows.

Theorem 1.1. *Let k and r be positive integers. Then there exists a positive integer $M = M(k,r)$ such that, however the set $\{1,2,\dots,M\}$ is partitioned into r subsets, at least one of the subsets contains an arithmetic progression of length k.*

It is natural to wonder how quickly the least such M grows as a function of k and r, but this has turned out to be a surprisingly difficult question. The original proof of van der Waerden bounds M above by an Ackermann-type function in k, even when $r = 2$, and it was a major advance when Shelah, in 1987, gave the first primitive recursive upper bound (with a beautifully transparent proof). His bound can be described as follows. Define a tower function T inductively by letting $T(1) = 2$ and $T(k) = 2^{T(k-1)}$ for $k > 1$. Then define a function W by $W(1) = 2$ and $W(k) = T(W(k-1))$ for $k > 1$. Shelah obtained a bound of the form $M(k,2) \leqslant W(Ck)$ (with C an absolute constant). Although this was a huge improvement on the previous bound, it still left an enormous gap, as the best known lower bound was, and still is, exponential in k.

A strengthening of a completely different kind was conjectured by Erdős and Turán in 1936. They realised that it ought to be possible to find arithmetic progressions of length k in any sufficiently dense set of integers, which would show that the colouring in van der Waerden's theorem was, in a sense, a distraction. The translation-invariance of the notion of an arithmetic progression rules out simple counterexamples to this stronger statement. (One can contrast this situation with a theorem of Schur which states that in any finite colouring of $\mathbb{N}$ there are solutions of the equation $x+y=z$ with x, y and z all of the same colour. However, the set of all odd integers has density $1/2$ and contains no solutions.) The conjecture was proved by Szemerédi in 1974. Szemerédi's theorem, which we now state precisely, is one of the milestones of combinatorics.

Theorem 1.2. *Let k be a positive integer and let $\delta > 0$. There exists a positive integer $N = N(k,\delta)$ such that every subset of the set $\{1,2,\dots,N\}$*

of size at least δN contains an arithmetic progression of length k.

It is very simple to see that this result strengthens van der Waerden's theorem, and that $M(k,r)$ can be chosen to be $N(k,r^{-1})$.

A second proof of Szemerédi's theorem was given by Furstenberg in 1977, using ergodic theory, which provides an extremely useful conceptual framework for discussing the result. This proof was also a major breakthrough, partly because of the difficulty of Szemerédi's original proof, and partly because Furstenberg's techniques have since been extended to prove many natural generalizations of the theorem which do not seem to follow from Szemerédi's approach. These include a density version of the Hales-Jewett theorem [FK] and a "polynomial Szemerédi theorem" [BL].

Why then, if there are already two proofs of Szemerédi's theorem, should one wish to find a third? There are several related reasons.

First of all, it is likely that Erdős and Turán, when they made their original conjecture, hoped that it would turn out to be the "real" theorem underlying van der Waerden's theorem, and perhaps for that reason have an easier proof. If they did, then their hope has not been fulfilled, as all known proofs are long and complicated. Szemerédi's original paper runs to 47 pages, full of intricate combinatorial arguments, and it takes a few seconds even to check that the diagram near the beginning of the dependences between the various lemmas really does indicate a valid proof. Furstenberg's proof is considerably simpler (especially as presented in [FKO]), but requires a certain initial investment in learning the necessary definitions from ergodic theory, and is still significantly harder than the proof of van der Waerden's theorem. (On the positive side, some of the ideas of Szemerédi's proof, most notably the so-called regularity lemma, have turned out to be extremely useful in many other contexts, and, as mentioned above, Furstenberg's proof has been the starting point of a great deal of further research.)

Second, Erdős and Turán gave as the main motivation for their conjecture the likelihood that in order to prove it one would be forced not to use the sorts of arguments that led to such weak bounds for van der Waerden's theorem, and would therefore obtain far better estimates. However, this hope was not fulfilled by Szemerédi's proof because he used van der Waerden's theorem in his argument. He also used the regularity lemma just mentioned, which makes a tower-type contribution to the size of the bound from any argument that uses it. (See [G1] for a proof that this is necessary.) Furstenberg's proof gives no bound, even in principle, as it uses the axiom

of choice. Moreover, although van der Waerden's theorem is not directly applied, it is likely that any attempt to make the argument quantitative would lead to rapidly growing functions for similar reasons.

Third, there is a possibility left open by the first result in the direction of Szemerédi's theorem, the assertion for progressions of length three, which was proved by Roth [R1]. Roth gave a beautiful argument using exponential-sum estimates, but his approach seemed not to generalize. Indeed, progress was made on the problem only when Szemerédi found a different, more combinatorial argument for progressions of length three which was more susceptible to generalization. However, it is highly desirable to find an exponential-sums argument for the general case, because all the best bounds for similar problems have come from these techniques rather than purely combinatorial ones [Sz3], [H-B], [Bou]. (Although Roth used ideas from Szemerédi's proof for progressions of length four [S1] and combined them with analytic techniques to give a second proof for that case [R2], the argument is not really a direct generalization of his earlier proof, and relies on van der Waerden's theorem.)

Fourth, there are certain important conjectures related to Szemerédi's theorem, and the existing arguments get nowhere near to them. The most famous is Erdős's conjecture that every set X of positive integers such that $\sum_{x\in X} x^{-1}$ diverges contains arbitrarily long arithmetic progressions. Since the set of primes has this property, a positive solution to the conjecture would answer an old question in number theory using no more about the primes than the fact that they are reasonably dense. Even if the conjecture turns out to be too optimistic, there is a resemblance between Roth's proof and the result of van der Corput (adapting the proof of Vinogradov's three-primes theorem) that the primes contain infinitely many arithmetic progressions of length three, which suggests that generalizing Roth's proof to longer progressions could at least lead to a number-theoretic proof that the primes contain arbitrarily long arithmetic progressions.

In this paper, we show that Roth's argument *can* be generalized, and that this does indeed result in a significant improvement to the bounds, even for van der Waerden's theorem. Our main result (restated in equivalent form later as Theorem 18.2) is the following.

Theorem 1.3. *For every positive integer k there is a constant $c = c(k) > 0$ such that every subset of $\{1, 2, \dots, N\}$ of size at least $N(\log\log N)^{-c}$ contains an arithmetic progression of length k. Moreover, c can be taken to be $2^{-2^{k+9}}$.*

This immediately implies an estimate for $N(k, \delta)$ which is doubly exponential in δ^{-1} and quintuply exponential in k.

There are, however, some serious difficulties in carrying out the generalization, as we shall demonstrate with examples later in the paper. This perhaps explains why the generalization has not been discovered already. Very roughly, our strategy is to reduce the problem to what is known as an inverse problem in additive number theory (deducing facts about the structure of a set of numbers from properties of its set of sums or differences). We then apply a variant of a famous inverse result due to Freiman [F1,2]. Freiman's proof of his theorem is very complicated, though it has recently been considerably tidied up by Bilu [Bi]. A very much simpler proof of Freiman's theorem was recently given by Ruzsa [Ru1,2], and to him we owe a huge mathematical debt. His methods have inspired many parts of this paper, including several arguments where his results are not quoted directly.

It has to be admitted that this paper is actually longer than those of Szemerédi and Furstenberg, and less self-contained. This is partly because my overriding priority when writing it has been to make the basic ideas as clear as possible, even if this adds several pages. Many results are proved first in a special case and later in full generality. This is intended to make it as easy as possible to read about progressions of length four and five, which involve most of the interesting ideas but by no means all of the technicalities. (The special case of progressions of length four was covered in an earlier paper [G2] but it is treated here as well, and a better bound, claimed in the earlier paper, is here proved in full.) Sections 4 and 11 are devoted to examples showing that certain simpler arguments do not work. They are therefore not logically necessary. However, the whole of the rest of the paper is, in a sense, a response to those examples. Another priority has been to make the sections as independent as possible. Where it is essential that one section depends on another, we have tried to make it depend on a single clearly stated result, in the hope that readers will if they wish be able to understand the broad outline of the proof without following the details.

Despite these efforts, the quickest way to understand a proof of Szemerédi's theorem is probably still to read the paper of Furstenberg, Katznelson and Ornstein [FKO] mentioned earlier. However, the proof in this paper gives quantitative information, and I hope that at least some mathematicians, particularly those with a background in additive number theory, will find the approach a congenial one.

2 Uniform Sets and Roth's Theorem

It is not hard to prove that a random subset of the set $\{1, 2, \dots, N\}$ of cardinality δN contains, with high probability, roughly the expected number of arithmetic progressions of length k, that is, δ^k times the number of such progressions in the whole of $\{1, 2, \dots, N\}$. A natural idea is therefore to try to show that random sets contain the fewest progressions of length k, which would then imply Szemerédi's theorem. In view of many other examples in combinatorics where random sets are extremal, this is a plausible statement, but unfortunately it is false. Indeed, if random sets were the worst, then the value of δ needed to ensure an arithmetic progression of length three would be of order of magnitude $N^{-2/3}$, whereas in fact it is known to be at least $\exp(-c(\log N)^{1/2})$ for some absolute constant $c > 0$ [Be]. (The random argument suggested above is to choose δ so that the expected number of arithmetic progressions is less than one. Using a standard trick in probabilistic combinatorics, we can instead ask for the expected number to be at most $\delta N/2$ and then delete one point from each one. This slightly better argument lifts the density significantly, but still only to $cN^{-1/2}$.)

Despite this, it is tempting to try to exploit the fact that random sets contain long arithmetic progressions. Such a proof could be organized as follows.

(1) Define an appropriate notion of pseudorandomness.
(2) Prove that every pseudorandom subset of $\{1, 2, \dots, N\}$ contains roughly the number of arithmetic progressions of length k that you would expect.
(3) Prove that if $A \subset \{1, 2, \dots, N\}$ has size δN and is *not* pseudorandom, then there exists an arithmetic progression $P \subset \{1, 2, \dots, N\}$ with length tending to infinity with N, such that $|A \cap P| \geqslant (\delta + \epsilon)|P|$, for some $\epsilon > 0$ that depends on δ (and k) only.

If these three steps can be carried out, then a simple iteration proves Szemerédi's theorem. As we shall see, this is exactly the scheme of Roth's proof for progressions of length three.

First, we must introduce some notation. Throughout the paper we shall be considering subsets of $\mathbb{Z}_N$ rather than subsets of $\{1, 2, \dots, N\}$. It will be convenient (although not essential) to take N to be a prime number. We shall write ω for the number $\exp(2\pi i/N)$. Given a function $f : \mathbb{Z}_N \to \mathbb{C}$ and $r \in \mathbb{Z}_N$ we set

$$\hat{f}(r) = \sum_{s \in \mathbb{Z}_N} f(s)\omega^{-rs}\,.$$

The function $\hat{f}$ is the discrete Fourier transform of f. (In most papers in analytic number theory, the above exponential sum is written $\sum_{s=1}^{N} e(-rs/N)$, or possibly $\sum_{s=1}^{N} e_N(-rs)$.) Let us write $f * g$ for the function

$$f * g(s) = \sum_{t \in \mathbb{Z}_N} f(t)\overline{g(t-s)}\,.$$

(This is not standard notation, but we shall have no use for the convolution $\sum f(t)g(s-t)$ in this paper, so it is very convenient.) From now on, all sums will be over $\mathbb{Z}_N$ unless it is specified otherwise. We shall use the following basic identities over and over again in the paper.

$$(f * g)^{\wedge}(r) = \hat{f}(r)\overline{\hat{g}(r)}\,, \tag{1}$$

$$\sum_r \hat{f}(r)\overline{\hat{g}(r)} = N \sum_s f(s)\overline{g(s)}\,, \tag{2}$$

$$\sum_r |\hat{f}(r)|^2 = N \sum_s |f(s)|^2\,, \tag{3}$$

$$f(s) = N^{-1} \sum_r \hat{f}(r)\omega^{rs}\,. \tag{4}$$

Of these, the first tells us that convolutions transform to pointwise products, the second and third are Parseval's identities and the last is the inversion formula. To check them directly, note that

$$\begin{aligned}(f * g)(r) &= \sum_s (f * g)(s)\omega^{-rs} \\ &= \sum_{s,t} f(t)\overline{g(t-s)}\omega^{-rt}\omega^{r(t-s)} \\ &= \sum_{t,u} f(t)\omega^{-rt}\overline{g(u)\omega^{-ru}} \\ &= \hat{f}(r)\overline{\hat{g}(r)}\,,\end{aligned}$$

which proves (1). We may deduce (2), since

$$\sum_r \hat{f}(r)\overline{\hat{g}(r)} = \sum_r \sum_s f * g(s)\omega^{-rs} = N f * g(0) = N \sum_s f(s)\overline{g(s)}\,,$$

where for the second equality we used the fact that $\sum_s \omega^{-rs}$ is N if $r = 0$ and zero otherwise. Identity (3) is a special case of (2). Noting that the function $r \mapsto \omega^{-rs}$ is the Fourier transform of the characteristic function of the singleton $\{s\}$, we can deduce (4) from (2) as well (though it is perhaps more natural just to expand the right-hand side and give a direct proof).

There is one further identity, sufficiently important to be worth stating as a lemma.

LEMMA 2.1. *Let f and g be functions from $\mathbb{Z}_N$ to $\mathbb{C}$. Then*

$$\sum_r |\hat{f}(r)|^2|\overline{\hat{g}(r)}|^2 = N\sum_t \Big|\sum_s f(s)\overline{g(s-t)}\Big|^2 . \tag{5}$$

Proof. By identities (1) and (2),

$$\begin{aligned}\sum_r |\hat{f}(r)|^2|\overline{\hat{g}(r)}|^2 &= \sum_r \big|(f*g)^\wedge(r)\big|^2\\ &= N\sum_t \big|f*g(t)\big|^2\\ &= N\sum_{t,s,u} f(s)\overline{g(s-t)f(u)}g(u-t)\\ &= N\sum_t \Big|\sum_s f(s)\overline{g(s-t)}\Big|^2\end{aligned}$$

as required. □

Setting $f=g$ and expanding the right-hand side of (5), one obtains another identity which shows that sums of fourth powers of Fourier coefficients have an interesting interpretation.

$$\sum_r |\hat{f}(r)|^4 = N\sum_{a-b=c-d} f(a)\overline{f(b)f(c)}f(d)\,. \tag{6}$$

It is of course easy to check this identity directly.

Nearly all the functions in this paper will take values with modulus at most one. In such a case, one can think of Lemma 2.1 as saying that if f has a large inner product with a large number of rotations of g, then f and g must have large Fourier coefficients in common, where large means of size proportional to N. We shall be particularly interested in the Fourier coefficients of characteristic functions of sets $A\subset\mathbb{Z}_N$ of cardinality δN, which we shall denote by the same letter as the set itself. Notice that identity (6), when applied to (the characteristic function of) a set A, tells us that the sum $\sum_r |\hat{A}(r)|^4$ is N times the number of quadruples $(a,b,c,d)\in A^4$ such that $a-b=c-d$.

For technical reasons it is also useful to consider functions of mean zero. Given a set A of cardinality δN, let us define the *balanced* function of A to be $f_A:\mathbb{Z}^N\to[-1,1]$ where

$$f_A(s)=\begin{cases}1-\delta & s\in A\\ -\delta & s\notin A\,.\end{cases}$$

This is the characteristic function of A minus the constant function $\delta\mathbf{1}$. Note that $\sum_{s\in\mathbb{Z}_N} f_A(s)=\hat{f}_A(0)=0$ and that $\hat{f}_A(r)=\hat{A}(r)$ for $r\neq 0$.

We are now in a position to define a useful notion of pseudorandomness. The next lemma (which is not new) gives several equivalent definitions involving constants c_i. When we say that one property involving c_i implies another involving c_j, we mean that if the first holds, then so does the second for a constant c_j that tends to zero as c_i tends to zero. (Thus, if one moves from one property to another and then back again, one does not necessarily recover the original constant.) From the point of view of the eventual bounds obtained, it is important that the dependence is no worse than a fixed power. This is always true below.

In this paper we shall use the letter D to denote the closed unit disc in $\mathbb{C}$ (unless it obviously means something else).

LEMMA 2.2. *Let f be a function from $\mathbb{Z}_N$ to D. The following are equivalent.*

(i) $\sum_k \bigl|\sum_s f(s)\overline{f(s-k)}\,\bigr|^2 \leqslant c_1 N^3$.
(ii) $\sum_{a-b=c-d} f(a)\overline{f(b)}f(c)f(d) \leqslant c_1 N^3$.
(iii) $\sum_r |\hat{f}(r)|^4 \leqslant c_1 N^4$.
(iv) $\max_r |\hat{f}(r)| \leqslant c_2 N$.
(v) $\sum_k \bigl|\sum_s f(s)\overline{g(s-k)}\,\bigr|^2 \leqslant c_3 N^2 \|g\|_2^2$ *for every function* $g : \mathbb{Z}_N \to \mathbb{C}$.

Proof. The equivalence of (i) and (ii) comes from expanding the left-hand side of (i), and the equivalence of (i) and (iii) follows from identity (6) above. It is obvious that (iii) implies (iv) if $c_2 \geqslant c_1^{1/4}$. Since

$$\sum_r |\hat{f}(r)|^4 \leqslant \max_r |\hat{f}(r)|^2 \sum_r |\hat{f}(r)|^2 \leqslant N^2 \max_r |\hat{f}(r)|^2 ,$$

we find that (iv) implies (iii) if $c_1 \geqslant c_2^2$. It is obvious that (v) implies (i) if $c_1 \geqslant c_3$. By Lemma 2.1, the left-hand side of (v) is

$$N^{-1} \sum_r |\hat{f}(r)|^2 |\hat{g}(r)|^2 \leqslant N^{-1} \Bigl(\sum_r |\hat{f}(r)|^4\Bigr)^{1/2} \Bigl(\sum_r |\hat{g}(r)|^4\Bigr)^{1/2}$$

by the Cauchy-Schwarz inequality. Using the additional inequality

$$\Bigl(\sum_r |\hat{g}(r)|^4\Bigr)^{1/2} \leqslant \sum_r |\hat{g}(r)|^2 ,$$

we see that (iii) implies (v) if $c_3 \geqslant c_1^{1/2}$. □

A function $f : \mathbb{Z}_N \to D$ satisfying condition (i) above, with $c_1 = \alpha$, will be called *α-uniform*. If f is the balanced function f_A of some set $A \subset \mathbb{Z}_N$, then we shall also say that A is α-uniform. If $A \subset \mathbb{Z}_N$ is an α-uniform set of cardinality δN, and f is its balanced function, then

$$\sum_r |\hat{A}(r)|^4 = |A|^4 + \sum_r |\hat{f}(r)|^4 \leqslant |A|^4 + \alpha N^4 .$$

We noted earlier that $\sum_r |\hat{A}(r)|^4$ is N times the number of quadruples $(a,b,c,d) \in A^4$ such that $a-b=c-d$. If A were a random set of size δN, then we would expect about $\delta^4 N^3 = N^{-1}|A|^4$ such quadruples (which from the above is clearly a lower bound). Therefore, the number α is measuring how close A is to being random in this particular sense. Notice that quadruples (a,b,c,d) with $a-b=c-d$ are the same as quadruples of the form $(x, x+s, x+t, x+s+t)$.

We remark that our definition of an α-uniform set coincides with the definition of *quasirandom* subsets of $\mathbb{Z}_N$, due to Chung and Graham. They prove that several formulations of the definition (including those of this paper) are equivalent. They do not mention the connection with Roth's theorem, which we shall now explain. We need a very standard lemma, which we prove in slightly greater generality than is immediately necessary, so that it can be used again later. Let us define the *diameter* of a subset $X \subset \mathbb{Z}_N$ to be the smallest integer s such that $X \subset \{n, n+1, \dots, n+s\}$ for some $n \in \mathbb{Z}_N$.

LEMMA 2.3. *Let r, s and N be positive integers with $r, s \leqslant N$ and $rs \geqslant N$, and let $\phi : \{0,1,\dots,r-1\} \to \mathbb{Z}_N$ be linear (i.e., of the form $\phi(x) = ax+b$). Then the set $\{0,1,\dots,r-1\}$ can be partitioned into arithmetic progressions $P_1, \dots, P_M$ such that for each j the diameter of $\phi(P_j)$ is at most s and the length of P_j lies between $(rs/4N)^{1/2}$ and $(rs/N)^{1/2}$.*

Proof. Let $t = \lceil (rN/4s)^{1/2} \rceil$. Of the numbers $\phi(0), \phi(1), \dots, \phi(t)$, at least two must be within N/t. Therefore, by the linearity of ϕ, we can find a non-zero $u \leqslant t$ such that $|\phi(u) - \phi(0)| \leqslant N/t$. Split $\{0,1,\dots,r-1\}$ into congruence classes mod u. Each congruence class is an arithmetic progression of cardinality either $\lfloor r/u \rfloor$ or $\lceil r/u \rceil$. If P is any set of at most st/N consecutive elements of a congruence class, then $\operatorname{diam} \phi(P) \leqslant s$. It is easy to check first that $st/N \leqslant r/3t \leqslant (1/2)\lfloor r/u \rfloor$, next that this implies that the congruence classes can be partitioned into sets P_j of consecutive elements with every P_j of cardinality between $\lceil st/2N \rceil$ and $\lfloor st/N \rfloor$, and finally that this proves the lemma. □

COROLLARY 2.4. *Let f be a function from the set $\{0,1,\dots,r-1\}$ to the closed unit disc in $\mathbb{C}$, let $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ be linear and let $\alpha > 0$. If*

$$\Big| \sum_{x=0}^{r-1} f(x) \omega^{-\phi(s)} \Big| \geqslant \alpha r \,,$$

then there is a partition of $\{0,1,\dots,r-1\}$ into $m \leqslant (8\pi r/\alpha)^{1/2}$ arithmetic

progressions $P_1, \dots, P_m$ such that

$$\sum_{j=1}^{m} \Big| \sum_{x \in P_j} f(x) \Big| \geqslant (\alpha/2) r$$

and such that the lengths of the P_j all lie between $(\alpha r/\pi)^{\frac{1}{2}}/4$ and $(\alpha r/\pi)^{\frac{1}{2}}/2$.

Proof. Let $s \leqslant \alpha N/4\pi$ and let $m = (16\pi r/\alpha)^{1/2}$. By Lemma 2.3 we can find a partition of $\{0, 1, \dots, r-1\}$ into arithmetic progressions $P_1, \dots, P_m$ such that the diameter of $\phi(P_j)$ is at most s for every j and the length of each P_j lies between r/m and $2r/m$. By the triangle inequality,

$$\sum_{j=1}^{m} \Big| \sum_{x \in P_j} f(x) \omega^{-\phi(x)} \Big| \geqslant \alpha r \,.$$

Let $x_j \in P_j$. The estimate on the diameter of $\phi(P_j)$ implies that $|\omega^{-\phi(x)} - \omega^{-\phi(x_j)}|$ is at most $\alpha/2$ for every $x \in P_j$. Therefore

$$\begin{aligned} \sum_{j=1}^{m} \Big| \sum_{x \in P_j} f(x) \Big| &= \sum_{j=1}^{m} \Big| \sum_{x \in P_j} f(x) \omega^{-\phi(x_j)} \Big| \\ &\geqslant \sum_{j=1}^{m} \Big| \sum_{x \in P_j} f(x) \omega^{-\phi(x)} \Big| - \sum_{j=1}^{m} (\alpha/2) |P_j| \\ &\geqslant \alpha r/2 \end{aligned}$$

as claimed. □

COROLLARY 2.5. *Let $A \subset \mathbb{Z}_N$ and suppose that $|\hat{A}(r)| \geqslant \alpha N$ for some $r \neq 0$. Then there exists an arithmetic progression $P \subset \{0, 1, \dots, N-1\}$ of length at least $(\alpha^3 N/128\pi)^{1/2}$ such that $|A \cap P| \geqslant (\delta + \alpha/8)|P|$.*

Proof. Define $\phi(x) = rx$ and let f be the balanced function of A (regarded as a function on $\{0, 1, \dots, N-1\}$). By Corollary 2.4 we can partition the set $\{0, 1, \dots, N-1\}$ into $m \leqslant (16\pi N/\alpha)^{1/2}$ arithmetic progressions $P_1, \dots, P_m$ of lengths between N/m and $2N/m$ such that

$$\sum_{j=1}^{m} \Big| \sum_{x \in P_j} f(x) \Big| \geqslant \alpha N/2 \,.$$

Since $\sum_{x \in P_j} f(x)$ is real for all j, and since $\sum_{j=1}^{m} \sum_{x \in P_j} f(x) = 0$, if we define J to be the set of j with $\sum_{x \in P_j} f(x) \geqslant 0$, we have

$$\sum_{j \in J} \sum_{x \in P_j} f(x) \geqslant \alpha N/4 \,.$$

Therefore, we can find j such that $\sum_{x\in P_j} f(x)\geqslant\alpha N/4m$. But $|P_j|\leqslant 2N/m$, so $\sum_{x\in P_j} f(x)\geqslant\alpha|P_j|/8$, which implies that $|A\cap P_j|\geqslant(\delta+\alpha/8)|P_j|$. □

We can now give Roth's proof of his theorem on arithmetic progressions of length three.

Theorem 2.6. *Let $\delta>0$, let $N\geqslant\exp\exp(C\delta^{-1})$ (where C is an absolute constant) and let $A\subset\{1,2,\dots,N\}$ be a set of size at least δN. Then A contains an arithmetic progression of length three.*

Proof. Since we are passing to smaller progressions and iterating, we cannot simply assume that N is prime, so we shall begin by dealing with this small technicality. Let N_0 be a positive integer and let A_0 be a subset of $\{1,2,\dots,N_0\}$ of size at least $\delta_0 N_0$.

By Bertrand's postulate (which is elementary – it would be a pity to use the full strength of the prime number theorem in a proof of Roth's theorem) there is a prime p between $N_0/3$ and $2N_0/3$. Write q for N_0-p. If $|A_0\cap\{1,2,\dots,p\}|\leqslant\delta_0(1-\delta_0/160)p$, then we know that

$$\begin{aligned}\big|A_0\cap\{p+1,\dots,N_0\}\big|&\geqslant\delta_0\big(N_0-(1-\delta_0/160)p\big)=\delta_0(q+\delta_0 p/160)\\&\geqslant\delta_0(1+\delta_0/320)q\,.\end{aligned}$$

Let us call this situation case 0.

If case 0 does not hold, then let N be the prime p obtained above, let $A=A_0\cap\{1,\dots,N\}$ and let $\delta=\delta_0(1-\delta_0/160)$. Let $B=A\cap[N/3,2N/3)$. If $|B|\leqslant\delta N/5$, then either $A\cap[0,N/3)$ or $A\cap[2N/3,N)$ has cardinality at least $2\delta N/5=(6\delta/5)(N/3)$. This situation we shall call case 1.

Next, let $\alpha=\delta^2/10$ and suppose that $|\hat A(r)|>\alpha N$ for some non-zero r. In this case, by Corollary 2.5 there is an arithmetic progression P of cardinality at least $(\alpha^3N/128\pi)^{1/2}$ such that $|A\cap P|\geqslant(\delta+\delta^2/80)|P|$. This situation will be case 2.

If case 2 does not hold, then $|\hat A(r)|\leqslant\alpha N$ for every non-zero r, which says that A satisfies condition (iv) of Lemma 2.2. The number of triples $(x,y,z)\in A\times B^2$ such that $x+z=2y$ is then

$$\begin{aligned}N^{-1}\sum_{x\in A}\sum_{y\in B}\sum_{z\in B}\sum_r\omega^{r(2y-x-z)}&=N^{-1}\sum_r\hat A(r)\hat B(-2r)\hat B(r)\\&\geqslant N^{-1}|A||B|^2-N^{-1}\max_{r\neq0}|\hat A(r)|\Big(\sum_{r\neq0}|\hat B(-2r)|^2\Big)^{1/2}\Big(\sum_{r\neq0}|\hat B(r)|^2\Big)^{1/2}\\&\geqslant\delta|B|^2-\alpha|B|N\,.\end{aligned}$$

If in addition case 1 does not hold, then this quantity is minimized when $|B|=\delta N/5$, and the minimum value is $\delta^3N^2/50$, implying the existence of

at least this number of triples $(x, y, z) \in A \times B^2$ in arithmetic progression mod N. Since B lives in the middle third, these are genuine progressions in $\{1, 2, \ldots, N\}$, and since there are only N degenerate progressions (i.e., with difference zero) we can conclude that A contains an arithmetic progression of length three as long as $N \geqslant 50\delta^{-3}$. This we shall call case 3.

To summarize, if case 3 holds and $N \geqslant 50\delta^{-3}$, then A contains an arithmetic progression of length three. In case 2, we can find a subprogression P of $\{1, \ldots, N\}$ of cardinality at least $(\alpha^3 N/128\pi)^{1/2}$ such that $|A \cap P| \geqslant \delta(1 + \delta/80)|P|$. Since $\{1, \ldots, N\}$ is a subprogression of $\{1, \ldots, N_0\}$, $A = A_0 \cap \{1, \ldots, N\}$ and one can easily check that $\delta(1+\delta/80) \geqslant \delta_0(1 + \delta_0/320)$, we may conclude that in case 2 there is a subprogression P of $\{1, \ldots, N_0\}$ of cardinality at least $(\alpha^3 N_0/384\pi)^{1/2}$ such that $|A_0 \cap P| \geqslant \delta_0(1 + \delta_0/320)|P|$. As for cases 0 and 1, it is easy to see that the same conclusion also holds, and indeed a much stronger one as P has a length which is linear in N_0.

This gives us the basis for an iteration argument. If A_0 does not contain an arithmetic progression of length three, then we drop down to a progression P where the density of A is larger, and repeat. If the density at step m of the iteration is δ_m, then at each subsequent iteration the density increases by at least $\delta_m^2/320$. It follows that the density reaches $2\delta_m$ after at most $320\delta_m^{-1}$ further steps. It follows that the total number of steps cannot be more than $320(\delta^{-1} + (2\delta)^{-1} + (4\delta)^{-1} + \ldots) = 640\delta^{-1}$. At each step, the size of the progression in which A lives is around the square root of what it was at the previous step. The result now follows from a simple calculation (left to the reader). □

3 Higher-degree Uniformity

There seems to be no obvious way of using α-uniformity to obtain progressions of length greater than three. (Of course, the truth of Szemerédi's theorem makes it hard to formalize this statement, but in the next section we show that α-uniformity does not give strong information about the number of arithmetic progressions of length k if $k > 3$.) The aim of this section is to define a notion of pseudo-randomness which is more suitable for the purpose. The next definition is once again presented as a series of approximately equivalent statements. In order to simplify the presentation for the case of progressions of length four, we shall prove two lemmas, even though the second implies the first. Given a function $f : \mathbb{Z}_N \to \mathbb{Z}_N$, we shall define,

for any k, the *difference function* $\Delta(f;k)$ by $\Delta(f;k)(s) = f(s)\overline{f(s-k)}$. The reason for the terminology is that if, as will often be the case, $f(s) = \omega^{\phi(s)}$ for some function $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$, then $\Delta(f;k)(s) = \omega^{\phi(k)-\phi(s-k)}$.

Now let us define iterated difference functions in two different ways as follows. The first is inductive, setting $\Delta(f;a_1,\dots,a_d)(s)$ to be $\Delta(\Delta(f;a_1,\dots,a_{d-1});a_d)(s)$. The second makes explicit the result of the inductive process. Let C stand for the map from $\mathbb{C}^N$ to $\mathbb{C}^N$ which takes a function to its pointwise complex conjugate. Given a function $f : \mathbb{Z}_N \to \mathbb{C}$, we define

$$\Delta(f;a_1,\dots,a_d)(s) = \prod_{\epsilon_1,\dots,\epsilon_d} (C^{\epsilon_1+\dots+\epsilon_d}f)\Big(s - \sum_{i=1}^d a_i\epsilon_i\Big)$$

where the product is over all sequences $\epsilon_1,\dots,\epsilon_d$ with $\epsilon_i \in \{0,1\}$. When $d = 3$, for example, this definition becomes

$$\begin{aligned}\Delta(f;a,b,c)(s) = f(s)&\overline{f(s-a)f(s-b)f(s-c)}\\ &\times f(s-a-b)f(s-a-c)f(s-b-c)\overline{f(s-a-b-c)}\,.\end{aligned}$$

We now define a function f from $\mathbb{Z}_N$ to the closed unit disc $D \subset \mathbb{C}$ to be *α-uniform of degree d* if

$$\sum_{a_1,\dots,a_d}\Big|\sum_s \Delta(f;a_1,\dots,a_d)(s)\Big|^2 \leqslant \alpha N^{d+2}\,.$$

When d equals two or three, we say that f is quadratically or cubically α-uniform respectively. As with the definition of α-uniformity (which is the same as α-uniformity of degree one) this definition has several useful reformulations.

LEMMA 3.1. *Let f be a function from $\mathbb{Z}_N$ to D. The following are equivalent.*

(i) *f is c_1-uniform of degree d.*
(ii) *$\sum_s \sum_{a_1,\dots,a_{d+1}} \Delta(f;a_1,\dots,a_{d+1})(s) \leqslant c_1N^{d+2}$.*
(iii) *There is a function $\alpha : \mathbb{Z}_N^{d-1} \to [0,1]$ such that $\sum_{a_1,\dots,a_{d-1}} \alpha(a_1,\dots,a_{d-1}) \leqslant c_1N^{d-1}$ and $\Delta(f;a_1,\dots,a_{d-1})$ is $\alpha(a_1,\dots,a_{d-1})$-uniform for every $(a_1,\dots,a_{d-1})$.*
(iv) *There is a function $\alpha : \mathbb{Z}_N \to [0,1]$ such that $\sum_r \alpha(r) = c_1N$ and $\Delta(f;r)$ is $\alpha(r)$-uniform of degree $d-1$ for every r.*
(v) *$\sum_{a_1,\dots,a_{d-1}} \sum_r \big|\Delta(f;a_1,\dots,a_{d-1})^\wedge(r)\big|^4 \leqslant c_1N^{d+3}$.*
(vi) *For all but c_2N^{d-1} choices of $(a_1,\dots,a_{d-1})$ the function $\Delta(f;a_1,\dots,a_{d-1})$ is c_2-uniform.*

(vii) There are at most c_3N^{d-1} values of $(a_1,\dots,a_{d-1})$ for which there exists some $r\in\mathbb{Z}_N$ with $\bigl|\Delta(f;a_1,\dots,a_{d-1})^\wedge(r)\bigr|\geqslant c_3N$.

Proof. The equivalence of (i) and (ii) is easy, as the left-hand sides of the relevant expressions are equal. It is also obvious that (ii) and (iii) are equivalent. A very simple inductive argument shows that (ii) is equivalent to (iv). The equivalence of (i) and (v) follows, as in the proof of the equivalence of (i) and (iii) in Lemma 2.1, by expanding the left-hand side of (v). Alternatively, it can be deduced from Lemma 2.1 by applying that equivalence to each function $\Delta(f;a_1,\dots,a_{d-1})$ and adding.

Averaging arguments show that (iii) implies (vi) as long as $c_1\leqslant c_2^2$, and that (vi) implies (iii) as long as $c_1\geqslant 2c_2$. Finally, the equivalence of (i) and (ii) in Lemma 2.1 shows that in this lemma (vi) implies (vii) if $c_3\geqslant c_2^{1/4}$ and (vii) implies (vi) if $c_2\geqslant c_3$. □

Notice that properties (i) and (ii) above make sense even when $d=0$. Therefore, we shall define a function $f:\mathbb{Z}_N\to D$ to be α-uniform of degree zero if $\bigl|\sum_s f(s)\bigr|^2\leqslant\alpha N^2$. Property (iv) now makes sense when $d=1$. This definition will allow us to begin an inductive argument at an earlier and thus easier place.

The next result is the main one of this section. Although it will not be applied directly, it easily implies the results that are needed for later.

Theorem 3.2. *Let $k\geqslant 2$ and let $f_1,\dots,f_k$ be functions from $\mathbb{Z}_N$ to D such that f_k is α-uniform of degree $k-2$. Then*

$$\Bigl|\sum_r\sum_s f_1(s)f_2(s-r)\dots f_k(s-(k-1)r)\Bigr|\leqslant\alpha^{1/2^{k-1}}N^2\,.$$

Proof. When $k=2$, we know that

$$\Bigl|\sum_r\sum_s f_1(s)f_2(s-r)\Bigr|=\Bigl|\Bigl(\sum_s f_1(s)\Bigr)\Bigl(\sum_t f_2(t)\Bigr)\Bigr|\leqslant\alpha^{1/2}N^2\,,$$

since $\bigl|\sum_s f_1(s)\bigr|\leqslant N$ and $\bigl|\sum_t f_2(t)\bigr|\leqslant\alpha^{1/2}N$.

When $k>2$, assume the result for $k-1$, let f_k be α-uniform of degree $k-2$ and let $\alpha:\mathbb{Z}_N\to[0,1]$ be a function with the property that $\Delta(f_k;r)$ is $\alpha(r)$-uniform of degree $k-3$ for every $r\in\mathbb{Z}_N$. Then

$$\begin{aligned}
&\Big|\sum_r\sum_s f_1(s)\dots f_k\big(s-(k-1)r\big)\Big|^2\\
&\leqslant N\sum_s\Big|\sum_r f_1(s)f_2(s-r)\dots f_k\big(s-(k-1)r\big)\Big|^2\\
&\leqslant N\sum_s\Big|\sum_r f_2(s-r)f_3(s-2r)\dots f_k\big(s-(k-1)r\big)\Big|^2\\
&= N\sum_s\sum_r\sum_t f_2(s-r)\overline{f_2(s-t)}\dots f_k\big(s-(k-1)r\big)\overline{f_k\big(s-(k-1)t\big)}\\
&= N\sum_s\sum_r\sum_u f_2(s)\overline{f_2(s-u)}\dots f_k\big(s-(k-2)r\big)\overline{f_k\big(s-(k-2)r-(k-1)u\big)}\\
&= N\sum_s\sum_r\sum_u \Delta(f_2;u)(s)\Delta(f_3;2u)(s-r)\dots\Delta\big(f_k;(k-1)u\big)\big(s-(k-2)r\big)\,.
\end{aligned}$$

Since $\Delta(f_k;(k-1)u)$ is $\alpha((k-1)u)$-uniform of degree $k-3$, our inductive hypothesis implies that this is at most $N\sum_u\alpha((k-1)u)^{1/2^{k-2}}N^2$, and since $\sum_u\alpha((k-1)u)\leqslant\alpha N$, this is at most $\alpha^{1/2^{k-2}}N^4$, which proves the result for k. □

The interest in Theorem 3.2 is of course that the expression on the left-hand side can be used to count arithmetic progressions. Let us now define a set $A\subset\mathbb{Z}_N$ to be α-uniform of degree d if its balanced function is. (This definition makes sense when $d=0$, but only because it applies to all sets.) The next result implies that a set A which is α-uniform of degree $d-2$ for some small α contains about the number of arithmetic progressions of length d that a random set of the same cardinality would have, where this means arithmetic progressions mod N. We shall then show how to obtain genuine progressions, which turns out to be a minor technicality, similar to the corresponding technicality in the proof of Roth's theorem.

COROLLARY 3.3. *Let $A_1,\dots,A_k$ be subsets of $\mathbb{Z}_N$, such that A_i has cardinality δ_iN for every i, and is $\alpha^{2^{i-1}}$-uniform of degree $i-2$ for every $i\geqslant 3$. Then*

$$\Big|\sum_r\big|(A_1+r)\cap\dots\cap(A_k+kr)\big|-\delta_1\dots\delta_kN^2\Big|\leqslant 2^k\alpha N^2\,.$$

Proof. For each i, let f_i be the balanced function of A_i. Then

$$\big|(A_1+r)\cap\dots\cap(A_k+kr)\big|=\sum_s\big(\delta_1+f_1(s-r)\big)\dots\big(\delta_k+f_k(s-kr)\big)\,,$$

so we can rewrite $|(A_1+r)\cap\dots\cap(A_k+kr)|-\delta_1\dots\delta_kN$ as

$$\sum_{B\subset[k],B\neq\emptyset}\prod_{i\notin B}\delta_i\sum_s\prod_{i\in B}f_i(s-ir)\,.$$

Now if $j = \max B$, then $\sum_r \sum_s \prod_{i\in B} f_i(s-ir)$ is at most $\alpha^{2^{j-1}/2^{j-1}}N^2$, by Theorem 3.2. It follows that

$$\Big|\sum_r \big|(A_1+r)\cap\cdots\cap(A_k+kr)\big| - \delta_1\dots\delta_k N^2\Big| \leqslant \sum_{B\subset[k],B\neq\emptyset}\prod_{i\notin B}\delta_i.\alpha N^2$$
$$= \alpha N^2\Big(\prod_{i=1}^{k}(1+\delta_i)-1\Big),$$

which is at most $2^k\alpha N^2$, as required. □

We now prove two simple technical lemmas.

LEMMA 3.4. *Let $d \geqslant 1$ and let $f : \mathbb{Z}_N \to D$ be α-uniform of degree d. Then f is $\alpha^{1/2}$-uniform of degree $d-1$.*

Proof. Our assumption is that

$$\sum_{a_1,\dots,a_d}\Big|\sum_s \Delta(f;a_1,\dots,a_d)(s)\Big|^2 \leqslant \alpha N^{d+2}.$$

By the Cauchy-Schwarz inequality, this implies that

$$\Big|\sum_{a_1,\dots,a_d}\sum_s \Delta(f;a_1,\dots,a_d)(s)\Big| \leqslant \alpha^{1/2}N^{d+1},$$

which, by the equivalence of properties (i) and (ii) in Lemma 3.1, proves the lemma. □

LEMMA 3.5. *Let A be an α-uniform subset of $\mathbb{Z}_N$ of cardinality δN, and let P be an interval of the form $\{a+1,\dots,a+M\}$, where $M=\beta N$. Then $\big||A\cap P| - \beta\delta N\big| \leqslant \alpha^{1/4}N$.*

Proof. First, we can easily estimate the Fourier coefficients of the set P. Indeed,

$$|\hat{P}(r)| = \Big|\sum_{s=1}^{M}\omega^{-r(a+s)}\Big|$$
$$= \big|(1-\omega^{rM})/(1-\omega^r)\big| \leqslant N/2r.$$

(We also know that it is at most M, but will not need to use this fact.)

This estimate implies that $\sum_{r\neq 0}|\hat{P}(r)|^{4/3}\leqslant N^{4/3}$. Therefore,

$$\begin{aligned}\big||A\cap P|-\beta\delta N\big| &= N^{-1}\Big|\sum_{r\neq 0}\hat{A}(r)\hat{P}(r)\Big|\\ &\leqslant N^{-1}\Big(\sum_{r\neq 0}|\hat{A}(r)|^4\Big)^{1/4}\Big(\sum_{r\neq 0}|\hat{P}(r)|^{4/3}\Big)^{3/4}\\ &\leqslant \Big(\sum_{r\neq 0}|\hat{A}(r)|^4\Big)^{1/4}\leqslant \alpha^{1/4}N\,,\end{aligned}$$

using property (iv) of Lemma 3.1. □

COROLLARY 3.6. *Let $A\subset\mathbb{Z}_N$ be α-uniform of degree $k-2$ and have cardinality δN. If $\alpha\leqslant(\delta/2)^{k2^k}$ and $N\geqslant 32k^2\delta^{-k}$, then A contains an arithmetic progression of length k.*

Proof. Let $A_1=A_2=A\cap[(k-2)N/(2k-3),(k-1)N/(2k-3)]$, and let $A_3=\cdots=A_k=A$. By Lemma 3.4 A is $\alpha^{1/2^{k-3}}$-uniform (of degree one), so by Lemma 3.5 the sets A_1 and A_2 both have cardinality at least $\delta N/4k$ since, by the first inequality we have assumed, we know that $\alpha^{1/2^{k-1}}\leqslant\delta/4k$.

Therefore, by Corollary 3.3, A contains at least $\big(\big(\frac{\delta^k}{16k^2}\big)-2^k\alpha^{1/2^{k-1}}\big)N^2$ arithmetic progressions modulo N with the first two terms belonging to the interval $[(k-2)N/(2k-3),(k-1)N/(2k-3)]$. The only way such a progression can fail to be genuine is if the common difference is zero, and there are at most δN such degenerate progressions. Thus the corollary is proved, since the two inequalities we have assumed imply that $(\delta^k/16k^2)-2^k\alpha^{1/2^{k-1}}\geqslant\delta^k/32k^2$ and $\delta^kN^2/32k^2>\delta N$. □

REMARK. Notice that the proof of Corollary 3.6 did not use Fourier coefficients. This shows that in the proof of Theorem 2.6, the Fourier analysis was not really needed for the analysis of case 3. However, it was used in a more essential way for case 2.

In order to prove Szemerédi's theorem, it is now enough to prove that if $A\subset\mathbb{Z}_N$ is a set of size δN which is not $(\delta/2)^{k2^k}$-uniform of degree $d-2$, then there is an arithmetic progression $P\subset\mathbb{Z}_N$ of length tending to infinity with N, such that $|A\cap P|\geqslant(\delta+\epsilon)|P|$, where $\epsilon>0$ depends on δ and d only. Thus, we wish to deduce a structural property of A from information about its differences. We do not quite have an inverse problem, as usually defined, of additive number theory, but it is certainly in the same spirit, and we shall relate it to a well-known inverse problem, Freiman's theorem, later in the paper. For the rest of this section we shall give a combinatorial

characterization of α-uniform sets of degree d. The result will not be needed for Szemerédi's theorem but gives a little more insight into what is being proved. Also, Lemma 3.7 below will be used near the end of the paper.

Let A be a subset of $\mathbb{Z}_N$ and let $d \geqslant 0$. By a *d-dimensional cube in A* we shall mean a function $\phi : \{0,1\}^d \to A$ of the form

$$\phi : (\epsilon_1, \dots, \epsilon_d) \mapsto a_0 + \epsilon_1 a_1 + \cdots + \epsilon_d a_d \,,$$

where $a_0, a_1, \dots, a_d$ all belong to $\mathbb{Z}_N$. We shall say that such a cube is contained in A, even though it is strictly speaking contained in $A^{\{0,1\}^d}$.

Let $A \subset \mathbb{Z}_N$ have cardinality δN. Then A obviously contains exactly δN cubes of dimension zero and $\delta^2 N^2$ cubes of dimension one. As remarked after Lemma 2.2, the number of two-dimensional cubes in A can be written as $N^{-1}\sum_r |\hat{A}(r)|^4$, so A is α-uniform if and only if there are at most $(\delta^4 + \alpha)N^3$ of them. We shall now show that A contains at least $\delta^{2^d} N^{d+1}$ cubes of dimension d, and that equality is nearly attained if A is α-uniform of degree $d-1$ for some small α. The remarks we have just made prove this result for $d = 1$. Notice that equality is also nearly attained (with high probability) if A is a random set of cardinality δN. This is why we regard higher-degree uniformity as a form of pseudorandomness.

LEMMA 3.7. *Let A be a subset of $\mathbb{Z}_N$ of cardinality δN and let $d \geqslant 0$. Then A contains at least $\delta^{2^d} N^{d+1}$ cubes of dimension d.*

Proof. We know the result for $d = 0$ or 1 so let $d > 1$ and assume that the result is known for $d-1$. The number of d-dimensional cubes in A is the sum over all r of the number of $(d-1)$-dimensional cubes in $A \cap (A + r)$. Write $\delta(r)N$ for the cardinality of $A \cap (A + r)$. Then by induction the number of d-dimensional cubes in A is at least $\sum_r \delta(r)^{2^{d-1}} N^d$. Since the average value of $\delta(r)$ is exactly δ^2, this is at least $\delta^{2^d} N^{d+1}$ as required. □

The next lemma is little more than the Cauchy-Schwarz inequality and some notation. It will be convenient to use abbreviations such as x for $(x_1, \dots, x_k)$ and $x.y$ for $\sum_{i=1}^k x_i y_i$. If $\epsilon \in \{0,1\}^k$ then we shall write $|\epsilon|$ for $\sum_{i=1}^k \epsilon_i$. Once again, C is the operation of complex conjugation.

