

AN INVERSE THEOREM FOR THE GOWERS $U^3(G)$ NORM

BEN GREEN^{1*} AND TERENCE TAO²

¹*Department of Mathematics, University of Bristol, University Walk,
Bristol BS8 1TW, UK (b.j.green@bristol.ac.uk)*

²*Department of Mathematics, University of California, Los Angeles,
CA 90095-1555, USA (tao@math.ucla.edu)*

(Received 11 March 2005)

Abstract There has been much recent progress in the study of arithmetic progressions in various sets, such as dense subsets of the integers or of the primes. One key tool in these developments has been the sequence of Gowers uniformity norms $U^d(G)$, $d = 1, 2, 3, \dots$, on a finite additive group G ; in particular, to detect arithmetic progressions of length k in G it is important to know under what circumstances the $U^{k-1}(G)$ norm can be large.

The $U^1(G)$ norm is trivial, and the $U^2(G)$ norm can be easily described in terms of the Fourier transform. In this paper we systematically study the $U^3(G)$ norm, defined for any function $f : G \rightarrow \mathbb{C}$ on a finite additive group G by the formula

$$\|f\|_{U^3(G)} := |G|^{-4} \sum_{x,a,b,c \in G} (f(x)\overline{f(x+a)}f(x+b)\overline{f(x+c)}f(x+a+b) \\ \times f(x+b+c)\overline{f(x+c+a)}\overline{f(x+a+b+c)})^{1/8}.$$

We give an inverse theorem for the $U^3(G)$ norm on an arbitrary group G . In the finite-field case $G = \mathbb{F}_5^n$ we show that a bounded function $f : G \rightarrow \mathbb{C}$ has large $U^3(G)$ norm if and only if it has a large inner product with a function $e(\phi)$, where $e(x) := e^{2\pi i x}$ and $\phi : \mathbb{F}_5^n \rightarrow \mathbb{R}/\mathbb{Z}$ is a quadratic phase function. In a general G the statement is more complicated: the phase ϕ is quadratic only locally on a Bohr neighbourhood in G .

As an application we extend Gowers's proof of Szemerédi's theorem for progressions of length four to arbitrary abelian G . More precisely, writing $r_4(G)$ for the size of the largest $A \subseteq G$ which does not contain a progression of length four, we prove that

$$r_4(G) \ll |G|(\log \log |G|)^{-c},$$

where c is an absolute constant.

We also discuss links between our ideas and recent results of Host, Kra and Ziegler in ergodic theory.

In future papers we will apply variants of our inverse theorems to obtain an asymptotic for the number of quadruples $p_1 < p_2 < p_3 < p_4 \leq N$ of primes in arithmetic progression, and to obtain significantly stronger bounds for $r_4(G)$.

Keywords: Gowers norm; inverse theorem; nilsequence

2000 *Mathematics subject classification:* Primary 11Bxx

* Present address: Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK (b.j.green@dpmms.cam.ac.uk).

1. Background and motivation

A famous and deep theorem of Szemerédi asserts that any set of integers of positive upper density contains arbitrarily long arithmetic progressions. More precisely, we have the following theorem.

Theorem 1.1 (Szemerédi’s theorem, infinitary version [60]). *Let A be a subset of the integers \mathbb{Z} whose upper density $\limsup_{N \rightarrow \infty} (2N + 1)^{-1} |A \cap [-N, N]|$ is strictly positive. Then, for any $k \geq 1$, the set A contains infinitely many arithmetic progressions $\{a, a + r, \dots, a + (k - 1)r\}$, $r \neq 0$, of length k .*

The first non-trivial case of this theorem is when $k = 3$, which was treated by Roth [56] using a Fourier-analytic argument. The case of higher k was more resistant to Fourier-analytic methods, and the first full proof of this theorem was achieved by Szemerédi [60] using combinatorial methods. Later, Furstenberg [18, 24] introduced an ergodic-theoretic proof of this theorem. More recently, Gowers [27] gave a proof that was both combinatorial and Fourier-analytic in nature, and which is substantially closer in spirit to Roth’s original argument than the other proofs. Even more recently there have been a number of other proofs of this theorem by other methods, such as hypergraph regularity [28, 48–50, 53–55] or ‘discrete ergodic theory’ [62]. This theorem and its various proofs have in turn generated many other mathematical developments. For instance, in [34], we were able to apply Theorem 1.1 to demonstrate that the primes contain arbitrarily long arithmetic progressions.

In this paper we shall be interested primarily in the Fourier-analytic approach to this theorem, specifically in the $k = 4$ case, which was treated separately by Gowers in [25] and then again in [27]. This latter paper will be our key reference. However, as we shall see later there are some strong connections between this approach and the ergodic one, especially after the work on characteristic factors by Host and Kra [41, 43] and Ziegler [63, 64] and on the connection to nilsequences by Bergelson et al. [5]. Before we give our main new results, however, we first give some further historical background and motivation.

Gowers’s proof of the full Szemerédi theorem in [27] is quite lengthy and involves many deep new ideas. However, it is possible to split it up into a number of simpler steps, all but one of which are straightforward. First, it is easy to show that for any fixed k , Theorem 1.1 is equivalent to the following finitary version.

Theorem 1.2 (Szemerédi’s theorem, finitary version [60]). *Let $\delta > 0$ and $k \geq 1$. Then there exists an integer $N_0 = N_0(\delta, k)$ such that whenever $N \geq N_0$ and $A \subseteq [1, N]$ is such that $|A|/[1, N] \geq \delta$, then A contains at least one proper arithmetic progression of length k .*

The next observation, due to Roth, is that one can hope to prove this theorem by downwards inducting on the density parameter δ (the case $\delta \geq 1$ being trivial or vacuous). In particular, for any fixed k , Theorem 1.2 is equivalent to the following assertion.

Theorem 1.3 (lack of progressions implies density increment). *Let $\delta > 0$ and $k \geq 1$. Let $N \geq 1$ and let $A \subseteq [1, N]$ be such that $|A|/[1, N] \geq \delta$ and such that A contains*

no proper arithmetic progressions of length k . Then, if N is sufficiently large depending on k and δ , there exists an arithmetic progression $P \subseteq [1, N]$ with $|P| \geq \omega(N, \delta)$ for some function $\omega(N, \delta)$ of N that goes to infinity as $N \rightarrow \infty$ for each fixed δ , such that we have the density increment $|A \cap P|/|P| \geq \delta + c(\delta)$, where $c(\delta) > 0$ is a function of δ that is bounded away from zero whenever δ is bounded away from zero.

The deduction of Theorem 1.2 from Theorem 1.3 is a straightforward induction argument. For details of arguments of this type any of [25–27, 31, 56] may be consulted, or indeed §§ 7 or 11 of this paper. Of course, the final bound $N_0(\delta, k)$ obtained in Theorem 1.2 will depend on the explicit bounds $\omega(N, \delta)$, $c(\delta)$ obtained in Theorem 1.3. In Roth's $k = 3$ argument in [56], $c(\delta)$ was roughly δ^2 and $\omega(N, \delta)$ was roughly $N^{1/2}$, which led to a final bound of the form $N_0(\delta, 3) \leq \exp(\exp(C/\delta))$. In Gowers's extension of Roth's argument in [27], $c(\delta)$ was roughly δ^{C_k} and $\omega(N, \delta)$ was roughly N^{c_k} for some $C_k, c_k > 0$ depending only on k , which led to a final bound of the form $N_0(\delta, k) \leq \exp(\exp(C_k/\delta^{C_k}))$ (see [27] for a more precise statement). These are the best known bounds for $N_0(\delta, k)$ except in the $k = 3$ case, where the current record is $N_0(\delta, 3) \leq (C/\delta)^{C/\delta^2}$, due to Bourgain [10].

The next step is to pass from the interval $[1, N]$ to a cyclic group $\mathbb{Z}/N\mathbb{Z}$ for some prime N . Indeed, by using Bertrand's postulate* and a simple covering argument to split progressions in $\mathbb{Z}/N\mathbb{Z}$ into progressions in $[1, N]$, one can show that Theorem 1.3 is in turn equivalent for each fixed k (up to minor changes in the bounds $\omega(N, \delta)$ and $c(\delta)$) to the following statement.

Theorem 1.4 (lack of progressions implies density increment). *Let $\delta > 0$ and $k \geq 1$. Let $N \geq 1$ be a prime, and let Q be a proper progression in $\mathbb{Z}/N\mathbb{Z}$ such that $|Q| \geq c_0 N$ for some $0 < c_0 \leq 1$. Let $A \subseteq Q$ be such that $|A| \geq \delta N$, and such that A contains no proper arithmetic progressions of length k . Then, if N is sufficiently large depending on k and δ , there exists a proper arithmetic progression $P \subseteq \mathbb{Z}/N\mathbb{Z}$ with $|P| \geq \omega(N, \delta, c_0, k)$ for some function $\omega(N, \delta, c_0, k)$ of N that goes to infinity as $N \rightarrow \infty$ for each fixed δ, c_0, k , such that we have the density increment*

$$\frac{|A \cap P|}{|P|} \geq \frac{|A \cap Q|}{|Q|} + c(\delta, c_0, k),$$

where $c(\delta, c_0, k) > 0$ is bounded away from zero whenever δ, c_0 are bounded away from zero and k is fixed.

The deduction of Theorem 1.3 from Theorem 1.4 is not difficult (see [25, 27, 56]). Of course, it remains to prove Theorem 1.4. This was achieved in the $k = 3$ case by Roth using Fourier-analytic methods. To extend these arguments to the case of higher k , Gowers introduced a collection of tools which form part of a theory which might be termed 'higher-order Fourier analysis' for reasons which will become clear later. In particular, handling the $k = 4$ case required 'quadratic Fourier analysis'.

While Gowers's original argument takes place in a cyclic group $\mathbb{Z}/N\mathbb{Z}$ of prime order, we will work in the more general setting of an arbitrary finite additive group. This might

* That is, there is always a prime between X and $2X$.

seem unnecessary but is consistent with what we call the *finite-field philosophy*. This is the observation that many questions concerning the integers $\{1, \dots, N\}$ or the cyclic group $\mathbb{Z}/N\mathbb{Z}$ may be asked very naturally for an arbitrary finite abelian group G , and they may be answered there by modifying the proof for $\mathbb{Z}/N\mathbb{Z}$ in a straightforward way. Thus it is often the case that the passage

$$\mathbb{Z}/N\mathbb{Z} \xrightarrow{\text{generalization}} G$$

is rather straightforward.