LEMMA 3.8. *For every $\epsilon \in \{0,1\}^k$ let f_ϵ be a function from $\mathbb{Z}_N$ to D. Then*

$$\Big|\sum_{x \in \mathbb{Z}_N^d} \sum_s \prod_{\epsilon \in \{0,1\}^d} C^{|\epsilon|} f_\epsilon(s - \epsilon.x)\Big| \leqslant \prod_{\epsilon \in \{0,1\}^d} \Big|\sum_{x \in \mathbb{Z}_N^d} \sum_s \prod_{\eta \in \{0,1\}^d} C^{|\eta|} f_\epsilon(s - \eta.x)\Big|^{\frac{1}{2^d}}.$$

Proof.

$$\Big|\sum_{x\in\mathbb{Z}_N^d}\sum_s\prod_{\epsilon\in\{0,1\}^d}C^{|\epsilon|}f_\epsilon(s-\epsilon.x)\Big|$$

$$=\Big|\sum_{x\in\mathbb{Z}_N^{d-1}}\Big(\sum_s\prod_{\epsilon\in\{0,1\}^{d-1}}C^{|\epsilon|}f_{\epsilon,0}(s-\epsilon.x)\Big)\Big(\sum_t\prod_{\epsilon\in\{0,1\}^{d-1}}C^{|\epsilon|}f_{\epsilon,1}(t-\epsilon.x)\Big)\Big|$$

$$\leqslant\Big(\sum_{x\in\mathbb{Z}_N^{d-1}}\Big|\sum_s\prod_{\epsilon\in\{0,1\}^{d-1}}C^{|\epsilon|}f_{\epsilon,0}(s-\epsilon.x)\Big|^2\Big)^{\frac12}$$

$$\cdot\Big(\sum_{x\in\mathbb{Z}_N^d}\Big|\sum_s\prod_{\epsilon\in\{0,1\}^{d-1}}C^{|\epsilon|}f_{\epsilon,1}(s-\epsilon.x)\Big|^2\Big)^{\frac12}.$$

Let us write $P_d(\epsilon)$ and $Q_d(\epsilon)$ for the sequences $(\epsilon_1,\dots,\epsilon_{d-1},0)$ and $(\epsilon_1,\dots,\epsilon_{d-1},1)$. Then

$$\sum_{x\in\mathbb{Z}_N^{d-1}}\Big|\sum_s\prod_{\epsilon\in\{0,1\}^{d-1}}C^{|\epsilon|}f_{\epsilon,0}(s-\epsilon.x)\Big|^2=\sum_{x\in\mathbb{Z}_N^d}\sum_s\prod_{\epsilon\in\{0,1\}^d}C^{|\epsilon|}f_{P_d(\epsilon)}(s-\epsilon.x)$$

and similarly for the second bracket with Q_d, so the two parts are square roots of expressions of the form we started with, except that the function f_ϵ no longer depends on ϵ_d. Repeating this argument for the other coordinates, we obtain the result. □

If we regard Lemma 3.8 as a modification of the Cauchy-Schwarz inequality, then the next lemma is the corresponding modification of Minkowski's inequality.

LEMMA 3.9. *Given any function $f:\mathbb{Z}_N\to\mathbb{C}$ and any $d\geqslant 2$, define $\|f\|_d$ by the formula*

$$\|f\|_d=\Big|\sum_{x\in\mathbb{Z}_N^d}\sum_s\prod_{\epsilon\in\{0,1\}^d}C^{|\epsilon|}f(s-\epsilon.x)\Big|^{1/2^d}.$$

Then $\|f+g\|_d\leqslant\|f\|_d+\|g\|_d$ for any pair of functions $f,g:\mathbb{Z}_N\to\mathbb{C}$. In other words, $\|.\|_d$ is a norm.

Proof. If we expand $\|f+g\|^{2^d}$, we obtain the sum

$$\sum_{x\in\mathbb{Z}_N^d}\sum_s\prod_{\epsilon\in\{0,1\}^d}C^{|\epsilon|}(f+g)(s-\epsilon.x)\,.$$

If we expand the product we obtain 2^{2^d} terms of the form $\prod_{\epsilon\in\{0,1\}^d}C^{|\epsilon|}f_\epsilon(s-\epsilon.x)$, where each function f_ϵ is either f or g. For each one of these terms, if we take the sum over $x_1,\dots,x_d$ and s and apply

Lemma 3.8, we have an upper estimate of $\|f\|_d^k \|g\|_d^l$, where k and l are the number of times that f_ϵ equals f and g respectively. From this it follows that

$$\|f+g\|^{2^d} \leqslant \sum_{k=0}^{2^d} \binom{2^d}{k} \|f\|_d^k \|g\|_d^{2^d-k} = (\|f\|_d + \|g\|_d)^{2^d},$$

which proves the lemma. □

It is now very easy to show that equality is almost attained in Lemma 3.7 for sets that are sufficiently uniform.

LEMMA 3.10. *Let A be α-uniform of degree $d-1$. Then A contains at most $(\delta + \alpha^{1/2^d})^{2^d} N^{d+1}$ cubes of dimension d.*

Proof. Write $A = \delta + f$, where $|A| = \delta N$ and f is the balanced function of A. Then $\|A\|_d \leqslant \|\delta\|_d + \|f\|_d$. It is easy to see that $\|A\|_d^{2^d}$ is the number of d-dimensional cubes in A and that $\|\delta\|_d^{2^d} = \delta^{2^d} N^{d+1}$. Moreover, the statement that A is α-uniform of degree $d-1$ is equivalent to the statement that $\|f\|_d^{2^d} \leqslant \alpha N^{d+1}$. Therefore, Lemma 3.9 tells us that A contains at most $(\delta + \alpha^{1/2^d})^{2^d} N^{d+1}$ cubes of dimension d. □

REMARK. In a sense, the normed spaces just defined encapsulate all the information we need about the arithmetical properties of the functions we consider. In their definitions they bear some resemblance to Sobolev spaces. Although I cannot think of any potential applications, I still feel that it would be interesting to investigate them further.

4 Two Motivating Examples

We now know that Szemerédi's theorem would follow from an adequate understanding of higher-degree uniformity. A natural question to ask is whether degree-one uniformity *implies* higher-degree uniformity (for which it would be enough to show that it implied quadratic uniformity). To make the question precise, if A has density δ and is α-uniform, does it follow that A is quadratically β-uniform, for some β which, for fixed δ, tends to zero as α tends to zero? If so, then the same result for higher-degree uniformity can be deduced, and Szemerédi's theorem follows easily, by the method of §2.

The first result of this section is a simple counterexample showing that uniformity does *not* imply quadratic uniformity. Let A be the set $\{s \in \mathbb{Z}_N : |s^2| \leqslant N/10\}$. If $s \in A \cap (A+k)$, then $|s^2| \leqslant N/10$ and $|(s-k)^2| \leqslant N/10$

as well, which implies that $|2sk - k^2| \leqslant N/5$, or equivalently that s lies inside the set $(2k)^{-1}\{s : |s - k/2| \leqslant N/5\}$. It follows that $A \cap (A + k)$ is not uniform for any $k \neq 0$.

It is possible, but not completely straightforward, to show that A itself is uniform. Rather than go into the details, we prove a closely related fact which is in some ways more natural. Let $f(s) = \omega^{s^2}$. We shall show that f is a very uniform function, while $\Delta(f;k)$ fails badly to be uniform for any $k \neq 0$. For the uniformity of f, notice that

$$|\hat{f}(r)| = \Big|\sum_s \omega^{s^2 - rs}\Big| = \Big|\sum_s \omega^{(s-r/2)^2}\Big| = \Big|\sum_s \omega^{s^2}\Big|$$

for every r. Therefore, $|\hat{f}(r)| = N^{1/2}$ for every $r \in \mathbb{Z}_N$, so f is as uniform as a function into the unit circle can possibly be. On the other hand, $\Delta(f;k)(s) = \omega^{2ks - k^2}$, so that

$$\Delta(f;k)^\wedge(r) = \begin{cases} N & \text{r=2k} \\ 0 & \text{otherwise}\,, \end{cases}$$

which shows that $\Delta(f;k)$ is, for $k \neq 0$, as non-uniform as possible.

More generally, if $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ is a quadratic polynomial and $f(s) = \omega^{\phi(s)}$, then f is highly uniform, but there is some $\lambda \in \mathbb{Z}_N$ such that, for every k,

$$\Delta(f;k)^\wedge(r) = \begin{cases} N & r = \lambda k \\ 0 & \text{otherwise}\,. \end{cases}$$

This suggests an attractive conjecture, which could perhaps replace the false idea that if A is uniform then so are almost all $A \cap (A + k)$. Perhaps if there are many values of k for which $A \cap (A + k)$ fails to be uniform, then there must be a quadratic function $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ such that $\big|\sum_{s \in A} \omega^{-\phi(s)}\big|$ is large. We shall see in the next section that such "quadratic bias" would actually imply the existence of a long arithmetic progression P_j such that $|A \cap P_j|/|P_j|$ was significantly larger than $|A|/N$. This would give a proof of Szemerédi's theorem for progressions of length four, and one can see how the above ideas might be generalized to higher-degree polynomials and longer arithmetic progressions.

The second example of this section shows that such conjectures are still too optimistic. As with the first example, we shall consider functions that are more general than characteristic functions of subsets of $\mathbb{Z}_N$. However, this should be enough to convince the reader not to try to prove the conjectures.

Let r be about $\sqrt{N}$ and for $0 \leqslant a, b < r/2$ define $\phi(ar + b)$ to be $a^2 + b^2$.

Now define

$$f(s) = \begin{cases} \omega^{\phi(s)} & s = ar + b \text{ for some } 0 \leqslant a, b < r/2 \\ 0 & \text{otherwise}. \end{cases}$$

The function f is not quadratic, but it resembles a quadratic form in two variables (with the numbers 1 and r behaving like a basis of a two-dimensional space).

Suppose $s = ar + b$ and $k = cr + d$ are two numbers in $\mathbb{Z}_N$, where all of a, b, $a - c$ and $b - d$ lie in the interval $[0, r/2)$. Then

$$f(s)\overline{f(s-k)} = \omega^{2ac-c^2+2bd-d^2} = \omega^{\phi_k(ar+b)},$$

where ϕ_k depends linearly on the pair (a, b). The property that will interest us about ϕ_k is that, at least when c and d are not too close to $r/2$, there are several pairs (a, b) such that the condition on (a, b, c, d) applies, and therefore several quadruples $\big((a_i, b_i)\big)_{i=1}^4$ such that

$$(a_1, b_1) + (a_2, b_2) = (a_3, b_3) + (a_4, b_4)$$

and

$$\phi_k(a_1 r + b_1) + \phi_k(a_2 r + b_2) = \phi_k(a_3 r + b_3) + \phi_k(a_4 r + b_4).$$

Here, "several" means a number proportional to N^3, which is the maximum it could be.

Let B be the set of all $s = ar + b$ for which a, b, c and d satisfy the conditions above. (Of course, B depends on k.) Then

$$\sum_q \Big|\sum_{s\in B} \omega^{\phi_k(s)-qs}\Big|^4 = N \sum \{\omega^{\phi_k(s)+\phi_k(t)-\phi_k(u)-\phi_k(v)} : s, t, u, v \in B,\ s + t = u + v\}.$$

Now the set B has been chosen so that if $s, t, u, v \in B$ and $s + t = u + v$, then $\phi_k(s) + \phi_k(t) = \phi_k(u) + \phi_k(v)$. Therefore, the right-hand side above is N times the number of quadruples $(s, t, u, v) \in B^4$ such that $s + t = u + v$. It is not hard to check that if c and d are smaller than $r/4$, say, then B has cardinality proportional to N^3, and therefore that the right-hand side above is proportional to N^4. Lemma 2.1 now tells us that ϕ_k has a large Fourier coefficient. Thus, at the very least, we have shown that, for many values of k, $\Delta(f; k)$ fails to be uniform.

If we could find a *genuinely* quadratic function $\phi(s) = as^2 + bs + c$ such that $\big|\sum_s f(s)\omega^{-\phi(s)}\big|^2$ was proportional to N^2, then, expanding, we would have

$$\sum_{s,k} f(s)\overline{f(s-k)}\omega^{-\phi(s)+\phi(s-k)} = \sum_{s,k} f(s)\overline{f(s-k)}\omega^{-2ask-bk}$$

proportional to N^2, which would imply that the number of k for which $\Delta(f;k)\tilde{}(2ak)$ was proportional to N was proportional to N. A direct calculation (left to the interested reader) shows that such a phenomenon does not occur. That is, there is no value of λ such that $\Delta(f;k)\hat{}(\lambda k)$ is large for many values of k.

There are of course many examples like the second one above. One can define functions that resemble d-dimensional quadratic forms, and provided that d is small the same sort of behaviour occurs. Thus, we must accept that the ideas of this paper so far do not lead directly to a proof of Szemerédi's theorem, and begin to come to terms with these "multi-dimensional" examples. It is for this purpose that our major tool, an adaptation of Freiman's theorem, is used, as will be explained later in the paper.

Returning to the first example of this section, it should be noted that the set $A = \{s \in \mathbb{Z}_N : |s^2| \leqslant N/10\}$ also serves to show that a uniform set need not have roughly the same number of arithmetic progressions of length four as a random set. Indeed, it is not hard to show that if x, $x+d$ and $x+2d$ all lie in A, then it is a little 'too likely' that $x+3d$ will also lie in A, which shows that A contains 'too many' progressions of length four.

Until recently, I was confident that a modificiation of this example could be constructed with too *few* progressions of length four. However, I have recently changed my mind, after a conversation with Gil Kalai in which he challenged me actually to produce such a modification. In fact, there are convincing heuristic arguments in support of the following conjecture, even though at first it seems very implausible.

CONJECTURE 4.1. *Let $A \subset \mathbb{Z}_N$ be a set of size δN. Then, if A is α-uniform, the number of quadruples $(x, x+d, x+2d, x+3d)$ in A^4 is at least $(\delta^4 - \beta)N^2$, where β tends to zero as α tends to zero.*

In other words, uniform sets always contain *at least* the expected number of progressions of length four.

It can be shown that quadratically uniform sets sometimes contain significantly fewer progressions of length five than random sets of the same cardinality. However, the example depends in an essential way on 5 being odd, and the following extension of Conjecture 4.1 appears to be true as well.

CONJECTURE 4.2. *Let $A \subset \mathbb{Z}_N$ be a set of size δN and let k be an even number. Then, if A is α-uniform of degree $k-1$, the number of sequences $(x, x+d, ..., x+(k+1)d)$ belonging to A^{k+2} is at least $(\delta^{k+2} - \beta)N^2$, where β tends to zero as α tends to zero.*

5 Consequences of Weyl's Inequality

In this section we shall generalize Lemma 2.2 and Corollary 2.3 from linear functions to general polynomials. Most of the results of the section are well known. Since the proofs are short, we shall give many of them in full, to keep the paper as self-contained as possible. The main exception is Weyl's inequality itself: there seems little point in reproducing the proof when it is well explained in many places. Once we have generalized these two results, we will have shown that for the proof of Szemerédi's theorem it is enough to prove that a set which fails to be uniform of degree d exhibits "polynomial bias", rather than "linear bias" as we showed in the case $d = 1$. We shall not try to define the notion of bias precisely. If a set A has balanced function f and there is a polynomial $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ of degree d such that $\left|\sum_s f(s)\omega^{-\phi(s)}\right|$ is large, then A exhibits polynomial bias in the required sense. However, the second example in the previous section showed that this is too much to ask for, so a precise definition would have to be somewhat weaker.

First, we give some simple estimates for certain Fourier coefficients. We shall write $[-M, M)$ for the set $\{-M, -(M-1), \dots, M-1\}$.

LEMMA 5.1. *Let $I \subset \mathbb{Z}_N$ be the interval $[-M, M)$. Then $|\hat{I}(r)| \leqslant \min\{2M, N/2|r|\}$.*

Proof. This is a simple direct calculation. The upper bound of $2M$ is trivial. To obtain the bound of $N/2|r|$, note that for θ in the range $[-\pi, \pi]$ one has

$$|1 - e^{i\theta}| \geqslant 2|\theta|/\pi\,.$$

Applying this estimate with $\theta = 2\pi r/N$ gives

$$|\hat{I}(r)| = \left|\sum_{s=-M}^{M-1} \omega^{rs}\right| = \left|\frac{\omega^{-rM} - \omega^{rM}}{1 - \omega^r}\right| \leqslant \frac{2}{|1 - \omega^r|} \leqslant \frac{N}{2|r|}\,,$$

as was wanted. □

Given an integer $r \in \mathbb{Z}_N$, we shall use the notation $|r|$ to stand for the modulus of the unique representative of r that lies in the interval $[-N/2, N/2)$ (i.e., the distance from r to zero).

LEMMA 5.2. *Let A be a subset of $\mathbb{Z}_N$ of cardinality t, let M be an even integer and suppose that $A \cap [-M, M) = \emptyset$. Then there exists r with $0 < |r| \leqslant N^2M^{-2}$ such that $|\hat{A}(r)| \geqslant tM/2N$.*

Proof. Let $I = [-M/2, M/2)$. Then $A \cap (I - I) = \emptyset$. It follows that $\langle A, I * I\rangle = 0$, which is the same as saying that $\sum_s A(s)I * I(s) = 0$.

By identities (1) and (2) of §2 (transforms of convolutions and Parseval's identity) it follows that $\sum_r \hat{A}(r)|\hat{I}(r)|^2 = 0$. Since $\hat{I}(0) = M$ and $\hat{A}(0) = t$, it follows that

$$\sum_{r\neq 0} |\hat{A}(r)||\hat{I}(r)|^2 \geqslant tM^2 .$$

By Lemma 5.1, we know, for each r, that $|\hat{I}(r)| \leqslant \min\{M, N/2|r|\}$. It follows that

$$\begin{aligned}\sum_{r\neq 0} |\hat{A}(r)||\hat{I}(r)|^2 &\leqslant \max_{0<|r|\leqslant N^2M^{-2}} |\hat{A}(r)| \sum_r |\hat{I}(r)|^2 + t \sum_{|r|\geqslant N^2M^{-2}} N^2/4|r|^2 \\ &\leqslant MN \max_{0<|r|\leqslant N^2M^{-2}} |\hat{A}(r)| + (3/4)tN^2(N^2M^{-2})^{-1} \\ &= MN \max_{0<|r|\leqslant N^2M^{-2}} |\hat{A}(r)| + (3/4)tM^2 .\end{aligned}$$

Therefore, there exists r with $|r| \leqslant N^2M^{-2}$ and $|\hat{A}(r)| \geqslant M^2t/4MN = tM/4N$, which proves the lemma. □

Remark. A more obvious approach to proving the above result would be to use I instead of $I * I$. That is, one would consider the sum $\sum_r \hat{A}(r)\hat{I}(r)$. It turns out, however, that the estimates that one obtains are not strong enough. The trick of using $I * I$ instead is basically the familiar device of replacing the Dirichlet kernel by the Féjer kernel.

The next lemma is a special case of Weyl's inequality. (To obtain the inequality in its full generality, replace s^k below by an arbitrary monic polynomial of degree k. The proof is unaffected.) We shall make a fairly standard deduction from it, so it seems appropriate to use standard notation as well. Thus, $e(x)$ means $\exp(2\pi ix)$.

Lemma 5.3. *Let a and q be integers with $(a, q) = 1$. Let α be a real number such that $\left|\alpha - a/q\right| \leqslant q^{-2}$. Then, for all $\epsilon > 0$,*

$$\left|\sum_{s=1}^{t} e(\alpha s^k)\right| \leqslant C_\epsilon t^{1+\epsilon}(q^{-1} + t^{-1} + qt^{-k})^{1/2^{k-1}} .$$

Moreover, if $t \geqslant 2^{2^{32k^2}}$, then the above inequality is valid with $\epsilon = 1/k2^{k+1}$ and $C_\epsilon = 1000$. □

The above estimate for C_ϵ is important because we wish to use the inequality to obtain explicit bounds. Unfortunately, I have not managed to find in the literature any presentation of Weyl's inequality that bothers to estimate C_ϵ. If one follows the proof given by Vaughan [V] and keeps

track of everything that is swallowed up by the t^ϵ, one can replace the $C_\epsilon t^\epsilon$ in the right-hand side of the inequality by

$$500(2t)^{8k/2^{k-1}\log\log t}(\log t)^{1/2^{k-1}}.$$

It is from this that we deduced the final part of the lemma. Note that, although C_ϵ became an absolute constant, we paid for it with the assumption that t was sufficiently large. Since we are stating this estimate rather than giving a detailed proof, the reader may be reassured to know that for what follows it would not matter if t was required to be far larger – quadruply exponential in k, say. Moreover, Weyl's inequality does not give the best known estimate for the exponential sum in question. It is used here because its proof is reasonably simple, which makes checking the estimate above relatively straightforward.

The next lemma is very standard, and is due to Dirichlet.

LEMMA 5.4. *Let α be a real number. For every integer $u \geqslant 1$ there exist integers a and q with $(a,q) = 1$, $1 \leqslant q \leqslant u$ and $\bigl|\alpha - a/q\bigr| \leqslant 1/qu$.* □

The next lemma is also due to Weyl. Since it is again hard to find in the literature in the quantitative form we need, we give a complete proof.

LEMMA 5.5. *Let $k \geqslant 2$, let $t \geqslant 2^{2^{32k^2}}$, let $N \geqslant t$ and and let $a \in \mathbb{Z}_N$. Then there exists $p \leqslant t$ such that $|p^k a| \leqslant t^{-1/k2^{k+1}}N$.*

Proof. Let $A = \{a, 2^k a, 3^k a, \dots, t^k a\}$. By Lemma 5.2, if the result is false then there exists r such that $0 < |r| \leqslant t^{1/k2^k}$ and $|\hat{A}(r)| \geqslant \frac{1}{2}t^{1-1/k2^{k+1}}$. Setting $\alpha = -ar/N$, we have

$$\hat{A}(r) = \sum_{u\in A}\omega^{-ru} = \sum_{s=1}^{t}\omega^{-rs^k a} = \sum_{s=1}^{t} e(\alpha s^k).$$

Lemma 5.4 gives us integers b and q with $(b,q) = 1$, $1 \leqslant q \leqslant t$ and $\bigl|\alpha - b/q\bigr| \leqslant 1/qt$. By Lemma 5.3 we know that

$$|\hat{A}(r)| \leqslant 1000t^{1+1/k2^{k+1}}(q^{-1} + t^{-1} + t^{1-k})^{1/2^{k-1}}.$$

By the lower bound for $|\hat{A}(r)|$, we may deduce that

$$2000t^{1/k2^k}(q^{-1} + t^{-1} + q^{1-k})^{1/2^{k-1}} \geqslant 1$$

which implies, after a small calculation (using the assumption that $t \geqslant 2^{2^{32k^2}}$), that $q \leqslant 2t^{1/2k}$.

We may now argue directly. We know that $\bigl|\alpha - b/q\bigr| \leqslant 1/qt$. Multiplying both sides by $(rq)^k N/r$ we find that

$$|-a(rq)^k - b(rq)^{k-1}N| \leqslant (|r|q)^{k-1}N/t \leqslant t^{-1/2}N,$$

so we can set $p = rq$ (contradicting the initial assumption that the result was false). □

COROLLARY 5.6. *Let $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ be any polynomial of degree k, let $K = (k!)^2 2^{(k+1)^2}$ and let r be an integer exceeding $2^{2^{40k^2}}$. Then for every $m \geqslant r^{1-1/K}$ the set $\{0, 1, 2, \dots, r-1\}$ can be partitioned into arithmetic progressions $P_1, \dots, P_m$ such that the diameter of $\phi(P_j)$ is at most $r^{-1/K}N$ for every j and the lengths of any two P_j differ by at most 1.*

Proof. The case $k = 1$ follows immediately from Lemma 2.2. Given $k > 1$, let us write $\phi(x) = ax^k + \psi_1(x)$, in such a way that ψ_1 is a polynomial of degree $k-1$. By Lemma 5.5 we can find $p \leqslant r^{1/2}$ such that $|ap^k| \leqslant r^{-1/k2^{k+2}}N$. Then for any s we have

$$\begin{aligned}\phi(x + sp) &= a(x + sp)^k + \psi_1(x + sp)\\ &= s^k(ap^k) + \psi_2(x, p)\,,\end{aligned}$$

where ψ_2 is, for any fixed x, a polynomial of degree at most $k-1$ in p.

For any u, the diameter of the set $\{s^k(ap^k) : 0 \leqslant s < u\}$ is at most $u^k|ap^k| \leqslant u^k r^{-1/k2^{k+2}}N$. Therefore, for any $u \leqslant r^{1/4}$, we can partition the set $\{0, 1, \dots, r-1\}$ into arithmetic progressions of the form

$$Q_j = \big\{x_j, x_j + p, \dots, x_j + (u_j - 1)p\big\}\,,$$

such that, for every j, $u - 1 \leqslant u_j \leqslant u$ and there exists a polynomial ϕ_j of degree at most $k-1$ such that, for any subset $P \subset Q_j$,

$$\operatorname{diam}(\phi(P)) \leqslant u^k r^{-1/k2^{k+1}}N + \operatorname{diam}(\phi_j(P))\,.$$

Let us choose $u = r^{1/k^2 2^{k+2}}$, with the result that $u^k r^{-1/k2^{k+1}} = r^{-1/k2^{k+2}}$. It is easy to check that $u \geqslant 2^{2^{40(k-1)^2}}$. Therefore, by induction, if $v \leqslant u^{1/L}$, where $L = ((k-1)!)^2 2^{k^2}$, then every Q_j can be partitioned into arithmetic progressions P_{jt} of length $v-1$ or v in such a way that $\operatorname{diam}(\phi_j(P_{jt})) \leqslant u^{-1/L}N$ for every t. It is not hard to check that this, with our choice of u above, gives us the inductive hypothesis for k. □

COROLLARY 5.7. *Let $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ be a polynomial of degree k, let $K = (k!)^2 2^{(k+1)^2}$, let $\alpha > 0$ and let r be an integer exceeding $\max\{2^{2^{40k^2}}, (4\pi/\alpha)^K\}$. Then, for any $m \geqslant r^{1-1/K}$, there is a partition of the set $\{0, 1, \dots, r-1\}$ into arithmetic progressions $P_1, \dots, P_m$ such that the sizes of the P_j differ by at most one, and if $f : \mathbb{Z}_N \to D$ is any function such that*

$$\left|\sum_{s=0}^{r-1} f(s)\omega^{-\phi(s)}\right| \geqslant \alpha r\,,$$

then

$$\sum_{j=1}^{m}\Big|\sum_{s\in P_j} f(s)\Big| \geqslant (\alpha/2)r\,.$$

Proof. By Corollary 5.6 we can choose $P_1,\dots,P_m$ such that $\operatorname{diam}(\phi(P_j)) \leqslant Nr^{-1/K}$ for every j. By the second lower bound for r, this is at most $\alpha N/4\pi$. Exactly as in the proof of Corollary 2.3, this implies the result. □

COROLLARY 5.8. *Let $A \subset \mathbb{Z}_N$ be a set of cardinality δN with balanced function f. Suppose that we can find disjoint arithmetic progressions $P_1,\dots,P_M$ such that $A \subset \bigcup_i P_i$, and polynomials $\phi_1,\dots,\phi_M$ of degree at most k such that*

$$\sum_{i=1}^{M}\Big|\sum_{s\in P_i} f(s)\omega^{-\phi_i(s)}\Big| \geqslant \alpha N\,.$$

Suppose also that $|P_i| \leqslant 2|P_j|$ for all i,j. Then there is an arithmetic progression Q of cardinality at least $(N/M)^{1/K}/8$ such that $|A\cap Q| \geqslant (\delta+\alpha/8)|Q|$.

Proof. We know that no P_i has cardinality more than $2N/M$. By Corollary 5.7, if $m \leqslant C(2N/M)^{1-1/K}$, each P_i can be partitioned into arithmetic progressions $P_{i1},\dots,P_{im}$ such that

$$\sum_{j=1}^{m}\Big|\sum_{s\in P_{ij}} f(s)\Big| \geqslant \frac{1}{2}\Big|\sum_{s\in P_i} f(s)\omega^{-\phi_i(s)}\Big|\,.$$

Summing over i, we find that

$$\sum_{i=1}^{M}\sum_{j=1}^{m}\Big|\sum_{s\in P_{ij}} f(s)\Big| \geqslant \alpha N\,.$$

Since A is contained in the union of the P_{ij} we also know that

$$\sum_{i=1}^{M}\sum_{j=1}^{m}\sum_{s\in P_{ij}} f(s) = 0\,.$$

Let $F_{ij} = \sum_{s\in P_{ij}} f(s)$ and let J be the set of (i,j) such that $F_{ij} \geqslant 0$. Then the inequalities above imply that $\sum_{(i,j)\in J} F_{ij} \geqslant \alpha N/4$, so we can find P_{ij} with $\sum_{s\in P_{ij}} f(s) \geqslant \alpha N/4Mm$. Since $|P_{ij}| \leqslant 4N/Mm$, this shows that $|A\cap P_{ij}| \geqslant (\delta+\alpha/16)|P_{ij}|$. □

We have now finished one of the key stages in the proof. As promised in the introduction to this section, if we want to generalize Roth's argument, we may now look for "polynomial bias", rather than the "linear bias" which

arises there, since polynomial bias implies linear bias on small subprogressions.

We continue the section with three results that generalize Lemma 5.5 and Corollary 5.6 to statements dealing with several polynomials at once. These generalizations will not be needed for progressions of length four, but they are very important for progressions of length six or more, and the next lemma is needed for progressions of length five (in the case $k = 2$). Our methods of proof are extremely crude, and it is quite likely that much better bounds are known. However, we have not been able to find them and the poor bounds here do not greatly affect the estimate we shall eventually obtain for Szemerédi's theorem.

LEMMA 5.9. *Let $\phi_1, \dots, \phi_q$ be polynomials from $\mathbb{Z}_N$ to $\mathbb{Z}_N$ of degree at most k, let $K = (k!)^2 2^{(k+1)^2}$ and let r be an integer exceeding $2^{2^{40k^2}K^{q-1}}$. Then for every $m \geqslant r^{1-1/2K^q}$ the set $\{0, 1, 2, \dots, r-1\}$ can be partitioned into arithmetic progressions $P_1, \dots, P_m$ such that the diameter of $\phi_i(P_j)$ is at most $r^{-1/K^q}N$ for every i and every j, and the lengths of any two P_j differ by at most 1.*

Proof. First we prove by induction that for every $p \leqslant q$ we can partition the set $\{0, 1, \dots, r-1\}$ into arithmetic progressions $P_1, \dots, P_m$ of size at least r^{1/K^p} such that $\operatorname{diam} \phi_i(P_j)$ is at most $r^{-1/K^p}N$ for every $i \leqslant p$ and $j \leqslant m$. When $p = 1$ this follows immediately from Corollary 5.6. If we know it for $p-1$, let $Q_1, \dots, Q_l$ be the arithmetic progressions obtained. The size of each Q_i is at least $r^{1/K^{p-1}} \geqslant 2^{2^{40k^2}}$, so by Corollary 5.6 each Q_i can be partitioned into further arithmetic progressions P_j of cardinality at least $(r^{1/K^{p-1}})^{1/K} = r^{1/K^p}$, such that, for every j, the diameter of $\phi_q(P_j)$ is at most $(r^{-1/K^{p-1}})^{1/K}N = r^{-1/K^p}N$. This is clearly enough to give us the statement for p.

In particular, we have the statement when $p = q$. To obtain the lemma, notice that if $k^2 \leqslant m$, then an arithmetic progression of length m can be partitioned into subprogressions each of which has length k or $k+1$. □

We are now going to prove a similar result for multilinear functions, which in this context means functions of the form

$$\mu(x_1, \dots, x_k) = \sum_{A \subset [k]} c_A \prod_{j \in A} x_j \,.$$

Define a *box in $\mathbb{Z}_N^k$ of common difference d* to be a product $P = Q_1 \times \dots \times Q_k$, where each Q_i is an arithmetic progression in $\mathbb{Z}_N$ (even when $\mathbb{Z}_N$ is

embedded into $\mathbb{Z}$) of common difference d. The *width* of P is defined to be $\min |Q_i|$.

LEMMA 5.10. *Let $k \geqslant 2$, let $K = k^2 2^{k+3}$, let $m \geqslant 2^{K^{2^k} 2^{32k^2+1}}$, let P be a box in $\mathbb{Z}_N^k$ of width at least m and let μ be a k-linear function from P to $\mathbb{Z}_N$. Then P can be partitioned into boxes $P_1, \dots, P_M$, such that each P_j has width at least $m^{K^{-2^k}}$ and the diameter of $\mu(P_j)$ is at most $2m^{-K^{-2^k}}N$ for every j.*

Proof. As noted above, μ can be written as a sum of terms of the form $c_A \prod_{j \in A} x_j$. Take any total ordering on the subsets of $[k]$ which extends the partial ordering by inclusion, and define the *height* of μ to be the largest position in this ordering of a set A such that the coefficient c_A is non-zero. We shall prove the result by induction on the height. The precise inductive hypothesis is that if μ has height at most p, then any box Q of width $t \geqslant 2^{K^p 2^{32k^2+1}}$ can be partitioned into boxes Q_j of width at least $t^{K^{-p}}$ such that for every j the diameter of $\mu(Q_j)$ is at most $(1 + 2^{-k}p)t^{-K^{-p}}N$.

First, if the height is zero or one, then μ is constant and the result is trivial. Now let Q be a box of width t and common difference d_0, let $\mu : Q \to \mathbb{Z}_N$ be a k-linear function of height p and suppose that the result is true for all multilinear functions of height less than p. Let A be the p^{th} set in the ordering on the subsets of $[k]$, and let c_A be the corresponding coefficient of μ.

By Lemma 5.5, we can find $r \leqslant t^{1/2}$, such that, setting $d = rd_0$, we have the inequality $|c_A d^{|A|}| \leqslant t^{-1/k2^{k+2}}N$. Now, for any $(x_1, \dots, x_k) \in Q$ we can define a function ν by

$$\nu(b_1, \dots, b_k) = \mu(x_1 + db_1, \dots, x_k + db_k)$$

and write it in the form

$$\nu(b_1, \dots, b_k) = \sum_{B \subset [k]} c'_B d^{|B|} \prod_{j \in B} b_j \,.$$

It is not hard to see that, because μ has height p, so does ν, and also that $c'_A = c_A$, whatever the choice of $(x_1, \dots, x_k)$. Therefore, we can write

$$\nu(b_1, \dots, b_k) = c_A d^{|A|} \prod_{j \in A} b_j + \nu'(b_1, \dots, b_k) \,,$$

where ν' has height at most $p - 1$. If $\max\{b_1, \dots, b_k\} \leqslant m^{1/k^2 2^{k+3}}$, then our estimate for $c_A d^{|A|}$ implies that $\left|c_A d^{|A|} \prod_{j \in A} b_j\right| \leqslant t^{-1/k2^{k+3}}N$.

Now we are almost finished. Since $r \leqslant t^{1/2}$, there is no problem in partitioning Q into boxes of common difference d and width $t^{1/k^2 2^{k+3}} = t^{1/K}$. In

each such box, we have shown that μ can be written as a sum $\nu_1+\nu_2$ of multilinear functions such that ν_1 is bounded above in size by $t^{-1/k2^{k+3}}N$ and ν_2 has height at most $p-1$ (with the functions ν_1 and ν_2 depending on the box). By induction, each such box can be further partitioned into boxes of width at least t^{1/K^p} such that ν_2 has diameter at most $(1+(p-1)2^{-k})t^{-K^{-p}}N$. Since

$$t^{-1/k2^{k+2}}N+\big(1+2^{-k}(p-1)\big)t^{-K^{-p}}N\leqslant(1+2^{-k}p)t^{-K^{-p}}N\,,$$

we have proved the inductive hypothesis for p and hence the whole lemma. □

It is now not hard to deduce a multiple version of the above lemma.

COROLLARY 5.11. *Let $k\geqslant 2$, let $K=k^22^{k+3}$, let P be a box in $\mathbb{Z}_N^k$ of width at least $m\geqslant 2^{K^{2^kq}2^{32k^2+1}}$ and let $\mu_1,\dots,\mu_q$ be k-linear functions from P to $\mathbb{Z}_N$. Then P can be partitioned into boxes $P_1,\dots,P_M$, such that each P_j has width at least $m^{K^{-2^kq}}$ and the diameter of $\mu_i(P_j)$ is at most $2m^{-K^{-2^kq}}N$ for every i and j.*

Proof. We can apply Lemma 5.10 q times, obtaining a sequence of finer and finer partitions into boxes, such that for each refinement another of the μ_i satisfies the conclusion of that lemma. The width of the boxes at the final stage of this process is at least the number obtained by raising m to the power K^{-2^k} q times, which is $m^{K^{-2^kq}}$. The worst estimate for the diameter comes at the last refinement, and gives $2m^{-K^{-2^kq}}N$. □

To end this section, we now give four simple lemmas, all of which are closely related to results that have already appeared in this paper (such as Lemma 2.3, Lemma 2.4 and Corollary 2.5). It will be convenient to have them stated explicitly.

LEMMA 5.12. *Let $Q\subset\mathbb{Z}_N$ be a mod-N arithmetic progression of size m. Then Q can be partitioned into $4m^{1/2}$ proper arithmetic progressions.*

Proof. Let $Q=\{a,a+d,\dots,a+(m-1)d\}$. By the pigeonhole principle we can find distinct integers l_1 and l_2 lying in the interval $[0,m^{1/2}]$ such that $|l_1d-l_2d|\leqslant m^{-1/2}N$ and hence l lying in the interval $(0,m^{1/2})$ such that $|ld|\leqslant m^{-1/2}N$. We can partition Q into l mod-N arithmetic progressions $R_1,\dots,R_l$ each of which has common difference ld and length at least $m^{1/2}$. Each R_i can be partitioned into mod-N arithmetic progressions S_j of common difference ld and length between $m^{1/2}$ and m. Of these there can be at most $2m^{1/2}$. Finally, each S_j can be split into at most two parts, each of which is a proper arithmetic progression. □

LEMMA 5.13. *Let $Q_1, \dots, Q_M$ be mod-N arithmetic progressions that form a partition of $\mathbb{Z}_N$. There is a refinement of this partition consisting of at most $4\sqrt{NM}$ proper arithmetic progressions.*

Proof. Let Q_i have cardinality m_i. By Lemma 18.1, one can partition Q_i into at most $4m_i^{1/2}$ proper arithmetic progressions. Since $m_1 + \dots + m_M = N$, the Cauchy-Schwarz inequality tells us that $4(m_1^{1/2} + \dots + m_M^{1/2}) \leqslant 4\sqrt{MN}$. □

LEMMA 5.14. *Let $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ be a polynomial of degree k and let $K = (k!)^2 2^{k^2}$. Let $f : \mathbb{Z}_N \to [-1, 1]$ and let $Q_1, \dots, Q_M$ be arithmetic progressions such that*

$$\sum_{i=1}^{M} \Big| \sum_{s \in Q_i} f(s)\omega^{-\phi(s)} \Big| \geqslant \alpha N \,.$$

There is a refinement of $Q_1, \dots, Q_M$ consisting of arithmetic progressions $R_1, \dots, R_L$ such that $L \leqslant CM^{1/K} N^{1-1/K}$ and

$$\sum_{j=1}^{L} \Big| \sum_{s \in R_j} f(s) \Big| \geqslant (\alpha/2)N \,.$$

Proof. Let m_i be the cardinality of Q_i and let α_i be defined by the equation

$$\Big| \sum_{s \in Q_i} f(s)\omega^{-\phi(s)} \Big| = \alpha_i |Q_i| \,.$$

Our assumption is that $\sum_{i=1}^{M} \alpha_i |Q_i| \geqslant \alpha N$. By Corollary 5.7, each Q_i can be partitioned into at most $Cm_i^{1-1/K}$ subprogressions $Q_{i1}, \dots, Q_{iM_i}$ such that

$$\sum_{j=1}^{M_i} \Big| \sum_{s \in Q_{ij}} f(s) \Big| \geqslant (\alpha_i/2)|Q_i| \,,$$

so, summing over i, we have the inequality

$$\sum_{i=1}^{M} \sum_{j=1}^{M_i} \Big| \sum_{s \in Q_{ij}} f(s) \Big| \geqslant (\alpha/2)N \,.$$

The number of sets we have used is at most $C \sum_{i=1}^{M} m_i^{1-1/K}$. Since $\sum_{i=1}^{M} m_i = N$, this is at most $CM^{1/K} N^{1-1/K}$, by Hölder's inequality. □

LEMMA 5.15. *Let $f : \mathbb{Z}_N \to [-1, 1]$, suppose that $\sum_s f(s) = 0$ and let $P_1, \dots, P_M$ be sets partitioning $\mathbb{Z}_N$ such that*

$$\sum_{j=1}^{M} \Big| \sum_{s \in P_j} f(s) \Big| \geqslant \alpha N \,.$$

Then there exists j such that $\sum_{s\in P_j} f(s) \geqslant \alpha|P_j|/4$ and $|P_j| \geqslant \alpha N/4M$.

Proof. For each j let $a_j = \max\Big\{0, \sum_{s\in P_j} f(s)\Big\}$. The hypotheses about the function f imply that $\sum_{j=1}^M a_j \geqslant \alpha N/2$. However,

$$\sum\big\{a_j : a_j < \alpha|Q_j|/4\big\} < \alpha N/4$$

and

$$\sum\big\{a_j : |Q_j| < \alpha N/4m\big\} < \alpha N/4$$

(as $a_j \leqslant |Q_j|$) so there must be other values of j contributing to $\sum_{j=1}^M a_j$. This proves the lemma. □

6 Somewhat Additive Functions

We saw in §4 that it is possible for a set A to have small Fourier coefficients, but for $A\cap(A+k)$ to have at least one non-trivial large Fourier coefficient for every k. Moreover, the obvious conjecture concerning such sets, that they correlate with some function of the kind $\omega^{q(s)}$ where q is a quadratic polynomial, is false. The aim of the next three sections is to show that such a set A must nevertheless exhibit quadratic bias of some sort. We will then be able to use the results of the last section to find linear bias, which will complete the proof for progressions of length four. The generalization to longer progressions will use similar ideas, but involves one extra important difficulty.

Notice that what we are trying to prove is very natural. If we replace A by a function on $\mathbb{Z}_N$ of the form $f(s) = \omega^{\phi(s)}$, where $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$, then we are trying to prove that if, for many k, the function $\phi_k(s) = \phi(s) - \phi(s-k)$ has some sort of linearity property, resulting in a large Fourier coefficient for the difference function $\Delta(f;k) = \omega^{\phi(s)-\phi(s-k)}$, then ϕ itself must in some way be quadratic. Many arguments in additive number theory (in particular Weyl's inequality) use the fact that taking difference functions reduces the degree of, and hence simplifies, a polynomial. We are trying to do something like the reverse process, "integrating" rather than "differentiating" and showing that the degree goes up by one. This is another sense in which we are engaged in an inverse problem.

This section contains a simple but crucial observation, which greatly restricts the possibilities for the Fourier coefficients of $A\cap(A+k)$ that are large. Let A be a set which is not quadratically α-uniform and let f be the balanced function of A. Then there are at least αN values of k such that

we can find r for which

$$\Big|\sum_s f(s)f(s-k)\omega^{-rs}\Big| \geqslant \alpha N\,.$$

Letting B be the set of k for which such an r exists, we can find a function $\phi : B \to \mathbb{Z}_N$ such that

$$\sum_{k\in B}\Big|\sum_s f(s)f(s-k)\omega^{-\phi(k)s}\Big|^2 \geqslant \alpha^3 N^3\ .$$

We shall show that the function ϕ has a weak-seeming property which we shall call γ-additivity, for a certain constant $\gamma > 0$ to be defined later. Using a variant of Freiman's theorem proved in the next section, we shall show that this property gives surprisingly precise information about ϕ.

PROPOSITION 6.1. *Let $\alpha > 0$, let $f : \mathbb{Z}_N \to D$, let $B \subset \mathbb{Z}_N$ and let $\phi : B \to \mathbb{Z}_N$ be a function such that*

$$\sum_{k\in B}\big|\Delta(f;k)^\wedge\big(\phi(k)\big)\big|^2 \geqslant \alpha N^3\,.$$

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. Expanding the left-hand side of the inequality we are assuming gives us the inequality

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

If we now introduce the variable $u = s-t$ we can rewrite this as

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

Since $|f(x)| \leqslant 1$ for every x, it follows that

$$\sum_u\sum_s\Big|\sum_{k\in B}\overline{f(s-k)}f(s-k-u)\omega^{-\phi(k)u}\Big| \geqslant \alpha N^3\,,$$

which implies that

$$\sum_u\sum_s\Big|\sum_{k\in B}\overline{f(s-k)}f(s-k-u)\omega^{-\phi(k)u}\Big|^2 \geqslant \alpha^2 N^4\,.$$

For each u and x let $f_u(x) = \overline{f(-x)}f(-x-u)$ and let $g_u(x) = B(x)\omega^{\phi(x)u}$. The above inequality can be rewritten

$$\sum_u\sum_s\Big|\sum_k f_u(k-s)\overline{g_u(k)}\Big|^2 \geqslant \alpha^2 N^4\,.$$

By Lemma 2.1, we can rewrite it again as

$$\sum_u \sum_r |\hat{f}_u(r)|^2 |\hat{g}_u(r)|^2 \geqslant \alpha^2 N^5 .$$

Since $\sum_r |\hat{f}(r)|^4 \leqslant N^4$, the Cauchy-Schwarz inequality now implies that

$$\sum_u \Big(\sum_r |\hat{g}_u(r)|^4 \Big)^{1/2} \geqslant \alpha^2 N^3 .$$

Applying the Cauchy-Schwarz inequality again, we can deduce that

$$\sum_{u,r} |\hat{g}_u(r)|^4 = \sum_{u,r} \Big| \sum_{k \in B} \omega^{\phi(s)u - rs} \Big|^4 \geqslant \alpha^4 N^5 .$$

Expanding the left-hand side of this inequality we find that

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

But now the left-hand side is exactly N^2 times the number of quadruples $(a,b,c,d) \in B^4$ for which $a+b=c+d$ and $\phi(a)+\phi(b)=\phi(c)+\phi(d)$. This proves the proposition. □

If G is an Abelian group and a,b,c,d are elements of G such that $a+b=c+d$, we shall say that (a,b,c,d) is an *additive quadruple*. Given a subset $B \subset \mathbb{Z}_N$ and a function $\phi : B \to \mathbb{Z}_N$, let us say that a quadruple $(a,b,c,d) \in B^4$ is *ϕ-additive* if it is additive and in addition $\phi(a)+\phi(b) = \phi(c)+\phi(d)$. Let us say also that ϕ is *γ-additive* if there are at least γN^3 ϕ-additive quadruples. It is an easy exercise to show that if $\gamma = 1$ then B must be the whole of $\mathbb{Z}_N$ and $\phi : \mathbb{Z}_N \to \mathbb{Z}_N$ must be of the form $\phi(x) = \lambda x + \mu$, i.e., linear. Notice that the property of γ-additivity appeared, undefined, in §4 during the discussion of the function ϕ_k. Let us now give a simple but useful reformulation of the concept of γ-additivity.