However, it may be that the question is significantly easier to answer when G is some specific group, typically a vector space over a finite field such as \mathbb{F}_2^n , \mathbb{F}_3^n or \mathbb{F}_5^n . This observation was made in such papers as [46, 58]. In this paper we will take a particular interest in \mathbb{F}_5^n since this is the smallest characteristic field for which arithmetic progressions of length four are a sensible thing to discuss. Now the passage

$$\mathbb{F}_5^n \xrightarrow{\text{generalization}} G$$

might not be at all easy. However, in attempting such a route one has split the problem into two presumably easier subproblems, and furthermore there is now a library of tools available for effecting the generalization. This started with the work of Bourgain [10] (though he did not phrase it this way) and has continued with various works such as [30, 32]. The present paper, particularly §§ 8 and 9, is another example in this vein. For a longer discussion of the finite-field philosophy, see [31].

Definition 1.5 (additive groups). Define an *additive group* to be a group $G = (G, +)$ with a commutative group operation $+$; if $x \in G$ and $n \in \mathbb{Z}$, we can define the product $nx \in G$ in the usual manner. If $f : G \rightarrow H$ is a function from one additive group to another, and $h \in H$, we define the shift* operator T^h applied to f by the formula $T^h f(x) := f(x + h)$, and the difference operator $h \cdot \nabla := T^h - 1$ applied to f by the formula $(h \cdot \nabla)f(x) := f(x + h) - f(x)$. We extend these definitions to functions of several variables by subscripting the variable to which the operator is applied: thus, for instance, if $f(x, y)$ is a function of two variables, we define $T_x^h f(x, y) = f(x + h, y)$ and $h \cdot \nabla_x f(x, y) = f(x + h, y) - f(x, y)$ if x, h range inside an additive group G , and we do the same thing for the y variable.

Remark 1.6. Throughout the paper, we will write $N := |G|$ for the cardinality of G .

Remark 1.7. The notation above is of course designed to mimic that of several-variable calculus. We caution, however, that we do not assign any independent meaning to the symbol ∇ , unless it is prepended with a shift h to create a difference operator $h \cdot \nabla$, which is of course a discrete analogue of a directional derivative operator.

We now introduce a multilinear form $\Lambda_k(f_1, \dots, f_k)$ which is useful for counting arithmetic progressions. Here, and throughout the paper it is convenient to adopt the notation of conditional expectation, which allows one to hide some distracting normalizing factors such as $1/N$ in our arguments. Thus if $f : G \rightarrow \mathbb{C}$ is a complex-valued

* This ‘ergodic’ notation corresponds to the backwards shift $T^h x := x - h$ on the underlying group G . We will discuss further connections with ergodic theory in § 12.

function on a finite set G and if $B \subseteq G$ is a non-empty subset of G , we will use $\mathbb{E}_{x \in B} f(x) := (1/|B|) \sum_{x \in B} f(x)$ to denote the average of f over B . We will abbreviate $\mathbb{E}_{x \in G} f(x)$ as $\mathbb{E}(f)$ when the domain G of f is clear from the context.

Now if G is a finite additive group and $f_0, \dots, f_{k-1} : G \rightarrow \mathbb{C}$ are complex-valued functions, we define the k -linear form $\Lambda_k(f_0, \dots, f_{k-1}) \in \mathbb{C}$ by

$$\Lambda_k(f_0, \dots, f_{k-1}) := \mathbb{E}_{x, r \in G} f_0(x) T^r f_1(x) \cdots T^{(k-1)r} f_{k-1}(x).$$

Observe that if $A \subseteq G$ and $f_0 = \dots = f_{k-1} = 1_A$, where $1_A : G \rightarrow \{0, 1\}$ denotes the indicator function of A , then $\Lambda_k(1_A, \dots, 1_A)$ is just the number of progressions of length k (including those with common difference 0), divided by the normalizing factor of N^2 . In particular, if $(N, (k-1)!) = 1$ and A contains no proper progressions of length k , then we see that $\Lambda_k(1_A, \dots, 1_A) = |A|/N^2$, which will be quite small when N is large.

It is thus of interest to determine under what conditions $\Lambda_k(1_A, \dots, 1_A)$ is small or large. To this end, Gowers introduced (what are now known as) the *Gowers uniformity norms* $\|f\|_{U^d(G)}$ for any complex function $f : G \rightarrow \mathbb{C}$, whose definition we now recall.

Definition 1.8 (Gowers uniformity norm). Let $d \geq 0$ and let $f : G \rightarrow \mathbb{C}$ be a function. We define the Gowers uniformity norm $\|f\|_{U^d(G)} \geq 0$ of f to be the quantity

$$\|f\|_{U^d(G)} := \left(\mathbb{E}_{x \in G, h \in G^d} \prod_{\omega \in \{0,1\}^d} \mathcal{C}^{|\omega|} T^{\omega \cdot h} f(x) \right)^{1/2^d},$$

where $\omega = (\omega_1, \dots, \omega_d)$, $h = (h_1, \dots, h_d)$, $\omega \cdot h := \omega_1 h_1 + \dots + \omega_d h_d$, $|\omega| := \omega_1 + \dots + \omega_d$, and \mathcal{C} is the conjugation operator $\mathcal{C}f(x) := \overline{f(x)}$.

Remark 1.9. An equivalent definition of the $U^d(G)$ norms is given by the recursive formulae

$$\|f\|_{U^0(G)} = \mathbb{E}(f), \quad \|f\|_{U^1(G)} = |\mathbb{E}(f)|, \quad \|f\|_{U^d} := (\mathbb{E}_{h \in G} \|T^h f \bar{f}\|_{U^{d-1}(G)}^{2^{d-1}})^{1/2^d} \tag{1.1}$$

for all $d \geq 1$.

Remark 1.10. A configuration of the form $(x + \omega \cdot h)_{\omega \in \{0,1\}^d}$ is called a *cube of dimension d* . Thus $\|f\|_{U^d(G)}^{2^d}$ is a weighted average of f over cubes; for instance, $\|1_A\|_{U^d(G)}^{2^d}$ is equal to the number of cubes contained in A , divided by the normalizing factor of N^{d+1} . The cases $d = 0, 1$ are rather degenerate, and indeed $U^d(G)$ is not a norm in these cases. However, for $d > 1$, one can show that $\|\cdot\|_{U^d(G)}$ is indeed a norm, i.e. it is homogeneous, non-negative, non-degenerate, and obeys the triangle inequality (see [27, Lemma 3.9]). These norms have also appeared recently in ergodic theory (see, for example, [43]) and (together with the dual norms $U^d(G)^*$) played a key role in [34]. It thus seems of interest to study these norms more systematically; the results here can be viewed as a step in that direction.

We will study these norms in detail later, but for now let us give an example to illustrate what they are trying to capture. Suppose f has the form $f(x) := e(\phi(x))$ for some phase

function $\phi : G \rightarrow \mathbb{R}/\mathbb{Z}$, where $e : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{C}$ is the exponential map $e(x) := e^{2\pi i x}$. Then a simple calculation shows that

$$\|f\|_{U^d}^{2^d} = \mathbb{E}_{x, h_1, \dots, h_d \in G} e((h_1 \cdot \nabla_x) \cdots (h_d \cdot \nabla_x) \phi(x)).$$

Thus the U^d norm is in some sense measuring the oscillation present in the d th ‘derivative’ of the phase. In particular, we expect the U^d norm to be large if the phase behaves like a ‘polynomial’ of degree $d - 1$ or lower, but small if the phase is behaving like a polynomial of degree d or higher.

We observe that as an immediate consequence of (1.1) and induction we have the monotonicity property

$$\|f\|_{U^d(G)} \leq \|f\|_{U^{d+1}(G)} \quad \text{for } d = 0, 1, 2, \dots \quad (1.2)$$

The relevance of the Gowers uniformity norms to arithmetic progressions lies in the following result, which was stated explicitly in [27, Theorem 3.2] (in the case of cyclic groups, $G = \mathbb{Z}/N\mathbb{Z}$) but has been implicit in the ergodic-theory literature for some time. Write $\mathcal{D} := \{z \in \mathbb{C} : |z| \leq 1\}$ for the unit disc.

Proposition 1.11 (generalized von Neumann theorem). *Let G be a finite abelian group with $(N, (k - 1)!) = 1$. Let $f_0, \dots, f_{k-1} : G \rightarrow \mathcal{D}$ be functions. Then we have*

$$|A_k(f_0, \dots, f_{k-1})| \leq \min_{1 \leq j \leq k} \|f_j\|_{U^{k-1}(G)}.$$

It is instructive to continue with the phase example given earlier. If $f_j = e(\phi_j)$, then

$$A_k(f_0, \dots, f_{k-1}) = \mathbb{E}_{x, r \in G} e(\phi_0(x) + \phi_1(x+r) + \cdots + \phi_{k-1}(x + (k-1)r)).$$

Thus $A_k(f_0, \dots, f_{k-1})$ is measuring the oscillation present in the expression $\phi_0(x) + \phi_1(x+r) + \cdots + \phi_{k-1}(x + (k-1)r)$. Proposition 1.11 can then be viewed as a statement that if this expression does not oscillate, then neither do the expressions $(h_1 \cdot \nabla_x) \cdots (h_{k-1} \cdot \nabla_x) \phi_j(x)$ for any $1 \leq j \leq k$. Note that such a fact morally follows by ‘differentiating’ the expression $\phi_0(x) + \phi_1(x+r) + \cdots + \phi_k(x + (k-1)r)$ in $k - 1$ different directions to eliminate all but one of the terms in this series. For completeness we give a proof of this proposition in §4.