LEMMA 6.2. *Let $\gamma > 0$, let $B \subset \mathbb{Z}_N$, let $\phi : B \to \mathbb{Z}_N$ be a γ-additive function and let $\Gamma \subset \mathbb{Z}_N^2$ be the graph of ϕ. Then Γ contains at least γN^3 additive quadruples (in the group $\mathbb{Z}_N^2$).* □

As we have just remarked, a 1-additive function must be a linear. We finish this section with an important (and, in the light of the second example of §4, natural) example of a γ-additive function which cannot be approximated by a linear function even though γ is reasonably large. Let $x_1, \dots, x_d \in Z_N$ and $r_1, \dots, r_d \in \mathbb{N}$ be such that all the numbers $\sum_{i=1}^d a_i x_i$ with $0 \leqslant a_i < r_i$ are distinct. Let $y_1, \dots, y_d \in \mathbb{Z}_N$ be arbitrary, and define

$$\phi\Big(\sum_{i=1}^d a_i x_i \Big) = \sum_{i=1}^d a_i y_i .$$

Let $\phi(s)$ be arbitrary for the other values of s. Then a simple calculation shows that the number of additive quadruples is at least $(2/3)^d r_1^3 \dots r_d^3$. If $r_1 \dots r_d = \beta N$, then ϕ is $(2/3)^d \beta^3$-additive.

The function ϕ resembles a linear map between vector spaces, and the number d can be thought of as the dimension of the domain of the ϕ. In the next two sections we shall show that all γ-additive functions have, at least in part, something like the above form, with d not too large and $r_1 \dots r_d$ an appreciable fraction of N (both depending, of course, on γ).

7 Variations on a Theorem of Freiman

Let A be a subset of $\mathbb{Z}$ of cardinality m. It is easy to see that $A+A = \{x+y : x, y \in A\}$ has cardinality between $2m-1$ and $m(m+1)/2$. Suppose that $|A+A| \leqslant Cm$ for some constant C. What information does this give about the set A? This problem is called an *inverse* problem of additive number theory, since it involves deducing the structure of A from the behaviour of $A + A$ – in contrast to a *direct* problem where properties of A give information about $A + A$.

It is clear that $A + A$ will be small when A is a subset of an arithmetic progression of length not much greater than m. After a moment's thought, one realises that there are other examples. For instance, one can take a "progression of progressions" such as $\{aM + b : 0 \leqslant a < h, 0 \leqslant b < k\}$ where $M \gg k$ and $hk = m$. This example can then be generalized to a large subset of a "d-dimensional" arithmetic progression, provided that d is reasonably small. A beautiful and famous result of Freiman asserts that these simple examples exhaust all possibilities. A precise statement of the theorem is as follows.

Theorem 7.1.. *Let C be a constant. There exist constants d_0 and K depending only on C such that whenever A is a subset of $\mathbb{Z}$ with $|A| = m$ and $|A + A| \leqslant Cm$, there exist $d \leqslant d_0$, an integer x_0 and positive integers $x_1, \dots, x_d$ and $k_1, \dots, k_d$ such that $k_1 k_2 \dots k_d \leqslant Km$ and*

$$A \subset \left\{ x_0 + \sum_{i=1}^{d} a_i x_i : 0 \leqslant a_i < k_i \ (i = 1, 2, \dots, d) \right\}.$$

The same is true if $|A - A| \leqslant Cm$.

It is an easy exercise to deduce from Theorem 7.1 the same result for subsets of $\mathbb{Z}^n$, where $x_0, x_1, \dots, x_d$ are now points in $\mathbb{Z}^n$. We shall in fact be interested in the case $n = 2$, since we shall be applying Freiman's theorem

to a graph coming from Proposition 6.1 and Lemma 6.2.

The number $k_1k_2\dots k_d$ is called the *size* of the d-dimensional arithmetic progression. Note that this is not necessarily the same as the cardinality of the set since there may be numbers (or more generally points of $\mathbb{Z}^D$) which can be written in more than one way as $x_0+\sum_{i=1}^d a_ix_i$. When every such representation is unique, we shall call the set a *proper* d-dimensional arithmetic progression. (This terminology is all standard.)

Freiman's original proof of Theorem 7.1 was long and very difficult to understand. Although a simplified version of his argument now exists [Bi], an extremely important breakthrough came a few years ago with a new and much easier proof by Ruzsa, which also provided a reasonable bound. This improved bound is very important for the purposes of our bound for Szemerédi's theorem. Full details of Ruzsa's proof can be found in [Ru1,2,3] or in a book by Nathanson [N], which also contains all necessary background material.

We shall in fact need a modification of Freiman's theorem, in which the hypothesis and the conclusion are weakened. In its qualitative form, the modification is a result of Balog and Szemerédi. However, they use Szemerédi's uniformity lemma, which for us is too expensive. Our argument will avoid the use of the uniformity lemma and thereby produce a much better bound than the bound of Balog and Szemerédi. It will be convenient (though not essential) to consider the version of Freiman's theorem where $A-A$, rather than $A+A$ is assumed to be small. Our weaker hypothesis concerns another parameter associated with a set A, which has several descriptions, and which appeared at the end of the previous section in connection with the graph of the function ϕ. It is

$$\|A*A\|_2^2=\sum_{k\in\mathbb{Z}}\big|A\cap(A+k)\big|^2=\big|\{(a,b,c,d)\in A^4:a-b=c-d\}\big|.$$

(Freiman calls this invariant M' in his book [F2 p. 41].) It is a straightforward exercise to show that

$$\|A*A\|_2^2\leqslant m^2+2\big(1^2+\dots+(m-1)^2\big)$$

with equality if and only if A is an arithmetic progression of length m. The Balog-Szemerédi theorem is the following result.

Theorem 7.2. *Let A be a subset of $\mathbb{Z}^D$ of cardinality m and suppose that $\|A*A\|_2^2\geqslant c_0m^3$. Then there are constants c, K and d_0 depending only on c_0 and an arithmetic progression P of dimension $d\leqslant d_0$ and size at most Km such that $|A\cap P|\geqslant cm$.*

This result states that if $\|A*A\|_2^2$ is, to within a constant, as big as possible, then A has a proportional subset satisfying the conclusion of Freiman's theorem. Notice that, qualitatively at least, the conclusion of Theorem 7.2 cannot be strengthened, since if A has a proportional subset B with $\|B*B\|_2^2$ large, then $\|A*A\|_2^2$ is large whatever $A\setminus B$ is. To see that the new hypothesis is weaker, notice that if $|A-A|\leqslant Cm$, then $A\cap(A+k)$ is empty except for at most Cm values of k, while $\sum_{k\in\mathbb{Z}}|A\cap(A+k)|=m^2$. It follows from the Cauchy-Schwarz inequality that $\sum_{k\in\mathbb{Z}}|A\cap(A+k)|^2\geqslant m^3/C$.

The most obvious approach to deducing Theorem 7.2 from Theorem 7.1 is to show that a set satisfying the hypothesis of Theorem 7.2 has a large subset satisfying the hypothesis of Theorem 7.1. This is exactly what Balog and Szemerédi did and we shall do as well.

PROPOSITION 7.3. *Let A be a subset of $\mathbb{Z}^n$ of cardinality m such that $\|A*A\|_2^2\geqslant c_0m^3$. Then there are constants c and C depending only on c_0 and a subset $A''\subset A$ of cardinality at least cm such that $|A''-A''|\leqslant Cm$. Moreover, c and C can be taken as $2^{-20}c_0^{12}$ and $2^{38}c_0^{-24}$ respectively.*

We shall need the following lemma for the proof.

LEMMA 7.4. *Let V be a set of size m, let $\delta>0$ and let $A_1,\dots,A_n$ be subsets of V such that $\sum_{x=1}^n\sum_{y=1}^n|A_x\cap A_y|\geqslant\delta^2mn^2$. Then there is a subset $K\subset[n]$ of cardinality at least $2^{-1/2}\delta^5n$ such that for at least 90% of the pairs $(x,y)\in K^2$ the intersection $A_x\cap A_y$ has cardinality at least $\delta^2m/2$. In particular, the result holds if $|A_x|\geqslant\delta m$ for every x.*

Proof. For every $j\leqslant m$ let $B_j=\{i:j\in A_i\}$ and let $E_j=B_j^2$. Choose five numbers $j_1,\dots,j_5\leqslant m$ at random (uniformly and independently), and let $X=E_{j_1}\cap\dots\cap E_{j_5}$. The probability p_{xy} that a given pair $(x,y)\in[n]^2$ belongs to E_{j_r} is $m^{-1}|A_x\cap A_y|$, so the probability that it belongs to X is p_{xy}^5. By our assumption we have that $\sum_{x,y=1}^n p_{xy}\geqslant\delta^2n^2$, which implies (by Hölder's inequality) that $\sum_{x,y=1}^n p_{xy}^5\geqslant\delta^{10}n^2$. In other words, the expected size of X is at least $\delta^{10}n^2$.

Let Y be the set of pairs $(x,y)\in X$ such that $|A_x\cap A_y|<\delta^2m/2$, or equivalently $p_{xy}<\delta^2/2$. Because of the bound on p_{xy}, the probability that $(x,y)\in Y$ is at most $(\delta^2/2)^5$, so the expected size of Y is at most $\delta^{10}n^2/32$.

It follows that the expectation of $|X|-16|Y|$ is at least $\delta^{10}n^2/2$. Hence, there exist $j_1,\dots,j_5$ such that $|X|\geqslant16|Y|$ and $|X|\geqslant\delta^{10}n^2/2$. It follows that the set $K=B_{j_1}\cap\dots\cap B_{j_5}$ satisfies the conclusion of the lemma. □

Proof of Proposition 7.3. The function $f(x) = A * A(x)$ (from $\mathbb{Z}^n$ to $\mathbb{Z}$) is non-negative and satisfies $\|f\|_\infty \leqslant m$, $\|f\|_2^2 \geqslant c_0 m^3$ and $\|f\|_1 = m^2$. This implies that $f(x) \geqslant c_0 m/2$ for at least $c_0 m/2$ values of x, since otherwise we could write $f = g + h$ with g and h disjointly supported, g supported on fewer than $c_0 m/2$ points and $\|h\|_\infty \leqslant c_0 m/2$, which would tell us that

$$\|f\|_2^2 \leqslant \|g\|_2^2 + \|h\|_\infty \|h\|_1 < (c_0 m/2)m^2 + (c_0 m/2).m^2 = c_0 m^3 .$$

Let us call a value of x for which $f(x) \geqslant c_0 m/2$ a *popular difference* and let us define a graph G with vertex set A by joining a to b if $b - a$ (and hence $a - b$) is a popular difference. The average degree in G is at least $c_0^2 m/4$, so there must be at least $c_0^2 m/8$ vertices of degree at least $c_0^2 m/8$. Let $\delta = c_0^2/8$, let $a_1, \dots, a_n$ be vertices of degree at least $c_0^2 m/8$, with $n \geqslant \delta m$, and let $A_1, \dots, A_n$ be the neighbourhoods of the vertices $a_1, \dots, a_n$. By Lemma 7.4 we can find a subset $A' \subset \{a_1, \dots, a_n\}$ of cardinality at least $\delta^5 n/\sqrt{2}$ such that at least 90% of the intersections $A_i \cap A_j$ with $a_i, a_j \in A'$ are of size at least $\delta^2 m/2$. Set $\alpha = \delta^6/\sqrt{2}$ so that $|A'| \geqslant \alpha m$.

Now define a graph H with vertex set A', joining a_i to a_j if and only if $|A_i \cap A_j| \geqslant \delta^2 m/2$. The average degree of the vertices in H is at least $(9/10)|A'|$, so at least $|A'|/2$ vertices have degree at least $4|A'|/5$. Define A'' to be the set of all such vertices.

We claim now that A'' has a small difference set. To see this, consider any two elements $a_i, a_j \in A''$. Since the degrees of a_i and a_j are at least $(4/5)|A'|$ in H, there are at least $(3/5)|A'|$ points $a_k \in A'$ joined to both a_i and a_j. For every such k we have $|A_i \cap A_k|$ and $|A_j \cap A_k|$ both of size at least $\delta^2 m/2$. If $b \in A_i \cap A_k$, then both $a_i - b$ and $a_k - b$ are popular differences. It follows that there are at least $c_0^2 m^2/4$ ways of writing $a_i - a_k$ as $(p - q) - (r - s)$, where $p, q, r, s \in A$, $p - q = a_i - b$ and $r - s = a_k - b$. Summing over all $b \in A_i \cap A_k$, we find that there are at least $\delta^2 c_0^2 m^3/8$ ways of writing $a_i - a_k$ as $(p - q) - (r - s)$ with $p, q, r, s \in A$. The same is true of $a_j - a_k$. Finally, summing over all k such that a_k is joined in H to both a_i and a_j, we find that there are at least $(3/5)|A'|\delta^4 c_0^4 m^6/64 \geqslant \alpha\delta^4 c_0^4 m^7/120$ ways of writing $a_i - a_j$ in the form $(p - q) - (r - s) - \big((t - u) - (v - w)\big)$ with $p, q, \dots, w \in A$.

Since there are at most m^8 elements in A^8, the number of differences of elements of A'' is at most $120m/\alpha\delta^4 c_0^4 \leqslant 2^{38} m/c_0^{24}$. Note also that the cardinality of A'' is at least $(1/2)\alpha m \geqslant c_0^{12} m/2^{20}$. The proposition is proved. □

It is possible to apply Theorem 7.2 as it stands in order to prove Szemerédi's theorem for progressions of length four (and quite possibly in

general). Instead, we shall combine Proposition 7.3 with a weaker version of Freiman's theorem that gives less information about the structure of a set A with small difference set. There are three advantages in doing this. The first is that with our weaker version we can get a much better bound. The second is that using the weaker version is cleaner, particularly when we come to the general case. The third is that the weaker version is easier to prove than Freiman's theorem itself, as it avoids certain arguments from the geometry of numbers.

We shall not be concerned in this paper with arbitrary sets A such that $|A-A| \leqslant C|A|$, but rather with graphs of functions from subsets of $\mathbb{Z}_N$ to $\mathbb{Z}_N$. We now prove a result for such functions. An important concept introduced by Freiman is that of a Freiman homomorphism (as it is now called). Let A and B be two subsets of Abelian groups. A function $\phi : A \to B$ is a *Freiman homomorphism of order k* if, whenever $a_1, \dots, a_{2k} \in A$ and

$$a_1 + \cdots + a_k = a_{k+1} + \cdots + a_{2k}\,,$$

we have also

$$\phi(a_1) + \cdots + \phi(a_k) = \phi(a_{k+1}) + \cdots + \phi(a_{2k})\,.$$

Equivalently, ϕ induces a well-defined function from kA to kB, where kA denotes the sum of k copies of the set A. When $k = 2$ one speaks simply of a *Freiman homomorphism*. Note that a Freiman homomorphism of order $2k$ also induces a well-defined function from $kA - kA$ to $kB - kB$. If ϕ has an inverse which is also a Freiman homomorphism of order k, then ϕ is said to be a *Freiman isomorphism of order k*. The next lemma shows that a function $\phi : B \subset \mathbb{Z}_N \to \mathbb{Z}_N$ for which the graph has a small difference set can be restricted to a large subset of B on which it is a Freiman homomorphism of order k. This lemma plays the role in our proof that Theorem 2 of [Ru1] did in Ruzsa's proof, and the proof is in a very similar spirit. Indeed, the whole scheme of our proof in the rest of this section is based on his ideas.

LEMMA 7.5. *Let $B \subset \mathbb{Z}_N$ and let $\phi : B \to \mathbb{Z}_N$ be a function with graph Γ. Suppose that $|\Gamma - \Gamma| \leqslant C|\Gamma|$. Then there is a subset $B' \subset B$ of size at least $|B|/8kC^{4k}$ such that the restriction of ϕ to C is a Freiman homomorphism of order k.*

Proof. First, a theorem of Ruzsa [Ru2] (deduced from a result of Plünnecke [P] for which Ruzsa discovered a simpler proof) implies that $|4k\Gamma - 4k\Gamma| \leqslant C^{4k}|\Gamma|$. If for some x we could find more than C^{4k} distinct values of y such

that $(x,y) \in 2k\Gamma - 2k\Gamma$, then for every $(z,w) \in 2k\Gamma - 2k\Gamma$ there would be more than C^{4k} distinct values of u such that $(z-x,u) \in 4k\Gamma - 4k\Gamma$. But the number of z such that $(z,w) \in 2k\Gamma - 2k\Gamma$ for some w is certainly at least $|\Gamma|$, so this would contradict the upper bound for $|4k\Gamma - 4k\Gamma|$.

Therefore, there are in particular at most C^{4k} distinct values of y such that $(0,y) \in 2k\Gamma - 2k\Gamma$. If $(x,y),(x,y') \in k\Gamma - k\Gamma$ then $(0,y-y') \in 2k\Gamma - 2k\Gamma$. Hence, there is a set K of size at most C^{4k} such that, writing K_x for the set $\{y : (x,y) \in k\Gamma - k\Gamma\}$, we have $K_x - K_x \subset K$ for every x.

Now let $0 \leqslant M < N/2$ be even. For every $w \in \mathbb{Z}_N$, there are exactly $2M$ non-zero values of d such that

$$w \in \big\{ -Md, -(M-1)d, \dots, -2d, -d\} \cup \{d, 2d, \dots, Md\big\},$$

since the equation $ad = w$ has a unique solution whenever $a \neq 0$. Therefore, the number of values of d for which $K \cap \{dy : -M \leqslant y \leqslant M\} \neq \{0\}$ is at most $2MC^{4k}$.

Let d be such that if we define P to be $\{dy : -M \leqslant y \leqslant M\}$ then $K \cap P = \{0\}$. Let $P' = \{dy : -M/2 \leqslant y \leqslant M/2\}$, let $L \leqslant M/2k$ and let $Q = \{dy : 0 \leqslant y \leqslant L\}$. Define Γ_a to be the set $\{(x,y) \in \Gamma : y \in a + Q\}$.

We claim that Γ_a is the graph of a homomorphism of order k. If not, then we can find $(x_1,y_1),\dots,(x_{2k},y_{2k})$ and $(x'_1,y'_1),\dots,(x'_{2k},y'_{2k}) \in \Gamma_a$ such that

$$x_1 + \cdots + x_k - x_{k+1} - \cdots - x_{2k} = x'_1 + \cdots + x'_k - x'_{k+1} - \cdots - x'_{2k},$$

but

$$y_1 + \cdots + y_k - y_{k+1} - \cdots - y_{2k} \neq y'_1 + \cdots + y'_k - y'_{k+1} - \cdots - y'_{2k},$$

and hence x, y, y' such that $y \neq y'$ and $(x,y),(x,y') \in k\Gamma_a - k\Gamma_a$. However, $k\Gamma_a - k\Gamma_a$ is the set of all points of the form

$$(x_1 + \cdots + x_k - x_{k+1} - \cdots - x_{2k}, y_1 + \cdots + y_k - y_{k+1} - \cdots - y_{2k})$$

such that $(x_i,y_i) \in \Gamma$ and $y_i \in a + Q$ for every i. This is a subset of $\{(x,y) \in k\Gamma - k\Gamma : y \in P'\} = \{(x,y) : y \in K_x \cap P'\}$. It follows that $(K_x - K_x) \cap (P' - P')$ is non-empty and hence that $K \cap P$ is non-empty, which is a contradiction.

Therefore, as long as $2MC^{4k} < N - 1$, we can find a value of d such that Γ_a is the graph of a homomorphism for every a. The average size of Γ_a is $(L+1)|\Gamma|/N$, so if we choose M to be at least $N/4C^{4k}$ and L to be at least $(M/2k) - 1$, as we may, then we can find a such that the size of Γ_a is at least $|\Gamma|/8kC^{4k}$. □

Let us now collect what we have done so far into a single result, specialized to the case $k = 8$.

COROLLARY 7.6. *Let $B_0 \subset \mathbb{Z}_N$ have cardinality αN, and let $\phi : B_0 \to \mathbb{Z}_N$ have $\gamma(\alpha N)^3$ additive quadruples. Then there is a subset $B \subset B_0$ of cardinality at least $2^{-1882}\gamma^{1164}\alpha N$ such that the restriction of ϕ to B is a homomorphism of order 8.*

Proof. By Proposition 7.3 we can find a subset $B_1 \subset B_0$ of cardinality at least $2^{-20}\gamma^{12}\alpha N$ such that, letting Γ be the graph of ϕ restricted to B_1, we have $|\Gamma - \Gamma| \leqslant 2^{58}\gamma^{-36}|\Gamma|$. Let $C = 2^{58}\gamma^{-36}$. By Lemma 7.5 we can restrict ϕ to a subset $B \subset B_1$ of cardinality at least $|B_1|/64C^{32} \geqslant 2^{-1882}\gamma^{1164}\alpha N$ such that it becomes a homomorphism of order 8. □

The next lemma is a variant of an argument of Bogolyubov [B]. The original argument was used by Ruzsa in his proof of Freiman's theorem. Given a subset $K \subset \mathbb{Z}_N$ and $\delta > 0$, let us define the *Bohr neighbourhood* $B(K, \delta)$ to be the set of all $d \in \mathbb{Z}_N$ such that $|sd| \leqslant \delta N$ for every $s \in K$. An elementary fact about Bohr neighbourhoods is contained in the next lemma, which is another well-known application of Dirichlet's "box" principle.

LEMMA 7.7. *Let K be a subset of $\mathbb{Z}_N$ and let $\delta > 0$. Then the cardinality of the Bohr neighbourhood $B(K, \delta)$ is at least $(\delta/2)^{|K|}N$. In particular, if $\delta > (N/2)^{-1/|K|}$ then $B(K, \delta)$ contains a non-zero element.*

Proof. Let the elements of K be $r_1, \dots, r_k$, and let ϕ be the mapping from $\mathbb{Z}_N$ to $\mathbb{Z}_N^k$ defined by $\phi : x \mapsto (r_1 x, \dots, r_k x)$. Let $m = \lceil \delta^{-1} \rceil$ and for $1 \leqslant j \leqslant m$ let $I_j = \{x \in \mathbb{Z}_N : (j-1)N/m \leqslant x < jN/m\}$. There are exactly m^k possible products of k of the intervals I_j, so one of them, Q say, must contain $\phi(x)$ for at least $m^{-k}N$ values of $x \in \mathbb{Z}_N$. Let C be the set of x such that $\phi(x) \in Q$. Then it is easy to see that $C - C \subset B$. Clearly also $|C - C| \geqslant |C|$. The lemma now follows from the observation that $m^{-1} \geqslant \delta/2$. □

Another useful remark about Bohr neighbourhoods is that $B(K, \delta_1) + B(K, \delta_2) \subset B(K, \delta_1 + \delta_2)$. Further facts about them will be proved in §10.

LEMMA 7.8. *Let $A \subset \mathbb{Z}_N$ be a set of size αN and let $\phi : A \to \mathbb{Z}_N$ be a Freiman homomorphism of order 8. Let $K = \{r \in \mathbb{Z}_N : |\hat{A}(r)| \geqslant \alpha^{3/2}N/4\}$. Then K has cardinality at most $16\alpha^{-2}$, and there is a homomorphism $\psi : B(K, \alpha/32\pi) \to \mathbb{Z}_N$ such that $\phi(x) - \phi(y) = \psi(x - y)$ whenever $x, y \in A$ and $x - y \in B(K, \alpha/32\pi)$.*

Proof. Let g be the function $A * A * A * A$. Then $\hat{g}(r) = |\hat{A}(r)|^4$ and $g(r) = N^{-1}\sum_r |\hat{A}(r)|^4 \omega^{rx}$ for every $r \in \mathbb{Z}_N$. Let $\lambda = \alpha^{3/2}/4$ so that

$K = \{r : |\hat{A}(r)| \geqslant \lambda N\}$. Since $\|\hat{A}\|^2 = \alpha N^2$, we have $\lambda^2 N^2 |K| \leqslant \alpha N^2$ and hence $|K| \leqslant \alpha\lambda^{-2} = 16\alpha^{-2}$ as stated. We also know that

$$\sum_{r \notin K} |\hat{A}(r)|^4 < \lambda^2 N^2 \sum_{r \notin K} |\hat{A}(r)|^2 \leqslant \alpha\lambda^2 N^4 .$$

Therefore, if we define $h(x)$ to be $N^{-1}\sum_{r\in K} |\hat{A}(r)|^4 \omega^{rx}$, we find that $|g(x) - h(x)| \leqslant \alpha\lambda^2 N^3$ for every x.

Now choose d such that $|rd| \leqslant \alpha N/32\pi$ for every $r \in K$. Then, for every x,

$$\begin{aligned} |h(x+d) - h(x)| &= N^{-1}\Big|\sum_{r\in K} |\hat{A}(r)|^4(\omega^{r(x+d)} - \omega^{rx})\Big| \\ &\leqslant N^{-1}\sum_{r\in K} |\hat{A}(r)|^4 |\omega^{rd} - 1| \\ &\leqslant 2\pi(\alpha/32\pi)\alpha^3 N^3 = \alpha\lambda^2 N^3 , \end{aligned}$$

where for the last inequality we used the fact that $\sum_r |\hat{A}(r)|^4 \leqslant (\alpha N)^2 \sum_r |\hat{A}(r)|^2 = \alpha^3 N^4$. It follows that, under the same condition on d, we have

$$\big|g(x+d) - g(x)\big| \leqslant 3\alpha\lambda^2 N^3$$

for every x.

Since $g(0) \geqslant N^{-1}|\hat{A}(0)|^4 = \alpha^4 N^3 = 4\alpha\lambda^2 N^3$, it follows that $g(d) > 0$ for every $d \in B = B(K, \alpha/32\pi)$, so $B \subset 2A - 2A$. Now ϕ induces a homomorphism ψ_0 (of order 2) on $2A - 2A$, which therefore restricts to a homomorphism ψ on B. If $x, y \in A$ with $x - y = d \in B$, then $\psi(d) = \phi(x) + \phi(x) - \phi(x) - \phi(y) = \phi(x) - \phi(y)$. □

Remark. Notice that the same result holds, with an almost identical proof, if ϕ maps A into a general Abelian group G rather than $\mathbb{Z}_N$.

Given that ϕ was already a homomorphism of order 8 in the statement of Lemma 7.8, the reader may be excused for wondering what has been gained in the conclusion. The answer is that $B = B(K, \alpha/32\pi)$ has so much structure, in particular containing many long arithmetic progressions, that much more can be said about homomorphisms defined on B than about homomorphisms on arbitrary sets. The next corollary illustrates this.

Corollary 7.9. *Let A and K be as in Lemma 7.8 and let m be a positive integer. For every $d \in B(K, \alpha/32\pi m)$ there exists c such that $\phi(x) - \phi(y) = c(x - y)$ whenever $x - y$ belongs to the set $\{jd : -m \leqslant j \leqslant m\}$.*

Proof. This follows from Lemma 7.8 together with the observations that $\{jd : -m \leqslant j \leqslant m\} \subset B(K, \alpha/32\pi)$, that the restriction of any homomorphism to $\{jd : -m \leqslant j \leqslant m\}$ is linear and that $\psi(0) = 0$. □

We now give a useful definition which arises naturally out of the statement of Lemma 7.8. Let $A, B \subset \mathbb{Z}_N$ and let $\phi : A \to \mathbb{Z}_N$. We shall say that ϕ is a *B-homomorphism* if there is a homomorphism $\psi : B \to \mathbb{Z}_N$ such that, whenever $x, y \in A$ and $x - y = z$ with $z \in B$, we have $\phi(x) - \phi(y) = \psi(z)$. In other words, ϕ induces a homomorphism on $(A - A) \cap B$.

The last two results of this section are once again simply a putting together of earlier results.

COROLLARY 7.10.. *Let N be sufficiently large, let $B_0 \subset \mathbb{Z}_N$ have cardinality αN and let $\phi : B_0 \to \mathbb{Z}_N$ have $\gamma(\alpha N)^3$ additive quadruples. Then there exist a mod-N arithmetic progression P of length at least $N^{2^{-3770}\gamma^{2328}\alpha^2}$, a subset $H \subset P$ of cardinality at least $2^{-1849}\gamma^{1164}\alpha|P|$ and constants $\lambda, \mu \in \mathbb{Z}_N$ such that $\phi(s) = \lambda s + \mu$ for every $s \in H$.*

Proof. Corollary 7.6 says that there is a subset $B \subset B_0$ of cardinality at least βN, where $\beta = 2^{-1882}\gamma^{1164}\alpha$, such that the restriction of ϕ to B is a Freiman homomorphism of order 8. To this pair (B, ϕ) we apply Corollary 7.9. Let K be the set of size at most $16\beta^{-2}$ coming from Corollary 7.8. By Lemma 7.7, the Bohr neighbourhood $B(K, \beta/32\pi m)$ has a non-zero element if $\beta/32\pi m > (N/2)^{-\beta^2/16}$. Assume that m is chosen so that this inequality is satisfied and let d be a non-zero element of $B(K, \beta/32\pi m)$. Let P_0 be the mod-N arithmetic progression $(d, 2d, \dots, md)$. By an easy averaging argument, there exists $k \in \mathbb{Z}_N$ such that $|(P_0 + k) \cap B| \geqslant \beta m$. Choose such a k and let $P = P_0 + k$ and $H = P \cap B$. Since $x - y \in \{jd : -m \leqslant j \leqslant m\}$ whenever $x, y \in P$, Corollary 7.9 gives us a constant $c \in \mathbb{Z}_N$ such that $\phi(x) - \phi(y) = c(x - y)$ for every $x, y \in H$. It remains only to check that if N is sufficiently large then there exists an integer $m \geqslant N^{2^{-3770}\gamma^{2328}\alpha^2}$ such that $\beta/32m > (N/2)^{-\beta^2/16}$. This is a calculation left to the reader, but we state here for later reference that N can be taken to be $(2\gamma^{-1}\alpha^{-1})^{2^{4000}\gamma^{-2328}\alpha^{-2}}$. □

The final result will be used when $q = 1$ in the proofs of Lemmas 13.7 and 13.9 and for general q in the proof of Lemma 16.3. Unlike our previous results, it applies to subsets of arithmetic progressions rather than subsets of $\mathbb{Z}_N$.

COROLLARY 7.11. *Let R be an arithmetic progression in $\mathbb{Z}$, for $1 \leqslant i \leqslant q$ let $A_i \subset R$ be a set of cardinality at least $\alpha|R|$ and for each i let $\phi_i : A_i \to \mathbb{Z}_N$*

be a homomorphism of order 8. As long as $m \leqslant |R|^{2^{-14}\alpha^2 q^{-1}}$ *it is possible to partition* R *into arithmetic progressions* $S_1, \dots, S_M$*, all of size* m *or* $m+1$ *and all with the same common difference, such that the restriction of any* ϕ_i *to any* $A_i \cap S_j$ *is linear.*

Proof. Let $R = \{a, a+h, \dots, a+(l-1)h\}$. We can embed R 8-isomorphically into $\mathbb{Z}_p$ for a prime $p < 16l$ using the map $\iota : a + jh \mapsto j$. Let $A_i' = \iota A_i$ and let $\phi_i' = \phi_i \iota^{-1}$. (In other words, let us regard each A_i as a subset of $\mathbb{Z}_p$.) We know that $|A_i'| \geqslant \alpha p/16$ for every i. We shall now apply Lemma 7.8, with α replaced by $\alpha/16$, to $A_i' \subset \mathbb{Z}_p$ and ϕ_i', which maps A_i' to $\mathbb{Z}_N$ (see the remark following Lemma 7.8).

Let $L = \{1\} \cup \{r \in \mathbb{Z}_p : |\hat{A}_i'(r)| \geqslant \alpha^{3/2} p/256 \text{ for some } i\}$. By Lemma 7.8 we know that $|L| \leqslant 2^{12}\alpha^{-2}q+1$ and that for each i there is a homomorphism $\psi_i : B(L, \alpha/512\pi) \to \mathbb{Z}_N$ such that $\phi_i'(x) - \phi_i'(y) = \psi(x-y)$ whenever $x - y \in B(L, \alpha/512\pi)$. By Lemma 7.7, $B(L, \alpha/512\pi m^2)$ contains a non-zero element d. Because $1 \in L$, we know that $|d| \leqslant \alpha p/512\pi m^2$, which implies that $\mathbb{Z}_p$ can be partitioned into (genuine) arithmetic progressions each of which has common difference d and length at least m^2. We can then partition these progressions into further subprogressions of length m or $m+1$. As in the proof of Corollary 7.9, for each i there exists c_i such that if S is one of these subprogressions and $x, y \in S$, then $\phi_i'(x) - \phi_i'(y) = c_i(x-y)$. The corollary follows on using ι^{-1} to transfer us back to R, A_i and ϕ_i. □

8 Progressions of Length Four

We have now shown that if $A \cap (A+k)^{\sim}(\phi(k))$ is large for many values of k then ϕ resembles a linear function. If ϕ *is* linear, then the rest of the argument is simple. Indeed, suppose that $\phi(k) = 2ck$ for every k, for some constant $c \in \mathbb{Z}_N$. Then inequality (6.1) becomes

$$\sum_k \sum_{s,u} A(s)A(s-k)A(s-u)A(s-k-u)\omega^{-2cku} \geqslant \alpha^3 N^3 .$$

Using the identity

$$2ku = s^2 - (s-k)^2 - (s-u)^2 + (s-k-u)^2 ,$$

we can deduce that

$$\sum_r \sum_{a,b,c,d} A(a)A(b)A(c)A(d)\omega^{-r(a-b-c+d)}\omega^{-c(a^2-b^2-c^2+d^2)} \geqslant \alpha^3 N^4 ,$$

or in other words that

$$\sum_r \Big| \sum_s A(s)\omega^{-cs^2}\omega^{-rs} \Big|^4 \geqslant \alpha^3 N^4 .$$

By the implication of (iii) from (iv) in Lemma 2.2, this tells us that for some value of r we have the lower bound

$$\Big|\sum_s A(s)\omega^{-cs^2}\omega^{-rs}\Big| \geqslant \alpha^{3/2}N\,,$$

or in other words that A exhibits quadratic bias of a particularly strong kind. The aim of this section is to give a similar argument that shows the existence of quadratic bias under the weaker assumption that ϕ has a reasonably large linear part, such as is guaranteed by Corollary 7.10.

Let us remind ourselves why this is needed. We are examining sets $A \subset \mathbb{Z}_N$ that fail to be quadratically α-uniform. Let A be such a set and let f be the balanced function of A. Then there is a subset $B \subset \mathbb{Z}_N$ of cardinality at least αN, and a function $\phi : B \to \mathbb{Z}_N$ such that $|\Delta(f;k)\hat{}(\phi(k))| \geqslant \alpha N$ for every $k \in B$. By Proposition 6.1 we know that B contains at least $\alpha^{12}N^3$ additive quadruples for the function ϕ. Corollary 7.10 then implies that ϕ can be restricted to a large arithmetic progression P where it often agrees with a linear function $s \mapsto as + b$. This provides the motivation for the next proposition.

PROPOSITION 8.1. *Let $A \subset \mathbb{Z}_N$ have balanced function f. Let P be an arithmetic progression (in $\mathbb{Z}_N$) of cardinality T. Suppose that there exist λ and μ such that $\sum_{k\in P}|\Delta(f;k)\hat{}(\lambda k+\mu)|^2 \geqslant \beta N^2T$. Then there exist quadratic polynomials $\psi_0, \psi_1, \dots, \psi_{N-1}$ such that*

$$\sum_s \Big|\sum_{z\in P+s} f(z)\omega^{-\psi_s(z)}\Big| \geqslant \beta NT/\sqrt{2}\,.$$

Proof. Expanding the assumption we are given, we obtain the inequality

$$\sum_{k\in P}\sum_{s,t} f(s)f(s-k)f(t)f(t-k)\omega^{-(\lambda k+\mu)(s-t)} \geqslant \beta N^2T\,.$$

Substituting $u = s - t$, we deduce that

$$\sum_{k\in P}\sum_{s,u} f(s)f(s-k)f(s-u)f(s-k-u)\omega^{-(\lambda k+\mu)u} \geqslant \beta N^2T\,.$$

Let $P = \{x + d, x + 2d, \dots, x + Td\}$. Then we can rewrite the above inequality as

$$\sum_{i=1}^{T}\sum_{s,u} f(s)f(s-x-id)f(s-u)f(s-x-id-u)\omega^{-(\lambda x+\lambda id+\mu)u} \geqslant \beta N^2T\,. \quad (*)$$

Since there are exactly T ways of writing $u = y + jd$ with $y \in \mathbb{Z}_N$ and $1 \leqslant j \leqslant T$, we can rewrite the left-hand side above as

$$\frac{1}{T}\sum_s\sum_{i=1}^T\sum_y\sum_{j=1}^T f(s)f(s-x-id)f(s-y-jd)$$
$$\cdot f(s-x-id-y-jd)\omega^{-(\lambda x+\lambda id+\mu)(y+jd)}\,.$$

Let us define $\gamma(s,y)$ by the equation

$$\Bigl|\sum_{i=1}^T\sum_{j=1}^T f(s-x-id)f(s-y-jd)f(s-x-id-y-jd)\omega^{-(\lambda x+\mu+\lambda id)(y+jd)}\Bigr|$$
$$=\gamma(s,y)T^2\,.$$

Since $|f(s)|\leqslant 1$, $(*)$ tells us that the average value of $\gamma(s,y)$ is at least β.

In general, suppose we have real functions f_1, f_2 and f_3 such that

$$\Bigl|\sum_{i=1}^T\sum_{j=1}^T f_1(i)f_2(j)f_3(i+j)\omega^{-(ai+bj-2cij)}\Bigr|\geqslant cT^2\,.$$

Since $2cij=c((i+j)^2-i^2-j^2)$, we can rewrite this as

$$\Bigl|\sum_{i=1}^T\sum_{j=1}^T f_1(i)\omega^{-(ai+ci^2)}f_2(i)\omega^{-(bj+cj^2)}f_3(i+j)\omega^{c(i+j)^2}\Bigr|\geqslant cT^2$$

and then replace the left-hand side by

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

If we now set $g_1(r)=\sum_{i=1}^T f_1(i)\omega^{-(ai+ci^2)}\omega^{-ri}$, $g_2(r)=\sum_{j=1}^T f_2(j)\omega^{-(bj+cj^2)}\omega^{-rj}$ and $g_3(r)=\sum_{k=1}^{2T}f_3(k)\omega^{-ck^2}\omega^{-rk}$, then we have

$$\Bigl|\sum_r g_1(r)g_2(r)g_3(r)\Bigr|\geqslant cT^2N\,,$$

which implies, by the Cauchy-Schwarz inequality, that $\|g_1\|_\infty\|g_2\|_2\|g_3\|_2\geqslant cT^2N$. Since $\|g_2\|_2^2\leqslant NT$ and $\|g_3\|_2^2\leqslant 2NT$ (by identity (3) of §2), this tells us that $|g_1(r)|\geqslant cT/\sqrt{2}$ for some r. In particular, there exists a quadratic polynomial ψ such that $\bigl|\sum_{i=1}^T f_1(i)\omega^{-\psi(i)}\bigr|\geqslant cT/\sqrt{2}$.

Let us apply this general fact to the functions $f_1(i)=f(x-s-id)$, $f_2(j)=f(s-y-jd)$ and $f_3(k)=f(s-x-y-kd)$. It gives us a quadratic polynomial $\psi_{s,y}$ such that

$$\Bigl|\sum_{i=1}^T f(s-x-id)\omega^{-\psi_{s,y}(i)}\Bigr|\geqslant\gamma(s,y)T/\sqrt{2}\,.$$

Let $\gamma(s)$ be the average of $\gamma(s,y)$, and choose ψ_s to be one of the $\psi_{s,y}$ in such a way that

$$\Big|\sum_{i=1}^{T} f(s-x-id)\omega^{-\psi_s(i)}\Big| \geqslant \gamma(s)T/\sqrt{2}\,.$$

If we now sum over s, we have the required statement (after a small change to the definition of the ψ_s). □

Theorem 8.2. *There is an absolute constant C with the following property. Let A be a subset of $\mathbb{Z}_N$ with cardinality δN. If $N \geqslant \exp\exp\big((1/\delta)^C\big)$, then A contains an arithmetic progression of length four.*

Proof. Our assumption certainly implies that $N \geqslant 32k^2\delta^{-k}$. Suppose now that the result is false. Then Corollary 3.6 implies that A is not α-quadratically uniform, where $\alpha = (\delta/2)^{64}$. By Lemma 3.1 (in particular the implication of (i) from (v)) there is a set $B \subset \mathbb{Z}_N$ of cardinality at least $\alpha N/2$ together with a function $\phi : B \to \mathbb{Z}_N$, such that $|\Delta(f;k)^\wedge(\phi(k))| \geqslant \alpha N/2$ for every $k \in B$. In particular,

$$\sum_{k\in B} \big|\Delta(f;k)^\wedge(\phi(k))\big|^2 \geqslant (\alpha/2)^3 N^3\,.$$

Hence, by Proposition 6.1, B contains at least $(\alpha/2)^{12}N^3$ ϕ-additive quadruples.

By Corollary 7.10, we can find a mod-N arithmetic progression P of size at least $N^{2^{-32000}\alpha^{30000}}$ and constants $\lambda, \mu \in \mathbb{Z}_N$ such that

$$\sum_{k\in P} \big|\Delta(f;k)^\wedge(\lambda k+\mu)\big|^2 \geqslant 2^{-16000}\alpha^{15000}|P|N^2\,.$$

Therefore, by Proposition 8.1, we have quadratic polynomials $\psi_0, \psi_1,...,\psi_{N-1}$ such that

$$\sum_s \Big|\sum_{z\in P+s} f(z)\omega^{-\psi_s(z)}\Big| \geqslant \beta N|P|/\sqrt{2}$$

where $\beta = 2^{-16000}\alpha^{15000}$.

By a simple averaging argument we can find a partition of $\mathbb{Z}_N$ into mod-N arithmetic progressions $P_1, \dots, P_M$ of length $|P|$ or $|P|+1$ and also a sequence $\psi_1, \dots, \psi_M$ (after renaming) of quadratic polynomials such that

$$\sum_{j=1}^{M} \Big|\sum_{z\in P_j} f(z)\omega^{-\psi_j(z)}\Big| \geqslant \beta N/2\,.$$

(Each P_j is either a translate of P or a translate of P extended by one point. Because of the small extensions we have changed $\sqrt{2}$ to 2.) By Lemma 5.13

we can refine this partition and produce a partition into genuine arithmetic progressions $Q_1, \dots, Q_L$, which automatically satisfy an inequality of the form

$$\sum_{j=1}^{M} \Big| \sum_{z \in Q_j} f(z) \omega^{-\psi_j(z)} \Big| \geqslant \beta N/2 \,.$$

Once again, we have renamed the functions ψ_j. Lemma 5.13 allows us to take $L \leqslant N^{1-2^{-32002}\alpha^{30000}}$. Next, Lemma 5.14 gives us a further refinement of $Q_1, \dots, Q_L$ into arithmetic progressions $R_1, \dots, R_H$ such that

$$\sum_{i=1}^{H} \Big| \sum_{s \in R_i} f(s) \Big| \geqslant \beta N/4$$

and H is at most $N^{1-2^{-32010}\alpha^{30000}}$ Finally, Lemma 5.15 gives us an arithmetic progression R of cardinality at least $\beta N^{2^{-32010}\alpha^{30000}}$ such that $\sum_{s \in R} f(s) \geqslant \beta |R|/16$. This implies that the cardinality of $A \cap R$ is at least $|R|(\delta + 2^{-16004}\alpha^{15000})$. Recalling that $\alpha = (\delta/2)^{64}$, we find that the density of A has gone up from δ in $\mathbb{Z}_N$ to at least $\delta(1 + (\delta/2)^{980000})$ inside the arithmetic progression R.

We now iterate this argument. The iteration can be performed at most $(\delta/2)^{-1000000}$ times, and at each step the value of N is raised to a power which exceeds $(\delta/2)^{2000000}$. It is not hard to check that N will always remain sufficiently large for the argument to work, as long as the initial value of N is at least $\exp\exp(\delta^{-C})$, where C can be taken to be 2000000. □

An alternative formulation of the condition on N and δ is that δ should be at least $(\log\log N)^{-c}$ for some absolute constant $c > 0$. We have the following immediate corollary.

COROLLARY 8.3. *There is an absolute constant $c > 0$ with the following property. If the set $\{1, 2, \dots, N\}$ is coloured with at most $(\log\log N)^c$ colours, then there is a monochromatic arithmetic progression of length four.* □

9 Obtaining Approximate Homomorphisms

The results of this section and the next can be combined to give an alternative proof of Corollary 7.9. The approach is longer, and the bound worse, but it does not make use of Plünnecke's inequality, so the comparison is less unfavourable than it seems. Our reason for giving it is that later in

the paper we shall come across functions that are almost Freiman homomorphisms, but not quite, and we have not found a quick way of turning them into genuine homomorphisms without losing important information about their Fourier coefficients. Instead, therefore, we have been forced to examine these approximate homomorphisms and produce a version of Corollary 7.9 for them directly. It is quite possible that there *is* an argument for obtaining genuine homomorphisms in the later contexts. This would result in a significant simplification of the paper.

The later applications all need results that are more complicated than those proved in this section (see §12 and §15). Therefore, this section is another one which is not strictly necessary. However, the reader may find it useful to see the method of proof at work in a simpler case. Recall that we showed in Corollary 7.6 that if $B \subset \mathbb{Z}_N$ and $\phi : B \to \mathbb{Z}_N$ is a somewhat additive function, then ϕ has a restriction to a large subset of B which is an isomorphism of order eight. In this section we shall give an alternative approach which yields what we shall call an approximate isomorphism. Because the isomorphism is approximate rather than exact, it is harder to apply Bogolyubov-type techniques to it, and that will be the task of the next section.

Let $B \subset \mathbb{Z}_N$. We shall call a function $\phi : B \to \mathbb{Z}_N$ a *γ-homomorphism of order k* if, of the sequences $(x_1, \dots, x_{2k}) \in B^{2k}$ such that

$$x_1 + \cdots + x_k = x_{k+1} + \cdots + x_{2k}\,,$$

the proportion that also satisfy

$$\phi(x_1) + \cdots + \phi(x_k) = \phi(x_{k+1}) + \cdots + \phi(x_{2k})$$

is at least γ. If γ is close to 1, then we shall say that ϕ is an *approximate homomorphism* of order k.

LEMMA 9.1. *Let $a_1, \dots, a_n$ be non-negative real numbers. Then*

$$\sum_{i=1}^n a_i^4 \leqslant \Big(\sum_{i=1}^n a_i^2\Big)^{6/7} \Big(\sum_{i=1}^n a_i^{16}\Big)^{1/7}\,.$$

Proof. The result follows from Hölder's inequality if one writes $a_i^4 = a_i^{12/7} a_i^{16/7}$ and takes $p = 7/6$, $q = 7$. □

LEMMA 9.2. *Let $B \subset \mathbb{Z}_N$ and let $\phi : B \to \mathbb{Z}_N$ be γ-additive. Then there are at least $\gamma^7 N^{15}$ sequences $a_1, \dots, a_{16}$ such that*

$$a_1 + \cdots + a_8 = a_9 + \cdots + a_{16}$$

and

$$\phi(a_1) + \cdots + \phi(a_8) = \phi(a_9) + \cdots + \phi(a_{16})\,.$$

Proof. Given $u \in \mathbb{Z}_N$, define $f_u(a)$ to be $\omega^{u\phi(a)}$ if $a \in B$, and zero otherwise. Then $\sum_a |f_u(a)|^2 = |B| \leqslant N$, so $\sum_{u,r} |\hat{f}_u(r)|^2 \leqslant N^3$.

Next, we look at fourth powers. We have

$$\sum_{u,r} |\hat{f}_u(r)|^4 = \sum_{u,r} \Big| \sum_a \omega^{u\phi(a)-ra} \Big|^4 ,$$

which is exactly N^2 times the number of additive quadruples (a_1, a_2, a_3, a_4), and thus, by hypothesis, at least γN^5.

Finally, we look at sixteenth powers. A similar argument shows that $\sum_{u,r} |\hat{f}_u(r)|^{16}$ counts N^2 times the number of sequences $(a_1, \dots, a_{16})$ such that

$$a_1 + \cdots + a_8 = a_9 + \cdots + a_{16}$$

and

$$\phi(a_1) + \cdots + \phi(a_8) = \phi(a_9) + \cdots + \phi(a_{16}) .$$

Lemma 9.1 implies that

$$\sum_{u,r} |\hat{f}_u(r)|^{16} \geqslant (\gamma N^5 . N^{-18/7})^7 = \gamma^7 N^{17} .$$

Hence, the number of sequences with the desired properties is, as stated, at least $\gamma^7 N^{15}$. □

LEMMA 9.3. *Let $\eta > 0$, let $B \subset \mathbb{Z}_N$ be a set of size βN and let $\phi : B \to \mathbb{Z}_N$ be a function with at least $\alpha\beta^{15}N^{15}$ sequences $a_1, \dots, a_{16}$ such that*

$$a_1 + \cdots + a_8 = a_9 + \cdots + a_{16} \tag{1}$$

and

$$\phi(a_1) + \cdots + \phi(a_8) = \phi(a_9) + \cdots + \phi(a_{16}) . \tag{2}$$

Then, as long as N is sufficiently large (in terms of α, β and η), there is a subset $B' \subset B$ with at least $(\alpha\eta/4)^{2^{19}}\beta^{15}N^{15}$ sequences $(a_1, \dots, a_{16})$ satisfying condition (1), such that the proportion of them that satisfy condition (2) as well is at least $1 - \eta$. In other words, B' is reasonably large and the restriction of ϕ to B' is a $(1 - \eta)$-homomorphism of order eight.

Proof. The basic idea is that if we let M be a suitable fraction of N and P be the interval $[-M, M] \subset \mathbb{Z}_N$, and if we choose r and s randomly, then the set of all $x \in A$ such that $rx + s\phi(x)$ belongs to P tends to have a larger proportion of sequences satisfying condition (2) than A itself. This is because the events that we choose a_i for $i = 1, 2, \dots, 16$ are better correlated if $(a_1, \dots, a_{16})$ satisfies condition (2) than if it does not. Repeating the process, one can make the proportion as close as one likes to 1. Note that this is a natural approach to try, given the proof of Lemma 7.5.

The calculations are, however, enormously simplified if one uses Riesz products (that is, products of the form $2^{-k}\prod_{i=1}^{k}(1+\cos\theta_i)$) and a small modification of the above idea. Choose $r_1,\dots,r_k$, $s_1,\dots,s_k$ uniformly and independently at random from $\mathbb{Z}_N$. Once the choice is fixed, let a point $x\in B$ go into B' with probability

$$2^{-k}\prod_{i=1}^{k}\bigl(1+\cos\tfrac{2\pi}{N}\bigl(r_ix+s_i\phi(x)\bigr)\bigr)\ ,$$

and let these probabilities be independent.

It must be stressed that this independence occurs only *after* we have conditioned on the choice of $r_1,\dots,r_k$, $s_1,\dots,s_k$. The whole point of the proof is that in total there is a dependence which favours sequences satisfying condition (2). To see that this is true, let $a_1,\dots,a_{16}$ be sixteen points in $\mathbb{Z}_N$. The probability that they are all chosen is

$$N^{-2k}\sum_{r_1,\dots,r_k}\sum_{s_1,\dots,s_k}2^{-16k}\prod_{i=1}^{k}\prod_{j=1}^{16}\bigl(1+\cos\tfrac{2\pi}{N}\bigl(r_ia_j+s_i\phi(a_j)\bigr)\bigr)$$

which equals

$$N^{-2k}2^{-16k}\biggl(\sum_{r,s}\prod_{j=1}^{16}\bigl(1+\cos\tfrac{2\pi}{N}\bigl(ra_j+s\phi(a_j)\bigr)\bigr)\biggr)^k,$$

which we shall rewrite as

$$N^{-2k}2^{-16k}\biggl(2^{-16}\sum_{r,s}\prod_{j=1}^{16}\bigl(1+1+\omega^{ra_j+s\phi(a_j)}+\omega^{-(ra_j+s\phi(a_j))}\bigr)\biggr)^k.$$

The product over j is a sum of 4^{16} terms, each of which is of the form

$$\prod_{j=1}^{16}\omega^{\epsilon_j(ra_j+s\phi(a_j))}=\omega^{r\sum_j\epsilon_ja_j+s\sum_j\epsilon_j\phi(a_j)}\ ,$$

where $\epsilon_1,\dots,\epsilon_{16}$ all belong to the set $\{-1,0,1\}$. Such a term contributes zero to the sum over r and s, unless $\sum_{j=1}^{16}\epsilon_ja_j=\sum_{j=1}^{16}\epsilon_j\phi(a_j)=0$, in which case it contributes N^2.

Let us now consider sequences $(a_1,\dots,a_{16})\in\mathbb{Z}_N^{16}$ satisfying condition (1). The set of such sequences is a fifteen-dimensional subspace of the vector space $\mathbb{Z}_N^{16}$. Given $(\epsilon_1,\dots,\epsilon_{16})\in\{-1,0,1\}^{16}$, the set of sequences $(a_1,\dots,a_{16})$ in this subspace satisfying the additional condition that $\epsilon_1a_1+\cdots+\epsilon_{16}a_{16}=0$ is a fourteen-dimensional subspace of $\mathbb{Z}_N^{16}$ and hence has cardinality N^{14}, except if $(\epsilon_1,\dots,\epsilon_{16})$ is a multiple of $(1,\dots,1,-1,\dots,-1)$

(eight 1s followed by eight -1s). Let us call a sequence $(a_1,\dots,a_{16})$ satisfying condition (1) *degenerate* if it also satisfies a genuinely distinct linear condition with coefficients in $\{-1,0,1\}$, and otherwise *non-degenerate*. The number of degenerate sequences is clearly at most $3^{16}N^{14}$. Let us call a non-degenerate sequence *good* if it satisfies condition (2) and *bad* otherwise. (It is part of the definition of non-degeneracy that both good and bad sequences satisfy condition (1).)

Our arguments above show that a bad sequence is chosen with probability 2^{-16k}, since the only terms that contribute are the 2^{16} terms with $\epsilon_j=0$ for every j. A good sequence, on the other hand, is chosen with probability $2^{-16k}\big(2^{-16}(2^{16}+2)\big)^k=2^{-16k}(1+2^{-15})^k$, because there are two further terms making a contribution, namely those with $\epsilon_1=\dots=\epsilon_8=-\epsilon_9=\dots=-\epsilon_{16}=\pm1$. Let X and Y be the numbers of good and bad sequences chosen. Then the expected value of X is, from our hypothesis, at least $(1+2^{-15})^k2^{-16k}\alpha\beta^{15}N^{15}$, and the expected value of Y is at most $2^{-16k}\beta^{15}N^{15}$. Using the fact that $2^{2^{-15}}\leqslant 1+2^{-15}$, we can deduce that if $2^{2^{-15}k}\geqslant 2/\alpha\eta$, then

$$\eta\mathbb{E}X-\mathbb{E}Y\geqslant \eta(2/\alpha\eta)2^{-16k}\alpha\beta^{15}N^{15}-2^{-16k}\beta^{15}N^{15}=2^{-16k}\beta^{15}N^{15}\,.$$

Now $2^{2^{-15}k}\geqslant(2/\alpha\eta)$ if and only if $2^{-16k}\leqslant(\alpha\eta/2)^{2^{19}}$. Let k be an integer such that

$$2(\alpha\eta/4)^{2^{19}}\leqslant 2^{-16k}\leqslant(\alpha\eta/2)^{2^{19}}\,.$$

If N is large enough that $(\alpha\eta/4)^{2^{19}}\beta^{15}N\geqslant 3^{16}$, then the values for the above expectations and the upper estimate for the number of degenerate configurations imply that there exists a set B' such that $\eta X\geqslant Y$ and $X\geqslant(\alpha\eta/4)^{2^{19}}\beta^{15}N^{15}$, as was claimed. □

Lemmas 9.2 and 9.3 combined show that a somewhat additive function can be restricted to an approximate homomorphism of order eight.

COROLLARY 9.4. *Let $B\subset\mathbb{Z}_N$ have size βN, let $\phi:B\to\mathbb{Z}_N$ be $\gamma\beta^3$-additive and let $\eta>0$. There is a subset $B'\subset B$ containing at least $(\gamma^7\beta^6\eta/4)^{2^{19}}\beta^{15}N^{15}$ sequences $(a_1,\dots,a_{16})$ with $a_1+\dots+a_8=a_9+\dots+a_{16}$, such that the restriction of ϕ to B' is a $(1-\eta)$-homomorphism of order eight.*

Proof. Lemma 9.2 allows us to take $\alpha=(\gamma\beta^3)^7\beta^{-15}=\gamma^7\beta^6$ in Lemma 9.3. □

10 Properties of Approximate Homomorphisms

Let $A\subset\mathbb{Z}_N$ be a set of size αN and let $\phi:A\to\mathbb{Z}_N$ be a $(1-\epsilon)$-homomorphism. Since A contains at least α^4N^3 additive quadruples, it

also contains at least $(1-\epsilon)\alpha^4 N^3$ ϕ-additive quadruples. Corollary 7.6 allows us to pass to a large subset A' of A such that the restriction of ϕ to A' is a Freiman homomorphism of order 8. Lemma 7.8 then provides a large Bohr neighbourhood B such that the restriction of ϕ to A' is a B-homomorphism.

Later in the paper approximate homomorphisms will arise in a context where we wish to restrict them to exact B-homomorphisms, but are unable to use the above argument. This may seem surprising, as the argument is perfectly valid: the reason it is inadequate is that the Bohr neighbourhood B that it gives is defined in terms of the set A', so by using it we lose information about the large Fourier coefficients of A. This will matter later, because then we shall have a collection of sets A_h and approximate homomorphisms ϕ_h indexed by a set $H \subset \mathbb{Z}_N^k$. The large Fourier coefficients associated with each set A_h will be related, and we shall exploit this. Therefore, in this section our aim is to obtain a theorem similar to Lemma 7.8, but the Bohr neighbourhood will be defined in terms of the Fourier coefficients of the original set A rather than those of the subset A'.

This seems to make the proof harder, although it is based on similar ideas, and in particular uses Bogolyubov's method. Most of the proofs in this section are simple averaging arguments. However, there are so many of them that when put together they are not particularly simple. It is likely that there is a shorter proof of the main result, but we have been unable to find one.

To complicate matters further, it is necessary to consider objects that are slightly more general than functions from $\mathbb{Z}_N$ to $\mathbb{Z}_N$, to allow for multisets that occur naturally in later sections. By a *multifunction from $\mathbb{Z}_N$ to $\mathbb{Z}_N$*, we shall mean a function from a set X to $\mathbb{Z}_N$, together with a partition $X = \bigcup_{r\in\mathbb{Z}_N} X_r$. Equivalently, it is simply a pair of functions from X to $\mathbb{Z}_N$, and indeed it will be useful to write $r(x)$ for the function that takes $x \in X$ to the unique r such that $x \in X_r$. We shall call a set X together with such a partition (or function) a *domain*, and if $\phi : X \to \mathbb{Z}_N$, we shall call X *the domain of ϕ*.

Given a domain $X = (X, r)$, we shall define $X - X$ to be the set $X \times X$ together with the function $(x, y) \mapsto r(y) - r(x)$, or equivalently the partition $X \times X = \bigcup_d Y_d$, where Y_d is the set of pairs (x, y) such that $x \in X_r$ and $y \in X_{r+d}$ for some r. More generally, by $kX - lX$ we mean the set X^{k+l} with the function

$$(x_1, \dots, x_{k+l}) \mapsto r(x_1) + \cdots + r(x_k) - r(x_{k+1}) - \cdots - r(x_{k+l})\,.$$

A function $\phi : X \to \mathbb{Z}_N$ will be called a $(1-\eta)$-*homomorphism of order* k if, out of the $2k$-tuples $(x_1, \dots, x_{2k}) \in X^{2k}$ such that

$$r(x_1) + \cdots + r(x_k) = r(x_{k+1}) + \cdots + r(x_{2k}),$$

the proportion such that

$$\phi(x_1) + \cdots + \phi(x_k) = \phi(x_{k+1}) + \cdots + \phi(x_{2k})$$

is at least $1-\eta$. Note that this definition is not vacuous when $k = 1$.

We shall define an *additive quadruple* to be a quadruple $(a, b, c, d) \in X^4$ such that $r(a) - r(b) = r(c) - r(d)$ and we shall say that it is ϕ-*additive* if in addition $\phi(a) - \phi(b) = \phi(c) - \phi(d)$. Then a $(1-\eta)$-homomorphism of order two is a function ϕ such that the proportion of additive quadruples that are ϕ-additive is at least $1-\eta$, just as when $X = \mathbb{Z}_N$.

We shall now investigate the extent to which these more general approximate homomorphisms resemble exact ones. The arguments are more complicated than one might expect, and the reason for the complication is the existence of examples of the following kind. Let A and B be subsets of $\mathbb{Z}_N$, with $A = \{a, a+r, \dots, a+(M-1)r\}$ and $B = \{b, b+s, \dots, b+(M-1)s\}$, where $M = \alpha N$ for some small $\alpha > 0$. If there are no small linear relations between r and s (i.e., pairs u, v of small elements of $\mathbb{Z}_N$ such that $ru + sv = 0$) then the intersection of A and B will have cardinality roughly $\alpha^2 N$. Moreover, almost all the additive quadruples in $A \cup B$ will lie entirely in A or entirely in B. (These facts are easy to check.) Hence, if we define a function ϕ to be linear on A and also linear, but with a different gradient, on $B \setminus A$, then ϕ will be a $(1-\eta)$-homomorphism for some small η (depending on α). In fact, ϕ will even be a $(1-\eta)$-homomorphism of high order (for a larger η, but still small). Most of the effort of this section is devoted to showing how to "pick out" A or B in an example such as the above, in order to obtain a well-defined and well-behaved difference function for the restriction of ϕ.

Let $X = \bigcup_r X_r$ be a domain, let B be a set and let L be a non-negative real number. We shall say that X is (B, L)-*invariant* if, given any $r \in \mathbb{Z}_N$ and any $d \in B$, the sizes of X_{r+d} and X_r differ by at most L. If L is small (compared, for example, with the average size of the X_r) we shall say that X is *almost* B-*invariant*.

We shall now prove several lemmas under the same set of hypotheses. To save repetition later, we state the hypotheses once and for all here. Let $X = (X, r)$ be a domain such that X has cardinality αMN and X_r has cardinality at most M for every r. Let $\sigma > 0$ be a parameter to be

chosen later and let $B \subset \mathbb{Z}_N$ be some set such that $B = -B$ and X is $(B, \sigma M)$-invariant. Let $\phi : X \to \mathbb{Z}_N$ be a $(1-\eta)$-homomorphism.

For every $(x, y) \in X^2$ let us define $q(x, y)$ to be the number of pairs $(z, w) \in X^2$ such that $r(w) - r(z) = r(y) - r(x)$. Let $b(x, y)$ be the number of pairs $(u, v) \in X^2$ such that $r(u) - r(x) = r(v) - r(y) \in B$. One can also write these as

$$q(x, y) = \sum_d |X_{r(x)+d}||X_{r(y)+d}|$$

and

$$b(x, y) = \sum_{d \in B} |X_{r(x)+d}||X_{r(y)+d}| .$$

We shall also let $e(x, y)$ be the number of pairs (u, v) such that $r(u) - r(x) = r(v) - r(y) \in B$ and $\phi(u) - \phi(x) \neq \phi(v) - \phi(y)$.

In words, $q(x, y)$ is the number of additive quadruples starting with (x, y), $b(x, y)$ is the number of such quadruples (x, y, z, w) such that $(r(z), r(w))$ is $(r(x), r(y))$ translated by some $d \in B$ and $e(x, y)$ (the error) is the number of those special additive quadruples that fail to be ϕ-additive. Finally, let $\epsilon(x, y)$ be the proportionate error, i.e., $e(x, y)/b(x, y)$.

Our first lemma collects together some simple facts about the function q.

LEMMA 10.1. *$q(x, y) \leqslant \alpha M^2 N$ for every x, y, $\sum_{y \in X} q(x, y) \leqslant \alpha^2 M^3 N^2$ for every x and $\sum_{x,y \in X} q(x, y) \geqslant \alpha^4 M^4 N^3$.*

Proof. For the first estimate we wish to count the number of pairs (z, w) such that (x, y, z, w) is an additive quadruple. There are at most $|X| = \alpha MN$ ways of choosing z. Once z is chosen, $r(w)$ is determined so there are at most M choices for w. The second estimate follows immediately.

As for the third, notice that the left-hand side is equal to $\sum_{r-s=t-u} |X_r||X_s||X_t||X_u|$. By §2 identity (6) applied to the function $f(s) = |X_s|$, this is at least $N^{-1}|X|^4$, which is the estimate claimed. □

One can think of the numbers $q(x, y)$ as defining a weighted graph, where the weight of the edge (x, y) measures the popularity of the difference $r(y) - r(x)$ in X. Roughly speaking, our aim will be to show that ϕ is well behaved on "components" of this weighted graph – that is, highly connected subsets which are not highly connected to the rest of the graph. In the example discussed earlier of two "unrelated" arithmetic progressions A and B, the components can be taken as A and $B \setminus A$, since $q(x, y)$ tends to be large if x and y both belong to A or both belong to B, and small otherwise. The pairs (x, y) that contribute a significant error $e(x, y)$ tend to be those for which x and y belong to different sets, and therefore for

which the weight $q(x,y)$ is small. Our next lemma shows that this is true in general. That is, most of the error occurs, if at all, on edges with small weight.

LEMMA 10.2. *If $\sigma \leqslant \eta\alpha^2$, then $\sum_{x,y\in X}\epsilon(x,y)q(x,y) \leqslant 15\eta\sum_{x,y\in X}q(x,y)$.*

Proof. Let $X' \subset X$ be the union of all X_r of size at least $5\eta\alpha^2 M$. We begin by estimating $\sum_{x,y\in X'}\epsilon(x,y)q(x,y)$. Let $x \in X_r$ and $y \in X_s$ and let $X_r \cup X_s \subset X'$. Then $b(x,y) = \sum_{d\in B}|X_{r+d}||X_{s+d}|$. Let $L = \eta\alpha^2 M$. Since X is (B,L)-invariant, we can deduce from this expression for $b(x,y)$ that

$$|B|\big(|X_r|-L\big)\big(|X_s|-L\big) \leqslant b(x,y) \leqslant |B|\big(|X_r|+L\big)\big(|X_s|+L\big)\,.$$

Furthermore, if u and v are such that $r(u)-r = r(v)-s \in B$, the (B,L)-invariance also implies that

$$|B|\big(|X_r|-2L\big)\big(|X_s|-2L\big) \leqslant b(u,v) \leqslant |B|\big(|X_r|+2L\big)\big(|X_s|+2L\big)\,.$$

Since both $|X_r|$ and $|X_s|$ are at least $5\eta\alpha^2 M = 5L$, the above estimates imply that $b(u,v)/b(x,y) < 4$.

Now let S be the set of all sextuples $(x,y,z,w,u,v) \in X^6$ with $x,y \in X'$ satisfying the following conditions:

$$r(w)-r(z) = r(y)-r(x) \tag{1}$$
$$r(u)-r(x) = r(v)-r(y) \in B \tag{2}$$
$$r(w)-r(z) = r(v)-r(u) \tag{3}$$
$$\phi(u)-\phi(x) \neq \phi(v)-\phi(y)\,. \tag{4}$$

Of course, (1) and (2) imply (3), and (2) and (3) imply (1).) Then

$$\sum_{(x,y,z,w,u,v)\in S} b(x,y)^{-1} = \sum_{x,y\in X'} b(x,y)^{-1}q(x,y)e(x,y) = \sum_{x,y\in X'}\epsilon(x,y)q(x,y)\,.$$

Condition (4) implies that either $\phi(x)-\phi(y) \neq \phi(z)-\phi(w)$ or $\phi(u)-\phi(v) \neq \phi(z)-\phi(w)$. Therefore, we can write $S = E \cup F$, where

$$E = \big\{(x,y,z,w,u,v) \in S : \phi(x)-\phi(y) \neq \phi(z)-\phi(w)\big\}$$

and

$$F = \big\{(x,y,z,w,u,v) \in S : \phi(u)-\phi(v) \neq \phi(z)-\phi(w)\big\}\,.$$

We now estimate the sum over S by splitting it into E and F.

For any fixed quadruple (x,y,z,w) satisfying condition (1), the number of pairs (u,v) such that (x,y,z,w,u,v) satisfies condition (2) is exactly $b(x,y)$. It follows that $\sum\{b(x,y)^{-1} : (x,y,z,w,u,v) \in E\}$ is at most the number of additive quadruples (x,y,z,w) with $x,y \in X'$ that fail to be

ϕ-additive, which is by hypothesis at most η times the total number of additive quadruples. That is,

$$\sum \big\{b(x,y)^{-1} : (x,y,z,w,u,v) \in E\big\} \leqslant \eta \sum_{x,y\in X} q(x,y)\,.$$

Since $B = -B$, for every quadruple (z,w,u,v) satisfying condition (3) the number of pairs (x,y) satisfying condition (2) is $b(u,v)$. For each such pair, we have shown that $b(u,v) < 4b(x,y)$, so the sum of $b(x,y)^{-1}$ over all of them is less than 4. Therefore, $\sum\{b(x,y)^{-1} : (x,y,z,w,u,v) \in F\}$ is less than 4 times the number of additive quadruples (z,w,u,v) that fail to be ϕ-additive. So this time we have

$$\sum \big\{b(x,y)^{-1} : (x,y,z,w,u,v) \in F\big\} < 4\eta \sum_{x,y\in X} q(x,y)\,.$$

Putting the two estimates together, we find that

$$\sum_{x,y\in X'} \epsilon(x,y)q(x,y) \leqslant 5\eta \sum_{x,y\in X} q(x,y)\,.$$

We must also count the additive quadruples (x,y,z,w) such that either $x \notin X'$ or $y \notin X'$, which means that either $|X_{r(x)}|$ or $|X_{r(y)}|$ is less than $5\eta\alpha^2 M$. There are easily seen to be at most $2(5\eta\alpha^2 MN)(\alpha MN)(\alpha MN)M = 10\eta\alpha^4 M^4 N^4$ of them. Since there are at least $\alpha^4 M^4 N^3$ additive quadruples, this number is at most $10\eta\sum_{x,y\in X} q(x,y)$. This estimate, together with the earlier one, proves the lemma. □

We are aiming to find a large subset of X where the error $\epsilon(x,y)$ is almost always small. The above lemma suggests that we can achieve this by choosing a subset of a "component" of the weighted graph given by q. Roughly speaking, we do this by picking a random point $x \in X$ and taking the set of all y in a neighbourhood of x (in an appropriate weighted sense). Such a set will be a union of sets $|X_r|$. For technical reasons it will be very convenient to have all the X_r that we choose of approximately the same size, and to have other properties of a similar kind. These properties will be obtained by somewhat messy averaging arguments.

To make these ideas more precise, let us define some more functions and prove another lemma. For every $x \in X$, let $R(x) = |X_{r(x)}|$, let $Q(x) = \sum_{y\in X} q(x,y)$ and let $E(x) = \sum_{y\in X} \epsilon(x,y)q(x,y)$.

LEMMA 10.3. *There exists $x \in X$ such that $R(x) \geqslant \alpha^2 M/2$, $Q(x) \geqslant \alpha^3 M^3 N^2/4$ and $E(x) \leqslant 60\eta Q(x)$.*

Proof. Lemma 10.1 tells us that $\sum_{x\in X} Q(x) \geqslant \alpha^4M^4N^3$ and that $Q(x) \leqslant \alpha^2M^3N^2$ for every x. Let X' be the set of $x \in X$ such that $R(x) \geqslant \alpha^2M/2$. Clearly, $|X \setminus X'| \leqslant \alpha^2MN/2$, so

$$\sum_{x\in X'} Q(x) \geqslant \alpha^4M^4N^3 - (\alpha^2MN/2)(\alpha^2M^3N^2) = \alpha^4M^4N^3/2 \geqslant \tfrac{1}{2}\sum_{x\in X} Q(x)\,.$$

Let us now choose $x \in X'$ uniformly at random. The expected value of $Q(x)$ is at least $|X|^{-1}\alpha^4M^4N^3/2 = \alpha^3M^3N^2/2$. By Lemma 10.2 the expectation of $E(x)$ is at most 15η times the expectation of $Q(x)$ over X, which is at most 30η times the expectation of $Q(x)$ over X'. It follows that the expectation of $Q(x) - (1/60\eta)E(x)$ over X' is at least $\alpha^3M^3N^2/4$, so we can find $x \in X'$ such that $Q(x) \geqslant \alpha^3M^3N^2/4$ and $E(x) \leqslant 60\eta Q(x)$. This proves the lemma. □

Let us now fix an x satisfying the conclusion of Lemma 10.3 and write $q(y)$ for $q(x,y)$, $\epsilon(y)$ for $\epsilon(x,y)$ and S for $Q(x)$.

LEMMA 10.4. *If $r(z) - r(y) \in B$, then $|q(z) - q(y)| \leqslant \sigma\alpha M^2N$.*

Proof. Let $r(z) - r(y) = d \in B$. Then

$$q(y) = \sum_{t-s=r(y)-r(x)} |X_s||X_t|$$

and

$$q(z) = \sum_{t-s=r(y)-r(x)} |X_s||X_{t+d}|\,.$$

From the $(B,\sigma M)$-invariance of X we deduce that

$$\begin{aligned}|q(z) - q(y)| &\leqslant \sum_{t-s=r(y)-r(x)} |X_s|\big||X_{t+d}| - |X_t|\big| \\ &\leqslant \sigma M \sum_s |X_s| = \sigma\alpha M^2N\,,\end{aligned}$$

as stated. □

For the next lemma, we use the notation $W+d$ to stand for all elements $x \in X$ such that there exists $w \in W$ with $r(x) = r(w) + d$.

LEMMA 10.5. *There exists a subset $W \subset X$ with the following properties.*

(i) W is a union of sets of the form X_r.
(ii) The function q varies by a factor of at most two on W.
(iii) For at least $(1-5\eta^{1/2})|W|$ of the points $y \in W$ we have $\epsilon(y) \leqslant 300\eta^{1/2}$.
(iv) W has cardinality at least $\rho^2\alpha^2MN/16$.
(v) The function R varies by a factor of at most two on W, and is always at least $\alpha^2M/16$.

(vi) $|W \cap (W+d)| \geqslant (1-\eta)|W|$ *for every* $d \in B$.

Proof. By Lemmas 10.1 and 10.3 we know that $q(y) \leqslant \alpha M^2 N$ for every y, that $S = \sum_{y\in X} q(y) \geqslant \alpha^3 M^3 N^2/4$ and that $\sum_{y\in X} \epsilon(y)q(y) \leqslant 60\eta \sum_{y\in X} q(y) = 60\eta S$. Let X' be the set of all $y \in X$ such that $R(y) \geqslant \alpha^2 M/8$ and let $S' = \sum_{y\in X'} q(y)$. Then

$$\sum_{y\in X\setminus X'} q(y) \leqslant (\alpha^2 MN/8)(\alpha M^2 N) = \alpha^3 M^3 N^2/8\,,$$

from which it follows that $S' \geqslant S/2$.

We now choose λ and μ independently and uniformly from the interval $[-\rho, 1+\rho]$ and make the following definitions.

$$\begin{aligned} W_{\lambda,\mu} &= \big\{y : (\lambda-\rho)\alpha M^2 N \leqslant q(y) \leqslant (\lambda+\rho)\alpha M^2 N\big\} \\ &\qquad \cap \big\{y : (\mu-\rho)M \leqslant R(y) \leqslant (\mu+\rho)M\big\}\,; \\ V_{\lambda,\mu} &= \big\{y : (\lambda-\rho)\alpha M^2 N \leqslant q(y) \leqslant (\lambda-\rho+\sigma)\alpha M^2 N\big\} \\ &\qquad \cap \big\{y : (\lambda+\rho-\sigma)\alpha M^2 N \leqslant q(y) \leqslant (\lambda+\rho)\alpha M^2 N\big\} \\ &\qquad \cap \big\{y : (\mu-\rho)M \leqslant R(y) \leqslant (\mu-\rho+\sigma)M\big\} \\ &\qquad \cap \big\{y : (\mu+\rho-\sigma)M \leqslant R(y) \leqslant (\mu+\rho)M\big\}\,. \end{aligned}$$

We also set $S_{\lambda,\mu} = \sum_{y\in X'\cap W_{\lambda,\mu}} q(y)$ and $E_{\lambda,\mu} = \sum_{y\in W_{\lambda,\mu}} \epsilon(y)q(y)$. We shall now use an averaging argument to find λ and μ such that $S_{\lambda,\mu}$ is large, while $E_{\lambda,\mu}$, and also the sizes of $W_{\lambda,\mu}$ and $V_{\lambda,\mu}$, are small.

To do this, we simply calculate or estimate the expectations of all the quantities concerned. Since any fixed $y \in X$ has a probability of $\big(\frac{2\rho}{1+2\rho}\big)^2$ of belonging to $W_{\lambda,\mu}$ we find that the expectation of $S_{\lambda,\mu}$ is $\big(\frac{2\rho}{1+2\rho}\big)^2 S'$, which we know is at least $\frac{2\rho^2}{(1+2\rho)^2}S$. We also find that the expectation of $E_{\lambda,\mu}$ is at most $\big(\frac{2\rho}{1+2\rho}\big)^2 60\eta S$ and the expected size of $W_{\lambda,\mu}$ is $\big(\frac{2\rho}{1+2\rho}\big)^2 \alpha MN$. The probability of any given $y \in X$ belonging to $V_{\lambda,\mu}$ is at most $\big(\frac{4\sigma}{1+2\rho}\big)$, which implies that the expected size of $V_{\lambda,\mu}$ is at most $\big(\frac{4\sigma}{1+2\rho}\big)\alpha MN$.

By linearity of expectation, we may deduce that

$$\mathbb{E}\left(S_{\lambda,\mu} - \frac{E_{\lambda,\mu}}{720\eta} - -\frac{S|W_{\lambda,\mu}|}{12\alpha MN} - \frac{\rho^2 S|V_{\lambda,\mu}|}{12\sigma(1+2\rho)\alpha MN}\right) \geqslant \frac{\rho^2}{(1+2\rho)^2}S \geqslant \frac{\rho^2 S}{4}\,.$$

Therefore there exist λ and μ such that $S_{\lambda,\mu} \geqslant \rho^2 S/4$, $E_{\lambda,\mu} \leqslant 720\eta S_{\lambda,\mu}$, $|W_{\lambda,\mu}| \leqslant 12\alpha MNS_{\lambda,\mu}/S$ and $|V_{\lambda,\mu}| \leqslant 12\sigma\rho^{-2}(1+2\rho)\alpha MNS_{\lambda,\mu}/S$. Our aim is now to prove that $W = W_{\lambda,\mu}$ has the desired properties.

Property (i) follows immediately from the definition of $W_{\lambda,\mu}$. To prove (ii), notice that the average value of $q(y)$ over $W_{\lambda,\mu}$ is at $S_{\lambda,\mu}/|W_{\lambda,\mu}|$, which

is at least $S/12\alpha MN \geqslant \alpha^2M^2N/64$. It follows that $\lambda+\rho \geqslant \alpha/48$, and hence that $\lambda-\rho \geqslant (\lambda+\rho)/2$ (as $\rho \leqslant \alpha/192$).

The upper estimate for $E_{\lambda,\mu}$ tells us that $\sum_{y\in W}\epsilon(y)q(y) \leqslant 720\eta\sum_{y\in W}q(y)$. Since the function q varies over W by a factor of at most two, we obtain (iii), since otherwise we would have $\sum_{y\in W}\epsilon(y)q(y) > 1500\eta|W|\min_{y\in W}q(w)$, a contradiction.

Since $S_{\lambda,\mu} \geqslant \rho^2S/4 \geqslant \rho^2\alpha^3M^3N^2/16$ and $q(y) \leqslant \alpha M^2N$ for every $y \in X$, the cardinality of W must be at least $\rho^2\alpha^2MN/16$, which is property (iv).

Because $S_{\lambda,\mu}$ is non-zero, there exists $y \in W$ such that $R(y) \geqslant \alpha^2M/8$, from which it follows that $\mu+\rho \geqslant \alpha^2/8$ and therefore, as $\rho \leqslant \alpha^2/32$, that $\mu-\rho \geqslant (\mu+\rho)/2$ and $\mu-\rho \geqslant \alpha^2/16$. This gives us (v).

Let us now set $V = V_{\lambda,\mu}$ and choose $y \in W \setminus V$. If $d \in B$ then by the $(B,\sigma M)$-invariance of X and our lower bound for μ, we have $|X_{r(y)-d}| \geqslant |X_{r(y)}| - \sigma M \geqslant (\mu-\rho)M > 0$. Choosing any $z \in X_{r(y)-d}$, we then know that $|q(z)-q(y)| \leqslant \sigma\alpha M^2N$, by Lemma 10.4, from which it follows that $(\lambda-\rho)\alpha M^2N \leqslant q(z) \leqslant (\lambda+\rho)\alpha M^2N$. The $(B,\sigma M)$-invariance of X also gives us that $|R(z)-R(y)| \leqslant \sigma M$, and from this it follows that $(\mu-\rho)M \leqslant R(z) \leqslant (\mu+\rho)M$. We have therefore shown that if $y \in W\setminus V$, $d \in B$ and $z \in X_{r(y)-d}$, then $z \in W$. Moreover, such a z exists, so $y \in W+d$.

All this shows that $W\setminus V \subset W\cap(W+d)$. We have shown that the cardinality of W is at least $\rho^2\alpha^2MN/16$, while the cardinality of V is at most $12\sigma\rho^{-2}(1+2\rho)\alpha MNS_{\lambda,\mu}/S$, which is certainly at most $24\sigma\rho^{-2}\alpha MN$. Since $\sigma \leqslant \eta\rho^4\alpha/384$, we find that $|V| \leqslant \eta|W|$ and therefore obtain property (vi). □

Before we state the next lemma, it will be very useful to introduce the following shorthand notation. Given any finite set U and any proposition $P(u)$ involving the elements u of U, we shall say that *for* $(1-\epsilon)$*-almost every* $u \in U$, $P(u)$ if the set $\{u \in U : P(u)\}$ has cardinality at least $(1-\epsilon)|U|$. We shall further abbreviate this by writing $((1-\epsilon)\,\text{a.e. } u \in U)\ P(u)$.

LEMMA 10.6. *If $\sigma \leqslant \eta\rho\alpha^2/16$ then there exist a subset B' of B of cardinality at least $(1-10\eta^{1/5})|B|$ and a function $\psi : B' \to \mathbb{Z}_N$ such that, for every $d \in B'$,*

$$\big((1-10\eta^{1/5})\,\text{a.e. } w\in W\big)\ \big((1-10\eta^{1/5})\,\text{a.e. } z\in X_{r(w)+d}\big)\quad \phi(z)-\phi(w)=\psi(d)\,.$$

Proof. Let us define W' to be the set of all $w \in W\cap(W-d)$ such that $\epsilon(w) \leqslant 300\eta^{4/5}$ We know from Lemma 10.5 (iii) and (vi) (and the fact that $B=-B$) that W' has cardinality at least $(1-6\eta^{1/5})|W|$.

Given $d \in B$ and $w \in W'$, let us say that d is *good for* w if for $(1-35\eta^{2/5})$-almost every pair $(y,z) \in X_{r(x)+d}\times X_{r(w)+d}$ we have $\phi(y)-\phi(x)$

$= \phi(z) - \phi(w)$. Notice that $|X_{r(x)+d}| \geqslant |X_{r(x)}|/2$ by Lemma 10.3 and $(B, \sigma M)$-invariance, and $|X_{r(w)+d}| \geqslant |X_{r(w)}|/2$ by Lemma 10.5 (iv). Therefore, the number of $d \in B$ that fail to be good for (x, w) is at most $35\eta^{2/5}|B|$, because otherwise we would have

$$\begin{aligned} e(x,w) &\geqslant 1225\eta^{4/5}|B| \min_{d\in B} |X_{r(x)+d}||X_{r(w)+d}| \\ &\geqslant 306\eta^{4/5} \sum_{d\in B} |X_{r(x)+d}||X_{r(w)+d}| \\ &= 306\eta^{4/5} b(x,w)\,, \end{aligned}$$

which would imply that $\epsilon(w) \geqslant 306\eta^{4/5}$, contradicting the assumption that $w \in W'$.

So far we have shown that

$$(\forall w \in W')\big((1-35\eta^{2/5}) \text{ a.e. } d \in B\big) \quad d \text{ is good for } w\,. \tag{$*$}$$

It follows that

$$\big((1-9\eta^{1/5}) \text{ a.e. } d \in B\big)\big((1-4\eta^{1/5}) \text{ a.e. } w \in W'\big) \quad d \text{ is good for } w\,,$$

since otherwise there would be at least $36\eta^{2/5}$ pairs $(d, w) \in B \times W'$ such that d is not good for w, which contradicts $(*)$.

Let B' be the set of all $d \in B$ such that d is good for $(1-6\eta^{1/5})$-almost every $w \in W'$. We have shown that B' has cardinality at least $(1-9\eta^{1/5})|B|$, as is required in the statement of the lemma. We turn now to the definition of the function ψ.

If d is good for w, then another simple averaging argument shows that

$$\big((1-6\eta^{1/5}) \text{ a.e. } y \in X_{r(x)+d}\big)\big((1-6\eta^{1/5}) \text{ a.e. } z \in X_{r(w)+d}\big)$$
$$\phi(y) - \phi(x) = \phi(z) - \phi(w)\,.$$

Let $Y_{r(x)+d}$ be the set of such y, and for each $y \in Y_{r(x)+d}$, let Z_y be the set of $z \in X_{r(w)+d}$ such that $\phi(y) - \phi(x) = \phi(z) - \phi(w)$. Since $1 - 6\eta^{1/5} > 1/2$, any two of the sets Z_y overlap. It is also clear that ϕ is constant on any set Z_y. Therefore, it is constant on $Y_{r(x)+d}$ as well, taking a value a, say. This argument also implies that we can find a set $Y_{r(w)+d} \subset X_{r(w)+d}$ of size at least $(1-6\eta^{1/5})|X_{r(w)+d}|$ on which ϕ is constant, since we may choose $Y_{r(w)+d} = Z_y$ for some $y \in Y_{r(x)+d}$. Let this constant value be b. Then $a - \phi(x) = b - \phi(w)$, and this common value we shall call $\psi(d)$. Because ϕ is constant on $Y_{r(x)+d}$ which has size at least half that of $X_{r(x)+d}$, the value of $\psi(d)$ is well-defined (i.e., does not depend on w).

We have shown that, if d is good for w, then $\phi(z) - \phi(w) = b - \phi(w) = \psi(d)$ whenever z belongs to a set $Y_{r(w)+d}$ of cardinality at least

$(1-6\eta^{1/5})|X_{r(w)+d}|$. Therefore, for every $d \in B'$,

$$\big((1-4\eta^{1/5})\text{ a.e. } w\in W'\big)\ \big((1-6\eta^{1/5})\text{ a.e. } w'\in X_{r(w)+d}\big)\quad \phi(w')-\phi(w)=\psi(d)\,.$$

This, together with the fact that $|W'| \geqslant (1-6\eta^{1/5})|W|$, proves the lemma. (Of course, we have proved a slightly better result, but it is convenient to set all the errors equal to the worst one of $10\eta^{1/5}$.) □

Later, the following small modification of Lemma 10.6 will be useful.

LEMMA 10.7. *Let $\psi : B' \to \mathbb{Z}_N$ be the function constructed in Lemma 10.6 and let $\theta = 10\eta^{1/5}$. Then for $(1-\theta^{1/2})$-almost every $w \in W$,*

$$\big((1-\theta^{1/2})\text{ a.e. } d\in B'\big)\ \big((1-\theta)\text{ a.e. } w'\in X_{r(w)+d}\big)\quad \phi(w')-\phi(w)=\psi(d)\,.$$

Proof. This is another simple averaging argument. Let us write $P(w,d)$ for the statement

$$\big((1-\theta)\text{ a.e. } w'\in X_{r(w)+d}\big)\quad \phi(w')-\phi(w)=\psi(d)\,.$$

Lemma 10.6 states that

$$(\forall d\in B)\big((1-\theta)\text{ a.e. } w\in W\big)\quad P(w,d)\,. \qquad (*)$$

If what we wish to prove is false, then there are at least $\theta|W||B'|$ pairs $(w,d) \in W \times B'$ such that not $P(w,d)$. This contradicts $(*)$. □

Our next main task will be to prove that ψ is a homomorphism on B'. Before we do this, we prove a technical lemma which will allow us to condense what would otherwise be a very tedious argument. Roughly speaking, it tells us that we can "shift" statements by some $d \in B$, introducing only a small error.

LEMMA 10.8. *Let $d \in B$, let $\theta > 0$ and let P be a property of elements of W such that $P(w)$ for $(1-\theta)$ a.e. $w \in W$. Then*

$$\big((1-\theta^{1/2}-\eta)\text{ a.e. } w\in W\big)\ \big((1-2\theta^{1/2})\text{ a.e. } w'\in X_{r(w)+d}\big)\quad P(w')\,.$$

Proof. Let Δ be the set of pairs $(w,w') \in W^2$ such that $r(w')-r(w)=d$ and not $P(w')$. Because $P(w')$ for $(1-\theta)$ a.e. $w \in W$, the cardinality of Δ is at most $\theta|W|\max_{w\in W} R(w)$.

Now Lemma 10.5 (vi) and the symmetry of B imply that at most $\eta|W|$ elements of W fail to belong to $W \cap (W-d)$. Therefore, if the lemma is false, then there are more than $\theta^{1/2}|W|$ elements $w \in W \cap (W-d)$ such that not $P(w')$ for at least $2\theta^{1/2}|X_{r(w)+d}|$ elements of $X_{r(w)+d}$. Therefore, the cardinality of Δ is greater than $2\theta|W|\min_{w\in W} R(w)$. By Lemma 10.5 (v), this is a contradiction, so the lemma is proved. □

LEMMA 10.9. *Let $\theta = 10\eta^{1/5}$ and assume that $6\theta^{1/2} < 1$. Then the function $\psi : B' \to \mathbb{Z}_N$ constructed in Lemma 10.6 is a Freiman homomorphism.*

Proof. Suppose that $d_1, d_2, d_3, d_4 \in B'$ are such that $d_1 + d_2 = d_3 + d_4$. Lemma 10.6 tells us that

$$\big((1-\theta)\text{ a.e. } w\in W\big)\ \big((1-\theta)\text{ a.e. } w'\in X_{r(w)+d_1}\big)\quad \phi(w')-\phi(w)=\psi(d_1) \tag{1}$$

and

$$\big((1-\theta)\text{ a.e. } w\in W\big)\big((1-\theta)\text{ a.e. } w''\in X_{r(w)+d_1}\big)\quad \phi(w')-\phi(w)=\psi(d_2)\,. \tag{2}$$

Applying Lemma 10.8 to (2), with $d = d_1$, we deduce that for

$$\big((1-2\theta^{1/2})\text{ a.e. } w\in W\big)\big((1-2\theta^{1/2})\text{ a.e. } w'\in X_{r(w)+d_1}\big)$$
$$\big((1-\theta)\text{ a.e. } w''\in X_{r(w')+d_2}\big)$$

we have

$$\phi(w'') - \phi(w') = \psi(d_2)\,. \tag{3}$$

Noting that $r(w') + d_2$ is the same as $r(w) + d_1 + d_2$ when $w' \in X_{r(w)+d_1}$, we can deduce from (1) and (3) that for

$$\big((1-3\theta^{1/2})\text{ a.e. } w\in W\big)\big((1-3\theta^{1/2})\text{ a.e. } w'\in X_{r(w)+d_1}\big)$$
$$\big((1-\theta)\text{ a.e. } w''\in X_{r(w)+d_1+d_2}\big)$$

we have

$$\phi(w') - \phi(w) = \psi(d_1) \quad\text{and}\quad \phi(w'') - \phi(w') = \psi(d_2)\,. \tag{4}$$

Because $1 - 3\theta^{1/2} > 0$, it follows from (4) that in particular

$$\begin{gathered}\big((1-3\theta^{1/2})\text{ a.e. } w\in W\big)\big((1-\theta)\text{ a.e. } w''\in X_{r(w)+d_1+d_2}\big)\\ \phi(w'')-\phi(w)=\psi(d_1)+\psi(d_2)\,.\end{gathered} \tag{5}$$

An identical argument shows that

$$\begin{gathered}\big((1-3\theta^{1/2})\text{ a.e. } w\in W\big)\ \big((1-\theta)\text{ a.e. } w''\in X_{r(w)+d_3+d_4}\big)\\ \phi(w'')-\phi(w)=\psi(d_3)+\psi(d_4)\,.\end{gathered} \tag{6}$$

Since $1 - 6\theta^{1/2} > 0$ and $d_1 + d_2 = d_3 + d_4$, (5) and (6) imply that $\psi(d_1) + \psi(d_2) = \psi(d_3) + \psi(d_4)$, as required. □

We shall now specialize to the case where B is a Bohr neighbourhood. (For the definition and elementary facts, see §7.) First, we need some more easy results about such sets.

LEMMA 10.10. *Let K be a set of size k, let $B=B(K,\delta)$ and let $d\in B(K,\zeta)$. Then $|B \cap (B + d)| \geqslant (1 - 2^{k+1}\delta^{-k}k\zeta)|B|$.*

Proof. If $x \in B \setminus (B+d)$, then for some $r \in K$ we must have

$$(\delta - \zeta)N \leqslant |rd| \leqslant \delta N\,,$$

as otherwise x would belong to $B(K, \delta - \zeta)$, which would imply that $x - d \in B$. It follows that the cardinality of $B \setminus (B+d)$ is at most $2k\zeta N$. Since B has cardinality at least $(\delta/2)^k N$, the result follows. □

Corollary 10.11. *Let K be a set of size k, let $B = B(K, \delta)$, let $B' \subset B$ be a set of size at least $(7/8)|B|$, let $\zeta = 2^{-(k+4)}\delta^k k$ and let $C = B(K, \zeta)$. Then $C \subset B' - B'$ and any homomorphism ψ from B' to $\mathbb{Z}_N$ induces a homomorphism ψ_1 from C to $\mathbb{Z}_N$.*

Proof. If $d \in C$, then $|B \cap (B+d)| \geqslant (7/8)|B|$, by Lemma 10.10 and our choice of ζ. This implies that $|B' \cap (B'+d)| \geqslant (5/8)|B|$ and in particular that $d \in B' - B'$.

It follows that ψ induces a function ψ_1 on C. The content of the corollary is that ψ_1 is itself a homomorphism. To prove this, let $d_1, d_2, d_3, d_4 \in C$ with $d_1 + d_2 = d_3 + d_4$. By what we have just proved, we know that $B \cap (B+d_1)$ and $(B+d_1) \cap (B+d_1+d_2)$ both have cardinality at least $(7/8)|B|$. Therefore, $B \cap (B+d_1) \cap (B+d_1+d_2)$ has cardinality at least $(3/4)|B|$. We also know that $B \cap (B+d_3)$ has cardinality at least $(7/8)|B|$, so $B \cap (B+d_1) \cap (B+d_1+d_2) \cap (B+d_3)$ has cardinality at least $(5/8)|B|$. This implies that $B' \cap (B'+d_1) \cap (B'+d_1+d_2) \cap (B'+d_3)$ has cardinality at least $(1/8)|B|$. It follows that we can find $x \in B'$ such that $x - d_1$, $x - d_3$ and $x - d_1 - d_2 = x - d_3 - d_4$ all belong to B'. Hence, $\psi_1(d_1) + \psi_1(d_2) = \psi_1(d_3) + \psi_1(d_4)$ as was needed. □

Armed with these facts about Bohr neighbourhoods, let us return to the set W, now with the assumption that $B = B(K, \delta)$ is a Bohr neighbourhood. Let W_1 be the set of all $w \in W$ such that

$$\big((1-\theta^{1/2})\,\text{a.e. } d \in B'\big)\big((1-\theta)\,\text{a.e. } z \in X_{r(w)+d}\big) \quad \phi(z) - \phi(w) = \psi(d)\,.$$

If $\theta^{1/2} \leqslant 1/8$ (as we shall assume), then Lemma 10.7 implies that W_1 has cardinality at least $7|W|/8$.