Corollary 1.12 (lack of progressions implies large uniformity norm [27]). *Let $k \geq 3$, let G be a finite additive group with $(N, (k - 1)!) = 1$ and let $A \subseteq G$, $|A| = \alpha N$, be a non-empty set such that A has no proper arithmetic progressions of length k . If $N \geq 2/\alpha^{k-1}$, then we have $\|1_A - \alpha\|_{U^{k-1}(G)} \geq 2^{-k-1} \alpha^{k-1}$.*

More generally, let P be a proper arithmetic progression in G such that $|P| \geq c_0 N$. Let $A \subseteq P$, $|A| = \alpha|P|$, be a non-empty set which contains no proper arithmetic progressions. If $N > N_0(c_0, k, \alpha)$, then we have $\|1_A - \alpha 1_P\|_{U^{k-1}(G)} \geq c(c_0, \alpha, k) > 0$, where the quantity $c(c_0, \alpha, k)$ stays bounded away from zero when c_0, α are bounded away from zero and k is fixed.

Proof. We begin with the first claim. Since $1_A = \alpha + (1_A - \alpha)$, we can split the sum $\Lambda_k(1_A, \dots, 1_A)$ into 2^k expressions, one of which is $\Lambda_k(\alpha, \dots, \alpha)$ and the other 2^{k-1} of which can be bounded in magnitude by $\|1_A - \alpha\|_{U^{k-1}(G)}$ thanks to Proposition 1.11. In particular we conclude that

$$|\Lambda_k(\alpha, \dots, \alpha) - \Lambda_k(1_A, \dots, 1_A)| \leq 2^k \|1_A - \alpha\|_{U^{k-1}(G)}.$$

But clearly $\Lambda_k(\alpha, \dots, \alpha) = \alpha^k$, while since A has no proper arithmetic progressions we have $\Lambda_k(1_A, \dots, 1_A) = |A|/N^2 = \alpha/N \leq \alpha^k/2$. The first claim follows.

The second claim proceeds similarly but is based on the decomposition $1_A = \alpha 1_P + (1_A - \alpha 1_P)$, and the observation that $\Lambda_k(1_P, \dots, 1_P) \geq c(c_0, k) > 0$ for some positive quantity $c(c_0, k)$ depending on c_0 and k ; we leave the details to the reader. \square

Comparing this proposition with Theorem 1.4, we thus see that in order to prove Szemerédi’s theorem for a fixed k , it suffices to prove the following theorem.

Theorem 1.13 (large uniformity norm implies density increment [27]). *Let $\eta > 0$ and $k \geq 3$. Let $G = \mathbb{Z}/N\mathbb{Z}$ be a cyclic group of prime order and let $f : G \rightarrow \mathcal{D}$ be a real-valued bounded function such that $\mathbb{E}(f) = 0$ and $\|f\|_{U^{k-1}(G)} \geq \eta$. Then, if $N > N_0(k, \eta)$, there exists a proper arithmetic progression $P \subseteq G$ with $|P| \geq \omega(N, \eta, k)$ such that $\mathbb{E}_{x \in P} f(x) \geq c(\eta, k)$, where*

- (i) $\omega(N, \eta) \rightarrow \infty$ as $N \rightarrow \infty$ for fixed η ;
- (ii) $c(\eta, k) > 0$ is bounded away from zero when η is bounded away from zero and k is fixed.

Indeed, Theorem 1.4 then follows by applying Corollary 1.12 and then invoking Theorem 1.13 with $f := 1_A - (\mathbb{E}_{x \in Q} 1_A(x))1_Q$. Theorem 1.13 is in fact deduced in [27] from the following stronger theorem.

Theorem 1.14 (weak inverse theorem for $U^{k-1}(\mathbb{Z}/N\mathbb{Z})$ [27]). *Let $\eta > 0$ and $k \geq 3$. Let $G = \mathbb{Z}/N\mathbb{Z}$ be a cyclic group of prime order and let $f : G \rightarrow \mathcal{D}$ be a bounded function such that $\|f\|_{U^{k-1}(G)} \geq \eta$. Then, if $N \geq \exp(C_k \eta^{-C_k})$ for some sufficiently large $C_k > 0$, one can partition $\mathbb{Z}/N\mathbb{Z}$ into arithmetic progressions $(P_j)_{j \in J}$, each of size $|P_j| \geq c_k \eta^{-C_k} N^{c_k \eta^{-C_k}}$ for some $c_k, C_k > 0$, such that*

$$\sum_{j \in J} |\mathbb{E}(f 1_{P_j})| \geq c_k \eta^{C_k}$$

for some $c_k, C_k > 0$.

Theorem 1.13 (and hence Theorem 1.1) follows quickly from this and the mean zero hypothesis $\sum_{j \in J} \mathbb{E}(f 1_{P_j}) = \mathbb{E}(f) = 0$. Indeed one gets a fairly good quantitative result for Theorem 1.2, with $N_0 = \exp(\exp(C_k \delta^{-C_k}))$ for some explicit $C_k > 0$ (see [27]).

We refer to Theorem 1.14 as a *weak* inverse theorem because it gives a necessary criterion in order for a bounded function f to have large $U^{k-1}(G)$ norm, and hence a sufficient condition for the $U^{k-1}(G)$ norm to be small. As discussed above, this theorem

is strong enough to imply Szemerédi's theorem. Also, Theorem 1.14 could potentially be useful, when combined with such tools as Theorem 1.11, not only for demonstrating the existence of progressions of length k in a given set A , but in fact for providing an accurate count as to *how many* such progressions there are. For instance, one might hope to count the number of progressions of length k in the primes less than N by using Theorem 1.14 to show that a certain counting function f associated with the primes has small $U^{k-1}(\mathbb{Z}/N\mathbb{Z})$ norm and hence its contribution to the count of progressions in the primes could be controlled using Theorem 1.11. However, the sufficient condition for smallness of $U^{k-1}(\mathbb{Z}/N\mathbb{Z})$ given by Theorem 1.14 is very difficult to verify for sets such as the primes (being at least as difficult as the Elliott–Halberstam conjecture, which is not known to be implied even by the generalized Riemann hypothesis).

It is thus of interest to obtain a better inverse theorem for the $U^{k-1}(\mathbb{Z}/N\mathbb{Z})$ norm, which gives a more easily checkable condition for when this norm is small. Ideally we would like this condition to be both necessary and sufficient, at least up to constant losses. In this paper we shall achieve these objectives for $k = 4$.

In subsequent work we will give various applications of the results and methods of this paper. In [35] we obtain a new bound on the size of the largest subset of the vector space \mathbb{F}_5^n with no four-term arithmetic progression, and we hope to generalize that result to arbitrary abelian groups G . In another series of papers [36, 37] we obtain an asymptotic formula for the number of quadruples $p_1 < p_2 < p_3 < p_4 \leq N$ of primes in arithmetic progression.

2. Inverse theorems for U^{k-1} norms

We have now explained why we are interested in an inverse theorem for the U^{k-1} norms. Before we state our main theorems, let us give some other examples and results that will illustrate what the inverse theorem should be. Recall that the U^d norm of a function $e(\phi)$ measures the oscillation in the d th derivative of the phase ϕ . Also recall that a polynomial of degree at most $d - 1$ is a function whose d th derivative vanishes. We generalize this concept as follows.

Definition 2.1 (locally polynomial phase functions). If B is any non-empty subset of a finite additive group G and $d \geq 1$, we say that a function $\phi : B \rightarrow \mathbb{R}/\mathbb{Z}$ is a *polynomial phase function of order at most $d - 1$ locally on B* if we have

$$(h_1 \cdot \nabla_x) \cdots (h_{d+1} \cdot \nabla_x) \phi(x) = 0$$

whenever the cube $(x + \omega_1 h_1 + \cdots + \omega_d h_d)_{\omega_1, \dots, \omega_d \in \{0, 1\}}$ is contained in B . If $f : B \rightarrow \mathbb{C}$ is a function, we define the *local polynomial bias of order d on B* $\|f\|_{u^d(B)}$ to be the quantity

$$\|f\|_{u^d(B)} := \sup |\mathbb{E}_{x \in B}(f(x)e(-\phi(x)))|,$$

where ϕ ranges over all local polynomial phase functions of order at most $d - 1$ on B .

To begin with we will work in the global setting $B = G$, but as will become clear later we will need to also work in the local setting. We will refer to polynomial phase

functions of degree at most 1 as *linear phase functions* and to those of degree at most 2 as *quadratic phase functions*, with the modifiers ‘local’ or ‘global’ as appropriate.

The quantity $\|\cdot\|_{u^d(B)}$ is clearly a seminorm. It shares several features with the $U^d(G)$ norm. First of all, like the $U^d(G)$ norm, we have the monotonicity $\|f\|_{u^d(B)} \leq \|f\|_{u^{d+1}(B)}$, and when $B = G$ we also have the shift invariance $\|T^h f\|_{u^d(G)} = \|f\|_{u^d(G)}$. We also have the conjugation symmetry $\|\bar{f}\|_{u^d(B)} = \|f\|_{u^d(B)}$, and the phase invariance $\|fe(\phi)\|_{u^d(B)} = \|f\|_{u^d(B)}$ whenever ϕ is a locally polynomial phase of degree at most $d - 1$ on B . The latter invariance also extends to the $U^d(G)$ norm, thus

$$\|fe(\phi)\|_{U^d(G)} = \|f\|_{U^d(G)} \tag{2.1}$$

whenever $\phi : G \rightarrow \mathbb{R}/\mathbb{Z}$ is a global polynomial phase function of degree at most $d - 1$. Indeed, this invariance* can easily be seen from (1.1) and induction, using the fact that the derivative of a polynomial of degree at most $d - 1$ is a polynomial of degree $d - 2$.

From this invariance and (1.1), (1.2) we conclude that

$$\|f\|_{U^d(G)} = \|fe(-\phi)\|_{U^d(G)} \geq \|fe(-\phi)\|_{U^1(G)} = |\mathbb{E}_{x \in G}(f(x)e(-\phi(x)))|$$

whenever ϕ is a global polynomial phase of degree at most $d - 1$. Taking suprema over all ϕ , we obtain the inequality

$$\|f\|_{U^d(G)} \geq \|f\|_{u^d(G)} \tag{2.2}$$

for all $d \geq 1$, all additive groups G and all $f : G \rightarrow \mathbb{C}$.