Lemma 10.12. *Assume that $\theta^{1/2} \leqslant 1/8$. Let $B = B(K, \delta)$ and let B' and ψ be given by Lemma 10.6. Let C and ψ_1 be as in Corollary 10.11 and let $w_1, w_2 \in W_1$ with $r(w_1) - r(w_2) = c \in C$. Then $\phi(w_1) - \phi(w_2) = \psi_1(c)$.*

Proof. By the definition of W_1 and the assumption that $\theta^{1/2} \leqslant 1/8$, we have the statements

$$(7/8\ \text{ a.e. } d \in B')\big((1-\theta)\,\text{a.e. } z \in X_{r(w_1)+d}\big) \quad \phi(z) - \phi(w_1) = \psi(d) \qquad (1)$$

and

$$(7/8 \text{ a.e. } d \in B')\big((1-\theta)\,\text{a.e. } z \in X_{r(w_2)+d}\big) \quad \phi(z) - \phi(w_2) = \psi(d)\,. \quad (2)$$

Because $r(w_1) - r(w_2) \in C$, we know from the proof of Corollary 10.11 that

$$\big|B' \cap (B' - r(w_1) + r(w_2))\big| \geqslant (5/8)|B'|\,. \quad (3)$$

(2) and (3) imply that

$$\begin{gathered}(1/2 \text{ a.e. } d \in B')\big((1-\theta)\,\text{a.e. } z \in X_{d+r(w_1)}\big) \\ \phi(z) - \phi(w_2) = \psi\big(d + r(w_1) - r(w_2)\big)\,.\end{gathered} \quad (4)$$

From (1) and (4) it follows that for 3/8-almost every $d \in B'$, for $(1-\theta)$-almost every $z \in X_{r(w_1)+d}$ we have both

$$\phi(z) - \phi(w_1) = \psi(d) \text{ and } \phi(z) - \phi(w_2) = \psi\big(d + r(w_1) - r(w_2)\big)\,.$$

In particular, there exist d and z such that both equations hold, which implies that

$$\phi(w_2) - \phi(w_1) = \psi(d + r(w_1) - r(w_2)) - \psi(d) = \psi(d+c) - \psi(c) = \psi(c)\,.$$

We are now in a position to prove a new version of Lemma 7.7 in which the hypotheses are weaker. Before stating it, let us consider the constraints on the various parameters that have been introduced in this section. First of all, the strongest condition that we have placed on η is that $\theta^{1/2} < 1/6$, where $\theta = 10\eta^{1/5}$ (see Lemma 10.9). It can be checked that this condition is satisfied when $\eta = 2^{-43}$. We set $\rho = \min\{\alpha/192, \alpha^2/32\}$ and $\sigma = \eta\rho^4\alpha/384$. If $\alpha \leqslant 1/6$, then $\rho = \alpha^2/32$ and all the results of the section are satisfied (for our chosen value of η) if $\sigma = 2^{-72}\alpha^9$.

Theorem 10.13. *Let $\eta = 2^{-43}$ and let $X = X_0 \cup \cdots \cup X_{N-1}$ be the domain of a $(1-\eta)$-homomorphism ϕ of order eight. Suppose that $|X_i| \leqslant M$ for each i and that $|X| = \alpha MN$. Let $g(s)$ be the size of X_s for every s, let $\lambda = 2^{-37}\alpha^{11/2}$ and define K to be $\{r \in \mathbb{Z}_N : |\hat{g}(r)| \geqslant \lambda M\}$. Then $|K| \leqslant 2^{74}\alpha^{-10}$. Let $k = 2^{74}\alpha^{-10}$, let $\epsilon = \alpha^{-4}\lambda^4/\pi$ and let $\zeta = 2^{-155k}\alpha^{18k}k \leqslant 2^{-(k+4)}\epsilon^k k$. If $C = B(K,\zeta)$, then there is a homomorphism $\psi_1 : C \to \mathbb{Z}_N$ together with a subset $Y \subset X$ of size at least $\alpha^3|X|/1000$ such that, whenever $y, z \in Y$ and $r(y) - r(z) \in C$, we have $\phi(y) - \phi(z) = \psi_1(r(y) - r(z))$.*

Proof. Let $B = B(K,\epsilon)$ and let $L = \alpha^3M^4N^3$. We know that $|2X - 2X| = (\alpha MN)^4$ and that $|(2X-2X)_s| \leqslant \alpha^3M^4N^3 = L$ for every s. Now we shall show that $2X - 2X$ is $(B, \sigma L)$-invariant.

If we write $h(s) = |(2X-2X)_s|$, then $\hat{h}(r) = |\hat{g}(r)|^4$. We know also that $|\hat{g}(r)| \leqslant \alpha MN$ for every r. Since $\|\hat{g}\|^2 = N\|g\|^2 \leqslant \alpha M^2N^2$, we find

that $|K| \leqslant \lambda^{-2}\alpha = 2^{74}\alpha^{-10}$, as in the proof of Lemma 7.7. We also have the obvious inequality

$$\sum_{r\notin K} |\hat{g}(r)|^4 \leqslant \lambda^2 M^2 N^2 \|\hat{g}\|^2 \leqslant \alpha\lambda^2 M^4 N^4 .$$

Now $h(s) = N^{-1}\sum_r \hat{h}(r)\omega^{rs}$, so

$$\begin{aligned} h(s) - h(t) &= N^{-1}\sum_r |\hat{g}(r)|^4(\omega^{rs} - \omega^{rt}) \\ &= N^{-1}\sum_{r\in K} |\hat{g}(r)|^4\omega^{rt}(\omega^{r(s-t)}-1) + N^{-1}\sum_{r\notin K} |\hat{g}(r)|^4\omega^{rt}(\omega^{r(s-t)}-1). \end{aligned}$$

From the inequality above, the sum over $r \notin K$ is at most $2\alpha\lambda^2 M^4 N^3$ (after the multiplication by N^{-1}). As for the other part, if we make the additional assumption that $s - t \in B$, then $|\omega^{r(s-t)} - 1| \leqslant 2\pi\epsilon$ for each $r \in K$, so the sum is at most $N^{-1}2\pi\epsilon|K_\lambda|(\alpha MN)^4 \leqslant 2\pi\epsilon\lambda^{-2}\alpha^5 M^4 N^3 = 2\alpha\lambda^2 M^4 N^3$. The $(B, \sigma L)$-invariance of $2X - 2X$ follows.

We know that ϕ induces a $(1-\eta)$-homomorphism ϕ'' (of order two) on $2X - 2X$. Therefore, we can find a set W of cardinality at least $\rho\alpha^2 LN/16 = \alpha^7 M^4 N^4/512$ with the properties claimed in Lemma 10.5. Corollary 10.11, Lemma 10.12 and the definition in between then give us a set W_1 of cardinality at least $\alpha^7 M^4 N^4/1000 = \alpha^3|2X - 2X|/1000$ and a homomorphism $\psi_1 : W_1 \to \mathbb{Z}_N$ such that, whenever $w_1, w_2 \in W_1$ and $r(w_1) - r(w_2) \in C$, we have $\phi''(w_1) - \phi''(w_2) = \psi_1(w_1 - w_2)$.

Now choose $(x_2, x_3, x_4) \in X^3$ uniformly at random. The expected number of $y \in X$ such that $(y, x_2, x_3, x_4) \in W$ is at least $\alpha^3|X|/1000$, so let us fix (x_2, x_3, x_4) such that the set Y of all y such that $(y, x_2, x_3, x_4) \in W$ has cardinality at least $\alpha^3|X|/1000$. If $y, z \in Y$ and $r(y) - r(z) = c \in C$, then $r(y, x_2, x_3, x_4) - r(z, x_2, x_3, x_4) = c$, so

$$\phi(y) - \phi(z) = \phi''(y, x_2, x_3, x_4) - \phi''(z, x_2, x_3, x_4) = \psi(y - z) .$$

This proves the theorem. □

Corollary 10.14. *Let K be as in Theorem 10.13, let Y be the set obtained there and let m be a positive integer. For every $d \in B(K, \zeta/m)$ there exists c such that $\phi(x) - \phi(y) = c(r(x) - r(y))$ whenever $x, y \in Y$ and $r(x) - r(y)$ belongs to the set $\{jd : -m \leqslant j \leqslant m\}$.*

Proof. As with Corollary 7.8 this follows from the observations that $\{jd : -m \leqslant j \leqslant m\} \subset B(K, \zeta)$, that the restriction of any homomorphism to $\{jd : -m \leqslant j \leqslant m\}$ is linear and that $\psi_1(0) = 0$. □

11 The Problem of Longer Progressions

This section is a brief introduction to the rest of the paper and the difficulties that must be overcome before the proof can be extended from progressions of length four to progressions of arbitrary length. As with the other known proofs of Szemerédi's theorem, the new difficulties that arise with progressions of length greater than four are considerable. In our case, it is because we must extend Freiman's theorem (or, to be more accurate, our weaker version of Freiman's theorem) from "linear" functions to "multilinear" ones.

To see this, consider the case of progressions of length five. The main result of §3 suggests that we should go up a degree, and look at sets that fail to be uniform of degree three, or, as we shall say, cubically uniform. (Sets such as $\{x \in \mathbb{Z}_N : |x^3| \leqslant N/10000\}$ show that this is necessary as well as sufficient.) If A is such a set and f is the balanced function of A, then $\Delta(f; k, l)$ has a large Fourier coefficient for many values of k, l. In other words, we can find a large subset $B \subset \mathbb{Z}_N^2$ and a function $\phi : B \to \mathbb{Z}_N$ such that $\Delta(f; k, l)$ has a large Fourier coefficient at $\phi(k, l)$. By the main result of §6, for some reasonably large $\gamma > 0$ the function ϕ is γ-additive in both variables, and this is true for the restriction of ϕ to any large subset of B. That is, for many x we can fix x and $\phi(x, y)$ will be somewhat additive in y, and vice versa.

The object of the next few sections will be to look at such "somewhat bi-additive" functions, and show that there is a large subset $C \subset B$ such that the restriction of ϕ to C resembles a multidimensional bilinear function, rather as a somewhat additive function has a restriction resembling a multidimensional linear one. This involves showing that the multidimensional linearity of ϕ in x somehow "interacts" with the multidimensional linearity in y, which turns out to be harder than one might think, as we shall now explain.

First, it is important that the additivity property should hold for restrictions of ϕ. For example, let λ be an arbitrary function from $\mathbb{Z}_N$ to $\mathbb{Z}_N$, and define

$$\phi(x, y) = \begin{cases} \lambda(x)y & 0 \leqslant x \leqslant y < N \\ x\lambda(y) & 0 \leqslant y < x < N\,. \end{cases}$$

There are certainly many additive quadruples in each variable, but if λ does not have special additivity properties, then the quadruples with x fixed do not mix with those with y fixed and there is nothing more to say about ϕ, and in particular no restriction of ϕ that looks bilinear.

Let us informally call a function *quasilinear* if it resembles a low-dimensional linear function (see, for example, the function defined at the end of §6). A more serious complication arises even if we know for every x that $\phi(x,y)$ is quasilinear in y for every x and vice versa. It is tempting to suppose that one might be able to find a large subset $B' \subset B$, and numbers $x_0, x_1, \dots, x_d,\ r_1, \dots, r_d,\ y_0, y_1, \dots, y_d,\ s_1, \dots, s_d$ and $(c_{ij})_{i,j=0}^d$ such that the restriction of ϕ to B' was of the form

$$\phi\Bigl(x_0 + \sum_{i=1}^d a_i x_i, y_0 + \sum_{i=1}^d b_j y_j\Bigr) = \sum_{i,j=0}^d c_{ij} a_i b_j$$

for $0 \leqslant x_i < r_i$ and $0 \leqslant y_j < s_j$.

However, this would imply that one could find a small "common basis" for all the functions $y \mapsto \phi(x,y)$ (and similarly the other way round) and a simple example shows that such a statement is too strong. Indeed, let ψ be a non-trivial (i.e., non-linear) quasilinear function from $\mathbb{Z}_N$ to $\mathbb{Z}_N$. (For definiteness one could let $\psi(z) = z \pmod m$ for some m near $\sqrt{N}$.) Define $\phi(x,y)$ to be $\psi(xy)$. The natural bases for the functions $y \mapsto \psi(xy)$ are all completely different, and there is no small basis that can be used for all (or even a large proportion) of them. We shall not prove this here.

However, just as what we really used when proving Szemerédi's theorem for progressions of length four was Corollary 7.10, which told us that the function ϕ had a small (but not too small) linear restriction, the statement we actually need for progressions of length five is that one can find reasonably long arithmetic progressions (we obtain a power of N) P and Q with the same common difference and a bilinear function $\psi : P \times Q \to \mathbb{Z}_N$ such that ψ agrees with ϕ for a significant proportion of the points $(x,y) \in P \times Q$. If we wish to prove Szemerédi's theorem for progressions of length k, we need the obvious generalization of this to $(k-3)$-linear functions. In proving these statements, we shall obtain some insight into the form of a typical "quasimultilinear" function, but we avoid having to describe them precisely. It would be interesting to obtain a precise description, so this is an area where there is still work to be done.

It seems, then, that there is something objective about the problem which makes the difficulty increase sharply as the size of the desired progression goes from two to three to four to five, and then remain roughly constant from that point onwards. Three is the first non-trivial case, four involves quadratic functions rather than just linear ones and five involves bilinearity in the large Fourier coefficients rather than just linearity, but that is the last time that some parameter, which one has hardly noticed

because it equals one, suddenly and annoyingly changes to two.

12 Strengthening a Bihomomorphism

Although the proof of Szemerédi's theorem for progressions of length five is not significantly easier than it is for the result in general, the notation is cleaner and one or two complications can be avoided. Therefore, we shall treat this case separately. Let us take a non-cubically uniform function $f : \mathbb{Z}_N \to D$ and begin the longish process of finding bilinear behaviour in any function ϕ for which $\Delta(f;k,l)^\wedge(\phi(k,l))$ is often large.

In order to motivate some of the lemmas that follow, let us consider what the natural two-variable analogue of a Freiman homomorphism ought to be. That is, given a subset $A \subset \mathbb{Z}_N^2$, we ask what property of a function $\phi : A \to \mathbb{Z}_N$ relates to that of being a homomorphism in the way that bilinearity relates to linearity. In the last section, we discussed an analogous problem for quasilinear functions rather than homomorphisms, and saw that it was not easy to give a satisfactory definition. Giving a good definition of a "bihomomorphism" is not all that easy either.

The most obvious definition is that $\phi(x,y)$ should be a homomorphism in y for any fixed x, and vice versa. This property can indeed be shown to hold for the functions ϕ that will concern us. However, to see that it is natural to ask for more, consider the set $A = A_1 \cup A_2$, where A_1 is the set of all (x,y) such that $0 < x < N/2$ and $0 < y < N/2$, while A_2 is the set of all (x,y) such that $N/2 < x < N$ and $N/2 < y < N$. Define a function ϕ by letting $\phi(x,y)$ be xy if $(x,y) \in A_1$ and $2xy$ if $(x,y) \in A_2$. This function has the following undesirable property. Suppose we define a new function ψ by setting

$$\psi(x,d) = \phi(x,y+d) - \phi(x,y)$$

whenever y can be found such that both (x,y) and $(x,y+d)$ belong to A. This is a well-defined function and for fixed x it is an isomorphism in d. However, it can be checked very easily that for fixed d it is not an isomorphism in x. This suggests that a stronger property will probably be useful, and the suggestion turns out to be correct.

To simplify the discussion, let us introduce some terminology. A *vertical parallelogram* is a quadruple of points in $\mathbb{Z}_N^2$ of the form $((x,y),(x,y+h),(x+w,y'),(x+w,y'+h))$. We shall call w and h respectively the *width* and *height* of the parallelogram. If P is the above parallelogram, then we shall denote these by $w(P)$ and $h(P)$. If ϕ is a function from $A \subset \mathbb{Z}_N^2$ to

$\mathbb{Z}_N$ and all the points of P lie in A, then we set

$$\phi(P) = \phi(x,y) - \phi(x,y+h) - \phi(x+w,y') + \phi(x+w,y'+h)\,.$$

Ideally, we would like to find, given suitable conditions on ϕ, a large set such that, for any vertical parallelogram P lying in the set, $\phi(P)$ depends only on the width and height of P. This may be possible, and has the potential to simplify this paper considerably, but we have not managed to find an argument for or against it. Instead, we shall obtain a set where $\phi(P)$ is *nearly* independent of everything except for the width and height.

Our first main task will be to find many pairs P_1, P_2 of vertical parallelograms of the same width and height, such that $\phi(P_1) = \phi(P_2)$. For this, we shall need a slight generalization of Proposition 6.1, proved in exactly the same way.

PROPOSITION 12.1. *For each $k \in \mathbb{Z}_N$, let $\lambda_k \geqslant 0$. Let $f_1, \dots, f_p$ be functions from $\mathbb{Z}_N$ to D and let $\phi_1, \dots, \phi_p$ be functions from $\mathbb{Z}_N$ to $\mathbb{Z}_N$ such that*

$$\sum_k \lambda_k \prod_{i=1}^p |\Delta(f_i;k)^\wedge(\phi_i(k))|^2 \geqslant \alpha N^{2p+1}\,.$$

Call a quadruple $(a,b,c,d) \in \mathbb{Z}_N^4$ simultaneously additive if $a-b = c-d$ and $\phi_i(a)-\phi_i(b) = \phi_i(c)-\phi_i(d)$ for every $i \leqslant p$. Then the sum of $\lambda_a\lambda_b\lambda_c\lambda_d$ over all simultaneously additive quadruples (a,b,c,d) is at least $\alpha^4 N^3$.

Proof. Expanding the given inequality yields that

$$\sum_k \lambda_k \sum_{s_1,\dots,s_p} \sum_{t_1,\dots,t_p} \prod_{i=1}^p f_i(s_i)\overline{f_i(s_i-k)f_i(t_i)}f_i(t_i-k)\omega^{-\phi_i(k)(s_i-t_i)} \geqslant \alpha N^{2p+1}.$$

Substituting $u_i = s_i - t_i$ then gives

$$\sum_k \lambda_k \sum_{s_1,\dots,s_p} \sum_{u_1,\dots,u_p} \prod_{i=1}^p f_i(s_i)\overline{f_i(s_i-k)f_i(s_i-u_i)}f_i(s_i-k-u_i)\omega^{-\phi_i(k)u_i} \geqslant \alpha N^{2p+1}.$$

Since $|f_i(x)| \leqslant 1$ for every x and i, this implies that

$$\sum_{s_1,\dots,s_p} \sum_{u_1,\dots,u_p} \Big|\sum_k \lambda_k \prod_{i=1}^p \overline{f_i(s_i-k)}f(s_i-k-u_i)\omega^{-\phi_i(k)u_i}\Big| \geqslant \alpha N^{2p+1}$$

and hence, by the Cauchy-Schwarz inequality, that

$$\sum_{s_1,\dots,s_p} \sum_{u_1,\dots,u_p} \Big|\sum_k \lambda_k \prod_{i=1}^p \overline{f_i(s_i-k)}f(s_i-k-u_i)\omega^{-\phi_i(k)u_i}\Big|^2 \geqslant \alpha^2 N^{2p+2}\,.$$

Let us introduce a new variable s and write $v_i = s - s_i$. Then, multiplying both sides by N (in different ways) we obtain

$$\sum_s \sum_{u_1,\dots,u_p} \sum_{v_1,\dots,v_p} \Big|\sum_k \lambda_k \prod_{i=1}^p \overline{f_i(s-v_i-k)} f_i(s-v_i-k-u_i)\omega^{-\phi_i(k)u_i}\Big|^2 \geqslant \alpha^2 N^{2p+3}.$$

We now apply Lemma 2.1 to the functions

$$a_{u,v}(k) = \prod_{i=1}^p \overline{f_i(-v_i-k)} f_i(-v_i-k-u_i)$$

and

$$b_{u,v}(k) = \lambda_k \prod_{i=1}^p \omega^{\phi_i(k)u_i},$$

which tells us that

$$\sum_{u,v}\sum_r |\hat{a}_{u,v}(r)|^2 |\hat{b}_{u,v}(r)|^2 \geqslant \alpha^2 N^{2p+4}.$$

By the Cauchy-Schwarz inequality it follows that

$$\Big(\sum_{u,v}\sum_r |\hat{a}_{u,v}(r)|^4\Big)\Big(\sum_{u,v}\sum_r |\hat{b}_{u,v}(r)|^4\Big) \geqslant \alpha^4 N^{4p+8}.$$

Now $\sum_r |\hat{a}_{u,v}(r)|^4$ is, for every u, v, at most N^4 (e.g. by §2 (6)) so $\sum_{u,v}\sum_r |\hat{a}_{u,v}(r)|^4 \leqslant N^{2p+4}$. Since $\hat{b}_{u,v}(r) = \sum_k \lambda_k \prod_{i=1}^p \omega^{\phi_i(k)u_i - rk}$, which does not depend on $v = (v_1, \dots, v_k)$, it follows that

$$\sum_{u_1,\dots,u_p}\sum_r \Big|\sum_k \lambda_k \prod_{i=1}^p \omega^{\phi_i(k)u_i - rk}\Big|^4 \geqslant \alpha^4 N^{p+4}.$$

But the left-hand side above is easily seen to be N^{p+1} times the sum of $\lambda_a\lambda_b\lambda_c\lambda_d$ over all simultaneously additive quadruples (a, b, c, d). The result is proved. □

We shall now apply the above result to find many good pairs of parallelograms. Note that the number of pairs (P_1, P_2) of vertical parallelograms with the same width and height is N^8.

LEMMA 12.2. *Let $\gamma, \eta > 0$, let $f : \mathbb{Z}_N \to D$ and let $B \subset \mathbb{Z}_N$ be a set of cardinality βN^2 such that $|\Delta(f; k, l)\hat{}(\phi(k, l))| \geqslant \gamma N$ for every $(k, l) \in B$. Then there are at least $\beta^{16}\gamma^{48}N^8$ pairs (P_1, P_2) of vertical parallelograms such that P_1 and P_2 have the same width and height and such that $\phi(P_1) = \phi(P_2)$.*

Proof. The average size of a vertical cross-section of B (that is, a set of the form $B_x = \{y \in \mathbb{Z}_N : (x, y) \in B\}$) is βN. Hence, by Lemma 6.1

and Hölder's inequality, the average number of additive quadruples in a vertical cross-section of B is at least $(\beta\gamma^2)^4N^3$. We shall call a pair of points $\big((x,y),(x,y+h)\big)$ a *vertical edge of height* h. Given such a pair, define $q\big((x,y),(x,y+h)\big)$ to be the number of $y' \in \mathbb{Z}_N$ such that

$$\phi(x,y+h)-\phi(x,y)=\phi(x,y'+h)-\phi(x,y')\,,$$

where equality is deemed not to hold unless ϕ is defined at all four points. Letting $\zeta=(\beta\gamma^2)^4$, we have that the average value of $q(e)$ over all vertical edges e is at least ζN.

For each h, let $\zeta(h)$ be the average of $q(e)$ over vertical edges e of height h. We can find y such that, setting $\lambda_x = N^{-1}q\big((x,y),(x,y+h)\big)$, we have $\sum_x \lambda_x \geqslant \zeta(h)N$. Since λ_x is zero unless both (x,y) and $(x,y+h)$ lie in B, this tells us that

$$\sum_x \lambda_x \big|\Delta(f;x,y+h)^\wedge(\phi(x,y+h))\big|^2\big|\Delta(f;x,y)^\wedge(\phi(x,y))\big|^2 \geqslant \zeta(h)\gamma^4N^5\,.$$

Hence, by Proposition 12.1, the sum of $\lambda_a\lambda_b\lambda_c\lambda_d$ over all quadruples (a,b,c,d) such that $a-b=c-d$, $\phi(a,y)-\phi(b,y)-\phi(c,y)+\phi(d,y)$ and $\phi(a,y+h)-\phi(b,y+h)-\phi(c,y+h)+\phi(d,y+h)$ is at least $(\zeta(h)\gamma^4)^4N^3$. Each such quadruple gives rise to a set of $N^4\lambda_a\lambda_b\lambda_c\lambda_d$ pairs of parallelograms with the desired properties, and all these sets are disjoint. Summing over all h and using the fact that the average value of $\zeta(h)$ is ζ, we obtain from Hölder's inequality that the total number of pairs of parallelograms with the given properties is at least $\zeta^4\gamma^{16}N^8$, which proves the result. □

We shall in fact need many arrangements of *eight* parallelograms $(P_1,\dots,P_8)$, all of the same height, such that

$$w(P_1)-w(P_2)-w(P_3)+w(P_4)=w(P_5)-w(P_6)-w(P_7)+w(P_8)$$

and

$$\phi(P_1)-\phi(P_2)-\phi(P_3)-\phi(P_4)=\phi(P_5)-\phi(P_6)-\phi(P_7)+\phi(P_8)\,.$$

(It is not particularly natural to divide the resulting 32 points into parallelograms – we do this merely to provide a link to the discussion so far.) It turns out that this follows automatically from Lemma 12.2. First we need a result similar to Lemma 9.2.

LEMMA 12.3. *Let $B \subset \mathbb{Z}_N^2$ and let $\phi : B \to \mathbb{Z}_N$. Suppose that there are θN^8 pairs of parallelograms (P_1,P_2) in B such that $h(P_1)=h(P_2)$, $w(P_1)=w(P_2)$ and $\phi(P_1)=\phi(P_2)$. Then there are at least θ^7N^{32} sequences $(x_1,\dots,x_{16},y_1,\dots,y_{16},h)$ such that*

$$x_1+\cdots+x_8=x_9+\cdots+x_{16}$$

and

$$\phi_h(x_1,y_1)+\cdots+\phi_h(x_8,y_8)=\phi_h(x_9,y_9)+\cdots+\phi_h(x_{16},y_{16})\,,$$

where $\phi_h(x,y)$ *stands for* $\phi(x,y+h)-\phi(x,y)$.

Proof. Given $u\in\mathbb{Z}_N$, define $g_u(x,y)$ to be $\omega^{u\phi(x,y)}$ if $(x,y)\in B$, and zero otherwise. Let $f_{u,h}(x)=\sum_y g_u(x,y+h)\overline{g_u(x,y)}$. Adopting the convention that ω raised to an undefined power is zero, we can write

$$f_{u,h}(x)=\sum_y \omega^{u(\phi(x,y+h)-\phi(x,y))}\,.$$

Clearly, $|f_{u,h}(x)|\leqslant N$ for every u,h,x, from which it follows that $\sum_x|f_{u,h}(x)|^2\leqslant N^3$ for every u,h and therefore that $\sum_{u,r}|\hat{f}_{u,h}(r)|^2\leqslant N^5$ for every h.

Next, we look at fourth powers. We have

$$\sum_{u,r}|\hat{f}_{u,h}(r)|^4=\sum_{u,r}\Big|\sum_{x,y}\omega^{u(\phi(x,y+h)-\phi(x,y))-rx}\Big|^4$$

which works out as N^2 times the number of octuples $(x_1,x_2,x_3,x_4,y_1,y_2,y_3,y_4)$ such that $x_1-x_2=x_3-x_4$ (so that the sum over r is N rather than zero) and

$$\phi_h(x_1,y_1)-\phi_h(x_2,y_2)=\phi_h(x_3,y_3)-\phi_h(x_4,y_4),$$

where we take the equality to be false unless both sides are defined. In other words, $\sum_{u,r}|\hat{f}_{u,h}(r)|^4$ is N^2 times the number of parallelogram pairs of height h with the same width and same value of ϕ.

Finally, we look at sixteenth powers. It is not hard to check that $\sum_{u,r}|\hat{f}_{u,h}(r)|^{16}$ counts N^2 times the number of sequences $(x_1,\dots,x_{16}, y_1,\dots,y_{16})$ such that

$$x_1+\cdots+x_8=x_9+\cdots+x_{16}$$

and

$$\phi_h(x_1,y_1)+\cdots+\phi_h(x_8,y_8)=\phi_h(x_9,y_9)+\cdots+\phi_h(x_{16},y_{16})\,.$$

From our assumption and the above arguments, we know that $\sum_{u,r,h}|\hat{f}_{u,h}(r)|^2\leqslant N^6$ and $\sum_{u,r,h}|\hat{f}_{u,h}(r)|^4\geqslant\theta N^{10}$. It follows from Lemma 9.1 that $\sum_{u,r,h}|\hat{f}_{u,h}(r)|^{16}\geqslant(\theta N^{10}/N^{36/7})^7=\theta^7N^{34}$. Hence, the number of sequences with the desired properties is at least θ^7N^{32}, as stated. □

Next, we combine Lemmas 12.2 and 12.3 in the obvious way.

LEMMA 12.4. *Let* $\beta,\gamma>0$, *let* $f:\mathbb{Z}_N\to D$, *let* $B\subset\mathbb{Z}_N^2$ *be a set of cardinality* βN^2 *and let* $\phi:B\to\mathbb{Z}_N$. *Suppose that* $|\Delta(f;k,l)\hat{}(\phi(k,l))|\geqslant\gamma N$ *for every* $(k,l)\in B$. *Then there are at least* $\beta^{112}\gamma^{336}N^{32}$ *sequences* $(x_1,\dots,x_{16},y_1,\dots,y_{16},h)$ *such that*

$$x_1+\cdots+x_8=x_9+\cdots+x_{16}$$

and

$$\phi_h(x_1,y_1)+\cdots+\phi_h(x_8,y_8)=\phi_h(x_9,y_9)+\cdots+\phi_h(x_{16},y_{16})\,.$$

Proof. Lemma 12.2 allows us to take $\theta=\beta^{16}\gamma^{48}$ in Lemma 12.3. □

Let us define a *d-arrangement of height h* to be a sequence of points $\big((x_1,y_1),(x_1,y_1+h),(x_2,y_2),(x_2,y_2+h),\dots,(x_{2d},y_{2d}),(x_{2d},y_{2d}+h)\big)$ such that $x_1+\cdots+x_d=x_{d+1}+\cdots+x_{2d}$. Given a set $B\subset\mathbb{Z}_N^2$ and a function $\phi:B\to\mathbb{Z}_N$ we shall say that ϕ *respects* such a d-arrangement if

$$\phi_h(x_1,y_1)+\cdots+\phi_h(x_d,y_d)=\phi_h(x_{d+1},y_{d+1})+\cdots+\phi_h(x_{2d},y_{2d})\,.$$

Of course, for this to happen, all the points of the d-arrangement must lie in the set B, so that ϕ_h is defined where it needs to be. Our interest will be principally in 8-arrangements.

Lemma 12.4 gives us, under certain hypotheses on B and ϕ, a large collection of 8-arrangements in the set B that are respected by ϕ. Indeed, since the total number of 8-arrangements cannot possibly exceed $\beta^{15}N^{32}$ if $|B|=\beta N^2$, it shows that the proportion of 8-arrangements respected by ϕ is greater than zero (and independent of N). In the rest of this section, we shall show how to choose a large subset $B'\subset B$ such that ϕ respects *almost all* of the 8-arrangements in B'. As in §9, when we restricted to an approximate homomorphism of order eight, this is done by a random selection with suitable dependences, with Riesz products to define the probabilities.

LEMMA 12.5. *Let $\eta>0$, let $B\subset\mathbb{Z}_N^2$ be a set of size βN^2 and let $\phi:B\to\mathbb{Z}_N$ be a function that respects at least $\alpha\beta^{15}N^{32}$ 8-arrangements. If N is sufficiently large (depending on β and η) then there is a subset $B'\subset B$ containing at least $(\alpha\eta/4)^{2^{36}}\beta^{15}N^{32}$ 8-arrangements, such that the proportion of 8-arrangements respected by ϕ is at least $1-\eta$.*

Proof. Choose $r_1,\dots,r_k,\ s_1,\dots,s_k,\ t_1,\dots,t_k\in\mathbb{Z}_N$ uniformly and independently at random from $\mathbb{Z}_N$. Having made the choice, let each point $(x,y)\in B$ be in B' with probability

$$p(x,y)=2^{-k}\prod_{i=1}^{k}\big(1+\cos\tfrac{2\pi}{N}(r_iy+s_ixy+t_i\phi(x,y))\big)\,,$$

and let these choices be independent. Note once again that this independence exists only after we condition on the choice of $r_1,\dots,r_k,\ s_1,\dots,s_k,\ t_1,\dots,t_k$: it is very important that in total there *is* a dependence. Now consider a sequence of points $(a_1,b_1),\dots,(a_{32},b_{32})$. The probability that

they are all chosen is

$$N^{-3k}\sum_{r_1,\dots,r_k}\sum_{s_1,\dots,s_k}\sum_{t_1,\dots,t_k}2^{-32k}\prod_{i=1}^{k}\prod_{j=1}^{32}\bigl(1+\cos\tfrac{2\pi}{N}(r_ib_j+s_ia_jb_j+t_i\phi(a_j,b_j))\bigr)$$

which equals

$$N^{-3k}2^{-32k}\Bigl(\sum_{r,s,t}2^{-32}\prod_{j=1}^{32}\bigl(1+1+\omega^{rb_j+sa_jb_j+t\phi(a_j,b_j)}+\omega^{-(rb_j+sa_jb_j+t\phi(a_j,b_j))}\bigr)\Bigr)^k.$$

When the product over j is expanded, each term is of the form

$$\omega^{r\sum\epsilon_jb_j+s\sum\epsilon_ja_jb_j+t\sum\epsilon_j\phi(a_j,b_j)}\,,$$

where $\epsilon_1,\dots,\epsilon_{32}$ belong to the set $\{-1,0,1\}$. Each such term, when summed over r,s and t, gives zero, unless

$$\sum_{j=1}^{32}\epsilon_jb_j=\sum_{j=1}^{32}\epsilon_ja_jb_j=\sum_{j=1}^{32}\epsilon_j\phi(a_j,b_j)=0\,,$$

in which case it gives N^3.

Now let us suppose that our sequence of points (a_i,b_i) forms an 8-arrangement. Then we can write $a_{2i-1}=a_{2i}=x_i$ and $b_{2i-1}=b_{2i}-h=y_i$ for some $(x_1,\dots,x_{16},y_1,\dots,y_{16},h)$ such that $x_1+\cdots+x_8=x_9+\cdots+x_{16}$. If $\epsilon_1=1$ and $\epsilon_2\neq-1$ and the corresponding term does not make a zero contribution, then $\epsilon_1b_1+\epsilon_2b_2$ is either y_1 or $2y_1+h$ and this must be zero. The number of choices of $(x_1,\dots,x_{16},y_1,\dots,y_{16},h)$ for which this is true and $x_1+\cdots+x_8=x_9+\cdots+x_{16}$ is at most N^{31} in each case. Repeating this argument for each ϵ_{2i-1} shows that the number of 8-arrangements making a non-zero contribution to a term where we do not have $\epsilon_{2j-1}+\epsilon_{2j}=0$ for every i is at most $32N^{31}$.

If $\epsilon_{2j-1}+\epsilon_{2j}=0$ for every j, then

$$\sum_{j=1}^{32}\epsilon_ja_jb_j=h(\epsilon_2x_1+\epsilon_4x_2+\cdots+\epsilon_{32}x_{16})\,.$$

The number of 8-arrangements of height 0 is obviously at most N^{31}. Let $(\epsilon_2,\epsilon_4,\dots,\epsilon_{32})$ be a sequence which is not a multiple of the sequence $(1,\dots,1,-1,\dots,-1)$ (where 1 and -1 each occur eight times). The number of 8-arrangements such that $\epsilon_2x_1+\epsilon_4x_2+\cdots+\epsilon_{32}x_{16}=0$ is at most N^{31} because we are imposing two independent linear conditions on the sequence $(x_1,\dots,x_{16},y_1,\dots,y_{16},h)$ (in the vector space $\mathbb{Z}_N^{33}$). Hence, with the exception of at most $(33+3^{16})N^{31}$ of them, an 8-arrangement makes a non-zero contribution to the sum only for sequences $\epsilon_1,\dots,\epsilon_{32}$ such that

$\epsilon_{2j-1}+\epsilon_{2j}=0$ for every j, and $\epsilon_2=\epsilon_4=\cdots=\epsilon_{16}=-\epsilon_{18}=\cdots=-\epsilon_{32}$. Moreover, the contribution *will* be zero unless $\sum_{j=1}^{32}\epsilon_j\phi(a_j,b_j)=0$. (This last statement follows from considering the sum over t.)

Our argument has shown that, ignoring at most $(65+3^{16})N^{31}$ degenerate cases, given an 8-arrangement in B, the probability that it lies in B' is $2^{-32k}\big(2^{-32}(2^{32}+2)\big)^k$ if ϕ respects the 8-arrangement, but only 2^{-32k} if it does not. Hence, our hypotheses imply that the expected number X of 8-arrangements respected by ϕ is at least $2^{-32k}(1+2^{-31})^k\alpha\beta^{15}N^{32}$, and the expected number Y of bad but non-degenerate 8-arrangements is at most $2^{-32k}\beta^{15}N^{32}$. Using the fact that $2^{2^{-31}}\leqslant 1+2^{-31}$, we can deduce that if $2^{2^{-31}k}\geqslant 2/\alpha\eta$, then

$$\eta\mathbb{E}X-\mathbb{E}Y\geqslant\alpha\eta(2/\alpha\eta)2^{-32k}\beta^{15}N^{32}-2^{-32k}\beta^{15}N^{32}=2^{-32k}\beta^{15}N^{32}.$$

Now $2^{2^{-31}k}\geqslant(2/\alpha\eta)$ if and only if $2^{-32k}\leqslant(\alpha\eta/2)^{2^{36}}$. Let k be an integer such that

$$2(\alpha\eta/4)^{2^{36}}\leqslant 2^{-64k}\leqslant(\alpha\eta/2)^{2^{36}}.$$

If N is large enough that $(\alpha\eta/4)^{2^{36}}\beta^{15}N\geqslant 65+3^{16}$, then the values for the above expectations and the upper estimate for the number of degenerate 8-arrangements imply that there exists a set B' such that $\eta X\geqslant Y$ and $X\geqslant(\alpha\eta/4)^{2^{36}}\beta^{15}N^{32}$, as was claimed. □

If we combine Lemmas 12.4 and 12.5 we obtain the main result of this section.

LEMMA 12.6. *Let $\beta,\gamma,\eta>0$. Let $f:\mathbb{Z}_N\to D$, let $B\subset\mathbb{Z}_N^2$ be a set of cardinality at least βN^2 and let $\phi:B\to\mathbb{Z}_N$ be such that $|\Delta(f;k,l)^\wedge(\phi(k,l))|\geqslant\gamma N$ for every $(k,l)\in B$. Then there is a subset $B'\subset B$ containing at least $2^{-2^{37}}\beta^{2^{43}}\gamma^{2^{45}}\eta^{2^{36}}N^{32}$ 8-arrangements, such that the proportion of them respected by ϕ is at least $1-\eta$.*

Proof. By Lemma 12.4 there are at least $\beta^{112}\gamma^{336}N^{32}$ 8-arrangements respected by ϕ. This allows us to take $\alpha=\beta^{97}\gamma^{336}$ in Lemma 12.5. It is not hard to check that $(\beta^{97}\gamma^{336}\eta/4)^{2^{36}}\beta^{15}\geqslant 2^{-2^{37}}\beta^{2^{43}}\gamma^{2^{45}}\eta^{2^{36}}N^{32}$, so the lemma is proved. □

13 Finding a Bilinear Piece

We shall now use the results of the previous two sections to prove that if $A\subset\mathbb{Z}_N$ is a set with balanced function f, B is a large subset of A and $\phi:B\to\mathbb{Z}_N$ has the property that $\Delta(f;x,y)^\wedge(\phi(x,y))$ is large for every $(x,y)\in B$, then ϕ exhibits a small (but not too small) amount of bilinearity,

in the following sense: there are arithmetic progressions $P, Q \subset \mathbb{Z}_N$ of size a power of N and with the same common difference, and a large subset C of $B \cap (P \times Q)$ such that the restriction of ϕ to C is bilinear. This is the key to extending our proof from progressions of length four to progressions of length five.

What we prove in this section is sufficient for finding progressions of length five, but not as strong as the corresponding case of the inductive hypothesis we shall need when generalizing the argument. Then it becomes necessary to show that almost all of the graph of ϕ is contained in a small number of bilinear pieces, which is not a huge extra difficulty but it makes the argument look more complicated. Another way in which the argument of this section is slightly simpler than the argument for the general case (in §16) is that we can use Lemma 7.10 to allow us to assume that $\phi(x, y)$ is a homomorphism of order 8 in y for every fixed x and vice versa (see the proof of Theorem 13.10 for this).

The next lemma is another generalization of Proposition 6.1 with an almost identical proof. To recover the earlier proposition for the function $f : \mathbb{Z}_N \to D$, apply this coming lemma to the function $g(x, y) = f(x + y)$.

LEMMA 13.1. *Let $f : \mathbb{Z}_N^2 \to D$ be a function into the closed unit disc. For any h, define*

$$f_h(x) = \sum_y f(x, y + h)\overline{f(x, y)}\,.$$

Let $B \subset \mathbb{Z}_N$ and let $\sigma : B \to Z_N$ be a function such that

$$\sum_{h \in B} \left|\hat{f}_h(\sigma(h))\right|^2 \geqslant \alpha N^5\,.$$

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

Proof. Expanding what the hypothesis says, we find that

$$\sum_{h \in B} |\hat{f}_h(\sigma(h))|^2 = \sum_{h \in B} \sum_{x,x'} f_h(x)\overline{f_h(x')}\omega^{-(x-x')\sigma(h)}$$

$$= \sum_{h \in B} \sum_{x,u} f_h(x + u)\overline{f_h(x)}\omega^{-u\sigma(h)}$$

$$= \sum_{h \in B} \sum_{x,u} \sum_{y,y'} f(x + u, y' + h)\overline{f(x + u, y')f(x, y + h)}f(x, y)\omega^{-u\sigma(h)}$$

is at least αN^5. It follows that

$$\sum_{x,u} \sum_{y,y'} \Big| \sum_{h \in B} f(x + u, y' + h)\overline{f(x, y + h)}\omega^{-u\sigma(h)} \Big| \geqslant \alpha N^5$$

which implies that

$$\sum_{x,u}\sum_{y,y'}\Big|\sum_{h\in B} f(x+u,y'+h)\overline{f(x,y+h)}\omega^{-u\sigma(h)}\Big|^2 \geqslant \alpha^2N^6\,.$$

For each triple $t=(u,x,w)$, let $a_t(h)=f(x+u,w+h)\overline{f(x,h)}$ and $b_t(h)=B(h)\omega^{u\sigma(h)}$. Then we may rewrite the above inequality as

$$\sum_{t}\sum_{y}\Big|\sum_{h} a_t(h+y)\overline{b_t(h)}\Big|^2 \geqslant \alpha^2N^6\,.$$

As in the proof of Proposition 12.1, we may apply Lemma 2.1 and the Cauchy-Schwarz inequality to deduce that

$$\Big(\sum_t\sum_r|\hat{a}_t(r)|^4\Big)\Big(\sum_t\sum_r|\hat{b}_t(r)|\Big)^4 \geqslant \alpha^4N^{14}\,.$$

Since $\sum_r|\hat{a}_t(r)|^4\leqslant N^4$ for every t and $\hat{b}_t(r)=\sum_{h\in B}\omega^{u\sigma(h)-rh}$ for every t and r, we then find that

$$N^2\sum_u\sum_r\Big|\sum_{h\in B}\omega^{u\sigma(h)-rh}\Big|^4 \geqslant \alpha^4N^7\,.$$

But the left-hand side is exactly N^4 times the number of quadruples (a,b,c,d) that we wish to find. □

In fact, we shall need the above lemma only in the special case of 01-valued functions. The next corollary is a restatement of the lemma in the language of §10.

COROLLARY 13.2. *Let $A\subset\mathbb{Z}_N^2$. For any $h\in\mathbb{Z}_N$, define a domain $X_h=X_{h,0}\cup X_{h,1}\cup\cdots\cup X_{h,N-1}$ by letting $X_{h,x}$ be the set of all pairs $((x,y),(x,y+h))$, for which both (x,y) and $(x,y+h)$ belong to A. Let $f_h(x)$ be the cardinality of $X_{h,x}$. Let $B\subset\mathbb{Z}_N$ and let $\sigma:B\to\mathbb{Z}_N$ be any function such that $\sum_{h\in B}|\hat{f}_h(\sigma(h))|^2\geqslant\alpha N^5$. Then there are at least α^4N^3 quadruples $(a,b,c,d)\in B^4$ such that $a+b=c+d$ and $\sigma(a)+\sigma(b)=\sigma(c)+\sigma(d)$.*

Proof. This follows immediately from Lemma 13.1 applied to the characteristic function of A. □

COROLLARY 13.3. *Let $A\subset\mathbb{Z}_N^2$, and for $h\in\mathbb{Z}_N$ let X_h and f_h be as in Corollary 13.2. Let $\theta>0$. Then there exist Freiman homomorphisms $\sigma_1,\dots,\sigma_q$ of order eight, defined on subsets $B_1,\dots,B_q$ of $\mathbb{Z}_N$, and a set $G\subset\mathbb{Z}_N$ of cardinality at least $(1-\theta)N$, such that whenever $h\in G$ and $|\hat{f}_h(r)|\geqslant\theta N^2$ there exists $i\leqslant q$ such that $r=\sigma_i(h)$. The sets B_i have cardinality at least $2^{-1882}\theta^{10477}N$ and $q\leqslant 2^{1882}\theta^{-10479}$.*

Proof. Suppose that $B \subset \mathbb{Z}_N$ is a set of size θN and that $\sigma : B \to \mathbb{Z}_N$ is a function with the property that $|\hat{f}_h(\sigma(h))| \geqslant \theta N^2$ for every $h \in B$. Then $\sum_{h\in B} |\hat{f}_h(\sigma(h))|^2 \geqslant \theta^3 N^5$. Hence, by Corollary 13.2, B contains at least $\theta^{12}N^3$ σ-additive quadruples. It follows from Corollary 7.6 (with $\alpha = \theta$ and $\gamma = \theta^9$) that there is a subset C of B of cardinality at least $2^{-1882}\theta^{10477}N$ such that the restriction of σ to C is a Freiman homomorphism of order eight.

Now let Γ_0 be the set of all pairs (h, r) such that $|\hat{f}_h(r)| \geqslant \theta N^2$. If the projection of Γ to the h-axis has size less than θN, then we are done. Otherwise, we can choose B and σ satisfying the hypotheses of the previous paragraph and hence can find $B_1 \subset B$ of cardinality at least $2^{-1882}\theta^{10477}N$ such that the restriction of σ to B_1 is a homomorphism of order eight. Let σ_1 be this restriction and let $\Gamma_1 = \Gamma \setminus \{(h, \sigma_1(h)) : h \in B_1\}$.

If the projection of Γ_1 to the h-axis has cardinality less than θN, we are done. Otherwise, the above argument can be repeated. Continue the repetitions until it is no longer possible and we are then done. Now $\sum_{h,r} |\hat{f}_h(r)|^2 = N \sum_{h,s} f_h(s)^2$ which is clearly at most N^5, so Γ_0 has cardinality at most $\theta^{-2}N$. It follows that $q \leqslant 2^{1882}\theta^{-10479}$ as stated. □

We shall now prove several lemmas under the same set of hypotheses, so it is convenient to state the hypotheses first and not keep repeating them. Let A be a subset of $\mathbb{Z}_N^2$ of cardinality αN^2 and let $\phi : A \to \mathbb{Z}_N$ be a function with the following two properties. First, $\phi(x, y)$ is, for every fixed x, a homomorphism of order 8 in y and for every fixed y a homomorphism of order 8 in x. Second, the proportion of all 8-arrangements in A respected by ϕ is at least $1 - \eta$, where $\eta = 2^{-44}$. (This second property states that A satisfies the conclusion of Lemma 12.6.) For each $h \in \mathbb{Z}_N$, let us write $C(h)$ for the number of 8-arrangements in A of height h and $G(h)$ for the number of these 8-arrangements respected by ϕ. The domains X_h and the functions f_h are as defined in Corollary 13.2.

LEMMA 13.4. *Let $\theta > 0$, $\theta_1 = 2^{-1882}\theta^{10477}$, $q = 2^{1882}\theta^{-10479}$ and $m = \lfloor(\theta_1/64\pi)N^{\theta_1^2/16q}\rfloor$. Then there exist an arithmetic progression P of length $m_0 \in \{m-1, m\}$ and a subset $H \subset P$ such that*

$$\sum \big\{C(h) : h \in H, G(h) \geqslant (1-2\eta)C(h)\big\} \geqslant \alpha^{32}N^{31}m_0/8\,,$$

and there exist constants $a_1, \dots, a_q$ and $b_1, \dots, b_q$ such that, whenever $h \in H$ and $r \in \mathbb{Z}_N$ have the property that $|\hat{f}_h(r)| \geqslant \theta N$, we have $r = a_i h + b_i$ for some $i \leqslant q$.

Proof. Let G be the set and $\sigma_1,\dots,\sigma_q$ the Freiman homomorphisms of order 8 given by Corollary 13.3. For each $i \leqslant q$ the homomorphism σ_i is defined on a set B_i of size at least $\theta_1 N$, so by Corollary 7.9 there exist a set K_i of size at most $16\theta_1^{-2}$ and some $c_i \in \mathbb{Z}_N$ such that if m is a positive integer, d belongs to the Bohr neighbourhood $B(K_i,\theta_1/32\pi m)$ and $x,y \in B_i$ with $x-y=jd$ for some j with $|j| \leqslant m$, then $\sigma_i(x)-\sigma_i(y)=c_i(x-y)$.

By Lemma 7.7 and the definition of m we can find a non-zero d belonging to the Bohr neighbourhood

$$\bigcap_{i=1}^{q} B(K_i,\theta_1/32\pi m) = B\Big(\bigcup K_i,\theta_1/32\pi m\Big)\,.$$

Let d_0 be such a value of d, and partition $\mathbb{Z}_N$ into arithmetic progressions $P_1,\dots,P_M$ such that each P_j has common difference d_0 and the lengths of the P_j are all equal to $m-1$ or m. By the way d_0 was chosen, the restriction of any σ_i to any P_j (or more correctly to $B_i \cap P_j$) is linear.

Our arithmetic progression P will be one of the progressions P_j, chosen by an averaging argument. By our second assumption on ϕ, we know that

$$\sum_h G(h) \geqslant (1-\eta)\sum_h C(h)\,,$$

which implies that

$$\sum \big\{C(h) : G(h) \geqslant (1-2\eta)C(h)\big\} \geqslant \tfrac{1}{2}\sum_h C(h)\,.$$

This estimate says that at least half of the 8-arrangements in A have a height h for which the function ϕ_h is a $(1-2\eta)$-homomorphism of order 8. We also know that $\sum_{h\notin G} C(h) \leqslant \theta N^{32}$. Since the total number of 8-arrangements in A is $\sum_h C(h) \geqslant \alpha^{32}N^{32}$, we can deduce that

$$\sum \big\{C(h) : h\in G, G(h) \geqslant (1-2\eta)C(h)\big\} \geqslant \tfrac{1}{4}\sum_h C(h)\,.$$

By averaging and the above estimate, we can choose some j such that

$$\begin{aligned}\sum \big\{C(h) : G(h) \geqslant (1-2\eta)C(h), h\in P_j\cap G\big\} &\geqslant \alpha^{32}N^{31}(m-1)/4\\ &\geqslant \alpha^{32}N^{31}m_0/8\,.\end{aligned}$$

Let us set $P=P_j$ and $H=P_j\cap G$. We know that each σ_i, when restricted to P, is linear. Therefore, we can find the constants $a_1,\dots,a_q$ and $b_1,\dots,b_q$ required by the lemma. □

This fact, that the set of large Fourier coefficients for each $\hat{f}_h$ "varies linearly" in h, is the key to the whole argument. Our version of Bogolyubov's argument in §10 tells us that $2X_h - 2X_h$ is approximately d-invariant if

$(a_ih + b_i)d$ is small for $1 \leqslant i \leqslant q$. Our next aim is to find a further partition of P into arithmetic progressions in each of which we can choose the same value of d with this property. The linearity of $a_ih + b_i$ allows us to do so. It is to show this that we shall need the multiple recurrence result, Lemma 5.9.

LEMMA 13.5. *There exists an arithmetic progression $Q \subset P$ of size $m_1 \geqslant m_0^{1/2^{12q}}/2$ and common difference d, such that $|(a_ih+b_i)d| \leqslant m_0^{-1/2^{11q}}N$ for every $i \leqslant q$. Moreover, Q can be chosen so that there are at least $\alpha^{32}m_1/20$ values of $h \in Q \cap H$ for which $C(h) \geqslant \alpha^{32}N^{31}/16$ and $G(h) \geqslant (1-2\eta)C(h)$.*

Proof. For each i, let τ_i be the quadratic polynomial $a_ih^2/4 + b_ih/2$. The result is very simple if $m_0 < 2^{2^{12q}}$, as then the only restriction on m_1 is that it should be at least 1. Otherwise, m_0 satisfies the lower bound on r required in Lemma 5.9 when $k = 2$ (and therefore $K = 2^{11}$). That lemma therefore tells us that P can be partitioned into arithmetic progressions $Q_1, \dots, Q_L$ of sizes differing by one and at least $m_0^{1/2^{12q}}$ such that, for every i and j, the diameter of $\tau_i(Q_j)$ is at most $m_0^{-1/2^{11q}}N$. By averaging, we can find one of these progressions, which we shall call Q', such that $|Q'| = m_1 \geqslant cm_0^{1/2^{12q}}$ and

$$\sum \big\{C(h) : h \in Q' \cap H, G(h) \geqslant (1-2\eta)C(h)\big\} \geqslant \alpha^{32}N^{31}m_1/8\,.$$

Let the common difference of Q' be d. Choose any $h \in Q'$ which is not an end point. Then $\tau_i(h+d) - \tau_i(h-d) = (a_ih + b_i)d$, so the estimate on the diameter of $\tau_i(Q')$ implies that $|(a_ih + b_i)d| \leqslant m_0^{-1/2^{11q}}N$ for every i. Now let Q be Q' without the two end points.

By another averaging argument, there are at least $\alpha^{32}m_1/16$ values of $h \in Q' \cap H$ such that $C(h) \geqslant \alpha^{32}N^{31}/16$ and $G(h) \geqslant (1-2\eta)C(h)$. This certainly implies the slightly worse estimate for Q. □

It is vital for our later purposes that the common difference d of Q should be the same as the d for which the numbers $(a_ih+b_i)d$ are all small. It was to achieve this that we needed to use quadratic recurrence and not just linear recurrence.

Let us define I to be the set of all $h \in Q \cap H$ such that $C(h) \geqslant \alpha^{32}N^{15}/16$ and $G(h) \geqslant (1-2\eta)C(h)$. Notice that if $h \in I$ then ϕ_h is a $(1-2\eta)$-homomorphism of order 8 on the domain X_h. Lemma 13.5 asserts that I has cardinality at least $\alpha^{32}m_1/20$. For the next lemma, recall that a typical element of the domain X_h, which was defined in the statement of Corollary 13.2, is a pair $v = ((x, y), (x, y + h))$ and that $r(v)$ is defined to be x.

LEMMA 13.6. *Let $k = 2^{114}\alpha^{-320}$, $\zeta = 2^{-228k}\alpha^{576k}$ and $q = 2^{2^{20}}\alpha^{-2^{21}}$. Then there is an arithmetic progression $R \subset \mathbb{Z}_N$ of size $m_2 \geqslant (\zeta/2)N^{1/2^{13q}}$ and common difference d such that for every $h \in I$ there is a subset $Y_h \subset X_h$ of size at least $2^{-26}\alpha^{128}N^2$ such that the restriction of ϕ_h to $\{v \in Y_h : r(v) \in R\}$ is linear. Moreover, R can be chosen such that*

$$\sum_{h\in I} \big|\{v \in Y_h : r(v) \in R\}\big| \geqslant 2^{-26}\alpha^{128}m_2N|I| \geqslant 2^{-31}\alpha^{160}m_1m_2N\,.$$

Proof. We know that $|X_{h,r}| \leqslant N$ for every r. The lower bound on $C(h)$ for each $h \in I$ implies that $|X_h| = \sum_r |X_{h,r}| \geqslant \alpha^{32}N^2/16$. We are about to apply Corollary 10.14, which uses the hypotheses of Theorem 10.13, to the domain X_h. We may do so if we replace α in the statements of Theorem 10.13 and Corollary 10.14 by $\alpha^{32}/16$. (Note that η was defined in this section to be 2^{-44}, so for $h \in I$ the $(1-2\eta)$-homomorphism ϕ_h is a $(1-\eta)$-homomorphism in the sense of Theorem 10.13.) This allows us to take $\lambda = 2^{-59}\alpha^{176}$, $k = 2^{114}\alpha^{-320}$ and $\zeta = 2^{-228k}\alpha^{576k}$. Corollary 10.14 then states that if $K_h = \{r \in \mathbb{Z}_N : |\hat{f}_h(r)| \geqslant \lambda N\}$, then there exists a set $Y_h \subset X_h$ of cardinality at least $(\alpha^{32}/16)^3|X_h|/1000 \geqslant 2^{-26}\alpha^{128}N^2$ with the following property: for every positive integer m and every d in the Bohr neighbourhood $B(K_h, \zeta/m)$, there exists $c_h \in \mathbb{Z}_N$ such that $\phi_h(v)-\phi_h(w) = c_h(r(v)-r(w))$ whenever $v, w \in Y_h$ and $r(v)-r(w) \in \{jd : -m \leqslant j \leqslant m\}$.

Now, Lemmas 13.4 and 13.5 combined, with θ set equal to λ, tell us that the common difference d of the arithmetic progression Q satisfies the property that $|rd| \leqslant m_0^{-1/2^{11q}}N$ whenever $|\hat{f}_h(r)| \geqslant \theta N$, where q may be taken to be the number in the statement of this lemma. In other words, we are told that this d belongs to all the Bohr neighbourhoods $B(K_h, \zeta/m)$, provided that the inequality $m_0^{-1/2^{11q}} \leqslant \zeta/m$ holds, where m_0 is as given in the statement of Lemma 13.4. It is not hard to check that $m = (\zeta/2)N^{1/2^{13q}}$ satisfies the inequality. (In fact, we could replace 13 by 12, but it is convenient later for m_2 to be significantly less than m_1.)

Since $m < N^{1/2}$, we can partition $\mathbb{Z}_N$ into arithmetic progressions $R_1, \dots, R_L$ of common difference d and lengths m or $m+1$. For every R_i and for every $h \in I$ the restriction of ϕ_h to $\{v \in Y_h : r(v) \in R_i\}$ is linear. Since $|Y_h| \geqslant 2^{-26}\alpha^{128}N^2$ for every $h \in I$, we know that $\sum_{h\in I}|Y_h| \geqslant 2^{-26}\alpha^{128}N^2|I|$. An averaging argument therefore gives us one of the R_i, which we shall call R, such that

$$\sum_{h\in I} \big|\{v \in Y_h : r(v) \in R\}\big| \geqslant 2^{-26}\alpha^{128}|I|m_2N \geqslant 2^{-31}\alpha^{160}m_1m_2N$$

where m_2 equals either m or $m+1$. This proves the lemma. □

LEMMA 13.7. *There exist $y \in \mathbb{Z}_N$, an arithmetic progression $S \subset R$ of size $m_3 \geqslant m_2^{2^{-66}\alpha^{256}}$ and a set $B \subset S \times (I+y)$ of size at least $2^{-26}\alpha^{128}m_3|I| \geqslant 2^{-31}\alpha^{160}m_1m_3$ such that, for every $h \in I$, the restriction of $\phi(x, y+h)$ to B is linear in x.*

Proof. Choose $y \in \mathbb{Z}_N$ uniformly at random. The expected number of pairs x, h such that $x \in R$, $h \in I$ and $((x,y),(x,y+h)) \in Y_h$ is at least $2^{-26}\alpha^{128}m_2|I|$, so let us fix a value of y such that there are at least this many. The number of $x \in R$ such that $(x,y) \in A$ is then at least $2^{-26}\alpha^{128}m_2$. One of our main assumptions is that ϕ is a homomorphism of order 8 for each fixed y. Hence, by Corollary 7.11, applied to the single set $\{x \in R : (x,y) \in A\}$ we can find a partition of R into arithmetic progressions $S_1, \dots, S_M$ all of length at least $m_2^{2^{-66}\alpha^{256}}$ such that the restriction of $x \mapsto \phi(y,x)$ to any S_j is linear (where defined). By averaging, we can choose some S_j, which we shall call S, such that the number of pairs x, h with $x \in S$, $h \in I$ and $((x,y),(x,y+h)) \in Y_h$ is at least $2^{-26}\alpha^{128}m_3|I|$, where m_3 is the size of S.

Let $B \subset S \times (I+y)$ be the set of all points $(x, y+h)$ such that $h \in I$ and $((x,y),(x,y+h)) \in Y_h$. We have shown that B has cardinality at least $2^{-26}\alpha^{128}m_3|I| \geqslant 2^{-31}\alpha^{160}m_1m_3$ and found constants c and c_h ($h \in I$), such that, for any $x_1, x_2 \in S$ and any $h \in I$,

$$\phi(x_1, y) - \phi(x_2, y) = c(x_1 - x_2)$$

and

$$\phi(x_1, y+h) - \phi(x_1, y) - \phi(x_2, y+h) + \phi(x_2, y) = c_h(x_1 - x_2)\,.$$

It follows that, for every $h \in I$,

$$\phi(x_1, y+h) - \phi(x_2, y+h) = (c_h - c)(x_1 - x_2)\,,$$

which tells us that the restrictions of ϕ to the rows of B are all linear. □

We have now effectively reduced the dimension of our problem by one, as the next two lemmas will demonstrate. For each $h \in H$, let $a(h)$ and $c(h)$ be the unique constants such that $\phi(x, y+h) = a(h) + c(h)x$ for every x with $(x, y+h) \in B$.

LEMMA 13.8. *Assume that $m_3 \geqslant 2^{84}\alpha^{-416}$. Then there is a subset $J \subset I$ such that the map $h \mapsto (a(h), c(h))$ is a homomorphism of order 8 on J and the set C of $(x, y+h) \in B$ such that $h \in J$ has cardinality at least $2^{-84}\alpha^{416}m_1m_3$.*

Proof. We know that B has size at least $2^{-31}\alpha^{160}m_1m_3 = 2^{-31}\alpha^{160}m_1m_3|S||Q+y|$. We also know (from our main assumption) that

$\phi(x, y+h)$ is a homomorphism of order 8 in h for every fixed x. Suppose that x_1, x_2 and $h(1), \dots, h(16)$ are such that $(x_i, y+h(j)) \in B$ for every i, j. Because $\phi(x, y+h)$ is a homomorphism of order 8 in h, easy linear algebra shows that

$$a_{h(1)} + \cdots + a_{h(8)} = a_{h(9)} + \cdots + a_{h(16)}$$

and

$$c_{h(1)} + \cdots + c_{h(8)} = c_{h(9)} + \cdots + c_{h(16)}\,.$$

For any pair $(x_1, x_2) \in S^2$ let $J(x_1, x_2)$ be the set of all $h \in I$ such that $(x_1, y+h)$ and $(x_2, y+h)$ are in B, and let $C(x_1, x_2)$ be the set of all $(x, y+h) \in B$ such that $h \in J(x_1, x_2)$. We shall choose J to be one of the $J(x_1, x_2)$ and for that we need the corresponding set $C(x_1, x_2)$ to be large, which (needless to say) we do by averaging.

Notice first that $\sum_{x_1, x_2} |C(x_1, x_2)|$ counts all quadruples $(x_1, x_2, x_3, h) \in S^3 \times I$ such that $(x_i, y+h) \in B$ for $i = 1, 2, 3$. Therefore, letting $D_h = \{x \in S : (x, y+h) \in B\}$, we can write this sum as $\sum_{h \in I} |D_h|^3$. Since $\sum_{h \in I} |D_h| = |B|$, this is at least $|I|^{-2}|B|^3$, which our earlier estimates tell us is at least $2^{-83}\alpha^{416} m_1 m_3^3$. The contribution to the sum from sets $C(x_1, x_2)$ such that $x_1 = x_2$ is certainly no more than $m_1 m_3^2$, which, by our assumed lower bound for m_3, is at most half the total. Therefore, there exist $x_1 \neq x_2$ such that $C(x_1, x_2)$ has cardinality at least $2^{-84}\alpha^{416} m_1 m_3$.

We have shown that the map $h \mapsto (a_h, c_h)$ is a homomorphism of order 8 from $J = J(x_1, x_2)$ to $\mathbb{Z}_N^2$, so we may set $J = J(x_1, x_2)$ and the lemma is proved. □

At this point let us recall that the arithmetic progression S is a subset of R, which has the same common difference d as Q. Moreover, we fixed our numbers so that R would be considerably smaller than Q. It follows that S is a subset of a translate of Q and, writing d_1 for the common difference of S, that d_1 is a multiple of d. Recall also that the cardinalities of S and Q are m_3 and m_1 respectively and that J is a subset of Q.

LEMMA 13.9. *There exists an arithmetic progression $U \subset Q$ of common difference d_2, which is a multiple of d_1, and size $m_4 \geqslant m_3^{2^{-182}\alpha^{832}}$ such that the set D of all $(x, y+h) \in C$ such that $h \in U \cap J$ has cardinality at least $2^{-84}\alpha^{416} m_3 m_4 = 2^{-84}\alpha^{416}|S||U+y|$ and the restriction of ϕ to D is bilinear.*

Proof. Let us partition Q into maximal subprogressions $T_1, \dots, T_M$ of common difference d_1. By the remarks immediately preceding the statement of this lemma, each T_i has cardinality at least m_3. By averaging, we can

choose $T = T_j$ such that the set of all $(x, y + h) \in C$ with $h \in J \cap T$ has cardinality at least $2^{-84}\alpha^{416}m_3|T|$. Applying Corollary 7.11 to the homomorphism $h \mapsto (a(h), c(h))$ restricted to $J \cap T$, with α replaced by $2^{-84}\alpha^{416}$, we obtain a partition of T into arithmetic progressions $U_1, \dots, U_L$ of size at least $|T|^{2^{-182}\alpha^{832}}$ such that the restriction of the map $h \mapsto (a(h), c(h))$ to any U_i is linear. By averaging again we may choose $U = U_i$ of cardinality m_4 such that the set D defined in the statement has the required size. Then because the coefficients $a(h)$ and $c(h)$ vary linearly in h when $h \in U \cap J$, the restriction of ϕ to D is bilinear. □

COROLLARY 13.10. *There exist arithmetic progressions V and W with the same common difference and same cardinality $m_5 \geqslant m_4^{1/2} - 1$, and a subset $E \subset V \times W$ of size at least $2^{-86}\alpha^{416}|V||W|$, such that the restriction of ϕ to E is bilinear.*

Proof. We already have a comparable statement for $S \times (U + y)$. The common difference of S is d_1 and the common difference of $U + y$ is d_2, which is a multiple of d_1. All we do now is apply one further averaging argument to pass to subprogressions of the same size and same common difference.

Since U is a subset of a translate of S, a maximal subprogression of S with common difference d_2 has cardinality at least $m_4 - 1$. It is therefore not hard to show that $S \times (U + y)$ can be partitioned into sets of the form $V \times W$, where V and W are arithmetic progressions with common difference d_2 and size m or $m + 1$, where $m \geqslant m_4^{1/2} - 1$. By an averaging argument we can choose one of these sets $V \times W$ such that $D \cap (V \times W) \geqslant 2^{-84}\alpha^{416}|V||W|$. The slightly worse bound in the lemma comes from the fact that we may wish to remove end-points from V and W to make them the same size. □

Let us now show that we can achieve the hypotheses that we have been assuming in the last few lemmas.

LEMMA 13.11. *Let $f : \mathbb{Z}_N \to D$ be a function which fails to be cubically α-uniform. Then there exists a set $A \subset \mathbb{Z}_N^2$ of size at least $(\alpha/2)^{2^{66}}N^2$ and a function $\phi : A \to \mathbb{Z}_N$ such that, for every fixed x, $\phi(x, y)$ is a Freiman homomorphism of order 8 in y, for every fixed y it is a homomorphism of order 8 in x, the proportion of all 8-arrangements in A respected by ϕ is at least $1 - 2^{-73}$ and $|\Delta(f; k, l)\hat{}(\phi(k, l))| \geqslant \alpha N/2$ for every $(k, l) \in A$.*

Proof. By Lemma 3.1 (the easy implication of (ii) from (vi)) there is a set $A_0 \subset \mathbb{Z}_N^2$ of size at least $\alpha N^2/2$ and a function $\phi : A_0 \to \mathbb{Z}_N$ such

that $|\Delta(f;k,l)\hat{}(\phi(k,l))| \geqslant \alpha N/2$ for every $(k,l) \in A_0$. For each k, let $A_{0,k}$ be the cross-section $\{l : (k,l) \in A_0\}$, let $|A_{0,k}| = \alpha_k N$ and define $\phi_k : A_{0,k} \to \mathbb{Z}_N$ by $\phi_k(l) = \phi(k,l)$.

Fixing k and applying Proposition 6.1 to the functions $\Delta(f;k) : \mathbb{Z}_N \to D$ and $\phi_k : A_{0,k} \to \mathbb{Z}_N$, we find that there are at least $\alpha_k^4 N^3$ ϕ_k-additive quadruples in $A_{0,k}$. Applying Corollary 7.6 with $B_0 = A_{0,k}$, $\phi = \phi_k$ and $\alpha = \gamma = \alpha_k$, we obtain a subset $A_{1,k} \subset A_{0,k}$ of size at least $2^{-1882}\alpha_k^{1165}N$ such that the restriction of ϕ_k to $A_{1,k}$ is a Freiman homomorphism of order 8. Since the average of α_k is at least $\alpha/2$, the union of the sets $A_{1,k}$ is a set A_1 of cardinality ζN^2, where $\zeta \geqslant 2^{-1882}(\alpha/2)^{1165} = 2^{-3047}\alpha^{1165}$. The restriction of $\phi(k,l)$ to A_1 is a homomorphism of order 8 in l for any fixed k.

Repeating this argument for the second variable, we can pass to a further subset $A_2 \subset A_1$ of cardinality at least $2^{-3047}\zeta^{1165}N^2 \geqslant (\alpha/2)^{2^{22}}N^2$ such that the restriction of ϕ to A_2 is a homomorphism of order 8 in each variable separately. Let $\beta = (\alpha/2)^{2^{22}}$.

We now apply Lemma 12.6 with $B = A_2$, $\gamma = \alpha/2$ and $\eta = 2^{-44}$. This yields a set A with at least

$$2^{-2^{37}}\beta^{2^{43}}\gamma^{2^{45}}\eta^{2^{36}}N^{32} \geqslant (\alpha/2)^{2^{66}}N^{32}$$

8-arrangements, such that the proportion respected by ϕ is at least $1-2^{-44}$. Since the cardinality of such a set must be at least $(\alpha/2)^{2^{66}}N^2$, the lemma is proved. □

We are now ready for the main result of this section.

Theorem 13.12. *Let f be a function from $\mathbb{Z}_N$ to the closed unit disc. If f is not cubically α-uniform then there exist arithmetic progressions P and Q of size at least $N^{(1/2)^{(1/\alpha)^{2^{70}}}}$ and with the same common difference, a subset $B \subset P \times Q$ of size at least $(\alpha/2)^{2^{76}}|P||Q|$ and a bilinear function $\phi : P \times Q \to \mathbb{Z}_N$, such that $\Delta(f;k,l)\hat{}(\phi(k,l)) \geqslant \alpha N/2$ for every $(k,l) \in B$.*

Proof. By Lemma 13.11 we can find a set $A \subset \mathbb{Z}_N^2$ of size at least $(\alpha/2)^{2^{66}}N^2$ and a function $\phi : A \to \mathbb{Z}_N$ such that, for every fixed x, $\phi(x,y)$ is a Freiman homomorphism of order 8 in y, for every fixed y it is a homomorphism of order 8 in x, the proportion of all 8-arrangements in A respected by ϕ is at least $1-2^{-44}$ and $|\Delta(f;k,l)\hat{}(\phi(k,l))| \geqslant \alpha N/2$ for every $(k,l) \in A$. Apart from the last condition, these are the hypotheses stated just before Lemma 13.4, except that α has been replaced by $(\alpha/2)^{2^{66}}$. The results numbered 13.4 to 13.10 all hold under this set of hypotheses, so the theorem follows from Corollary 13.10 and a back-of-envelope estimate for m_5 when α is replaced by $(\alpha/2)^{2^{66}}$. □

Notice the relationship between the above theorem and Freiman's theorem. The hypotheses are somewhat different, but all we have used is that there are many 8-arrangements respected by ϕ, which is a fairly natural generalization of the hypotheses of the Balog-Szemerédi theorem to graphs of functions in two variables. The conclusion of the theorem is in some ways much weaker, since we find only a very small set with good structure. On the other hand, the structure obtained is stronger, as we have gone up from linearity to bilinearity. It is very likely that a development of the argument above could be used to give a complete description of functions $\phi : \mathbb{Z}_N^2 \to \mathbb{Z}_N$ that respect many 8-arrangements. This would deserve to be called a bilinear Freiman (or Balog-Szemerédi) theorem. Theorem 13.12 one could perhaps call a weak bilinear Freiman theorem.

The next three sections will generalize the above theorem from non-cubically uniform functions to functions that fail to be uniform of degree k, producing an appropriate $(k-1)$-linear piece. The generalization is long, but does not involve any significant new ideas. The reader who wishes to follow a proof of Szemerédi's theorem for progressions of length five can go straight to §17.

14 Obtaining Many Respected Arrangements

This section and the next consist of relatively routine generalizations of the results of §12 to functions of k variables. The reason we are presenting them separately is that the argument for two variables is notationally simpler and therefore easier to understand, while containing all the essential ideas.

We begin with a result which, in both its statement and its proof, is very similar to Proposition 12.1, but which seems to be hard to unify with that result. Recall that if $f : \mathbb{Z}_N^2 \to \mathbb{C}$, then $f_h(y)$ is defined as $\sum_x f(x+h,y)\overline{f(x,y)}$.

PROPOSITION 14.1. *For each $h \in \mathbb{Z}_N$ let $\lambda_h \geqslant 0$. Let $f^{(1)}, \dots, f^{(p)}$ be functions from $\mathbb{Z}_N^2$ to the closed unit disc D and let $\sigma_1, \dots, \sigma_p$ be functions from $\mathbb{Z}_N$ to $\mathbb{Z}_N$ such that*

$$\sum_h \lambda_h \prod_{i=1}^{p} |\hat{f}_h^{(i)}(\sigma_i(h))|^2 \geqslant \alpha N^{4p+1}.$$

Then the sum of $\lambda_a\lambda_b\lambda_c\lambda_d$ over all quadruples (a,b,c,d) such that $a+b=c+d$ and $\sigma_i(a)+\sigma_i(b)=\sigma_i(c)+\sigma_i(d)$ for every i is at least $\alpha^4 N^3$.

Proof. In the argument to follow, we shall often abbreviate $(x_1,\dots,x_p)$ by x, and similarly for other sequences of length p. The left-hand side of the inequality we are assuming is, when written out in full,

$$\sum_h \lambda_h \sum_{x,w,y,z} \prod_{i=1}^p f^{(i)}(x_i+h,y_i)\overline{f^{(i)}(x_i,y_i)f^{(i)}(w_i+h,z_i)} \cdot f^{(i)}(w_i,z_i)\omega^{-\sigma_i(h)(y_i-z_i)}\,.$$

Substituting $u_i=y_i-z_i$, this becomes

$$\sum_h \lambda_h \sum_{x,w,u,z} \prod_{i=1}^p f^{(i)}(x_i+h,z_i+u_i)\overline{f^{(i)}(x_i,z_i+u_i)f^{(i)}(w_i+h,z_i)} \cdot f^{(i)}(w_i,z_i)\omega^{-\sigma_i(h)u_i}\,.$$

Since this exceeds αN^{4p+1} and $|f^{(i)}(x,y)|\leqslant 1$ for every i,x,y we may deduce that

$$\sum_{x,w,u,z}\Big|\sum_h \lambda_h \prod_{i=1}^p f^{(i)}(x_i+h,z_i+u_i)\overline{f^{(i)}(w_i+h,z_i)}\omega^{-u_i\sigma_i(h)}\Big| \geqslant \alpha N^{4p+1}$$

and hence, by the Cauchy-Schwarz inequality, that

$$\sum_{x,w,u,z}\Big|\sum_h \lambda_h \prod_{i=1}^p f^{(i)}(x_i+h,z_i+u_i)\overline{f^{(i)}(w_i+h,z_i)}\omega^{-u_i\sigma_i(h)}\Big|^2 \geqslant \alpha^2 N^{4p+2}\,.$$

We now introduce a variable s and write $x_i=s+x_i'$ and $w_i=s+w_i'$. From the above, we can deduce that

$$\sum_s \sum_{x',w',u,z}\Big|\sum_h \lambda_h \prod_{i=1}^p f^{(i)}(s+x_i'+h,z_i+u_i)\overline{f^{(i)}(s+w_i'+h,z_i)}\omega^{-u_i\sigma_i(h)}\Big|^2 \geqslant \alpha^2 N^{4p+3}\,.$$

Applying Lemma 2.1 and the Cauchy-Schwarz inequality in the usual way (see for example the proof of Proposition 12.1) we deduce that

$$\sum_r \sum_{x',w',u,z}\Big|\sum_h \lambda_h \prod_{i=1}^p \omega^{\sigma_i(h)u_i-rh}\Big|^4 \geqslant \alpha^4 N^{4p+4}\,.$$

Since the left-hand side above is N^{4p+1} times the sum of $\lambda_a\lambda_b\lambda_c\lambda_d$ over all quadruples (a,b,c,d) such that $a+b=c+d$ and $\sigma_i(a)+\sigma_i(b)=\sigma_i(c)+\sigma_i(d)$ for every i, the result is proved. □

In the next section, we shall need to deal with functions defined on sets $B\subset\mathbb{Z}_N^k$ which will be k-dimensional generalizations of the somewhat

additive functions that appeared in §6. They arise in two different ways, but in both cases they have a property which we shall call the *product property*. To define this, suppose that B is a subset of $\mathbb{Z}_N^k$ and that $\phi : B \to \mathbb{Z}_N$. Given any $j \leqslant k$ and any $y \in \mathbb{Z}_N^k$, define $B(y,j)$ to be the set of all $x \in B$ such that $x_i = y_i$ whenever $i \neq j$. This is the one-dimensional cross-section of B that goes through y in the j-direction. Now define $C(y,j)$ to be the set of all $x \in \mathbb{Z}_N$ such that $(y_1, \dots, y_{j-1}, x, y_{j+1}, \dots, y_k) \in B(y,j)$, and define a function $\phi_{y,j} : C(y,j) \to \mathbb{Z}_N$ by

$$\phi_{y,j}(x) = \phi(y_1, \dots, y_{j-1}, x, y_{j+1}, \dots, y_k)\,.$$

This is the restriction of ϕ to $B(y,j)$, but for convenience regarded as a function defined on a subset of $\mathbb{Z}_N$. Let us define a *j-restriction* of ϕ to be any function of the form $\phi_{y,j}$ for some $y \in \mathbb{Z}_N^k$. We shall say that ϕ has the *product property with parameter* γ if, whenever $j \leqslant k$, $\psi_1, \dots, \psi_p$ are j-restrictions of ϕ, E is a subset of $\mathbb{Z}_N$ on which all the ψ_i are defined and $\theta : E \to \mathbb{R}_+$, the sum of $\theta(a)\theta(b)\theta(c)\theta(d)$ over all additive quadruples (a,b,c,d) that are ψ_i-additive for every i is at least $\gamma^{8p} N^{-1} \big(\sum_x \theta(x)\big)^4$.

LEMMA 14.2. *Let $f : \mathbb{Z}_N \to D$, let $B \subset \mathbb{Z}_N^k$ and let $\phi : B \to \mathbb{Z}_N$ be such that $|\Delta(f; r_1, \dots, r_k)^\sim(\phi(r_1, \dots, r_k))| \geqslant \gamma N$ for every $(r_1, \dots, r_k) \in B$. Then ϕ has the product property with parameter γ.*

Proof. Let $y_1, \dots, y_p$ be elements of $\mathbb{Z}_N^{k-1}$ and let E be the set of all $r \in \mathbb{Z}_N$ such that $(y_i, r) \in B$ for every i. Then if we are given a function $\theta : E \to \mathbb{R}_+$, we can set $\theta(k) = 0$ for $k \notin E$ and apply Proposition 12.1 to the functions $f_i = \Delta(f; y_i)$. Since $\Delta(f_i; r) = \Delta(f; (y_i, r))$ these functions satisfy the hypothesis of Lemma 12.1 with $\alpha = \gamma^{2p} \sum_k \theta(k) N^{-1}$ and $\sigma_i(r) = \phi(y_i, r)$. The conclusion of the lemma then gives us exactly what we want, at least for k-restrictions. By symmetry, the result is true for the other j-restrictions as well, and ϕ has the product property with parameter γ. □

The second case in which we wish to deduce the product property is similar to the first, but we shall use Proposition 14.1 instead of Proposition 12.1. Given a function $f : \mathbb{Z}_N^{k+1} \to \mathbb{C}$ and $h = (h_1, \dots, h_k) \in \mathbb{Z}_N^k$, define a function $f_h : \mathbb{Z}_N \to \mathbb{C}$ by

$$f_h(y) = \sum_{x \in \mathbb{Z}_N^k} \prod_{\epsilon \in \{0,1\}^k} C^{|\epsilon|+k} f(x_1 + \epsilon_1 h_1, \dots, x_k + \epsilon_k h_k, y)\,,$$

where once again C stands for complex conjugation and $|\epsilon| = \sum \epsilon_i$. For example when $k = 1$ we have $f_h(y) = \sum_x f(x+h, y)\overline{f(x,y)}$, as before. (We have taken $C^{|\epsilon|+k}$ rather than the simpler $C^{|\epsilon|}$ in the definition merely to make it consistent with the earlier one.)

LEMMA 14.3. *Let $f : \mathbb{Z}_N^{k+1} \to D$, let $B \subset \mathbb{Z}_N^k$ and let $\phi : B \to \mathbb{Z}_N$ be such that $|\hat{f}_z(\phi(z))| \geqslant \gamma N^{k+1}$ for every $z \in B$. Then ϕ has the product property with parameter γ.*

Proof. For any $y = (y_1, \dots, y_{k-1}) \in \mathbb{Z}_N^{k-1}$ we can define a function $g_y : \mathbb{Z}_N^2 \to D$ by the formula

$$g_y(a,b) = \sum_{u \in \mathbb{Z}_N^{k-1}} \prod_{\epsilon \in \{0,1\}^{k-1}} C^{|\epsilon|+k-1} f(u_1 + \epsilon_1 y_1, \dots, u_{k-1} + \epsilon_{k-1} y_{k-1}, a, b)\,.$$

It is then easy to check that for any $h \in \mathbb{Z}_N$ we have $(g_y)_h = f_{(y,h)}$.

The proof is now more or less the same as that of Lemma 14.2. Let $y_1, \dots, y_p$ be elements of $\mathbb{Z}_N^{k-1}$ (note that y_i is now a vector rather than a coefficient of y) and let E be the set of all $h \in \mathbb{Z}_N$ such that $(y_i, h) \in B$ for every i. Given a function $\theta : E \to \mathbb{R}_+$, set $\theta(k) = 0$ for $k \notin E$ and this time apply Proposition 14.1 to the functions $g^{(i)} = N^{-(k-1)} g_{y_i}$. We certainly have $g^{(i)} : \mathbb{Z}_N^2 \to D$. Since $g_h^{(i)} = N^{-(k-1)} f_{y_i,h}$, we find that $\hat{g}_h^{(i)}(r) = N^{-(k-1)} \hat{f}_{y_i,h}(r)$, which is at least γN^2 if $h \in E$ and $r = \phi(y_i, h)$. Therefore, the functions $g^{(i)}$ satisfy the hypothesis of Proposition 14.1 with $\alpha = \gamma^{2p} \sum_k \theta(k) N^{-1}$ and $\sigma_i(h) = \phi(y_i, h)$. The conclusion of the lemma then gives us exactly what we want for k-restrictions. Once again the result for j-restrictions follows by symmetry. □

Now we shall define, in two stages, an appropriate generalization of a parallelogram. Let B be a subset of $\mathbb{Z}_N^k$. By a *cube in B with sidelengths* $(h_1, \dots, h_k)$ we shall mean a function κ from $\{0,1\}^k$ to B of the form

$$\kappa : (\epsilon_1, \dots, \epsilon_k) \mapsto (r_1 + \epsilon_1 h_1, \dots, r_k + \epsilon_k h_k)\,.$$

We shall sometimes denote this cube $[r_1, \dots, r_k; h_1, \dots, h_k]$. For $k \geqslant 2$ it will later be convenient to think of $\mathbb{Z}_N^{k+1}$ as a product $\mathbb{Z}_N^k \times \mathbb{Z}_N$. Given a subset $B \subset \mathbb{Z}_N^{k+1}$, we shall mean by a *cross-section of B* a set of the form $B_r = \{(r_1, \dots, r_k, r_{k+1}) \in B : r_{k+1} = r\}$. A cube in B_r will simply mean a function from $\{0,1\}^k$ to B_r of the form $\epsilon \mapsto (\kappa(\epsilon), r)$, where κ is a cube in $\mathbb{Z}_N^k$. We shall sometimes denote this cube by (κ, r). Two cubes (not necessarily in the same cross-section) will be called *congruent* if they have the same sidelengths $(h_1, \dots, h_k)$. By a *parallelepiped in B* we shall mean an ordered pair of congruent cubes, both lying in cross-sections of B. A *parallelepiped pair* will mean an ordered quadruple $\big((\kappa_1, r_1), (\kappa_2, r_2), (\kappa_3, r_3), (\kappa_4, r_4)\big)$, where $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ are congruent and (r_1, r_2, r_3, r_4) is an additive quadruple.

In order to prove facts about parallelepiped pairs, it will be convenient to make two further definitions. If $B \subset \mathbb{Z}_N^k$, then by a *configuration in*

B we shall mean, roughly speaking, a product $Q_1 \times \cdots \times Q_k$ of additive quadruples. This is not quite an accurate description as additive quadruples are defined as ordered sets. The order matters here as well, and a precise definition is that a configuration in B is a function $\lambda : \{0,1\}^k \times \{0,1\}^k \to B$ of the form

$$\lambda : (\epsilon, \eta) \mapsto (r_1 + \epsilon_1 g_1 + \eta_1 h_1, r_2 + \epsilon_2 g_2 + \eta_2 h_2, \dots, r_k + \epsilon_k g_k + \eta_k h_k)\,.$$

We shall sometimes denote this configuration by $[r_1, \dots, r_k; g_1, \dots, g_k; h_1, \dots, h_k]$.

If we choose j and fix every ϵ_i and η_i for $i \neq j$, then we define a restriction of λ which gives an additive quadruple in the j-direction. If ϕ is a function from B to $\mathbb{Z}_N$ such that all the 4^{k-1} additive quadruples that arise in this way are ϕ-additive, then we shall say that ϕ *respects* the configuration λ.

Given $B \subset \mathbb{Z}_N^k$, a function $\phi : B \to \mathbb{Z}_N$ and a cube κ in B, we define

$$\phi(\kappa) = \sum_{\epsilon \in \{0,1\}^k} (-1)^{|\epsilon|} \phi(\kappa(\epsilon))\,.$$

(Here, as elsewhere, $|\epsilon|$ denotes $\sum_{i=1}^k \epsilon_i$.) Just to illustrate this definition, we note that

$$\phi[x, y; a, b] = \phi(x + a, y + b) - \phi(x + a, y) - \phi(x, y + b) + \phi(x, y)\,.$$

Define a *cube pair* in B to be an ordered pair (κ_1, κ_2) of congruent cubes. (The difference between this and a parallelepiped is that the cubes are full-dimensional.) We shall say that ϕ *respects* this pair if $\phi(\kappa_1) = \phi(\kappa_2)$.

LEMMA 14.4. *Let $B \subset \mathbb{Z}_N^k$ be a set of size βN^k, let $\phi : B \to \mathbb{Z}_N$ and suppose that ϕ has the product property with parameter γ. Then ϕ respects at least $\beta^{4^k} \gamma^{2k.4^k} N^{3k}$ configurations in B.*

Proof. When $k = 1$, a configuration is an additive quadruple and ϕ respects it if and only if it is ϕ-additive. Therefore, Proposition 6.1 gives us the result.

Now suppose that $k > 1$ and that the result is true for $k - 1$. Let $B \subset \mathbb{Z}_N^k$ be a set of cardinality βN^k and for each r let B_r be the cross-section $\{(x_1, \dots, x_k) \in B : x_k = r\}$. Write $\beta(r) N^{k-1}$ for the cardinality of B_r.

By our inductive hypothesis, B_r contains at least $\beta(r)^{4^{k-1}} \gamma^{2(k-1).4^{k-1}} N^{3(k-1)}$ configurations respected by ϕ. By Jensen's inequality, the average of this quantity over r is at least $\beta^{4^{k-1}} \gamma^{2(k-1).4^{k-1}} N^{3(k-1)}$. Therefore, if a random configuration λ is chosen in $\mathbb{Z}_N^{k-1}$, then the average number of values of r for

which ϕ respects the configuration (λ, r) (by which we mean the function from $\{0,1\}^{k-1}$ to B_r defined by $\epsilon \mapsto (\lambda(\epsilon), r)$) is at least $\beta^{4^{k-1}}\gamma^{2(k-1).4^{k-1}}N$. Let $E(\lambda)$ be the set of such r and let $\eta(\lambda)N$ be the size of $E(\lambda)$.

We now fix λ and apply the product property to the 4^{k-1} functions $x \mapsto \phi(\lambda(\epsilon_1, \epsilon_2), x)$, which are all defined on the set $E = E(\lambda)$. Taking θ to be identically 1, we obtain from the product property that there are at least $\gamma^{8.4^{k-1}}\eta(\lambda)^4N^3$ quadruples $a+b=c+d$ such that for every $(\epsilon_1, \epsilon_2) \in \{0,1\}^{k-1} \times \{0,1\}^{k-1}$ we have

$$\phi(\lambda(\epsilon_1, \epsilon_2), a) + \phi(\lambda(\epsilon_1, \epsilon_2), b) = \phi(\lambda(\epsilon_1, \epsilon_2), c) + \phi(\lambda(\epsilon_1, \epsilon_2), d)\,.$$

But, by the definition of E, each such quadruple gives us a configuration in B which is respected by ϕ. Since the average of $\eta(\lambda)$ is at least $\beta^{4^{k-1}}\gamma^{2(k-1).4^{k-1}}$, Jensen's inequality implies that the number of configurations in B that are respected by ϕ is at least $\gamma^{8.4^{k-1}}\beta^{4^k}\gamma^{2(k-1).4^k}N^{3k} = \beta^{4^k}\gamma^{2k.4^k}N^{3k}$, which proves the result. □

COROLLARY 14.5. *Let $B \subset \mathbb{Z}_N^k$ be a set of size βN^k, let $\phi : B \to \mathbb{Z}_N$ and suppose that ϕ has the product property with parameter γ. Then ϕ respects at least $\beta^{4^k}\gamma^{2k.4^k}N^{3k}$ cube pairs in B.*

Proof. Let $\lambda = [r_1, \dots, r_k; g_1, \dots, g_k; h_1, \dots, h_k]$ be a configuration in B which is respected by ϕ. We shall show that the cube pair

$$\big([r_1, \dots, r_k; h_1, \dots, h_k], [r_1 + g_1, \dots, r_k + g_k; h_1, \dots, h_k]\big)$$

is also respected by ϕ. Since distinct configurations give distinct cube pairs in this way, we will have proved the corollary. For every j between 0 and k let us define κ_j to be the cube

$$[r_1 + g_1, \dots, r_j + g_j, r_{j+1}, \dots, r_k; h_1, \dots, h_k]\,.$$

Because ϕ respects the configuration λ, we know that for all choices of η_j for $j \neq i$, the additive quadruple

$$(r_1 + g_1 + \eta_1 h_1, \dots, r_{j-1} + g_{j-1} + \eta_{j-1}, r_j + \epsilon_j g_j + \eta_j h_j, r_{j+1} + \eta_{j+1}h_{j+1}, \dots, r_k + \eta_k h_k)\,,$$

where ϵ_i and η_i take the values 0 or 1, is ϕ-additive. This implies that $\phi(\kappa_{j-1}) = \phi(\kappa_j)$, and the argument works for every j between 1 and k. Therefore, $\phi(\kappa_0) = \phi(\kappa_k)$, which is the required result. □

COROLLARY 14.6. *Let $B \subset \mathbb{Z}_N^{k+1}$ be a set of size βN^{k+1}, let $\phi : B \to \mathbb{Z}_N$ and suppose that ϕ has the product property with parameter γ. Then ϕ respects at least $\beta^{4^{k+1}}\gamma^{k.4^{k+1}}N^{5k+3}$ parallelepiped pairs in B.*

Proof. The proof of this result is similar to that of Lemma 14.4. For each $r \in \mathbb{Z}_N$ let $\beta(r)N^k$ be the size of the cross-section B_r of B. By Corollary 14.5, ϕ respects at least $\beta(r)^{4^k}\gamma^{2k.4^k}N^{3k}$ cube pairs in B_r. The average of this number over r is at least $\beta^{4^k}\gamma^{2k.4^k}N^{3k}$. Therefore, if we choose r at random and choose a cube κ in $\mathbb{Z}_N^k$ at random, the expected number of cubes κ' in $\mathbb{Z}_N^k$ for which $\big((\kappa,r),(\kappa',r)\big)$ is a cube pair in B_r respected by ϕ is at least $\beta^{4^k}\gamma^{2k.4^k}N^k$.

Now let κ be some fixed cube in $\mathbb{Z}_N^k$ and for each r let $\theta(r)$ be the number of cubes κ' in $\mathbb{Z}_N^k$ for which $\big((\kappa,r),(\kappa',r)\big)$ is a cube pair in B_r respected by ϕ. Let θ be the average of the $\theta(r)$. By the product property, the sum of $\theta(a)\theta(b)\theta(c)\theta(d)$ over all additive quadruples (a,b,c,d) such that

$$\phi(\kappa(\epsilon),a)+\phi(\kappa(\epsilon),b)=\phi(\kappa(\epsilon),c)+\phi(\kappa(\epsilon),d)$$

for every $\epsilon \in \{0,1\}^k$ is at least $\theta^4\gamma^{8.2^k}N^{4k+3}$. But this sum counts the number of parallelepiped pairs $\big((\kappa_i',r_i)\big)_{i=1}^4$ in B such that, for each i, (κ_i,r_i) lies in B_{r_i} and (κ_i',r_i) is congruent to it. It is certainly a lower bound for the number of parallelepiped pairs such that each of the four cubes is congruent to κ.