It is now natural to ask whether the inequality (2.2) can be reversed. When $d = 1$ it is easy to verify (using (1.1) and the fact that polynomials of degree at most 0 are constant) that we in fact have equality:

$$\|f\|_{U^1(G)} = \|f\|_{u^1(G)}.$$

Consider next the case $d = 2$. For this we need the Fourier transform. Let \hat{G} be the Pontryagin dual of G , in other words the space of homomorphisms $\xi : x \mapsto \xi \cdot x$ from G to \mathbb{R}/\mathbb{Z} . As is well known, \hat{G} is an additive group which is isomorphic to G . If $\xi \in \hat{G}$, we define the *Fourier coefficient* $\hat{f}(\xi)$ of f at the frequency ξ by the formula

$$\hat{f}(\xi) = \mathbb{E}_x f(x)e(-\xi \cdot x).$$

As is well known, we have the *Fourier inversion formula*

$$f(x) = \sum_{\xi \in \hat{G}} \hat{f}(\xi)e(\xi \cdot x)$$

* This polynomial phase invariance also indicates why Fourier analysis—which is essentially invariant under modulation by linear phase functions but not by quadratic or higher phases—is only able to effectively deal with the $U^2(G)$ norm and not with higher norms. To deal with the $U^3(G)$ norm thus requires some sort of ‘quadratic Fourier analysis’ which is insensitive to phase modulations by quadratic phases. The results here can be viewed as some preliminary steps towards establishing such a quadratic Fourier analysis theory.

and the *Plancherel identity*

$$\mathbb{E}(|f|^2) = \sum_{\xi \in \hat{G}} |\hat{f}(\xi)|^2. \quad (2.3)$$

One can then easily verify the pleasant identity

$$\|f\|_{U^2(G)}^4 = \sum_{\xi \in \hat{G}} |\hat{f}(\xi)|^4. \quad (2.4)$$

For instance, this can be achieved by first establishing the identity $\|f\|_{U^2}^4 = \mathbb{E}(|f * f|^2)^2$, where $f * f(x) := \mathbb{E}_y f(y) f(x - y)$ is the convolution of f with itself, and then using (2.3). Next, we make the easy observation that if $\phi : G \rightarrow \mathbb{R}/\mathbb{Z}$ is a global polynomial phase function of degree at most 1, then $x \mapsto \phi(x) - \phi(0)$ is a homomorphism from G to \mathbb{R}/\mathbb{Z} , and hence there exists $\xi \in \hat{G}$ such that $\phi(x) = \xi \cdot x + \phi(0)$. From this it is easy to see that

$$\|f\|_{u^2(G)} = \sup_{\xi \in \hat{G}} |\hat{f}(\xi)|. \quad (2.5)$$

Combining (2.3), (2.4) and (2.5) we readily conclude the following proposition.

Proposition 2.2 (inverse theorem for the $U^2(G)$ norm). *Let $f : G \rightarrow \mathcal{D}$ be a bounded function. Then*

$$\|f\|_{u^2(G)} \leq \|f\|_{U^2(G)} \leq \|f\|_{u^2(G)}^{1/2}.$$

We remark that this proposition easily implies the $k = 3$ case of Theorem 1.14 (using Dirichlet's theorem on approximation by rationals to cover G by progressions on which ϕ is close to constant), and hence also implies Szemerédi's theorem for $k = 3$. Indeed this is essentially Roth's original argument [56], albeit phrased in very modern language.

Based on evidence such as Proposition 2.2, one is tempted to conjecture that the $U^d(G)$ and $u^d(G)$ norms are also related for higher d , in the sense that if f is bounded and one of the two norms $\|f\|_{U^d(G)}$, $\|f\|_{u^d(G)}$ is small, then the other is as well. From (2.2) we already know that one direction is true: smallness of the $U^d(G)$ norm implies smallness of the $u^d(G)$ norm. Our first main result establishes a converse to this in the $d = 3$ case when $G = \mathbb{F}_5^n$, though with only partially satisfactory control on the constants.

Theorem 2.3 (inverse theorem for $U^3(\mathbb{F}_5^n)$). *Let $f : \mathbb{F}_5^n \rightarrow \mathcal{D}$ be a bounded function and let $0 < \eta \leq 1$.*

- (i) *If $\|f\|_{U^3(\mathbb{F}_5^n)} \geq \eta$, then there exists a subspace $W \leq \mathbb{F}_5^n$ of codimension at most $(2/\eta)^C$ such that*

$$\mathbb{E}_{y \in \mathbb{F}_5^n} \|f\|_{u^3(y+W)} \geq (\eta/2)^C, \quad (2.6)$$

where we can take $C = 2^{16}$. In particular, there exists $y \in G$ such that

$$\|f\|_{u^3(y+W)} \geq (\eta/2)^C.$$

- (ii) *Conversely, given any subspace $W \leq \mathbb{F}_5^n$ and any function $f : \mathbb{F}_5^n \rightarrow \mathbb{C}$ we have $\|f\|_{U^3(\mathbb{F}_5^n)} \geq \|f\|_{u^3(\mathbb{F}_5^n)} \geq 5^{-n} |W| \|f\|_{u^3(y+W)}$ for any $y \in \mathbb{F}_5^n$.*

Combining the two parts of the theorem together we see that

$$\|f\|_{u^3(\mathbb{F}_5^n)} \leq \|f\|_{U^3(\mathbb{F}_5^n)} \leq \frac{C}{\log^c(1 + 1/\|f\|_{u^3(\mathbb{F}_5^n)})} \tag{2.7}$$

for some absolute constants $c, C > 0$. Thus this does give a result which asserts that the smallness of the $U^3(\mathbb{F}_5^n)$ norm implies the smallness of the $u^3(\mathbb{F}_5^n)$ norm, and vice versa, although the dependencies between constants are poor.* Note, however, that the control is much better if one localizes the quadratic bias norm $u^3(\mathbb{F}_5^n)$ to cosets $y + W$ of W . We remark that there is nothing particularly special about the finite field \mathbb{F}_5^n , and one has similar results for any other finite field of odd characteristic, though the constants depend of course on the field.

We shall prove Theorem 2.3 (i) in § 6 ((ii) is easier, and we will prove it in § 3); it contains many of the main ideas of this paper, which combine the Fourier and combinatorial analysis of Gowers in [25] with an additional ‘symmetry argument’ that is necessary to obtain a strong inverse theorem instead of a weak inverse theorem. As a consequence we obtain, in § 7, a Szemerédi theorem for progressions of length four in \mathbb{F}_5^n .

Let us now discuss finite abelian groups G in general, particular importance being attached to $\mathbb{Z}/N\mathbb{Z}$ on account of potential applications. It is tempting to conjecture, in light of the preceding results, that in such groups any bounded function with small $u^d(G)$ norm must necessarily have small $U^d(G)$ norm. Unfortunately, such a statement is false, even for $G = \mathbb{Z}/N\mathbb{Z}$. This fact was essentially discovered by Furstenberg and Weiss [23] in the closely related context of determining characteristic factors for multiple recurrence in ergodic theory; a similar observation was also made on p. 487 of [27]. See § 12 for some further discussion of this connection. We give one instance of the Furstenberg–Weiss example as follows.

Example 2.4. Let N be a large prime number and let M be the largest integer less than \sqrt{N} . Let $G := \mathbb{Z}/N\mathbb{Z}$ and let $f : G \rightarrow \mathbb{C}$ be the bounded function defined by setting $f(yM + z) := e(yz/M)\psi(y/M)\psi(z/M)$ whenever $-M/10 \leq y, z \leq M/10$ and $f = 0$ otherwise; here $\psi : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ is a non-negative smooth cutoff function which equals 1 on the interval $[-\frac{1}{20}, \frac{1}{20}]$ and vanishes outside of $[-\frac{1}{10}, \frac{1}{10}]$. Then a direct calculation shows that $\|f\|_{U^3(G)} \geq c_0$ for some absolute constant $c_0 > 0$, basically because all the phases in the expression for $\|f\|_{U^3(G)}^8$ cancel out leaving only the non-negative cutoffs ψ , whereas a Weyl sum computation reveals that $\mathbb{E}(fe(-\phi)) = O(N^{-c})$ for any quadratic phase function ϕ and some explicit constant $c > 0$. (Note that when N is prime, the only quadratic phase functions ϕ are those of the form $\phi(x) = ax^2 + bx + c$ where $a, b, c \in \mathbb{R}/\mathbb{Z}$ with $Na = Nb = 0$ (see Lemma 3.1).) We omit the details.

The heart of the difficulty here is that the function $yM + z \mapsto yz/M$ is locally quadratic on the set $B := \{yM + z : -M/10 \leq y, z \leq M/10\}$, which is a fairly large subset of G , but does not extend (even approximately) to a *globally* quadratic phase function on all of G . These locally quadratic phase functions are thus a genuinely new class of obstructions to

* We conjecture that one can improve the upper bound in (2.7) to a polynomial dependence (bringing this estimate in line with Proposition 2.2; see § 13 for further discussion.

having a small $U^d(G)$ norm which must now also be accounted for in order to produce a genuine inverse theorem for the $U^3(G)$ norm. Similar considerations also apply, of course, to the $U^d(G)$ norms for $d > 3$.

We must therefore understand the proper generalization of sets such as $B = \{yM + z : -M/10 \leq y, z \leq M/10\}$, are two possibilities, which are in a sense dual to one another: generalized arithmetic progressions and Bohr sets. For technical reasons it is convenient to work in the first instance with the latter notion, but we will discuss generalized arithmetic progressions later in the paper.

Definition 2.5 (Bohr sets). Let G be a finite additive group and let $S \subseteq \hat{G}$, $|S| = d$, be a subset of the dual group. We define a sub-additive quantity $\|\cdot\|_S$ on G by setting

$$\|x\|_S := \sup_{\xi \in S} \|\xi \cdot x\|_{\mathbb{R}/\mathbb{Z}},$$

where $\|x\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance to the nearest integer, and we define the *Bohr set* $B(S, \rho) \subseteq G$ for any $\rho > 0$ to be the set

$$B(S, \rho) := \{x \in G : \|x\|_S < \rho \text{ for all } \xi \in S\}.$$

Note that the dependence of the Bohr set $B(S, \rho)$ on ρ can be rather discontinuous, as can be seen rather dramatically in finite-field geometries such as \mathbb{F}_3^n . This is inconvenient in applications, but fortunately it was noted by Bourgain [10]* that one may restrict attention to ‘regular’ Bohr sets, which enjoy some limited continuity properties in ρ .

Definition 2.6 (regular Bohr sets [10]). Let $S \subseteq \hat{G}$, $|S| = d$, be a set of characters, and suppose that $\rho \in (0, 1)$. A Bohr set $B(S, \rho)$ is said to be *regular* if one has

$$(1 - 100d|\kappa|)|B(S, \rho)| \leq |B(S, (1 + \kappa)\rho)| \leq (1 + 100d|\kappa|)|B(S, \rho)|$$

whenever $|\kappa| \leq 1/100d$.

Lemma 8.2 gives a plentiful supply of regular Bohr sets. The constant 100 can be lowered but this will not concern us here. With this definition in place, we can now give the generalization of Theorem 2.3 to arbitrary groups.

Theorem 2.7 (inverse theorem for $U^3(G)$). Let G be a finite additive group of odd order, let $f : G \rightarrow \mathcal{D}$ be a bounded function and let $0 < \eta \leq 1$.

- (i) If $\|f\|_{U^3(G)} \geq \eta$, then there exists a regular Bohr set $B := B(S, \rho)$ in G with $|S| \leq (2/\eta)^C$ and $\rho \geq (\eta/2)^C$ such that

$$\mathbb{E}_{y \in G} \|f\|_{u^3(y+B)} \geq (\eta/2)^C, \tag{2.8}$$

* In fact, rather earlier Gowers [27, Lemma 10.10] employed an argument which establishes that all Bohr sets in $\mathbb{Z}/N\mathbb{Z}$, N prime, are regular in a very weak sense.

where it is permissible to take $C = 2^{24}$. In particular, there exists $y \in G$ such that $\|f\|_{u^3(y+B)} \geq (\eta/2)^C$.

- (ii) Conversely, if $B = B(S, \rho)$ is a regular Bohr set, if $f : G \rightarrow \mathcal{D}$ is a bounded function and if $\|f\|_{u^3(y+B)} \geq \eta$, then we have

$$\|f\|_{U^3(G)} \geq (\eta^3 \rho^2 / C' d^3)^d$$

for some absolute constant C' .

Note that the $u^3(G)$ norm is no longer involved in this inverse theorem; this is necessary as demonstrated by Example 2.4, and has to do with the lack of extendibility of some local quadratic phases to global ones. In later sections we will prove other, related, inverse theorems for the $U^3(G)$ norm. In § 10 we will obtain a result in which the quadratic phases are given quite explicitly when $G = \mathbb{Z}/N\mathbb{Z}$. Then, in Theorem 12.9, we will provide a link to recent ergodic-theoretic work of Host, Kra and Ziegler.

In a future series of papers we will prove an enhanced version of Theorem 2.7 and use it to establish an asymptotic for the number of quadruples $p_1 < p_2 < p_3 < p_4 \leq N$ of primes in arithmetic progression. The enhancement required is that we must be able to deal with functions $f : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{C}$ which are not necessarily bounded: in particular, functions such as $f = \Lambda - 1$, where Λ is the von Mangoldt function. Once this is done, one may analyse the u^3 norm of f using what are, in essence, rather classical methods of analytic number theory, such as Vaughan’s decomposition of Λ .

A word of reassurance is perhaps in order for the reader interested in this result concerning primes. Although the present paper is long, only a few sections of it, namely §§ 4, 5, 8 and 9, are relevant to that work. In fact it is hoped that the subsequent papers on primes will be readable largely independently of the present work.

Let us briefly mention the connection between the results here and those in [62]. In that paper the second author introduced the concept of a *uniformly almost periodic function of order $k - 2$* , which generalized the concept of a polynomial phase of order at most $k - 2$ (and which incorporates the ‘locally polynomial phases’ discussed above). An inverse theorem for the U^{k-1} norms involving these uniformly almost periodic functions was also obtained (by very elementary means): see [62, Lemma 5.11]. However, because the uniformly almost periodic functions are a larger class than the locally polynomial phases, those results are weaker than the inverse theorems presented here. Nevertheless, with substantial additional effort (involving, for instance, the van der Waerden theorem) it is possible to use the inverse theorem for uniformly almost periodic functions to obtain another proof of Szemerédi’s theorem (see [62] for more details). We also remark that very similar objects (the *anti-uniform functions*) were also used in [34] in order to reduce the task of establishing arbitrarily long progressions in the primes to Szemerédi’s theorem.

Finally, let us offer a word of explanation for our policy concerning constants. For many of the arguments of this paper we have supplied exact constants, eschewing excessive use of the O -notation. This perhaps allows one to better see how bounds from different lemmas combine with one another to influence later bounds. Some readers may, however, prefer to replace such quantities as $2^{90}\eta^{-384}$ with $C\eta^{-C}$ when reading the paper.

3. A model problem: global quadratic phase functions

We now present a simple result, namely the classification of globally quadratic phase functions on an arbitrary additive group G of odd order, which we will need later, and which will serve to illustrate our strategy for the more advanced results we give below.

Let us call a homomorphism $M : G \mapsto \hat{G}$ *self-adjoint* if we have

$$Mx \cdot y - My \cdot x = 0 \quad \text{for all } x, y \in G.$$

Lemma 3.1 (inverse theorem for globally quadratic phase functions). *Let G be a finite additive group of odd order and let $\phi : G \rightarrow \mathbb{R}/\mathbb{Z}$ be a globally quadratic phase function. Then there exists $c \in \mathbb{R}/\mathbb{Z}$, $\xi \in \hat{G}$ and a self-adjoint homomorphism $M : G \rightarrow \hat{G}$ such that $\phi(x) = Mx \cdot x + \xi \cdot x + c$. Conversely, all such functions $x \mapsto Mx \cdot x + \xi \cdot x + c$ are globally quadratic phase functions.*

Proof. The converse is easy so we focus on the forward direction. It is convenient to adopt the notation $\mathbf{p}(x_1, \dots, x_n)$ to denote an arbitrary function of variables x_1, \dots, x_n that takes values in \mathbb{R}/\mathbb{Z} , where \mathbf{p} can vary from line to line or even within the same line. This is useful for handling expressions whose exact value is not important for the argument but whose functional dependencies on other variables needs to be recorded.

We shall give an argument which may seem a bit cumbersome (and is certainly not the shortest proof of this lemma), but it will serve to motivate the proof of Theorems 2.3 and 2.7. Indeed, this lemma can be thought of in some sense as the $\eta = 1$ case of those theorems. First observe that if ϕ is a quadratic phase function, then $(h \cdot \nabla_x)\phi$ is a linear phase function for each h . Thus for each $h \in G$ there exists $\xi_h \in \hat{G}$ such that

$$(h \cdot \nabla_x)\phi(x) = \phi(x + h) - \phi(x) = \xi_h \cdot x + \mathbf{p}(h) \quad \text{for all } x, h \in G. \quad (3.1)$$

The next step is to obtain some linearity on the map $h \mapsto \xi_h$ by using difference operators to eliminate various terms. Let $k \in G$ be arbitrary. If we apply the difference operator $(k \cdot \nabla_x)$ to (3.1), we can eliminate the $\mathbf{p}(h)$ term to obtain

$$(k \cdot \nabla)\phi(x + h) - (k \cdot \nabla)\phi(x) = \xi_h \cdot k \quad \text{for all } x, h, k \in G. \quad (3.2)$$

If we then apply the difference operator $(h_1 \cdot \nabla_h)$ to eliminate the $(k \cdot \nabla)\phi(x)$ term, where $h_1 \in G$ is arbitrary, we obtain

$$(h_1 \cdot \nabla)(k \cdot \nabla)\phi(x + h) = (h_1 \cdot \nabla_h)\xi_h \cdot k \quad \text{for all } x, h, k, h_1 \in G. \quad (3.3)$$

Making the substitution $y = x + h$ we obtain

$$(h_1 \cdot \nabla)(k \cdot \nabla)\phi(y) = (h_1 \cdot \nabla_h)\xi_h \cdot k \quad \text{for all } y, h, k, h_1 \in G. \quad (3.4)$$

If we then apply the difference operator $(h_2 \cdot \nabla_h)$ to eliminate the remaining ϕ term, where $h_2 \in G$ is arbitrary, we conclude that

$$0 = (h_2 \cdot \nabla_h)(h_1 \cdot \nabla_h)\xi_h \cdot k \quad \text{for all } y, h, k, h_1 \in G. \quad (3.5)$$

Since k is arbitrary, we conclude that

$$(h_2 \cdot \nabla_{h_1})(h_1 \cdot \nabla_{h_2})\xi_h = 0 \quad \text{for all } h_1, h_2 \in G. \tag{3.6}$$

Thus if we write

$$\xi_h = 2Mh + \xi_0, \tag{3.7}$$

then we see that $M : G \mapsto \hat{G}$ is a group homomorphism. Note that we can insert the 2 in front of M because $|G|$ is odd; this factor of 2 will be convenient later. Inserting this back into (3.1) we obtain

$$(h \cdot \nabla_x)\phi(x) = 2Mh \cdot x + \xi_0 \cdot x + \mathbf{p}(h) \quad \text{for all } x, h \in G. \tag{3.8}$$

This is almost what we want, but we must somehow ‘integrate’ the partial derivative $h \cdot \nabla_x$. To do this we must first establish that M is self-adjoint. Informally, this self-adjointness reflects the symmetry $(h' \cdot \nabla)(h \cdot \nabla)\phi = (h \cdot \nabla)(h' \cdot \nabla)\phi$ of the second derivative. To make this rigorous (and in a manner which will extend suitably to more general situations) we shall use the following ‘symmetry argument’. In order to focus on the $Mh \cdot x$ term, we shall write (3.8) as

$$\mathbf{p}(x+h) + \mathbf{p}(x) + \mathbf{p}(h) - 2Mh \cdot x = 0 \quad \text{for all } x, h \in G. \tag{3.9}$$

Substituting y for x and subtracting to eliminate $\mathbf{p}(h)$, we obtain

$$\mathbf{p}(y+h) + \mathbf{p}(y) + \mathbf{p}(x+h) + \mathbf{p}(x) - 2Mh \cdot (y-x) = 0 \quad \text{for all } x, y, h \in G. \tag{3.10}$$

Making the substitution $z = x + y + h$, we conclude that

$$\mathbf{p}(x, z) + \mathbf{p}(y, z) - 2M(z-x-y) \cdot (y-x) = 0 \quad \text{for all } x, y, z \in G. \tag{3.11}$$

Absorbing as many terms as possible into the unspecified functions \mathbf{p} , we conclude that

$$\mathbf{p}(x, z) + \mathbf{p}(y, z) + 2\{x, y\} = 0 \quad \text{for all } x, y, z \in G, \tag{3.12}$$

where $\{x, y\}$ is the antisymmetric form

$$\{x, y\} := Mx \cdot y - My \cdot x. \tag{3.13}$$

Freezing the value of z , we conclude that

$$\mathbf{p}(x) + \mathbf{p}(y) + 2\{x, y\} = 0 \quad \text{for all } x, y \in G. \tag{3.14}$$

Replacing y by y' and subtracting to eliminate the $\mathbf{p}(x)$ factor, we conclude that

$$\mathbf{p}(y) + \mathbf{p}(y') + 2\{x, y' - y\} = 0 \quad \text{for all } x, y, y' \in G. \tag{3.15}$$

Thus for each y, y' , the linear function $x \mapsto \{x, y' - y\}$ is independent of x and is hence always zero*:

$$\{x, y' - y\} = 0 \quad \text{for all } x, y, y' \in G. \quad (3.16)$$

Thus M is self-adjoint.