If we now choose randomly, for every $h=(h_1,\dots,h_k)$, some cube $\kappa(h)$ with sidelengths $(h_1,\dots,h_k)$ and apply the above argument, we shall obtain, on average, at least $(\beta^{4^k}\gamma^{2k.4^k})^4\gamma^{8.2^k}N^{5k+3}$ distinct parallelepiped pairs, since the average of θ (which still depends on κ) is at least $\beta^{4^k}\gamma^{2k.4^k}$. This proves the corollary (where, just for the sake of neatness, we have stated a weaker bound). □

Let $B \subset \mathbb{Z}_N^{k+1}$. By a *d-arrangement in B* we shall mean a sequence $C_1,\dots,C_{2d}$ of congruent cubes, where C_j lies in the cross-section B_{r_j}, and

$$r_1+\cdots+r_d=r_{d+1}+\cdots+r_{2d}\,.$$

Thus, a parallelepiped pair is simply a 2-arrangement, and when $k=1$ we recover the definition of d-arrangement given in §12. It is also convenient to think of a d-arrangement as a function $\rho:\{0,1\}^k\times\{1,2,\dots,2d\}\to\mathbb{Z}_N^{k+1}$ of the form

$$\rho:(\epsilon_1,\dots,\epsilon_k,j)\mapsto(y_1^j+\epsilon_1h_1,\dots,y_k^j+\epsilon_kh_k,r_j)\,.$$

Here, $r_1,\dots,r_{2d}$ are as above and each of the constituent cubes of the d-arrangement has sidelengths $(h_1,\dots,h_k)$ but is otherwise arbitrary. It is easy to see that the number of d-arrangements in $\mathbb{Z}_N^{k+1}$ is $N^{(2d+1)k+2d-1}$. The next lemma is a generalization of Lemma 12.4 and has a very similar proof.

LEMMA 14.7. *Let $B \subset \mathbb{Z}_N^{k+1}$ and let $\phi : B \to \mathbb{Z}_N$. Suppose that ϕ respects θN^{5k+3} parallelepiped pairs in B. Then ϕ respects at least $\theta^7 N^{17k+15}$ 8-arrangements in B.*

Proof. Given $u, x \in \mathbb{Z}_N$ and $h = (h_1, \dots, h_k) \in \mathbb{Z}_N^k$, let $f_{u,h}(x)$ be $\sum_\kappa \omega^{u\phi(\kappa,x)}$, where the sum is over all configurations κ in $\mathbb{Z}_N^k$ with sidelengths $(h_1, \dots, h_k)$, and we interpret $\omega^{u\phi(\kappa,x)}$ as zero when $\phi(\kappa, x)$ is not defined (which happens when (κ, x) does not live in B_x). Clearly $|f_{u,h}(x)|$ is at most N^k for every u, x, h, which implies that $\sum_x |f_{u,h}(x)|^2 \leqslant N^{2k+1}$ for every u, h, and therefore that $\sum_{u,r,h} |\hat{f}_{u,h}(r)|^2 \leqslant N^{3k+3}$.

We also have

$$\sum_{u,r} |\hat{f}_{u,h}(r)|^4 = \sum_{u,r} \Big| \sum_{\kappa,x} \omega^{u\phi(\kappa,x)-rx} \Big|^4$$

for every h, where once again κ ranges over all cubes in $\mathbb{Z}_N^k$ with sidelengths $(h_1, \dots, h_k)$. This is N^2 times the number of parallelepiped pairs respected by ϕ for which the sidelengths of the cubes are $(h_1, \dots, h_k)$. It follows from our assumptions that

$$\sum_{u,r,h} |\hat{f}_{u,h}(r)|^4 \geqslant \theta N^{5k+5}\,.$$

Similarly, $\sum_{u,r,h} |\hat{f}_{u,h}(r)|^{16}$ is N^2 times the number of 8-arrangements respected by ϕ. Therefore, by Lemma 9.1, the number of 8-arrangements is at least $N^{-2}(\theta N^{5k+5}/N^{6(3k+3)/7})^7 = \theta^7 N^{17k+15}$ as claimed. □

Combining Corollary 14.6 and Lemma 14.7 we obtain the main result of this section (which will be applied in conjunction with Lemmas 14.2 and 14.3).

LEMMA 14.8. *Let $B \subset \mathbb{Z}_N^{k+1}$ be a set of size βN^{k+1}, let $\phi : B \to \mathbb{Z}_N$ and suppose that ϕ has the product property with parameter γ. Then ϕ respects at least $\beta^{7.4^{k+1}} \gamma^{7k.4^{k+1}} N^{17k+15}$ 8-arrangements in B.* □

15 Increasing the Density of Respected Arrangements

We shall now use an argument similar to those of §9 and §12 to pass to a subset of B where ϕ respects almost all 8-arrangements. (We shall actually prove our results for general d-arrangements and then take d to be 8 later.) In order to do this, we shall need a brief discussion of a small number of degenerate cases where a later argument does not work. Just for the next lemma it will be convenient to consider sequences in $\{-1, 1\}^k$ rather than $\{0, 1\}^k$.

LEMMA 15.1. *Let $h_1, \dots, h_k$ be non-zero elements of $\mathbb{Z}_N$, and let $\eta : \{-1,1\}^k \to \{-1,0,1\}$ be a function such that the sum*

$$\sum_{\epsilon \in \{-1,1\}^k} \eta(\epsilon) \prod_{i \in A} (y_i + \epsilon_i h_i)$$

is independent of $y_1, \dots, y_k$ for every subset $A \subset \{1, 2, \dots, k\}$. Then η is a multiple of the function $\epsilon \mapsto \prod \epsilon_i$.

Proof. Throughout this lemma, any sum over ϵ will denote the sum over all ϵ in the set $\{-1,1\}^k$. The functions $\epsilon \mapsto \prod_{i \in A} \epsilon_i$ are orthogonal with respect to the symmetric bilinear form $\langle \eta_1, \eta_2 \rangle = \sum_\epsilon \eta_1(\epsilon)\eta_2(\epsilon)$. (Recall that N is prime. This bilinear form is defined on the vector space $\mathbb{Z}_N^{\{-1,1\}^k}$ and the functions are the Walsh basis for this space.) Therefore, it is enough to prove that $\sum_\epsilon \eta(\epsilon) \prod_{i \in A} \epsilon_i = 0$ for every proper subset $A \subset \{1, 2, \dots, k\}$. This we do by induction on A (with respect to containment).

First, let A be a proper subset of $\{1, 2, \dots, k\}$ and let $j \notin A$. From the assumption of the lemma, applied to the set $A \cup \{j\}$, we know that

$$y_j \sum_\epsilon \eta(\epsilon) \prod_{i \in A} (y_i + \epsilon_i h_i) + h_j \sum_\epsilon \eta(\epsilon)\epsilon_j \prod_{i \in A} (y_i + \epsilon_i h_i)$$

is independent of y_j. Since the second part of the sum does not involve y_j, this implies that

$$\sum_\epsilon \eta(\epsilon) \prod_{i \in A} (y_i + \epsilon_i h_i) = 0\,.$$

Now we give the inductive argument. When $A = \emptyset$, we have

$$\sum_\epsilon \eta(\epsilon) \prod_{i \in A} \epsilon_i = \sum_\epsilon \eta(\epsilon) = \sum_\epsilon \eta(\epsilon) \prod_{i \in A} (y_i + \epsilon_i h_i)\,,$$

which is zero by the above inequality. For general A, we have

$$\begin{aligned} 0 &= \sum_\epsilon \eta(\epsilon) \prod_{i \in A} (y_i + \epsilon_i h_i) \\ &= \sum_\epsilon \eta(\epsilon) \sum_{B \subset A} \prod_{i \in A \setminus B} y_i \prod_{i \in B} \epsilon_i h_i \\ &= \sum_{B \subset A} \prod_{i \in A \setminus B} y_i \prod_{i \in B} h_i \sum_\epsilon \eta(\epsilon) \prod_{i \in B} \epsilon_i \\ &= \prod_{i \in A} h_i \sum_\epsilon \eta(\epsilon) \prod_{i \in A} \epsilon_i \end{aligned}$$

where the last equality follows from the inductive hypothesis. Since the h_i are non-zero, the result is proved for A. □

If we now make the substitution $y_i' = y_i - h_i$ and $h_i' = 2h_i$, and then remove the dashes, we obtain the result for functions on $\{0,1\}^k$, which is what we actually want.

COROLLARY 15.2. *Let $h_1, \dots, h_k$ be non-zero elements of $\mathbb{Z}_N$, and let $\eta : \{0,1\}^k \to \{-1,0,1\}$ be a function such that the sum*

$$\sum_{\epsilon \in \{0,1\}^k} \eta(\epsilon) \prod_{i \in A} (y_i + \epsilon_i h_i)$$

is independent of $y_1, \dots, y_k$ for every subset $A \subset \{1,2,\dots,k\}$. Then η is a multiple of the parity function $\pi : \epsilon \mapsto (-1)^{\sum \epsilon_i}$. □

Define a function $\eta_0 : \{0,1\}^k \times \{1,2,\dots,2d\} \to \{-1,1\}$ by letting $\eta_0(\epsilon, j)$ be $\pi(\epsilon)$ if $1 \leqslant j \leqslant d$ and $-\pi(\epsilon)$ if $d+1 \leqslant j \leqslant 2d$. We shall say that a d-arrangement ρ is *degenerate* if there is a function $\eta : \{0,1\}^k \times \{1,2,\dots,2d\} \to \{-1,0,1\}$ which is not a multiple of η_0 but which nevertheless has the property that

$$\sum_{\epsilon, j} \eta(\epsilon, j) \prod_{i \in A} \rho(\epsilon, j)_i = 0$$

for every subset $A \subset \{1,2,\dots,k+1\}$. (Here, $\rho(\epsilon,j)_i$ denotes the i^{th} coordinate of $\rho(\epsilon,j)$.) We wish to show that there are very few degenerate d-arrangements. Let us give a simple lemma first.

LEMMA 15.3. *Let $\mu : \mathbb{Z}_N^k \to \mathbb{Z}_N$ be a multilinear function which is not constant. Then for any a the number of solutions of $\mu(y_1, \dots, y_k) = a$ is at most $N^k - (N-1)^k \leqslant kN^{k-1}$.*

Proof. The result is trivial when $k = 1$, so let $k > 1$ and assume the result for $k-1$. There are unique multilinear functions μ_1 and μ_2 such that

$$\mu(y_1, \dots, y_k) \equiv y_k \mu_1(y_1, \dots, y_{k-1}) + \mu_2(y_1, \dots, y_{k-1})\,.$$

If we can find two different elements r, s of $\mathbb{Z}_N$ such that the $(k-1)$-linear restrictions $\mu(y_1, \dots, y_{k-1}, r)$ and $\mu(y_1, \dots, y_{k-1}, s)$ of μ are both constant, then we can solve for μ_1 and μ_2 and show that they are both constant as well. Since μ is non-constant, μ_1 is not identically zero and there are exactly N^{k-1} solutions of the equation.

Otherwise, with the exception of at most one r, the function $\mu(y_1, \dots, y_{k-1}, r)$ is not constant. This allows us to apply our inductive hypothesis to conclude that the number of solutions of $\mu(y_1, \dots, y_k) = a$ is at most $N^{k-1} + (N-1)(N^{k-1} - (N-1)^{k-1}) = N^k - (N-1)^k$. □

The estimate above is sharp, since it gives the exact number of solutions of the equation $y_1 \dots y_k = 0$.

LEMMA 15.4. *The number of degenerate d-arrangements in $\mathbb{Z}_N^{k+1}$ is at most $3^{2d.2^k}kN^{(2d+1)k+2d-2}$.*

Proof. Let us fix non-zero sidelengths h_1,...,h_k and cross-sections r_1,...,r_{2d} and take a general d-arrangement

$$\rho : (\epsilon_1, \dots, \epsilon_k, j) \mapsto (y_1^j + \epsilon_1 h_1, \dots, y_k^j + \epsilon_k h_k, r_j)$$

with those sidelengths.

Suppose first that $\eta(\epsilon, j)$ fails, for some j, to be a multiple of the parity function π. Then Corollary 15.2 tells us that there exists a set $A \subset \{1, \dots, k\}$ such that $\sum_\epsilon \eta(\epsilon, j) \prod_{i\in A}(y_i + \epsilon_i h_i)$, when considered as a function of $y_1, \dots, y_k$, is non-constant. By Lemma 15.3, whatever the choice of $\rho(\epsilon, t)$ for $t \neq j$ there are at most kN^{k-1} choices of $(y_1^j, \dots, y_k^j)$ for which $\sum_{\epsilon,t} \eta(\epsilon, t) \prod_{i\in A} \rho(\epsilon, t)_i = 0$. Therefore, the number of d-arrangements with sidelengths $(h_1, \dots, h_k)$ and cross-sections $r_1, \dots, r_{2d}$ such that $\sum_{\epsilon,t} \eta(\epsilon) \prod_{i\in A} \rho(\epsilon, t)_i = 0$ for every $A \subset \{1, 2, \dots, k\}$ is at most $kN^{k-1}N^{(2d-1)k} = kN^{2dk-1}$.

If on the other hand $\eta(\epsilon, j)$ is a multiple of the parity function for every j, then let us write $\eta(\epsilon, j) = \eta_j \pi(\epsilon)$ and consider the set $A = \{1, 2, \dots, k+1\}$. We have

$$\sum_{\epsilon,t} \eta(\epsilon, t) \prod_{i\in A} \rho(\epsilon, t)_i = \sum_{\epsilon,t} \eta(\epsilon, t) \prod_{i=1}^{k+1} \rho(\epsilon, t)_i = (-1)^k h_1 \dots h_k \sum_j \eta_j r_j \,.$$

If η is not a multiple of η_0, then $(\eta_1, \dots, \eta_{2d})$ is not a multiple of the sequence $(1, \dots, 1, -1, \dots, -1)$ (d ones followed by d minus ones). Therefore the equation $\sum_j \eta_j r_j$ places a further linear restriction on the sequence $(r_1, \dots, r_{2d})$, meaning that the number of choices for this sequence is at most N^{2d-2}.

There are (strictly) fewer than $3^{2d.2^k}$ functions $\eta : \{0, 1\}^k \times \{1, \dots, 2d\} \to \{-1, 0, 1\}$ that are not multiples of η_0. For each such function, the arguments we have just given show that the proportion of d-arrangements such that $\sum_{\epsilon,t} \eta(\epsilon) \prod_{i\in A} \rho(\epsilon, t)_i = 0$ for every $A \subset \{1, 2, \dots, k+1\}$ is at most k/N. Finally, the proportion of d-arrangements for which at least one of the sidelengths h_i is zero is also at most k/N. The lemma is proved. □

Notice that, in the above proof, the only set A containing the element $k+1$ that we needed to consider was the set $\{1, 2, \dots, k+1\}$ itself. Thus, it would be possible to get away with a weaker definition of degeneracy. We are now ready for another random selection with dependences defined using Riesz products.

LEMMA 15.5. *Let $\beta, \eta > 0$, let $B \subset \mathbb{Z}_N^{k+1}$ be a set of size βN^{k+1} and let $\phi : B \to \mathbb{Z}_N$ be a function respecting at least $\alpha\beta^{15}N^{17k+15}$ 8-arrangements in B. Then there is a subset $B' \subset B$ containing at least $(\alpha\eta/4)^{2^{2^{k+4}+k+3}}\beta^{15}N^{17k+15}$ 8-arrangements, such that the proportion of them that are respected by ϕ is at least $1-\eta$.*

Proof. Let r be a positive integer to be determined later. For every set $A \subset \{1,\dots,k+1\}$ and every $1 \leqslant j \leqslant r$ choose elements t_j and $s_{A,j}$ uniformly and independently at random from $\mathbb{Z}_N$. Having made the choices of the t_j and the $s_{A,j}$, let each element $y \in B$ belong to B' with probability $p(y)$ given by the formula

$$2^{-r}\prod_{j=1}^{r}\Big(1+\cos\frac{2\pi}{N}\big(t_j\phi(y)+\sum_A s_{A,j}\prod_{i\in A}y_i\big)\Big)$$

and let these probabilities be independent (conditional on the choices for the t_j and $s_{A,j}$).

Here, and for the rest of the proof, any sum over A ranges over all subsets of $\{1,2,\dots,k+1\}$. Let us adopt the following similar conventions. Any sum over ϵ will range over $\{0,1\}^k$, any sum over h will be over $\{1,2,\dots,16\}$ and any sum over S or S_j will be over functions from the power set of $\{1,2,\dots,k+1\}$ to $\mathbb{Z}_N$. The idea of the last convention is that a sum over S or S_j is shorthand for a string of 2^{k+1} sums of the form $\sum_{s_A}$ or $\sum_{s_{A,j}}$.

The probability that an 8-arrangement $\lambda : \{0,1\}^k \times \{1,\dots,16\} \to B$ belongs to B' is

$$N^{-(2^{k+1}+1)r}\sum_{t_1,\dots,t_r}\sum_{S_1,\dots,S_r}\prod_{\epsilon,h}2^{-r}\prod_{j=1}^{r}\Big(1+\cos\frac{2\pi}{N}\big(t_j\phi(\lambda(\epsilon,h))+\sum_A s_{A,j}\lambda(\epsilon,h)_i\big)\Big)$$

which equals

$$N^{-(2^{k+1}+1)r}\Big(\sum_t\sum_S\prod_{\epsilon,h}2^{-r}\Big(1+\cos\frac{2\pi}{N}\big(t\phi(\lambda(\epsilon,h))+\sum_A s_A\lambda(\epsilon,h)_i\big)\Big)^r.$$

By rewriting $1+\cos\frac{2\pi}{N}\big(t\phi(\lambda(\epsilon,h))+\sum_A s_A\lambda(\epsilon,h)_i\big)$ as

$$\frac{1}{2}(1+1+\omega^{t\phi(\lambda(\epsilon,h))+\sum_A s_A\lambda(\epsilon,h)_i}+\omega^{-t\phi(\lambda(\epsilon,h))-\sum_A s_A\lambda(\epsilon,h)_i})$$

we see that the product over (ϵ,h) is a sum of $4^{2^{k+4}}$ terms of the form

$$2^{-2^{k+4}(r+1)}\prod_{\epsilon,h}\omega^{\eta(\epsilon,h)\left(t\phi(\lambda(\epsilon,h))+\sum_A s_A\lambda(\epsilon,h)_i\right)}$$

which equals

$$2^{-2^{k+4}(r+1)}\omega^{t\sum_{\epsilon,h}\eta(\epsilon,h)\phi(\lambda(\epsilon,h))+\sum_A s_A\sum_{\epsilon,h}\eta(\epsilon,h)\prod_{i\in A}\lambda(\epsilon,h)_i}.$$

A term contributes to the sum over all the s_A if and only if all the sums $\sum_{\epsilon,h}\eta(\epsilon,h)\prod_{i\in A}\lambda(\epsilon,h)_i$ are zero, and if λ is non-degenerate, then this can happen only when η is a multiple of η_0. If λ is non-degenerate, $\eta\neq 0$ and the term contributes to the sum over t, then we must have in addition that $\sum_{\epsilon,h}\eta_0(\epsilon,h)\phi(\lambda(\epsilon,h))=0$. It is not hard to see that this is precisely the definition of ϕ respecting the 8-arrangement λ. The contribution of a non-zero term to the sum over t and the s_A is $2^{-2^{k+4}(r+1)}N^{2^k+1}$ and the number of multiples of η_0, counted with multiplicity, is $2^{2^{k+4}}+2$, since 0 can be produced in $2^{2^{k+4}}$ ways, and $\pm\eta_0$ in one way each.

Therefore, if λ is non-degenerate, the sum over t and the s_A is $2^{-2^{k+4}r}N^{2^k+1}$ if ϕ does not respect λ and $2^{-2^{k+4}r}N^{2^k+1}(1+2.2^{-2^{k+4}})$ if it does. It follows that the probability that λ belongs to B' is $2^{-2^{k+4}r}$ if ϕ does not respect λ and $2^{-2^{k+4}r}(1+2.2^{-2^{k+4}})^r$ if it does. Therefore, our hypotheses imply that the expected number X of 8-arrangements respected by ϕ is at least $2^{-2^{k+4}r}(1+2.2^{-2^{k+4}})^r\alpha\beta^{15}N^{17k+15}$, and the expected number Y of non-degenerate 8-arrangements not respected by ϕ is at most $2^{-2^{k+4}r}\alpha\beta^{15}N^{17k+15}$. Using the fact that

$$1+2.2^{-2^{k+4}}\geqslant 2^{2^{-2^{k+4}+1}},$$

we can deduce that if $2^{2^{-2^{k+4}+1}r}\geqslant 2/\alpha\eta$, then

$$\begin{aligned}\eta\mathbb{E}X-\mathbb{E}Y&\geqslant\alpha\eta(2/\alpha\eta)2^{-2^{k+4}r}\beta^{15}N^{17k+15}-2^{-2^{k+4}r}\beta^{15}N^{17k+15}\\&\geqslant 2^{-2^{k+4}r}\beta^{15}N^{17k+15}.\end{aligned}$$

But $2^{2^{-2^{k+4}+1}r}\geqslant 2/\alpha\eta$ if and only if $2^r\geqslant(2/\alpha\eta)^{2^{2^{k+4}-1}}$ if and only if $2^{-r}\leqslant(\alpha\eta/2)^{2^{2^{k+4}-1}}$ if and only if

$$2^{-2^{k+4}r}\leqslant(\alpha\eta/2)^{2^{2^{k+4}-1}2^{k+4}}=(\alpha\eta/2)^{2^{2^{k+4}+k+3}}.$$

Let r be an integer such that

$$2(\alpha\eta/4)^{2^{2^{k+4}+k+3}}\leqslant 2^{-2^{k+4}r}\leqslant(\alpha\eta/2)^{2^{2^{k+4}+k+3}}.$$

If N is large enough that $(\alpha\eta/4)^{2^{2^{k+4}+k+3}}\beta^{15}N\geqslant 3^{2^{k+4}}k$, then Lemma 15.4 and the values of the above expectations imply that a set B' exists such that $X\geqslant(\alpha\eta/4)^{2^{2^{k+4}+k+3}}\beta^{15}N^{17k+15}$, $\eta X\geqslant 2Y$ and $Y\geqslant Z$, where Z is the number of degenerate 8-arrangements. This proves the lemma. □

The final lemma of this section is a combination of the previous one with Lemma 14.8 in the case $\eta=2^{-44}$, which is the value that will be used in applications.

LEMMA 15.6. *Let $\beta, \gamma > 0$. Let $B \subset \mathbb{Z}_N^{k+1}$ be a set of size βN^{k+1} and let $\phi : B \to \mathbb{Z}_N$ be a function satisfying the product property with parameter γ. Then B has a subset B' containing at least $(\beta\gamma/2)^{2^{2^{k+5}}} N^{17k+15}$ 8-arrangements such that the proportion of them that are respected by ϕ is at least $1 - 2^{-44}$.*

Proof. By Lemma 14.8, ϕ respects at least $\beta^{7.4^k}\gamma^{7k.4^{k+1}}N^{17k+15}$ 8-arrangements, and therefore, by Lemma 15.5, B has a subset B' containing at least $\big(2^{-46}\beta^{7.4^k}\gamma^{7k.4^{k+1}}\big)^{2^{2^{k+4}+k+3}}N^{17k+15}$ 8-arrangements, such that the proportion respected by ϕ is at least $1 - 2^{-44}$. The lemma follows from a simple numerical check. □

16 Finding a Multilinear Piece

This section is, as its title suggests, a generalization of §13. As with the last two sections, there will be no major new ideas over and above those needed for bilinearity (and hence progressions of length five) but it is not quite true that there is an obvious one-to-one correspondence between the lemmas that are needed. We begin with a simple consequence of Corollary 5.11. It is the appropriate generalization of Lemma 13.5.

LEMMA 16.1. *Let k be an integer, let $K = (k+1)^2 2^{k+4}$ and let $m \geqslant 2^{K^{2^{k+1}q}2^{32(k+1)^2+1}}$. Let P be a box in $\mathbb{Z}_N^k$ of width at least m, and let $\mu_1, \dots, \mu_q$ be k-linear functions defined on P. Then P can be partitioned into boxes $P_1, \dots, P_M$ of width at least $m^{K^{-2^{k+1}q}}$ with the following property. For every i and j and every $x \in P_j$ we have the inequality $|\mu_i(x)d_j| \leqslant 2m^{-K^{-2^{k+1}q}}N$, where d_j is the common difference of the box P_j.*

Proof. Let d be the common difference of P and let $I \subset \mathbb{Z}_N$ be an arithmetic progression (in $\mathbb{Z}$) of common difference d and size at least m. Then $Q = P \times I$ is a box in $\mathbb{Z}_N^{k+1}$ of gap d and width at least m. Define $(k+1)$-linear functions $\nu_i : Q \to \mathbb{Z}_N$ by $\nu_i(x, y) = \mu_i(x)y$. By Corollary 5.11, we can partition Q into boxes Q_j of width at least $m^{K^{-2^{k+1}q}}$ in such a way that the diameter of every set $\nu_i(Q_j)$ is at most $2C_{k+1}m^{-K^{-2^{k+1}q}}N$. Let y be the minimal element of I and define an equivalence relation on P by setting $x_1 \sim x_2$ if (x_1, y) and (x_2, y) lie in the same box Q_j. The equivalence classes are clearly boxes of width at least $m^{K^{-2^{k+1}q}}$, and the common difference d_j of one of these boxes P_j is the common difference of the box $Q_{j'}$ containing

$P_j \times \{y\}$. The result now follows from the observation that, given $x \in P_j$,

$$\left|\mu_i(x)d_j\right| = \left|\nu_i(x, y + d_j) - \nu_i(x, y)\right|,$$

which is at most the diameter of $\nu_i(Q_{j'})$. □

We are about to state a somewhat complicated inductive hypothesis (Theorem 16.2 below) which will be used to prove the main result of this section. First, let us extend slightly the definition of the product property from §14. Let Γ be any subset of $\mathbb{Z}_N^k \times \mathbb{Z}_N$ and let $\gamma > 0$. We shall say that Γ *has the product property with parameter* γ if, for every subset $B \subset \mathbb{Z}_N^k$ and every function $\phi : B \to \mathbb{Z}_N$ with graph contained in Γ (in other words, $(x, \phi(x)) \in \Gamma$ for every $x \in B$), ϕ has the product property with parameter γ. We shall sometimes abbreviate this as the γ-product property.

Before making the next definition, let us define three similar functions. We let $c(\theta, \gamma, k) = (\gamma\theta)^{2^{2^{k+8}}}$, $q(\theta, \gamma, k) = 1/c(\theta, \gamma, k)$ and $s(\theta, \gamma, k) = (2/\theta\gamma)^{2^{2^{k+6}}}$. We shall now define Γ to be (γ, r)-*multiply* k-*linear* if, for every $\theta > 0$ and every box P of width m, there exists a subset $H \subset P$ of cardinality at least $(1 - \theta)|P|$ together with a partition of P into boxes $P_1, \dots, P_M$ of width at least $m^{c(r^{-1}\theta,\gamma,k)^r}$ such that for each j there are k-linear functions $\mu_1, \dots, \mu_q$ defined on P_j, where $q \leqslant q(r^{-1}\theta, \gamma, k)^r$, such that, for every $x \in P_j \cap H$ and every y with $(x, y) \in \Gamma$, $y = \mu_i(x)$ for some i. Loosely speaking, this says that every box P can be partitioned into further boxes P_j such that, after a small bit of Γ has been thrown away, for every j, $\Gamma \cap (P_j \times \mathbb{Z}_N)$ is contained in the union of the graphs of not too many k-linear functions. If $r = 1$, we shall say simply that Γ is γ-multiply k-linear. If we do not wish to specify k, then we shall say that Γ is (γ, r)-multiply multilinear.

It is an immediate consequence of the definition that if Γ is a (γ, r)-multiply multilinear set and $\Gamma' \subset \Gamma$, then Γ' is also (γ, r)-multiply multilinear.

The next theorem is the main inductive statement we shall need in order to prove an appropriate generalization of Theorem 13.12 to functions that fail to be uniform of degree $k + 1$. (See Corollary 16.11 below.)

Theorem 16.2. *Let $\Gamma \subset \mathbb{Z}_N^k \times \mathbb{Z}_N$ have cardinality at most $\gamma^{-2}N^k$ and satisfy the product property with parameter γ. Then for every $\theta > 0$ there is a subset J of $\mathbb{Z}_N^k$ of size at least $(1 - \theta)N^k$, such that $\Gamma \cap (J \times \mathbb{Z}_N)$ is $(\gamma, \gamma^{-2}s(\theta, \gamma, k))$-multiply k-linear.*

We shall split the proof of Theorem 16.2 into a number of lemmas, most of them easy. First, we check that the induction starts.

LEMMA 16.3. *Theorem 16.2 is true in the case $k = 1$.*

Proof. Let $\theta > 0$. Either there is a set H of size at most θN such that $\Gamma \subset H \times \mathbb{Z}_N$ or we can find a set A of size at least θN and a function $\phi : A \to \mathbb{Z}_N$ such that $(x, \phi(x)) \in \Gamma$ for every $x \in A$. In the first case we can simply set $J = \mathbb{Z}_N \setminus H$ and the result is trivial. Otherwise, we know that ϕ has the product property with parameter γ, which implies that the number of ϕ-additive quadruples is at least $\gamma^8\theta^4N^3 = \gamma^8\theta(\theta N)^3$. Corollary 7.6 now gives us a subset $B \subset A$ of cardinality at least $2^{-1882}\gamma^{9312}\theta^{1165}N$ such that the restriction of ϕ to B is a homomorphism of order 8.

If we now remove from Γ all points $(x, \phi(x))$ with $x \in B$, we obtain a new set Γ_1 to which the above argument may be applied again. Continuing, we construct sets $B_1, \dots, B_q$ of cardinality at least $2^{-1882}\gamma^{9312}\theta^{1165}N$ and homomorphisms $\phi_i : B_i \to \mathbb{Z}_N$ of order 8, such that the graph of each ϕ_i is contained in Γ, these graphs are disjoint and there is a set $J \subset \mathbb{Z}_N$ of size at least $(1 - \theta)N$ such that $x \in J$ and $(x, y) \in \Gamma$ implies that $y = \phi_i(x)$ for some i. Moreover, the upper bound on the size of Γ implies that $q \leqslant 2^{1882}\gamma^{-9314}\theta^{-1165}$.

Now let P be an arithmetic progression (or a one-dimensional box). By Corollary 7.11, with $\alpha = 2^{-1882}\gamma^{9312}\theta^{1165}$ and q as above, we can partition P into subprogressions $Q_1, \dots, Q_M$, each of size at least $|P|^{2^{-14}\alpha^2q^{-1}} \geqslant |P|^{2^{-6000}\gamma^{30000}\theta^{15000}}$ such that the restriction of each ϕ_i to each $B_i \cap Q_j$ is linear. It is not hard to check that these numbers do indeed demonstrate that $\Gamma \cap (J \times \mathbb{Z}_N)$ is $(\gamma, \gamma^{-2}s(\theta, \gamma, 1))$-multiply linear. □

We are now ready to begin the inductive argument in earnest.

LEMMA 16.4. *Suppose that Theorem 16.2 is true for k. Let $\theta > 0$ and let $\Gamma \subset \mathbb{Z}_N^{k+1} \times \mathbb{Z}_N$ be a set of cardinality at most $\gamma^{-2}N^{k+1}$ satisfying the product property with parameter γ. Then either there is a set $H \subset \mathbb{Z}_N^{k+1}$ of cardinality less than θN^{k+1} such that $\Gamma \subset H \times \mathbb{Z}_N$, or one can find a set $B \subset \mathbb{Z}_N^{k+1}$ and a function $\phi : B \to \mathbb{Z}_N$ with the following properties:*

- *(i) the restriction of ϕ to any proper cross-section of B is $(\gamma, \gamma^{-2}s(2^{-(k+2)}\theta, \gamma, k))$-multiply multilinear;*
- *(ii) B contains at least $(\theta\gamma)^{2^{2^{k+5}}}N^{17k+15}$ 8-arrangements;*
- *(iii) of the 8-arrangements in B, the proportion respected by ϕ is at least $1 - 2^{-44}$.*

Proof. If the first alternative does not hold, then we can find a set $A \subset \mathbb{Z}_N^{k+1}$ of cardinality at least θN^{k+1} and a function $\phi : A \to \mathbb{Z}_N$ such that $(x, \phi(x)) \in \Gamma$ for every $x \in A$. Then ϕ has the γ-product property. In fact, so does the restriction of ϕ to any cross-section of A of the form $A_{X,z} = \{x \in A : x_i = z_i \text{ for every } i \in X\}$. (This follows directly from the definition.) Let $\zeta = 2^{-(k+2)}\theta$, let $l \leqslant k$ and let $A_{X,z}$ be an l-dimensional cross-section of A of cardinality βN^l. By the inductive hypothesis, there is a subset $A'_{X,z} \subset A_{X,z}$ of cardinality at least $(\beta - \zeta)N^l$ such that the restriction of ϕ to $A'_{X,z}$ is $(\gamma, \gamma^{-2}s(\zeta, \gamma, l))$-multiply l-linear.

For any given set $X \subset [k+1]$ of size $k+1-l$, there are N^{k+1-l} different cross-sections $A_{X,z}$, which partition A. Therefore, we can find a subset $A' \subset A$ of cardinality at least $(\theta - \zeta)N^{k+1}$, such that the restriction of ϕ to any cross-section of A' in direction X is $(\gamma, \gamma^{-2}s(\zeta, \gamma, k))$-multiply l-linear. Repeating this argument for all the $2^{k+1} - 1$ non-empty sets $X \subset [k+1]$, we can find a subset $A'' \subset A$ of cardinality at least $\theta N^{k+1}/2$ such that the restriction of ϕ to any proper cross-section of A'' is $(\gamma, \gamma^{-2}s(\zeta, \gamma, k))$-multiply multilinear of the appropriate dimension.

As remarked just before the statement of the lemma, this property is preserved if we pass to a subset of A''. We now do precisely that, using Lemma 15.6 to find a subset B of A'' containing at least $(\theta\gamma/2)^{2^{2^{k+5}}} N^{17k+15}$ 8-arrangements, such that the proportion of them respected by ϕ is at least $1 - 2^{-44}$. The lemma is now proved. □

For any $h_1, \dots, h_k, x$, let $X_{h_1,\dots,h_k}(x)$ be the set of all cubes with sidelengths $(h_1, \dots, h_k)$ in the cross-section B_x of B. Let $X_{h_1,\dots,h_k}$ be the union of these sets. Then $X_{h_1,\dots,h_k}$ is a domain (in the sense of §10, under the splitting into the sets $X_{h_1,\dots,h_k}(x)$). For the rest of this section, we shall frequently abbreviate $(h_1, \dots, h_k)$ by h, as we did in §14. If we let f be the characteristic function of B, then it is easy to see from the definition of the function f_h (given just before Lemma 14.3) that $f_h(y)$ is the number of cubes with sidelengths $(h_1, \dots, h_k) = h$ in the cross-section B_y, or in other words the cardinality of $X_h(y)$. For each $h = (h_1, \dots, h_k)$, let $C(h)$ be the number of 8-arrangements in B made out of cubes with sidelengths $(h_1, \dots, h_k)$. Let $G(h)$ be the number of these that are respected by ϕ. Recall from just before Lemma 14.4 that any function $\phi : \mathbb{Z}_N^{k+1} \to \mathbb{Z}_N$ induces a function (which we also call ϕ) on X_h.

We shall write θ_1 for the number $(\theta\gamma/2)^{2^{2^{k+5}}}$ that appeared in the last lemma. Recall that the Bohr neighbourhood $B(K, \zeta)$ is defined to be the set of all $s \in \mathbb{Z}_N$ such that $|rs| \leqslant \zeta N$ for every $r \in K$. Given a function ϕ

defined on a domain (Z,r) and a subset $B\subset\mathbb{Z}_N$ with $B=-B$, we shall say that ϕ is a *B-homomorphism* if there is a Freiman homomorphism $\psi:B\to\mathbb{Z}_N$ such that $\phi(x)-\phi(y)=\psi(r(x)-r(y))$ whenever $r(x)-r(y)\in B$. (This generalizes to multifunctions the definition given just after Corollary 7.9.) Thus, the conclusion of Theorem 10.13 is that ϕ restricted to Y is a C-homomorphism.

LEMMA 16.5. *Let (B,ϕ) be a pair satisfying conditions (i), (ii) and (iii) of Lemma 16.4 and let f be the characteristic function of B. Then there exists a set $H\subset\mathbb{Z}_N^k$ such that $\sum_{h\in H}C(h)\geqslant(\theta_1/4)N^{17k+15}$ with the following property. For every $h\in H$ there is a set $Y_h\subset X_h$ of cardinality at least $2^{-22}\theta_1^6|X_h|$ such that the restriction of ϕ to Y_h is a $B(K_h,\zeta)$-homomorphism, where*

$$K_h=\big\{r\in\mathbb{Z}_N:|\hat{f}_h(r)|\geqslant 2^{-37}(\theta_1/4)^{11/2}N^{k+1}\big\}$$

and $\zeta=2^{-s(\theta,\gamma,k)}$.

Proof. We continue to write η for the number 2^{-44}. Since ϕ respects a proportion of at least $1-\eta$ of the 8-arrangements of B, of which there are at least θ_1N^{17k+15}, we may deduce that

$$\sum\{C(h):G(h)\geqslant(1-2\eta)C(h)\}\geqslant\tfrac12\sum C(h)\geqslant(\theta_1/2)N^{17k+15}\,.$$

Another simple averaging argument shows that

$$\sum\big\{C(h):G(h)\geqslant(1-2\eta)C(h),C(h)\geqslant(\theta_1/4)N^{16k+15}\big\}\geqslant(\theta_1/4)N^{17k+15}\,.$$

Let H be the set of all $h\in\mathbb{Z}_N^k$ such that $G(h)\geqslant(1-2\eta)C(h)$ and $C(h)\geqslant(\theta_1/4)N^{16k+15}$, and note that our estimate for $\sum_{h\in H}C(h)$ implies that H has cardinality at least $\theta_1N^k/4$.

The statement that $G(h)\geqslant(1-2\eta)C(h)$ is equivalent to the statement that the function induced by ϕ on X_h is a $(1-2\eta)$-homomorphism of order eight. We know that $X_h(y)$ has cardinality at most N^k for each y, and it is not hard to show that if $C(h)\geqslant(\theta_1/4)N^{16k+15}$ then the cardinality of X_h is at least $\theta_1N^{k+1}/4$. Therefore, for every $h\in H$ we may apply Theorem 10.13 (with $\alpha=\theta_1/4$ and $g=f_h$) and find a subset $Y_h\subset X_h$ of cardinality at least $2^{-16}\theta_1^3|X_h|$ such that the restriction of ϕ to Y_h is a $B(K_h,\zeta)$-homomorphism, where ζ is determined (as a function of θ) by the equations $\theta_1=(\theta\gamma/2)^{2^{2^{k+5}}}$, $\alpha=\theta_1/4$, $k_0=2^{74}\alpha^{-10}$ and $\zeta=2^{-155k_0}\alpha^{18k_0}k_0$. It can be checked that the resulting number ζ exceeds $2^{-s(\theta,\gamma,k)}$. □

From the definition of a $B(K_h,\zeta)$-homomorphism, we know in particular that, for any $x\in\mathbb{Z}_N$ and any $h\in H$, the restriction of ϕ to $Y_h(x)$ is

constant. Let us write $\phi'(h,x)$ for this value, when it is defined. Corollary 10.14 tells us that if $d \in B(K_h, \zeta/l)$ then the restriction of $\phi'(h,.)$ to any arithmetic progression with common difference d and length at most l is linear. This fact will be used in the next lemma. We shall also adopt the convention that if a box in $\mathbb{Z}_N^{k+1}$ is written as a Cartesian product $A \times B$, then A and B are boxes in $\mathbb{Z}_N^k$ and $\mathbb{Z}_N$ respectively.

Let $\delta = 2^{-37}(\theta_1/4)^{11/2}$ and define $\Delta \subset \mathbb{Z}_N^k \times \mathbb{Z}_N$ to be the set of all (h,r) such that $|\hat{f}_h(r)| \geqslant \delta N^{k+1}$, that is, such that $r \in K_h$. Lemma 14.3 tells us that Δ has the product property with parameter δ. Therefore, if Theorem 16.2 is true for k, then there is a subset $J \subset \mathbb{Z}_N^k$ of size at least $(1-\theta_1/8)N^k$ such that $\Delta_1 = \Delta \cap (J \times \mathbb{Z}_N)$ is $(\delta, \delta^{-2}s(\theta_1/8, \delta, k))$-multiply k-linear. Let $H_1 = H \cap J$, where H is the set defined in the proof of Lemma 16.5. Since $\sum_{h \notin J} C(h) \leqslant (\theta_1/8)N^{17k+15}$, we find that $\sum_{h \in H_1} C(h) \geqslant (\theta_1/8)N^{17k+15}$.

Before we state and prove the next lemma, let us remark that the definition of the set Δ above is in a sense the moment where the induction takes place. For any given h, there are at most δ^{-2} values of r such that $(h,r) \in \Delta$, so Δ is the union of the graphs of at most δ^{-2} functions. We have therefore managed once again to reduce the number of variables by one by considering a new function which tells us where some Fourier coefficients related to the domain of the old function are large. The results of the previous two sections together with the inductive hypothesis have told us that the new function has a lot of structure; this will now be used to tell us about the old function, which will complete the inductive step.

LEMMA 16.6. *Let $P = Q \times I$ be a box in $\mathbb{Z}_N^{k+1}$ of width at least m, let $t = \delta^{-2}s(\theta_1/8, \delta, k)$, let $\sigma > 0$ and let $l = (\zeta/2)m^{c(t^{-1}\sigma,\delta,k)^t/2K^{2^{k+1}q}}$. Then there is a subset $G \subset Q$ of size at least $(1-\sigma)|Q|$ and a partition of P into boxes $S_u = T_u \times J_u$ of width at least l, such that, for every u and every $h \in G \cap H_1 \cap T_u$, the function from J_u to $\mathbb{Z}_N$ defined by $x \mapsto \phi'(h,x)$ is linear.*

Proof. Because Δ_1 is (δ, t)-multiply k-linear, we can find a subset $G \subset Q$ of size at least $(1-\sigma)|Q|$ and a partition of Q into subboxes Q_j of width at least $m_1 = m^{c(\sigma/t,\delta,k)^t}$ such that for each j there are k-linear functions $\mu_1, \dots, \mu_q$ from Q_j to $\mathbb{Z}_N$, with $q \leqslant q(\sigma/t, \delta, k)^t$ such that $\Delta_1 \cap \big((G \cap Q_j) \times \mathbb{Z}_N\big)$ is contained in the union of the graphs of the μ_i. This says that, for any $h \in G \cap Q_j$, the set $K_h = \{r \in \mathbb{Z}_N : |\hat{f}_h(r)| \geqslant \delta N^{k+1}\}$ is a subset of $\{\mu_1, \dots, \mu_q\}$.

By Lemma 16.1, each Q_j may be partitioned into subboxes R_t of width

at least $m_2 = m_1^{1/K^{2^{k+1}q}} \geqslant l^2$ and common difference d_t such that, given any $h \in R_t$ and any i, $|\mu_i(h)d_t| \leqslant 2m_1^{-1/K^{2^{k+1}q}} \leqslant \zeta N/(l+1)$. By the conclusion of the last paragraph, this implies that, whenever $h \in G \cap R_t$, d_t belongs to the Bohr neighbourhood $B(K_h, \zeta/(l+1))$. For each t, $R_t \times I$ can be partitioned into boxes $S_u = T_u \times J_u$ of width l or $l+1$. As remarked after the statement of Lemma 16.5, the restriction of $\phi'(h,.)$ to an arithmetic progression with common difference $d \in B(K_h, \zeta/(l+1))$ and length at most $l+1$ is linear. In other words, the function from J_u to $\mathbb{Z}_N$ defined by $x \mapsto \phi'(h,x)$ is linear, as stated. □

Notice that it was vital in the above lemma that Δ_1 should have good structure and that this should give us information, via Fourier coefficients, about the restriction of ϕ to the sets Y_h, even though the definition of Δ_1 was in terms of the X_h. It was to achieve this that we worked so hard in §10.

LEMMA 16.7. *Let $\theta_2 = 2^{-21}\theta_1^5$. Then there exist elements $x_1, \dots, x_k$ of $\mathbb{Z}_N$ such that, for at least $\theta_2 N^{k+1}$ choices of $(h_1, \dots, h_k, x)$ with $h \in H_1$, $\phi'(h,x)$ is defined and equals*

$$\sum_{\epsilon \in \{0,1\}^k} (-1)^{|\epsilon|} \phi(x_1 + \epsilon_1 h_1, x_2 + \epsilon_2 h_2, \dots, x_k + \epsilon_k h_k, x)\,.$$

Proof. The expression given for $\phi'(h,x)$ is valid whenever $Y_h(x)$ contains the cube $[x_1, \dots, x_k; h_1, \dots, h_k]$. Since $|H_1| \geqslant (\theta_1/8)N^k$ and $|X_h| \geqslant (\theta_1/4)N^{k+1}$ for every $h \in H_1$, we find that $\sum_{h \in H} |Y_h| \geqslant 2^{-21}\theta_1^5 N^{2k+1} = \theta_2 N^{2k+1}$ (by the estimate for the sizes of the sets Y_h in Lemma 16.5). Therefore, if we choose the x_j randomly, the expected number of choices of $(h_1, \dots, h_k, x)$ with $h \in H_1$ for which the equality holds is at least $\theta_2 N^{k+1}$. The lemma follows. □

LEMMA 16.8. *Suppose that $\Gamma_1, \dots, \Gamma_r$ are (γ, s)-multiply $(k+1)$-linear subsets of $\mathbb{Z}_N^{k+1} \times \mathbb{Z}_N$. Then $\Gamma_1 \cup \dots \cup \Gamma_r$ is (γ, rs)-multiply $(k+1)$-linear. If $\phi_1, \dots, \phi_r$ are (γ, s)-multiply $(k+1)$-linear functions defined on a subset $B \subset \mathbb{Z}_N^{k+1}$, then $\phi_1 + \dots + \phi_r$ is (γ, rs)-multiply $(k+1)$-linear.*

Proof. Let P be a box. We can find a subset $H_1 \subset P$ of size at least $(1 - \theta/rs)|P|$ and a partition of P into boxes Q of width at least $m^{c((rs)^{-1}\theta,\gamma,k)^s}$ such that Γ_1 restricted to any $Q \cap H_1$ is contained in the union of the graphs of $q((rs)^{-1}\theta, \gamma, k,)^s$ multilinear functions. Now repeat this argument inside each Q for the set Γ_2 and so on. At each of the r stages of this process, the width of the boxes is raised to the power

$c((rs)^{-1}\theta,\gamma,k)^s$, the number of new multilinear functions introduced inside each box is at most $q((rs)^{-1}\theta,\gamma,k)^s$ and $\theta|P|/rs$ points are thrown away. Therefore, at the end of the process we have a width of at least $m^{c((rs)^{-1}\theta,\gamma,k)^{rs}}$ and $rq((rs)^{-1}\theta,\gamma,k)^s$ multilinear functions for each set Γ_i. The result about unions follows (and in fact we have overestimated the number of multilinear functions needed). The result for sums of functions also follows, once we notice that there are $q((rs)^{-1}\theta,\gamma,k)^{rs}$ functions of the form $\mu_1+\cdots+\mu_r$, with each μ_i one of the multilinear functions chosen at the i^{th} stage. □

Let us now fix a choice of $x_1,\dots,x_k$ satisfying the conclusion of Lemma 16.7 and write $\phi_\epsilon(h,x)$ for $\phi(x_1+\epsilon_1h_1,\dots,x_k+\epsilon_kh_k,x)$. Write also $\phi_1(h,x)$ for the function $\phi_\epsilon(h,x)$ when $\epsilon=(1,1,\dots,1)$. Regard all these functions as being defined on the set B_1 of (h,x) that satisfy the conclusion of Lemma 16.7, which can be rephrased as $\phi_1(h,x)=\phi'(h,x)-\sum_{\epsilon\neq 1}(-1)^{|\epsilon|}\phi_\epsilon(h,x)$ and $h\in H_1$. We now show that something like Lemma 16.6, but weaker, holds for the function ϕ_1 as well.