We return now to (3.8), which we write as

$$\mathbf{p}(h) + \mathbf{p}(x+h) - \phi(x) - 2Mh \cdot x = 0 \quad \text{for all } x, h \in G. \quad (3.17)$$

From the self-adjointness of M we have

$$2Mh \cdot x = M(x+h) \cdot (x+h) - Mx \cdot x - Mh \cdot h \quad (3.18)$$

and hence

$$\mathbf{p}(h) + \mathbf{p}(x+h) - \phi(x) + Mx \cdot x = 0 \quad \text{for all } x, h \in G. \quad (3.19)$$

In particular, if we apply difference operators in the h and $x+h$ variables, we see that the phase function $-\phi(x) + Mx \cdot x$ is linear:

$$(h_1 \cdot \nabla_x)(h_2 \cdot \nabla_x)(-\phi(x) + Mx \cdot x) = 0 \quad \text{for all } x, h_1, h_2 \in G. \quad (3.20)$$

Hence there exists $\xi \in \hat{G}$ and a $c \in \mathbb{R}/\mathbb{Z}$ such that

$$\phi(x) - Mx \cdot x = \xi \cdot x + c. \quad (3.21)$$

The claim follows. \square

As one corollary of this classification, we obtain the following ‘quadratic extension theorem’.

Proposition 3.2 (quadratic extension theorem). *Let G be an additive group, let $H \leq G$ be a subgroup and suppose that $y \in G$. Then any quadratic phase function $\phi : y + H \rightarrow \mathbb{R}/\mathbb{Z}$ can be extended (non-uniquely in general) to a globally quadratic phase function on G .*

Remark 3.3. An important theme of this paper is that this behaviour is specific to cosets $y + H$, and breaks down for other sets such as Bohr sets.

As a consequence of Proposition 3.2, let us now establish the (easy) second part of Theorem 2.3. The first inequality $\|f\|_{U^3(G)} \geq \|f\|_{u^3(G)}$ follows from (2.2), so we focus on the second inequality $\|f\|_{u^3(G)} \geq (|W|/|G|)\|f\|_{u^3(y+W)}$. It suffices to show that

$$\|f\|_{u^3(G)} \geq \frac{|W|}{|G|} \left| \mathbb{E}_{x \in y+W} f(x) e(-\phi(x)) \right|$$

* There appear to be some intriguing parallels with symplectic geometry here. Roughly speaking, the vanishing (3.16) is an assertion that the graph $\{(h, Mh) : h \in G\}$ is a ‘Lagrangian manifold’ on the ‘phase space’ $G \times \hat{G}$. This graph can also be interpreted (essentially) as the ‘wavefront set’ $\{(x, \nabla \phi(x)) : x \in G\}$ of the original function $e(\phi)$. A similar interpretation persists in the proofs of Theorem 2.3 and 2.7 below. Thus we see hints of some kind of ‘combinatorial symplectic geometry’ emerging, though we do not see how to develop these possible connections further.

whenever ϕ is a locally quadratic function on $y + W$. But by the preceding discussion, we can extend ϕ to all of G and write

$$\frac{|W|}{|G|} |\mathbb{E}_{x \in y+W} f(x) e(-\phi(x))| = |\mathbb{E}_{x \in G} f(x) 1_W(x-y) e(-\phi(x))|.$$

But by Fourier inversion we may write

$$1_W(x-y) = \mathbb{E}_{\xi \in W^\perp} e(\xi \cdot (x-y))$$

where $W^\perp := \{\xi \in \hat{G} : \xi \cdot x = 0 \text{ for all } x \in W\}$. Thus we have

$$\mathbb{E}_{x \in G} f(x) 1_W(x-y) e(-\phi(x)) = \mathbb{E}_{\xi \in W^\perp} \mathbb{E}_{x \in G} f(x) e(\xi \cdot (x-y) - \phi(x)).$$

But since $\xi \cdot (x-y) - \phi(x)$ is globally quadratic in x , we have

$$|\mathbb{E}_{x \in G} f(x) e(\xi \cdot (x-y) - \phi(x))| \leq \|f\|_{u^3(G)},$$

and the claim follows from the triangle inequality. Note that this argument in fact works for arbitrary groups G (with W now being a subgroup of G rather than a subspace).

4. Averaging lemmas

In this section we collect some very simple averaging estimates that we shall rely on frequently below.

Lemma 4.1 (averaging on a subgroup). *Let G be an additive group, let H be a finite subgroup of G and let $A \subseteq H$ be non-empty. Let $f : H \rightarrow \mathbb{C}$ be a function. Then*

$$\mathbb{E}_{x \in H} \mathbb{E}_{y \in x+A} f(y) = \mathbb{E}_{y \in H} f(y).$$

In particular, by the pigeonhole principle there exists $x \in H$ such that

$$|\mathbb{E}_{y \in x+A} f(y)| \geq |\mathbb{E}_{y \in H} f(y)|.$$

Proof. Since $H + h = H$ for all $h \in A$ we have

$$\mathbb{E}_{x \in H} f(x+h) = \mathbb{E}_{y \in H} f(y) \quad \text{for all } h \in A.$$

Averaging this over all $h \in A$ we obtain the first claim, and the second claim then follows from the pigeonhole principle. \square

This lemma will be adequate for our purposes when we are in the finite-field geometry case, because we will have plenty of subgroups available. In the general group case, however, we will also need a more general type of averaging principle. The next lemma contains several rather similar formulations of such a result; all of them will be useful later on.

Lemma 4.2 (averaging on a Bohr set). *Let $S \subseteq \hat{G}$ be a set of d characters and let $0 < \rho < 1$ and $\varepsilon \leq 1/200d$ be parameters. Suppose that the Bohr set $B := B(S, \rho)$ is regular, and let $A \subseteq B(S, \varepsilon\rho)$ be any set. Finally, let $f : G \rightarrow \mathcal{D}$ be any function. Then*

- (i) $|\mathbb{E}_{x \in B} f(x) - \mathbb{E}_{x \in B} f(x+y)| \leq 200d\varepsilon$ if $y \in A$;
- (ii) $|\mathbb{E}_{x \in B} f(x) - \mathbb{E}_{x \in B} \mathbb{E}_{y \in x+A} f(x)| \leq 200d\varepsilon$;
- (iii) there is some $x \in B$ for which $\mathbb{E}_{y \in x+A} f(y) \geq \mathbb{E}_{y \in B} f(y) - 200d\varepsilon$;
- (iv) there is some $x \in B(S, (1 - \varepsilon)\rho)$ such that $\mathbb{E}_{y \in x+A} f(y) \geq \mathbb{E}_{y \in B} f(y) - 500d\varepsilon$.

Proof. To prove (i) we must check that

$$\left| \sum_{x \in B+y} f(x) - \sum_{x \in B} f(x) \right| \leq 200d\varepsilon|B|.$$

This follows from the fact that $B(S, \rho)$ and $y + B(S, \rho)$ differ in at most $200d\varepsilon$ elements. Indeed it is easy to see that

$$B(S, \rho) \Delta (y + B(S, \rho)) \subseteq B(S, (1 + \varepsilon)\rho) \setminus B(S, (1 - \varepsilon)\rho)$$

(‘ Δ ’ denotes symmetric difference), and the size of this latter set is no greater than $200d\varepsilon$ by regularity.

(ii) follows from (i) and the triangle inequality. Indeed $|\mathbb{E}_{x \in B} f(x+y) - \mathbb{E}_{x \in B} f(x)| \leq 200d\varepsilon$, whence

$$|\mathbb{E}_{x \in B} \mathbb{E}_{y \in A} f(x+y) - \mathbb{E}_{x \in B} f(x)| \leq 200d\varepsilon,$$

which is (ii).

(iii) is immediate from (ii) and the pigeonhole principle.

(iv) follows from the pigeonhole principle and the fact that

$$|\mathbb{E}_{x \in B} f(x) - \mathbb{E}_{x \in B(S, (1-\varepsilon)\rho)} \mathbb{E}_{y \in x+A} f(y)| \leq 500d\varepsilon,$$

which is in turn implied by (ii) and the bound

$$|\mathbb{E}_{x \in B} F(x) - \mathbb{E}_{x \in B(S, (1-\varepsilon)\rho)} F(x)| \leq 300\varepsilon d,$$

valid for any $F : G \rightarrow \mathcal{D}$. To confirm this, note that

$$\left| \sum_{x \in B} F(x) - \sum_{x \in B(S, (1-\varepsilon)\rho)} F(x) \right| \leq |B \setminus B(S, (1 - \varepsilon)\rho)| \leq 100\varepsilon d|B|.$$

Therefore,

$$\begin{aligned} |\mathbb{E}_{x \in B} F(x) - \mathbb{E}_{x \in B(S, (1-\varepsilon)\rho)} F(x)| &\leq \frac{1}{|B|} \left| \sum_{x \in B} F(x) - \sum_{x \in B(S, (1-\varepsilon)\rho)} F(x) \right| \\ &\quad + \left| \frac{1}{|B|} - \frac{1}{|B(S, (1-\varepsilon)\rho)|} \right| \left| \sum_{x \in B(S, (1-\varepsilon)\rho)} F(x) \right| \\ &\leq 100\varepsilon d + \left| \frac{|B|}{|B(S, (1-\varepsilon)\rho)|} - 1 \right| \\ &\leq 100\varepsilon d + \frac{1}{1 - 100\varepsilon d} - 1 \leq 300\varepsilon d. \end{aligned}$$

This completes the proof of Lemma 4.2. □

We adopt the following useful notation, analogous to the \mathbf{p} notation used in proving Lemma 3.1. When $g(x_1, \dots, x_n)$ is a complex-valued function of certain variables x_1, \dots, x_n with $\|g\|_\infty \leq 1$, we shall instead refer to $g(x_1, \dots, x_n)$ as $\mathbf{b}(x_1, \dots, x_n)$. The notation \mathbf{b} thus denotes a function with $\|\mathbf{b}\|_\infty \leq 1$, but the notation may refer to different functions from line to line, or even on the same line (similar to the O notation, or the use of unspecified constants C).

We now record a basic application of the Cauchy–Schwarz inequality, whose proof is immediate.

Lemma 4.3 (Cauchy–Schwarz). *Let X, Y be finite sets and let $f : X \times Y \rightarrow \mathbb{C}$ be a function. Then for any bounded function $\mathbf{b}(x)$ of X , we have*

$$|\mathbb{E}_{x,y} f(x,y)\mathbf{b}(x)| \leq \mathbb{E}_x |\mathbb{E}_y f(x,y)| \leq (\mathbb{E}_x |\mathbb{E}_y f(x,y)|^2)^{1/2} = (\mathbb{E}_{x,y,y'} f(x,y') \overline{f(x,y)})^{1/2}.$$

In the special case when $Y = G$ is a group, we conclude in particular the *van der Corput inequality*

$$|\mathbb{E}_{x \in X, y \in G} f(x,y)\mathbf{b}(x)| \leq |\mathbb{E}_{x \in X, y, h \in G} T_y^h f(x,y) \overline{f(x,y)}|^{1/2} \tag{4.1}$$

whenever $f : X \times G \rightarrow \mathbb{C}$ is a function, which follows by using the substitution $y' = y + h$. This inequality is very useful for eliminating unknown bounded functions $\mathbf{b}(x)$ in an expression to be estimated. Using the van der Corput inequality, we can now prove Proposition 1.11.

Proof of Proposition 1.11. It suffices to prove the more general statement

$$\left| \mathbb{E}_{x,h} \prod_{j \in J} T^{a_j h} f_j(x) \right| \leq \|f_{j_0}\|_{U^{|J|-1}(G)} \tag{4.2}$$

for all finite sets J with $|J| \geq 1$, all $j_0 \in J$, all bounded functions $(f_j)_{j \in J}$, and all distinct integers $(a_j)_{j \in J}$ such that $a_j - a_{j'}$ is coprime to $|G|$ for all distinct $j, j' \in J$.

We induct on $|J|$. When $|J| = 1$ the claim is trivial from (1.1), so suppose that $|J| \geq 2$ and that the claim has already been proven for smaller values of J . Let j_1 be an element

in $J \setminus \{j_0\}$. By making the change of variables $x \rightarrow x + a_{j_1}h$ if necessary we may assume that $a_{j_1} = 0$. Since f_{j_1} is bounded, we can then express the left-hand side of (4.2) as

$$\left| \mathbb{E}_{x,h \in G} \left(\mathbf{b}(x) \prod_{j \in J \setminus \{j_1\}} T^{a_j h} f_j(x) \right) \right|,$$

which by (4.1) can be bounded by

$$\mathbb{E}_{k \in G} \left(\mathbb{E}_{x,h \in G} \left(\prod_{j \in J \setminus \{j_1\}} T^{a_j h} (T^{a_j k} f_j \bar{f}_j)(x) \right) \right)^{1/2}.$$

Applying the inductive hypothesis (4.2) to the inner expectation, we can bound this in turn by

$$\mathbb{E}_{k \in G} (\|T^{a_j k} f_j \bar{f}_j\|_{U^{|J|-2}(G)})^{1/2},$$

which by Hölder’s inequality and the substitution $h := a_j k$ (noting that $(N, a_j - a_{j_1}) = (N, a_j) = 1$) is bounded by

$$\mathbb{E}_{h \in G} (\|T^h f_j \bar{f}_j\|_{U^{|J|-2}(G)})^{\frac{1}{2} |J|-1}.$$

The claim then follows from (1.1). □

Let us now consider averages of the form

$$\mathbb{E}_{z \in B', x \in B} (f(z) \mathbf{b}(x) \mathbf{b}(z + x)),$$

where B, B' are non-empty subsets of an additive group G and where f, g, h are bounded functions. If $B = B' = G$, then a variant of Proposition 1.11 shows that this quantity is bounded by $\|f\|_{U^2(G)}$, and hence (by Proposition 2.2) if the above average is large, then f must correlate with a linear phase function. It turns out that a similar statement is true for arbitrary B, B' , provided that $B + B'$ is only a little bit larger than B .

Lemma 4.4 (large trilinear form implies correlation with linear phase). *Let B, B' be two non-empty subsets of an additive group G . Then we have*

$$\|f\|_{u^2(B')} \geq \frac{\mathbb{E}(1_B)}{\mathbb{E}(1_{B+B'})} |\mathbb{E}_{z \in B', x \in B} (f(z) \mathbf{b}_1(x) \mathbf{b}_2(z + x))|$$

for any $f : G \rightarrow \mathbb{C}$ and any two bounded functions $\mathbf{b}_1, \mathbf{b}_2$.

Proof. Without loss of generality we may assume that $f, \mathbf{b}_1, \mathbf{b}_2$ vanish outside of $B', B, B + B'$, respectively. From Fourier expansion we have

$$\begin{aligned} \mathbb{E}_{z \in B', x \in B} (f(z) \mathbf{b}_1(x) \mathbf{b}_2(z + x)) &= \frac{1}{\mathbb{E}(1_{B'}) \mathbb{E}(1_B)} \mathbb{E}_{y \in G} (f * \mathbf{b}_1(y) \mathbf{b}_2(y)) \\ &= \frac{1}{\mathbb{E}(1_{B'}) \mathbb{E}(1_B)} \sum_{\xi \in \hat{G}} \hat{f}(\xi) \hat{\mathbf{b}}_1(\xi) \hat{\mathbf{b}}_2(-\xi). \end{aligned}$$

On the other hand, from Plancherel’s theorem we have

$$\sum_{\xi \in \hat{G}} |\hat{\mathbf{b}}_1(\xi)|^2 \leq \mathbb{E}(1_B) \quad \text{and} \quad \sum_{\xi \in \hat{G}} |\hat{\mathbf{b}}_2(-\xi)|^2 \leq \mathbb{E}(1_{B+B'}).$$

From Hölder’s inequality we thus conclude that

$$|\mathbb{E}_{z \in B', x \in B}(f(z)\mathbf{b}_1(x)\mathbf{b}_2(z+x))| \leq \frac{\mathbb{E}(1_{B+B'})}{\mathbb{E}(1_{B'})\mathbb{E}(1_B)} \sup_{\xi \in \hat{G}} |\hat{f}(\xi)|,$$

and hence there exists $\xi \in \hat{G}$ such that

$$|\mathbb{E}_{z \in B'}(f(z)e(-\xi \cdot z))| \geq |\mathbb{E}_{z \in B', x \in B}(f(z)\mathbf{b}_1(x)\mathbf{b}_2(z+x))| \frac{\mathbb{E}(1_B)}{\mathbb{E}(1_{B+B'})}.$$

The claim follows. □

5. An argument of Gowers

We now begin the proofs of the inverse theorems in Theorem 2.3 and Theorem 2.7. As we shall see, the arguments are analogous to those used to prove Lemma 3.1 and closely follow the treatment in [25]. The first part of the argument is to establish a ‘phase derivative’ $h \rightarrow \xi_h$ for the function f and to establish some additivity properties on this phase derivative. These arguments apply to arbitrary finite additive groups G ; later on we shall treat the finite-field case separately.

Proposition 5.1 (large U^3 norm gives many additive quadruples [25]). *Let G be an arbitrary finite additive group, and let $f : G \rightarrow \mathcal{D}$ be a bounded function such that $\|f\|_{U^3(G)} \geq \eta$ for some $\eta > 0$. Then there exists a set $H \subseteq G$, and a function $\xi : H \rightarrow \hat{G}$ whose graph $\Gamma := \{(h, \xi_h) : h \in H\} \subseteq G \times \hat{G}$ obeys the estimate*

$$|\{(z_1, z_2, z_3, z_4) \in \Gamma : z_1 + z_2 = z_3 + z_4\}| \geq 2^{-8}\eta^{64}N^3. \tag{5.1}$$

Furthermore, for each $(h, \xi_h) \in \Gamma$ we have

$$|\mathbb{E}_x T^h f(x) \bar{f}(x) e(-\xi_h \cdot x)| \geq \eta^4/2^{1/2}.$$

Proof. As we shall see, this proposition corresponds fairly closely with the first part of the proof of Lemma 3.1 (up to (3.6)). From (1.1) we have

$$\mathbb{E}_{h \in G} \|T^h f \bar{f}\|_{U^2(G)}^4 \geq \eta^8.$$

Applying Proposition 2.2 we conclude that

$$\mathbb{E}_{h \in G} \|T^h f \bar{f}\|_{u^2(G)}^2 \geq \eta^8.$$

Thus if we let

$$H := \{h \in G : \|T^h f \bar{f}\|_{u^2(G)}^2 \geq \eta^8/2\},$$

then we have

$$\mathbb{E}_{h \in G} \|T^h f \bar{f}\|_{u^2(G)}^2 1_{G \setminus H}(h) \geq \eta^8/2$$

and hence

$$\mathbb{E}_{h \in G} \|T^h f \bar{f}\|_{u^2(G)}^2 1_H(h) \geq \eta^8/2.$$

In particular we have

$$\mathbb{E}(1_H) \geq \eta^8/2. \tag{5.2}$$

By (2.5) and the definition of H we can find a map $h \mapsto \xi_h$ from H to \hat{G} such that

$$|\mathbb{E}_x T^h f(x) \bar{f}(x) e(-\xi_h \cdot x)| \geq \eta^4/2^{1/2} \quad \text{for all } h \in H. \tag{5.3}$$