LEMMA 16.9. *Let $t=\delta^{-2}s(\theta_1/8,\delta,k)$, $r=(2^k-1)\gamma^{-2}s(2^{-(k+2)}\theta,\gamma,k)$ and $q=q(\gamma,\sigma/2r,k)^r$. Let $P=Q\times I$ be any box in $\mathbb{Z}_N^{k+1}$ of width at least m and let $\sigma>0$. Then there is a subset $E\subset P$ of size at least $(1-\sigma)|P|$ and a partition of P into boxes $S_u=T_u\times J_u$ with the following property. Given u and $h\in T_u$ let $\psi_{u,h}$ be the function $x\mapsto\phi_1(h,x)$, where the domain is the set of all x such that $(h,x)\in B_1\cap E\cap S_u$. Then for every u and $h\in T_u$ the graph of $\psi_{u,h}$ is contained in the union of the graphs of at most q linear functions. The width of each box S_u is at least*

$$l=(\zeta/2C_{k+1})m^{c(\sigma/2r,\gamma,k)^rc(\sigma/2t,\delta,k)^t/2K^{2^{k+1}q}}.$$

Proof. Given any sequence $\epsilon\in\{0,1\}^k$ apart from $(1,1,\dots,1)$, let $X=\{j:\epsilon_j=0\}$ and let B_ϵ be the cross-section of B defined as the set of all $y\in B$ such that $y_j=x_j$ for every $j\in X$. By property (i) of Lemma 16.4, the restriction of ϕ to the cross-section B_ϵ is $(\gamma,\gamma^{-2}s(2^{-(k+2)}\theta,\gamma,k))$-multiply multilinear. It follows easily that ϕ_ϵ itself is $(\gamma,\gamma^{-2}s(2^{-(k+2)}\theta,\gamma,k))$-multiply $(k+1)$-linear, since ϕ_ϵ is obtained from the restriction of ϕ by introducing variables that make no difference, namely the h_i with $i\in X$.

Hence, by Lemma 16.8 and the expression for ϕ_1 just before the statement of this lemma, we can write

$$\phi_1(h,x)=\phi'(h,x)+\phi''(h,x),$$

where ϕ'' is (γ,r)-multiply $(k+1)$-linear. By the definition of (γ,r)-multiple multilinearity, we can find a subset $F\subset P$ of cardinality at least $(1-\sigma/2)|P|$

and a partition of P into boxes P_j of width at least $m_1 = m^{c((2r)^{-1}\sigma,\gamma,k)^r}$ such that for every j there are $(k+1)$-linear functions $\mu_1,\dots,\mu_q$ from P_j to $\mathbb{Z}_N$ with the property that $\phi''(h,x) = \mu_i(h,x)$ for some i, whenever $(h,x) \in B_1 \cap F \cap P_j$.

By Lemma 16.6, each of the boxes $P_j = Q_j \times I_j$ gives a subset $G_j \subset Q_j$ of size at least $(1-\sigma/2)|Q_j|$ and a further partition into boxes $S_{ju} = T_{ju} \times J_{ju}$ of width at least $m_2 = (\zeta/2C_{k+1})m_1^{c((2t)^{-1}\sigma,\delta,k)^t/2K^{2^{k+1}q}}$ such that for every $h \in H_1 \cap G_j \cap T_{ju}$ (recall that $(h,x) \in B_1$ implies that $h \in H_1$, which was defined just before the statement of Lemma 16.6), the restriction of $\phi'(h,x)$ to $B_1 \cap S_{ju}$ is linear in x. The lemma now follows on adding ϕ' and ϕ'' and taking E to be $F \cap \bigcup_j (G_j \times I_j)$. □

We have just shown that ϕ_1 has a property similar to multiple multilinearity but much weaker because it gives us linearity only in one of the variables. However, we also have information about the restriction of ϕ_1 to proper cross-sections, and this enables us to show that the linear functions in the final variable are related to each other in a multilinear way. The details are in the next lemma.

LEMMA 16.10. *The function ϕ_1 is itself $(\gamma,1)$-multiply $(k+1)$-linear.*

Proof. We begin by remarking that, since ϕ_1 is a translation of a restriction of ϕ, property (i) of Lemma 16.4 implies that the restriction of ϕ_1 to any cross-section of B_1 formed by fixing the final variable x is $(\gamma,\gamma^{-2}s(2^{-(k+2)}\theta,\gamma,k))$-multiply multilinear of the appropriate dimension.

Now let $\rho > 0$, let $\sigma = \rho/4$ and let $P = Q \times I$ be a box of width at least m. Applying Lemma 16.9, we can find a subset $E \subset P$ of size at least $(1-\sigma)|P|$ and a partition of P into boxes $S_u = T_u \times J_u$ of width at least l satisfying the conclusion of that lemma. Let $S = T \times J$ be one of these boxes, and write $B_1(h)$ for the set $\{(h',x) \in B_1 \cap E \cap S : h' = h\}$. Each set $B_1(h)$ can be partitioned into subsets $C_1(h),\dots,C_q(h)$ such that the restriction of ϕ_1 to any $C_t(h)$ is linear. An easy averaging argument shows that

$$\sum_{j=1}^{q} \sum_{h \in S} \big\{ |C_t(h)| : |C_t(h)| \geqslant \sigma|J|/q \big\} \geqslant (1-\sigma)|S|\,.$$

Hence, there is a subset $D \subset S$ of size at least $(1-\sigma)|S|$ such that, for every t, $C_t(h) \cap D$ is either empty or of size at least $\sigma|J|/q$.

Now let $r = q\sigma^{-2}$ and choose $x_1,\dots,x_r$ randomly from J. If $|C_t(h)| \geqslant \sigma|J|/q$, then the probability that $C_t(h)$ does not contain two distinct points (h,x_i) and (h,x_j) is at most $(1-\sigma/q)^r + r(\sigma/q)(1-\sigma/q)^{r-1}$, which is much

smaller than σ. If we discard every set $C_t(h)$ which does not contain such a distinct pair, then the expected number of points discarded is at most σ times the total number of points in the $C_t(h)$, which is certainly at most $\sigma|S|$. Hence, we can choose $x_1, \dots, x_r$ and find a set $F \subset S$ of size at least $(1-\sigma)|S|$ such that for every h, t and every $(h,x) \in C_t(h) \cap D \cap F$ there are x_i and x_j not the same with (h,x_i) and (h,x_j) both in $C_t(h) \cap D \cap F$ as well.

For any fixed h, t, there are constants $\lambda_t(h)$ and $\mu_t(h)$ such that $\phi_1(h,x) = \lambda_t(h)x + \mu_t(h)$ for every $(h,x) \in C_t(h)$. If in addition $x_i \neq x_j$ and (h,x_i) and (h,x_j) both belong to $C_t(h)$, then $\lambda_t(h)x_i + \mu_t(h) = \phi_1(h,x_i)$ and $\lambda_t(h)x_j + \mu_t(h) = \phi_1(h,x_j)$. These equations imply that

$$\lambda_t(h) = (x_i - x_j)^{-1}\big(\phi_1(h,x_i) - \phi_2(h,x_j)\big)\,,$$

which we shall denote by $\lambda_{ij}(h)$, and that

$$\mu_t(h) = \phi_1(h,x_i) - \lambda_{ij}(h)x_i\,,$$

which we shall denote by $\mu_{ij}(h)$. By Lemma 16.8 and the remark with which we opened the proof, the functions λ_{ij} and μ_{ij} are all $(\gamma, 2\gamma^{-2}s(2^{-(k+2)}\theta,\gamma,k))$-multiply multilinear, and for every $(h,x) \in B_1 \cap E \cap D \cap F$ we can find i,j such that $\phi_1(h,x) = \lambda_{ij}(h)x + \mu_{ij}(h)$.

It is not hard to see (using Lemma 16.8 again) that $\lambda_{ij}(h)x + \mu_{ij}$ is a $(\gamma, 4\gamma^{-2}s(2^{-(k+2)}\theta,\gamma,k))$-multiply multilinear function of (h,x), and, by one further application of Lemma 16.8, the union of the graphs of all these functions, which contains the graph of ϕ_1 restricted to $B_1 \cap E \cap D \cap F$, is $(\gamma, 4r^2\gamma^{-2}s(2^{-(k+2)}\theta,\gamma,k))$-multiply multilinear.

Let $p = 4r^2\gamma^{-2}s(2^{-(k+2)}\theta,\gamma,k)$. By what we have just shown, there is a subset $G \subset S$ of size at least $(1-\sigma)|S|$ and a partition of S into boxes V of width at least $l^{c(\gamma,\sigma/p,k)^p}$ such that for each one the graph of ϕ_1 restricted to $V \cap E \cap D \cap F \cap G$ is contained in the union of the graphs of $q(\sigma/p,\gamma,k)^p$ multilinear functions.

To complete the proof of the lemma, it is necessary only to check that $q(\sigma/p,\gamma,k)^p \leqslant q(\rho,\gamma,k+1)$ and that $l^{c(\sigma/p,\gamma,k)^p} \geqslant m^{c(\rho,\gamma,k+1)}$. This is a back-of-envelope calculation left to the reader. □

Proof of Theorem 16.2. Lemma 16.10 shows that if the result is true for k and $\Gamma \subset \mathbb{Z}_N^{k+1} \times \mathbb{Z}_N$ has the product property with parameter γ, then Γ has a $(\gamma,1)$-multiply $(k+1)$-linear subset of cardinality at least $\theta_2 N^{k+1} \geqslant N^{k+1}/s(\theta,\gamma,k)$, if its projection has size at least θN^{k+1}. Now apply this result repeatedly, removing such sets from Γ until it no longer has a projection of size at least θN^{k+1}. Since $|\Gamma| \leqslant \gamma^{-2}N^{k+1}$, the num-

ber of sets removed is at most $\gamma^{-2}s(\theta,\gamma,k)$. The result now follows from Lemma 16.8. □

COROLLARY 16.11. *If $f:\mathbb{Z}_N\to D$ fails to be α-uniform of degree $k+1$ then there is a box $P\subset\mathbb{Z}_N^k$ of width at least $N^{(\alpha/2)^{2^{2^{k+9}}}}$ and a multilinear function $\mu:P\to\mathbb{Z}_N$ such that, for at least $(\alpha/2)^{2^{2^{k+9}}}|P|$ values of $(y_1,\dots,y_k)$, we have $|\Delta(f;y_1,\dots,y_k)^\wedge(\mu(y_1,\dots,y_k))|\geqslant(\alpha/2)N$.*

Proof. Since f is not α-uniform of degree $k+1$, we find, using the implication of (ii) from (vi) in Lemma 3.1, that there is a set $B\subset\mathbb{Z}_N^k$ of size at least $(\alpha/2)N^k$ and a function $\phi:B\to\mathbb{Z}_N$ such that $|\Delta(f;a_1,\dots,a_k)^\wedge(\phi(a_1,\dots,a_k))|\geqslant(\alpha/2)N$ for every $(a_1,\dots,a_k)\in B$. Lemma 14.2 then implies that ϕ has the product property with parameter $\alpha/2$. Next, Theorem 16.2 implies that B has a subset C of size at least $(\alpha/4)N^k$ such that the restriction of ϕ to C is $(\alpha/2,r)$-multiply k-linear, where $r=4\alpha^{-2}s(\alpha/4,\alpha/2,k)$. Applying the definition of multiple multilinearity in the case where the box P is the whole of $\mathbb{Z}_N^k$ and $\theta=\alpha/8$, we find a set $H\subset\mathbb{Z}_N^k$ of size at least $(1-\alpha/8)N^k$ and partition of $\mathbb{Z}_N^k$ into boxes $P_1,\dots,P_M$ of width at least $N^{c(\alpha/8r,\alpha/2,k)^r}$ such that for every j the restriction of ϕ to $C\cap P_j\cap H$ is contained in the graph of at most $q(\alpha/8r,\alpha/2,k)^r$ multilinear functions. By averaging, we can find a box P_j such that $|C\cap P_j\cap H|\geqslant(\alpha/8)|P_j|$. By further averaging, we can find a subset $D\subset P_j$ of size at least $(q(\alpha/8r,\alpha/2,k)^r)^{-1}(\alpha/8)|P_j|$ such that the restriction of ϕ to D is multilinear. A straightforward calculation shows that this implies the corollary. □

17 The Main Inductive Step

We are finally ready to generalize the argument of §8, to complete a proof of Szemerédi's theorem for progressions of arbitrary length. It turns out that there is a second reason for this being harder than for progressions of length four, but fortunately it is much less serious than the difficulties we have dealt with in the last two sections.

To see the problem, let A be a set with balanced function f which fails to be cubically uniform. We know then that there are many pairs (k,l) such that $\Delta(f;k,l)$ has a large Fourier coefficient. The results of §9 show that the large Fourier coefficient has regions where it depends bilinearly on (k,l). As at the beginning of §8, let us imagine that we actually have the

best possible situation: that is, that we can find c such that

$$\sum_{k,l} |\Delta(f;k,l)^\wedge(6ckl)|^2 \geqslant \alpha N^4\,,$$

so that the dependence on (k,l) of where the large Fourier coefficient appears is genuinely bilinear.

Writing out the above inequality in full and making the usual substitution, we find that

$$\sum_s \sum_{k,l,m} \Delta(f;k,l,m)(s)\omega^{-6cklm} \geqslant \alpha N^4\,.$$

If we now use the identity

$$6klm = \sum_{\epsilon_1,\epsilon_2,\epsilon_3} (s-\epsilon_1 k-\epsilon_2 l-\epsilon_3 m)^3\,,$$

where the sum is over the eight triples $(\epsilon_1,\epsilon_2,\epsilon_3)$ with $\epsilon_i = 0$ or 1, then, writing C once again for the operation of complex conjugation, we can deduce that

$$\sum_s \sum_{k,l,m} \prod_{\epsilon_1,\epsilon_2,\epsilon_3} C^{\epsilon_1+\epsilon_2+\epsilon_3}\big(f(s-\epsilon_1 k-\epsilon_2 l-\epsilon_3 m)\omega^{-c(s-\epsilon_1 k-\epsilon_2 l-\epsilon_3 m)^3}\big) \geqslant \alpha N^4\,.$$

Unfortunately, the standard trick that we applied in §8 (and of course many other places in the paper) of inserting a term $\omega^{-r(a-b-c+d)}$ simply does not have an equivalent here. (Indeed, if it did, then the whole paper would be far simpler.) So have we gained anything at all with the above manipulations? The answer is that we have, because the above inequality tells us precisely that the function $g(s) = f(s)\omega^{-cs^3}$ is not *quadratically* α-uniform. Therefore, by the results of §8, g has plenty of quadratic bias, which tells us that there are many progressions P for which $|\sum_{s\in P} f(s)\omega^{\phi(s)}|$ is large for some cubic polynomial ϕ (depending on the progression). Finally, the results of §5 can be used to find a small progression where A is denser than it should be.

Of course, if we have only a small piece of bilinearity to work with, the argument above has to be modified a little, but the rough form of our inductive hypothesis, and indeed the rest of the proof, ought by now to be clear. Our first lemma is by no means new, but we state and briefly prove it, for the convenience of the reader.

LEMMA 17.1. *Let σ be any k-linear function (over $\mathbb{Z}_N$) in variables x_1,...,x_k. Then there are polynomials ϕ_ϵ ($\epsilon \in \{0,1\}^k$) of degree at most k giving the identity*

$$\phi(x_1,\dots,x_k) = \sum_{\epsilon\in\{0,1\}^k} (-1)^{|\epsilon|}\phi_\epsilon(s-\epsilon.x)\,.$$

Proof. It is enough to prove the result in the case $\sigma(x_1,\dots,x_k)=x_1\dots x_k$. Now $s^k-(s-x_1)^k$ is a polynomial in s of degree $k-1$ with leading term kx_1s^{k-1}. It follows that $s^k-(s-x_1)^k-(s-x_2)^k+(s-x_1-x_2)^k$ is a polynomial in s of degree $k-2$ with leading term $k(k-1)x_1x_2s^{k-2}$. Continuing, we find that

$$k!x_1\dots x_k=\sum_{\epsilon\in\{0,1\}^k}(-1)^{|\epsilon|}(s-\epsilon.x)^k$$

as we wanted. □

We now prove a proposition which is not exactly what we need later. Rather, it is a special case, which we give in the hope that the more general result, which is a bit complicated, will be easier to understand.

PROPOSITION 17.2. *Let $f:\mathbb{Z}_N\to D$. Suppose that there is a k-linear function $\sigma:\mathbb{Z}_N^k\to\mathbb{Z}_N$ such that*

$$\sum_{x\in\mathbb{Z}_N^k}\bigl|\Delta(f;x)^\wedge(\sigma(x))\bigr|^2\geqslant\alpha N^{k+2}\,.$$

Then there is a polynomial ϕ of degree at most k such that, setting $g(s)=f(s)\omega^{-\phi(s)}$, we have

$$\sum_{x\in\mathbb{Z}_N^k}\Bigl|\sum_s\Delta(g;x)(s)\Bigr|^2\geqslant\alpha N^{k+2}\,.$$

Proof. By Lemma 17.2 we may write

$$x_{k+1}\sigma(x)=\sum_{\epsilon\in\{0,1\}^{k+1}}\phi_\epsilon(s-\epsilon.x)\,.$$

We also have, for any $x\in\mathbb{Z}_N^k$,

$$\bigl|\Delta(f;x)^\wedge(\sigma(x))\bigr|^2=\sum_s\sum_y\Delta(f;x,y)(s)\omega^{-y\sigma(x)}\,.$$

Therefore,

$$\sum_{x\in\mathbb{Z}_N^k}\bigl|\Delta(f;x)^\wedge(\sigma(x))\bigr|^2=\sum_{x\in\mathbb{Z}_N^{k+1}}\sum_s\sum_{\epsilon\in\{0,1\}^{k+1}}C^{|\epsilon|}\bigl(f(s-\epsilon.x)\omega^{-\phi_\epsilon(s-\epsilon.x)}\bigr)\,.$$

By Lemma 17.1, we can find some ϵ such that the function $g(s)=f(s)\omega^{-\phi_\epsilon(s)}$ satisfies the inequality

$$\sum_{x\in\mathbb{Z}_N^{k+1}}\sum_s\Delta(g;x)(s)\geqslant\alpha N^{k+2}\,,$$

which is equivalent to the inequality we want. □

We must now deal with the fact that the results of the last two sections did not give us a k-linear function on the whole of $\mathbb{Z}_N^k$, so the above proposition cannot be applied directly. We shall use another very standard and well known lemma. It is a reflection of the fact that the set of half-spaces in $\mathbb{R}^k$ has VC-dimension at most k, but the proof is elementary and we very briefly sketch it. The next three lemmas are not essential to our main argument, as their purpose is to improve the bound coming from a trivial argument, when using the trivial bound would have a negligible effect on our eventual estimates.

LEMMA 17.4. *The number of distinct regions defined by a set of m hyperplanes in $\mathbb{R}^k$ is at most $\sum_{j=0}^k \binom{m}{j}$, with equality when the hyperplanes are in general position.*

Proof. Apply induction on m, by considering how many new regions are created when each new hyperplane is added to the arrangement. To calculate this, use induction on k. The result is trivial when $k = 1$. □

COROLLARY 17.5. *Given real numbers $\alpha_1, \dots, \alpha_k$, set $\alpha = (\alpha_1, \dots, \alpha_k)$ and define a function $f : \{0,1\}^k \to \mathbb{Z}$ by $f(\epsilon) = \lfloor \epsilon.\alpha \rfloor$. If r is an integer and the α_i are allowed to vary in the interval $(-r, r)$, then the number of distinct such functions that can result is at most 2^{2rk^3}.*

Proof. The possible values taken by f are the integers between $\lfloor -rk \rfloor$ and $\lfloor rk \rfloor$, and the set of ϵ such that $f(\epsilon) \leqslant j$ is the set of ϵ such that $\epsilon.\alpha < j+1$. Let us estimate how many distinct such sets can be obtained as α varies. Two real numbers α and α' give distinct sets if and only if there exists some ϵ such that $\epsilon.\alpha < j+1$ and $\epsilon.\alpha' \geqslant j+1$, that is, if and only if the hyperplane $\{\beta : \epsilon.\beta = j+1\}$ separates α from α'. There are 2^k different such hyperplanes, so the previous lemma tells us that the number of distinct sets of the given form is at most $k\binom{2^k}{k} \leqslant 2^{k^2}$. The function f is determined by the $2rk$ sets $\{\epsilon : f(\epsilon) \leqslant j\}$ with $\lfloor -k^2/2 \rfloor < j \leqslant \lfloor k^2/2 \rfloor$, so the result follows. □

COROLLARY 17.6. *Let $\alpha_0, \alpha_1, \dots, \alpha_k$ be real numbers, let $\alpha = (\alpha_1, \dots, \alpha_k)$ and define a function $f : \{0,1\}^k \to \mathbb{Z}_M$ by $f(\epsilon) = \lfloor \alpha_0 + \alpha.\epsilon \rfloor \pmod M$. If α_0 can be arbitrary and the α_i are allowed to vary in the interval $(-r, r)$, then the number of distinct such functions that can result is at most $M.2^{2r(k+1)^3}$.*

Proof. By Corollary 17.5, the number of functions that can result if the integer part of α_0 is $j \pmod M$ is at most $2^{2r(k+1)^3}$, since each such function

can be thought of as j added to the restriction of a function on $\{0,1\}^{k+1}$ of the given form (and reduced mod M). The result follows. □

The trivial bound in Corollary 17.5 is $k^{2^{k+1}}$, which gives a bound of $M.(k+1)^{2^{k+2}}$ in Corollary 17.6. As we commented above, this bound would be enough for our main result.

Before stating the next proposition, we define a concept which is similar to α-uniformity but designed for situations where we are given a multilinear function on a small domain. Let $f:\mathbb{Z}_N\to D$. Suppose that we can partition $\mathbb{Z}_N$ into mod-N arithmetic progressions $Q_1,\dots,Q_M$, each of length at most m, such that, defining $Q_if(s)$ to be $f(s)$ when $s\in Q_i$ and 0 otherwise, we have

$$\sum_{i=1}^M\sum_{x\in\mathbb{Z}_N^{k+1}}\sum_s\Delta(Q_if;x)(s)\leqslant\alpha m^{k+2}M\,.$$

We shall then say that f is *α-uniform of degree k with respect to the partition $Q_1,\dots,Q_M$*.

Notice that if $p\geqslant km$, then we can find an isomorphism γ from Q_i to an arithmetic progression of length $|Q_i|$ inside $\mathbb{Z}_p$ such that, defining a "copy" g of Q_if inside $\mathbb{Z}_p$ by setting $g(\gamma(s))=f(s)$ for $s\in Q_i$ and $g(t)=0$ for t not in the image of γ, we have

$$\sum_{x\in\mathbb{Z}_N^{k+1}}\sum_s\Delta(Q_if;x)(s)=\sum_{y\in\mathbb{Z}_p^{k+1}}\sum_t\Delta(g;y)(t)\,.$$

Hence, if

$$\sum_{x\in\mathbb{Z}_N^{k+1}}\sum_s\Delta(Q_if;x)(s)\geqslant\beta m^{k+2}\,,$$

we find that g is not $\beta(m/p)^{k+2}$-uniform (in $\mathbb{Z}_p$) of degree k.

PROPOSITION 17.7. *Let $f:\mathbb{Z}_N\to D$. Suppose that there is a product $P=P_1\times\cdots\times P_k$ of arithmetic progressions P_i of common difference d and odd length $m\leqslant N^{1/2}$, and a k-linear function $\sigma:P\to\mathbb{Z}_N$ such that*

$$\sum_{x\in P}\bigl|\Delta(f;x)^\wedge(\sigma(x))\bigr|^2\geqslant\alpha N^2m^k\,.$$

Then there exist a polynomial ϕ of degree at most $k+1$ and a partition of $\mathbb{Z}_N$ into mod-N arithmetic progressions $Q_1,\dots,Q_M$ of size at least $m/3k$ such that the function $g(x)=f(x)\omega^{-\phi(x)}$ is not $2^{-2(k+1)^3}\alpha$-uniform of degree k with respect to the partition $Q_1,\dots,Q_M$.

Proof. Without loss of generality $d=1$. Let $2l+1=m$ and let w be a real number such that $\lfloor l/k\rfloor-1<w\leqslant\lfloor l/k\rfloor$ and $M=N/w$ is an integer. For

$0 \leqslant j \leqslant M-1$ define Q_j to be the interval $\{x \in \mathbb{Z}_N : jw \leqslant x < (j+1)w\}$. Notice that the cardinality of Q_j is always $\lfloor l/k \rfloor - 1$ or $\lfloor l/k \rfloor$. Let a sequence $r = (r_1, \dots, r_k)$ be defined by $P_i = \{r_i - l, r_i - l + 1, \dots, r_i + l\}$ and let I be the interval $\{-l, -l+1, \dots, l\}$. Then any $x \in P$ can be written uniquely as $a + r$ for some $a \in I^k$. Given $\epsilon \in \{0,1\}^k$, we shall write f_ϵ for the function that takes s to $f(s - \epsilon.r)$. Let us also write $\tau(a)$ for $\sigma(a + r)$.

Then

$$\sum_{x \in P} |\Delta(f;x)^\wedge(\sigma(x))|^2 = \sum_{a \in I^k} \Big| \sum_s \omega^{-s\sigma(a+r)} \prod_\epsilon C^{|\epsilon|} f(s - \epsilon.a - \epsilon.r) \Big|^2$$
$$= \sum_{a \in I^k} \Big| \sum_s \omega^{-s\tau(a)} \prod_\epsilon C^{|\epsilon|} f_\epsilon(s - \epsilon.a) \Big|^2$$

where the products are over all ϵ in the set $\{0,1\}^k$.

If we now split each function f_ϵ up as $\sum_{j \in \mathbb{Z}_M} Q_j f_\epsilon$, this expression becomes

$$\sum_{a \in I^k} \Big| \sum_s \omega^{-s\tau(a)} \prod_{\epsilon \in \{0,1\}^k} \sum_{j=1}^{M} (Q_j C^{|\epsilon|} f_\epsilon)(s - \epsilon.a) \Big|^2 .$$

Interchanging the product over ϵ with the sum over j, we obtain

$$\sum_{a \in I^k} \Big| \sum_s \sum_j \omega^{-s\tau(a)} \prod_{\epsilon \in \{0,1\}^k} (Q_{j(\epsilon)} C^{|\epsilon|} f_\epsilon)(s - \epsilon.a) \Big|^2 ,$$

where now the sum over j stands for the sum over all functions $j : \{0,1\}^k \to \mathbb{Z}_M$. Let us estimate how many such functions can give rise to a non-zero contribution to the entire expression.

This we can do using Corollary 17.5. If $Q_{j(\epsilon)} f_\epsilon(s - \epsilon.a)$ is non-zero, then $s - \epsilon.a \in Q_{j(\epsilon)}$ which implies that $j(\epsilon)w \leqslant s - \epsilon.a < j(\epsilon)w$ and therefore that $j(\epsilon)$ is exactly the integer part of $w^{-1}s - \epsilon.(w^{-1}a)$. Since $-k-1 < w^{-1}a_i < k+1$ for every i, Corollary 17.5 implies that the number of functions j for which the product over ϵ can ever be non-zero is at most $M.2^{2(k+1)^4}$.

Let us define functions j_1 and j_2 (from $\{0,1\}^k$ to $\mathbb{Z}_M$) to be *equivalent* if they are translations of each other. Corollary 17.5 with $M = 1$ implies that the number of equivalence classes is at most $2^{k(k+1)^2}$. Let us call them $J_1, \dots, J_L$. By the Cauchy-Schwarz inequality, we can deduce from our calculations above that

$$\alpha N^2 m^k \leqslant L \sum_{r=1}^{L} \sum_{a \in I^k} \Big| \sum_s \sum_{j \in J_r} \omega^{-s\tau(a)} \prod_\epsilon (Q_{j(\epsilon)} C^{|\epsilon|} f_\epsilon)(s - \epsilon.a) \Big|^2 .$$

We can therefore find r such that, choosing some representative j of J_r, we have

$$L^{-2}\alpha N^2 m^k \leqslant \sum_{a\in I^k}\Big|\sum_s\sum_{i=1}^M \omega^{-s\tau(a)}\prod_\epsilon (Q_{j(\epsilon)+i}C^{|\epsilon|}f_\epsilon)(s-\epsilon.a)\Big|^2.$$

Applying the Cauchy-Schwarz inequality again, this is at most

$$M\sum_{i=1}^M\sum_{a\in I^k}\Big|\sum_s \omega^{-s\tau(a)}\prod_\epsilon (Q_{j(\epsilon)+i}C^{|\epsilon|}f_\epsilon)(s-\epsilon.a)\Big|^2.$$

Obviously this still exceeds $L^{-2}\alpha N^2 m^k$ if we replace the sum over $a\in I_k$ above by a sum over all of $\mathbb{Z}_N^k$. Expanding the modulus squared and substituting in the usual way, the resulting inequality can be rewritten

$$M\sum_{i=1}^M\sum_{a\in\mathbb{Z}_N^{k+1}}\sum_s \omega^{-\rho(a)}\prod_{\epsilon\in\{0,1\}^{k+1}}(Q_{j(\epsilon)+i}C^{|\epsilon|}f_\epsilon)(s-\epsilon.a)\geqslant L^{-2}\alpha N^2 m^k,$$

where $j(\epsilon)$ and f_ϵ now mean $j(\epsilon_1,\dots,\epsilon_k)$ and $f_{\epsilon_1,\dots,\epsilon_k}$ respectively, and $\rho(a)$ is defined to be $a_{k+1}\tau(a_1,\dots,a_k)$. Applying Lemma 17.1, we obtain for each $\epsilon\in\{0,1\}^{k+1}$ a polynomial ϕ_ϵ of degree at most $k+1$ in such a way that

$$\rho(a)=\sum_{\epsilon\in\{0,1\}^{k+1}}(-1)^{|\epsilon|}\phi_\epsilon(s-\epsilon.a)$$

for every $a\in\mathbb{Z}_N^{k+1}$. Using these, we can rewrite the inequality yet again, this time as

$$M\sum_{i=1}^M\sum_{a\in\mathbb{Z}_N^{k+1}}\sum_s\prod_{\epsilon\in\{0,1\}^{k+1}}(Q_{j(\epsilon)+i}C^{|\epsilon|}f_\epsilon)(s-\epsilon.a)\omega^{-\phi_\epsilon(s-\epsilon.a)}\geqslant L^{-2}\alpha N^2 m^k.$$

We shall now apply Lemma 3.8 to the functions $Q_{j(\epsilon)+i}C^{|\epsilon|}f_\epsilon$. By the AM-GM inequality, the lemma implies that, for every i, the sum

$$\sum_{a\in\mathbb{Z}_N^{k+1}}\sum_s\prod_{\epsilon\in\{0,1\}^{k+1}}(Q_{j(\epsilon)+i}C^{|\epsilon|}f_\epsilon)(s-\epsilon.a)\omega^{-\phi_\epsilon(s-\epsilon.a)}$$

is bounded above by the average over $\eta\in\{0,1\}^{k+1}$ of

$$\sum_{a\in\mathbb{Z}_N^{k+1}}\sum_s\prod_{\epsilon\in\{0,1\}^{k+1}}(Q_{j(\eta)+i}C^{|\epsilon|}f_\eta)(s-\epsilon.a)\omega^{-\phi_\eta(s-\epsilon.a)}.$$

It follows by an averaging argument that we may choose $\eta\in\{0,1\}^{k+1}$ such that, setting $g(s)=f_\eta(s)\omega^{-\phi_\eta(s)}$, we have

$$M\sum_{i=1}^M\sum_{a\in\mathbb{Z}_N^{k+1}}\sum_s\prod_{\epsilon\in\{0,1\}^{k+1}}(Q_{j(\epsilon)+i}C^{|\epsilon|}g)(s-\epsilon.a)\geqslant L^{-2}\alpha N^2 m^k,$$

and this may be rewritten

$$\sum_{i=1}^{M}\sum_{a\in\mathbb{Z}_N^{k+1}}\sum_{s}\prod_{\epsilon\in\{0,1\}^{k+1}}(C^{|\epsilon|}Q_ig)(s-\epsilon.a)\geqslant L^{-2}\alpha M^{-2}N^2m^kM\,,$$

or

$$\begin{aligned}\sum_{i=1}^{M}\sum_{a\in\mathbb{Z}_N^{k+1}}\sum_{s}\Delta(Q_ig;a)(s)&\geqslant L^{-2}\alpha M^{-2}N^2m^kM\\&\geqslant 2^{-2(k+1)^3}\alpha m^{k+2}M\,.\end{aligned}$$

This implies that the function g is not $2^{-2(k+1)^3}\alpha$-uniform with respect to the partition $Q_1,\dots,Q_M$.

This is not quite the statement of the proposition. To obtain it, recall that $f_\eta(s)=f(s-\eta.r)$. Therefore, the statement about g implies that the function $f(s)\omega^{-\phi_\eta(s+\eta.r)}$ is not $2^{-2(k+1)^3}\alpha$-uniform with respect to the partition $(Q_i+\eta.r)_{i=1}^m$. Since $\phi_\eta(s+\eta.r)$ is still a polynomial in s of degree at most $k+1$, the proposition is proved. □

18 Putting Everything Together

We are now ready for the proof of the main theorem. Indeed all we need to do is combine our earlier results in an obvious way. We shall divide the argument into two parts.

Theorem 18.1. *Let $\alpha\leqslant 1/2$ and let $A\subset\mathbb{Z}_N$ be a set which fails to be α-uniform of degree k. There exists a partition of $\mathbb{Z}_N$ into arithmetic progressions $P_1,\dots,P_M$ of average size at least $N^{\alpha^{2^{2^{k+10}}}}$ such that*

$$\sum_{j=1}^{M}\Big|\sum_{s\in P_j}f(s)\Big|\geqslant\alpha^{2^{2^{k+10}}}N\,.$$

Proof. The result will be proved by induction on k. First, Corollary 16.11 gives us a box $P\subset\mathbb{Z}_N^k$ of width at least $N^{(\alpha/2)^{2^{2^{k+8}}}}$ and a multilinear function $\mu:P\to\mathbb{Z}_N$ such that, for at least $(\alpha/2)^{2^{2^{k+8}}}|P|$ values of $(y_1,\dots,y_k)\in P$, we have

$$|\Delta(f;y_1,\dots,y_k)^\wedge(\mu(y_1,\dots,y_k))|\geqslant(\alpha/2)N\,.$$

Let $\beta=(\alpha^2/8)(\alpha/2)^{2^{2^{k+8}}}$ and let m be the largest odd number less than or equal to $N^{(\alpha/2)^{2^{2^{k+8}}}}$. Then the hypotheses for Proposition 17.7 are satisfied (with α replaced by β). We can therefore find a polynomial ϕ of

degree at most k and a partition of $\mathbb{Z}_N$ into mod-N arithmetic progressions $Q_1,\dots,Q_M$ of size l or $l+1$, where $l \geqslant m/3k$, such that the function $g(x) = f(x)\omega^{-\phi(x)}$ is not $2^{-2(k+3)^3}\beta$-uniform of degree $k-1$ with respect to the partition $Q_1,\dots,Q_M$.

For each i, define a (non-negative real) number β_i by the equation

$$\sum_{x\in\mathbb{Z}_N^k}\sum_s \Delta(Q_i g;x)(s) = \beta_i l^{k+1}\,.$$

Since the sets Q_i all have approximately the same size, the average value of β_i is at least $2^{-2(k+3)^3-1}\beta$. It follows that there is a set I of cardinality at least $2^{-2(k+4)^3}\beta M$ such that, for every $i\in I$, $\beta_i \geqslant 2^{-2(k+4)^3}\beta$. Let us now fix some $i\in I$.

As described in the remarks before Proposition 17.7, we may associate with $Q_i g$ an "isomorphic" function $h_i:\mathbb{Z}_p\to\mathbb{C}$ which fails to be $(\beta_i/2k)$-uniform of degree $k-1$. When this is done, the mod-N arithmetic progression Q_i corresponds to an interval of integers in $\mathbb{Z}_p$. By our inductive hypothesis, we can partition $\mathbb{Z}_p$ into proper arithmetic progressions $R_{i1},\dots,R_{iM_i}$ of average size at least $p^{(\beta_i/2k)^{2^{2^{k+9}}}}$ in such a way that

$$\sum_{j=1}^{M_i}\Big|\sum_{s\in R_{ij}} h_i(s)\Big| \geqslant (\beta_i/2k)^{2^{2^{k+9}}}p\,.$$

It follows that Q_i can be partitioned into mod-N arithmetic progressions $S_{i1},\dots,S_{iM_i}$ of average size at least $r_i = (2k)^{-1}p^{(\beta_i/2k)^{2^{2^{k+9}}}}$ such that

$$\sum_{j=1}^{M_i}\Big|\sum_{s\in S_{ij}} g(s)\Big| \geqslant (\beta_i/2k)^{2^{2^{k+9}}}|Q_i|\,.$$

The mod-N progressions S_{ij} with $i\in I$, together with those Q_i for which $i\notin I$, partition $\mathbb{Z}_N$. From the way we chose I, the average size of a cell in this partition is at least $r = (2k)^{-1}p^{(2^{-(k+4)^3}\beta/2k)^{2^{2^{k+9}}}}$. By Lemma 5.13 we can find a refinement of this partition into proper arithmetic progressions of average size at least $r^{1/2}/4$. Let us call these progressions $T_1,\dots,T_L$. Since $\sum_{i\in I}|Q_i| \geqslant 2^{-2(k+4)^3}\beta N$ we have the inequality

$$\sum_{j=1}^{L}\Big|\sum_{s\in T_j} g(s)\Big| \geqslant (2^{-2(k+4)^3}\beta/2k)^{2^{2^{k+9}}}2^{-2(k+4)^3}\beta N\,.$$

Let $\gamma = (2^{-2(k+4)^3}\beta/2k)^{2^{2^{k+9}}}2^{-2(k+4)^3}\beta$. We may now apply Lemma 5.14 to find a refinement of $T_1,\dots,T_L$ into arithmetic progressions $U_1,\dots,U_H$

such that $H \leqslant CL^{1/K}N^{1-1/K}$ and

$$\sum_{h=1}^{H}\Big|\sum_{s\in U_h} f(s)\Big| \geqslant \gamma N/2\,.$$

All that remains is to check that $H^{-1}N \geqslant N^{\alpha^{2^{2^{k+10}}}}$ and that $\gamma/2 \geqslant \alpha^{2^{2^{k+10}}}$. These are easy exercises for the reader (easy because $2^{2^{k+10}}$ is so much bigger than $2^{2^{k+9}}$ that estimates can be incredibly crude). □

For the statement of our main theorem we shall use the notation $a \uparrow b$ for a^b, with the obvious convention for bracketing, so that for example $a \uparrow b \uparrow c$ stands for $a \uparrow (b \uparrow c)$.

Theorem 18.2. *Let $0 < \delta \leqslant 1/2$, let k be a positive integer, let $N \geqslant 2 \uparrow 2 \uparrow \delta^{-1} \uparrow 2 \uparrow 2 \uparrow (k+9)$ and let A be a subset of the set $\{1, 2, \dots, N\}$ of size at least δN. Then A contains an arithmetic progression of length k.*

Proof. It is not hard to check that $N \geqslant 32k^2\delta^{-k}$. Therefore, Corollary 3.6 implies the result when A is $(\delta/2)^{k2^k}$-uniform of degree $k-2$.

Let $\alpha = (\delta/2)^{k2^k}$. If A is *not* α-uniform of degree $k-2$ then by Theorem 18.1 and Lemma 5.15 there is an arithmetic progression P of size at least $N^{\alpha^{2^{2^{k+8}}}}$ such that $|A \cap P| \geqslant (\delta + \alpha^{2^{2^{k+8}}})|P|$. We may then repeat the argument with the new density. After at most $\alpha^{-2^{2^{k+8}}}$ repetitions, we find an arithmetic progression of length k, as long as N is large enough. Since at each repetition we are raising N to a power at least as big as $\alpha^{2^{2^{k+8}}}$ and the argument works as long as $N \geqslant 32k^2\delta^{-k}$, a sufficient condition on the original N is that

$$N \uparrow \big(\alpha \uparrow 2 \uparrow 2 \uparrow (k+8)\big) \uparrow \big(\alpha^{-1} \uparrow 2 \uparrow 2 \uparrow (k+8)\big) \geqslant 32k^2\delta^{-k}\,.$$

It is not hard to check that this condition is satisfied when $N \geqslant 2 \uparrow 2 \uparrow \delta^{-1} \uparrow 2 \uparrow 2 \uparrow (k+9)$, and the theorem is proved. □

Notice that what matters for the bounds in the above proof is the number of times the iteration is performed. The fact that at each iteration we raise N to a very small power makes hardly any further difference.

COROLLARY 18.7. *Let k be a positive integer and let $N \geqslant 2 \uparrow 2 \uparrow 2 \uparrow 2 \uparrow 2 \uparrow (k+9)$. Then however the set $\{1, 2, \dots, N\}$ is coloured with two colours, there will be a monochromatic arithmetic progression of length k.* □

Ron Graham has conjectured in several places (see e.g. [GRS]) that the function $M(k,2)$ is bounded above by a tower of twos of height k. Corollary 18.7 proves this conjecture for $k \geqslant 9$, and indeed gives a much stronger bound. It looks as though more would be needed to prove it for $k = 7$ (for example) than merely tidying up our proof. For $k \leqslant 5$, the exact values of $M(k,2)$ are known and satisfy the conjecture.

Concluding Remarks and Acknowledgements

The arguments of this paper leave open many interesting questions. The most obvious one is whether the multidimensional version of Szemerédi's theorem follows from similar arguments. There is not even a good bound in the case of three points in a triangle. (The precise statement is that, for sufficiently large N, every subset of $[N]^2$ of size at least δN^2 contains a triple of the form $\{(a,b),(a+d,b),(a,b+d)\}$. Very recently, Jozsef Solymosi sent me an argument that proves this using a lemma of Ruzsa and Szemerédi, which itself uses Szemerédi's regularity lemma. Thus, at least a tower-type bound can be proved for this problem.) It would of course also be extremely interesting to have quantitative versions of the results of [BL] and [FK] mentioned in the introduction.

Some of the ideas in this proof turn out not to be new. In particular, the content of §4, that is, the relevance of exponentials in polynomials as well as the fact that they are not sufficient, was discovered in an ergodic-theoretic context, independently and earlier by Kazhdan in recent unpublished work. In general, there seem to be very interesting connections between the methods of this paper and a new ergodic-theoretic approach that is not yet complete.

A more obvious connection with the ergodic methods is that the arguments of §3 closely resemble the arguments used by Furstenberg for the case of weak-mixing measure-preserving dynamical systems. His argument is based on the fact that a system that is weak-mixing is sufficiently random to work, while one that is not can be decomposed in a useful way. This appears to be analogous in some way to the idea here of passing from a non-uniform set to a denser subset.

I am very grateful to Béla Bollobás for encouraging me to continue working on Szemerédi's theorem when an earlier attempt at proving it collapsed, and to Vitali Milman for making sure that I eventually finished this paper.

References

[BS] A. Balog, E. Szemerédi, A statistical theorem of set addition, Combinatorica 14 (1994), 263–268.

[Be] F.A. Behrend, On sets of integers which contain no three in arithmetic progression, Proc. Nat. Acad. Sci. 23 (1946), 331–332.

[BerL] V. Bergelson, A. Leibman, Polynomial extensions of van der Waerden's and Szemerédi's theorems, J. Amer. Math. Soc. 9 (1996), 725–753.

[Bi] Y. Bilu, Structure of sets with small sumset, in "Structure Theory of Set Addition", Astérisque 258 (1999), 77–108.

[Bo] N.N. Bogolyubov, Sur quelques propriétés arithmétiques des presque-périodes, Ann. Chaire Math. Phys. Kiev 4 (1939), 185–194.

[Bou] J. Bourgain, On triples in arithmetic progression, Geom. Funct. Anal. 9 (1999), 968–984.

[CGW] F.R.K. Chung, R.L. Graham, R.M. Wilson, Quasi-random graphs, Combinatorica 9 (1989), 345–362.

[ET] P. Erdős, P. Turán, On some sequences of integers, J. London Math. Soc. 11 (1936), 261–264.

[F1] G.R. Freiman, Foundations of a Structural Theory of Set Addition, Kazan Gos. Ped. Inst., Kazan, 1966 (in Russian).

[F2] G.R. Freiman, Foundations of a Structural Theory of Set Addition, Translations of Mathematical Monographs 37, Amer. Math. Soc., Providence, RI, USA, 1973.

[Fu] H. Furstenberg, Ergodic behaviour of diagonal measures and a theorem of Szemerédi on arithmetic progressions, J. Analyse Math. 31 (1977), 204–256.

[FuK] H. Furstenberg, Y. Katznelson, A density version of the Hales-Jewett theorem, J. d'Analyse Math. 57 (1991), 64–119.

[FuKO] H. Furstenberg, Y. Katznelson, D. Ornstein, The ergodic theoretical proof of Szemerédi's theorem, Bull. Amer. Math. Soc. 7 (1982), 527–552.

[G1] W.T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, Geometric And Functional Analysis 7 (1997), 322–337.

[G2] W.T. Gowers, A new proof of Szemerédi's theorem for arithmetic progressions of length four, Geometric And Functional Analysis 8 (1998), 529–551.

[GRS] R.L. Graham, B.L. Rothschild, J.H. Spencer, Ramsey Theory (2nd ed.), Wiley Interscience 1990.

[H] D.R. Heath-Brown, Integer sets containing no arithmetic progressions, J. London Math. Soc. (2) 35 (1987), 385–394.

[N] M.B. Nathanson, Additive Number Theory: Inverse Problems and the Geometry of Sumsets, Graduate Texts in Mathematics 165, Springer-Verlag, 1996.

[P] H. Plünnecke, Eigenschaften und Abschätzungen von Wirkingsfunktionen, Vol. 22, Berichte der Gesellschaft für Mathematik und Datenverarbeitung, Bonn, 1969.

[R1] K.F. Roth, On certain sets of integers, J. London Math. Soc. 28 (1953), 245–252.

[R2] K.F. Roth, Irregularities of sequences relative to arithmetic progressions, IV, Period. Math. Hungar. 2 (1972), 301–326.

[Ru1] I.Z. Ruzsa, Arithmetic progressions and the number of sums, Period. Math. Hungar. 25 (1992), 105–111.

[Ru2] I.Z. Ruzsa, An application of graph theory to additive number theory, Scientia, Ser. A 3 (1989), 97–109.

[Ru3] I. Ruzsa, Generalized arithmetic progressions and sumsets, Acta Math. Hungar. 65 (1994), 379–388.

[S] S. Shelah, Primitive recursive bounds for van der Waerden numbers, J. Amer. Math. Soc. 1 (1988), 683–697.

[Sz1] E. Szemerédi, On sets of integers containing no four elements in arithmetic progression, Acta Math. Acad. Sci. Hungar. 20 (1969), 89–104.

[Sz2] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, Acta Arith. 27 (1975), 299–345.

[Sz3] E. Szemerédi, Integer sets containing no arithmetic progressions, Acta Math. Hungar. 56 (1990), 155–158.

[V] R.C. Vaughan, The Hardy-Littlewood Method (2nd ed.), Cambridge Tracts in Mathematics 125, CUP 1997.

[W] H. Weyl, Über die Gleichverteilung von Zahlen mod Eins, Math. Annalen 77 (1913), 313–352.

W.T. Gowers, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK

Submitted: March 2000
Final version: March 2001