Let us fix this map $h \mapsto \xi_h$. We square sum the above expression in h and use (5.2) to conclude that

$$\mathbb{E}_h |\mathbb{E}_x T^h f(x) \bar{f}(x) e(-\xi_h \cdot x)|^2 1_H(h) \geq \eta^{16}/4.$$

But from the identity

$$|\mathbb{E}_x T^h f(x) \bar{f}(x) e(-\xi_h \cdot x)|^2 = \mathbb{E}_{x,k} T^k (T^h f)(x) \overline{T^h f(x)} \overline{T^k f(x)} T^k f(x) e(\xi_h \cdot k)$$

we conclude that

$$|\mathbb{E}_{x,h,k} T^k (T^h f)(x) \overline{T^h f(x)} e(\xi_h \cdot k) 1_H(h) \overline{T^k f(x)} T^k f(x)| \geq \eta^{16}/4. \tag{5.4}$$

At this point we suppress the explicit mention of the functions f and write this simply as

$$|\mathbb{E}_{x,h,k \in G} (\mathbf{b}(x+h, k) 1_H(h) e(\xi_h \cdot k) \mathbf{b}(x, k))| \geq \eta^{16}/4.$$

Applying (4.1) to eliminate the $\mathbf{b}(x, k)$ factor, we conclude that

$$\mathbb{E}_{h,h_1,x,k} \mathbf{b}(x+h+h_1, k) \mathbf{b}(x+h, k) e((h_1 \cdot \nabla_h) \xi_h \cdot k) 1_H(h+h_1) 1_H(h) \geq \eta^{32}/16. \tag{5.5}$$

Making the substitution $y = x + h$, we obtain

$$\mathbb{E}_{h,y,h_1,k} e((h_1 \cdot \nabla_h) \xi_h \cdot k) 1_H(h+h_1) 1_H(h) \mathbf{b}(y, h_1, k) \geq \eta^{32}/16. \tag{5.6}$$

Applying (4.1) again we conclude that

$$\mathbb{E}_{h,h_1,h_2,y,k} e((h_2 \cdot \nabla_h)(h_1 \cdot \nabla_h) \xi_h \cdot k) 1_H(h+h_1+h_2) 1_H(h+h_2) 1_H(h+h_1) 1_H(h) \geq 2^{-28} \eta^{64}. \tag{5.7}$$

Summing this in k using the Fourier inversion formula, and discarding the irrelevant y averaging, we infer that

$$\mathbb{E}_{h,h_1,h_2} 1_{(h_2 \cdot \nabla_h)(h_1 \cdot \nabla_h) \xi_h = 0} 1_H(h+h_1+h_2) 1_H(h+h_2) 1_H(h+h_1) 1_H(h) \geq \eta^{64}/256. \tag{5.8}$$

The claim now follows by substituting $z_1 := (h, \xi_h)$, $z_2 := (h+h_1, \xi_{h+h_1})$, $z_3 := (h+h_2, \xi_{h+h_2})$ and $z_4 := (h+h_1+h_2, \xi_{h+h_1+h_2})$. □

As remarked, the analogy between this argument and the first part of the proof of Lemma 3.1 is very close. In particular, equations (5.3), (5.4), (5.5), (5.6), (5.7) and (5.8) are analogues of (3.1), (3.2), (3.3), (3.4), (3.5) and (3.6), respectively.

In order to exploit the conclusion (5.1), we require two very useful results from additive combinatorics. The first result asserts that a set Γ' with some partial additive structure (in the sense that it contains many additive quadruples) can be refined to have a more complete additive structure (in the sense that its sum set is small). This type of result was first obtained by Balog and Szemerédi [1], but with very weak constants. A version of the theorem with polynomial dependencies between the constants was obtained by Gowers [25]. The version we quote below, with rather good powers in those polynomials, may be proved by a careful working of the argument in [13]. Of course for the purposes of this paper the precise values of the constants are somewhat unimportant.

Write $\Gamma' - \Gamma' := \{z - z' : z, z' \in \Gamma'\}$ for the difference set of Γ' .

Theorem 5.2 (Balog–Szemerédi–Gowers theorem [13]). *Let G be an additive group and let Γ be a finite non-empty subset of G such that*

$$|\{(z_1, z_2, z_3, z_4) \in \Gamma : z_1 + z_2 = z_3 + z_4\}| \geq |\Gamma|^3/K$$

for some $K \geq 1$. Then there exists a subset $\Gamma' \subseteq \Gamma$ such that

$$|\Gamma'| \geq 2^{-6}K^{-1}|\Gamma| \quad \text{and} \quad |\Gamma' - \Gamma'| \leq 2^{42}K^6|\Gamma'|.$$

The other tool we need is the Plünnecke inequality, a simple proof* of which can be found in [57].

Theorem 5.3 (Plünnecke inequalities [52, 57]). *Let G be an arbitrary additive group and let Γ' be a finite non-empty subset of G such that $|\Gamma' - \Gamma'| \leq K|\Gamma'|$ for some $K \geq 1$. Then we have $|k\Gamma' - l\Gamma'| \leq K^{k+l}|\Gamma'|$ for all $k, l \geq 1$, where $k\Gamma'$ denotes the k -fold sumset of Γ' .*

In our applications the integers k, l will be quite small—in fact they will not exceed 9. Combining Proposition 5.1 with Theorems 5.2 and 5.3, we conclude the following proposition.

Proposition 5.4 ($h \mapsto \xi_h$ is nearly affine-linear). *Let G be an arbitrary finite additive group and let $f : G \rightarrow \mathcal{D}$ be a bounded function such that $\|f\|_{U^3(G)} \geq \eta$ for some $\eta > 0$. Then there exist a set $H' \subseteq G$ and a function $\xi : H' \rightarrow \hat{G}$ whose graph $\Gamma' := \{(h, \xi_h) : h \in H'\} \subseteq G \times \hat{G}$ obeys the estimates*

$$|\Gamma'| \geq 2^{-14}\eta^{64}N \tag{5.9}$$

* Actually, since we only need this theorem for bounded k, l , and would not be concerned if the bound of K^{k+l} were worsened to $K^{C(k+l)}$ for some absolute constant C , it is possible to modify the proof of Theorem 5.2 in order to gain control on $|k\Gamma' - l\Gamma'|$ directly, without needing the Plünnecke inequalities. However, this would not simplify the remainder of the argument and so we do not give the details of this alternate approach here.

and

$$|k\Gamma' - l\Gamma'| \leq (2^{90}\eta^{-384})^{k+l}|\Gamma'| \quad \text{for all } k, l \geq 1.$$

Furthermore, for each $(h, \xi_h) \in \Gamma'$ we have

$$|\mathbb{E}_{x \in G}(T^h f(x)\overline{f(x)}e(-\xi_h \cdot x))| \geq \eta^4/2. \tag{5.10}$$

Thus far we have not used any property of the underlying group G other than it is finite and abelian. We could continue doing this, proving Theorem 2.7 and extracting Theorem 2.3 as a corollary. However, for expository reasons we will now restrict to the finite-field geometry case $G = \mathbb{F}_5^n$ and give a complete proof of Theorem 2.3. The arguments there serve as a simplified model which will help motivate the general case.

6. The finite-field case

We now restrict attention to a finite-field geometry setting $G = \mathbb{F}_5^n$ and prove Theorem 2.3 (i). In this section, then, $N = 5^n$. Thanks to Proposition 5.4, we have already isolated a phase derivative $h \mapsto \xi_h$ which exhibits some linear behaviour. The first step (following Gowers [25, 27]) is to show that ξ in fact matches up with a linear phase function on a large subspace; then we shall show that in fact ξ matches up with a *self-adjoint* linear phase function on a large subspace (this is the substantially new part of the argument). Finally, we shall again follow Gowers [25] and conjugate f by a quadratic phase to eliminate this linear phase derivative and conclude the argument.

6.1. Step 1: linearization of the phase derivative

In this subsection we establish the following proposition.

Proposition 6.1 (graphs have large linear component [25, 27]). *Let $H' \subseteq \mathbb{F}_5^n$ and let $\xi : H' \rightarrow (\mathbb{F}_5^n)^*$ be a function whose graph $\Gamma' := \{(h, \xi_h) : h \in H'\} \subseteq \mathbb{F}_5^n \times (\mathbb{F}_5^n)^*$ obeys the estimates*

$$K^{-1}N \leq |\Gamma'| \leq |9\Gamma' - 8\Gamma'| \leq KN \tag{6.1}$$

for some $K \geq 1$. Then there exists a linear subspace $V \subseteq \mathbb{F}_5^n$ with the codimension bound

$$n - \dim(V) \leq (2K)^6$$

and a translate $x_0 + V$ of this subspace, together with a linear transformation $M : V \rightarrow (\mathbb{F}_5^n)^*$ and an element $\xi_0 \in (\mathbb{F}_5^n)^*$ such that

$$\mathbb{E}_{h \in V}(1_H(x_0 + h)1_{\xi_{x_0+h}=2Mh+\xi_0}) \geq (2K)^{-4}. \tag{6.2}$$

Proof. Note that while Γ' is a graph, the slightly larger sets $k\Gamma' - l\Gamma'$ need not be graphs. The next lemma shows that this can be rectified by passing to an appropriate subset $\Gamma'' \subseteq \Gamma'$.

Lemma 6.2. *There exists a subset Γ'' of Γ' with*

$$|\Gamma''| \geq N/5K^3 \tag{6.3}$$

such that $4\Gamma'' - 4\Gamma''$ is a graph.

Remarks 6.3. When $G = \mathbb{Z}/N\mathbb{Z}$ is a cyclic group of prime order, this lemma is essentially [27, Lemma 7.5], and our arguments both here and in §9 are in a similar spirit. Curiously, the argument of [27] uses the fact that $\mathbb{Z}/N\mathbb{Z}$ is a field, and so this is a rare instance of the cyclic group case being somewhat easier than the general group case. The conclusion can also be rephrased as an assertion that the map $h \mapsto \xi_h$ is a Freiman 8-homomorphism on H'' .

Proof. Let $A \subseteq (\mathbb{F}_5^n)^*$ be the set of all ξ such that $(0, \xi) \in 8\Gamma' - 8\Gamma'$. Observe that since Γ' is a graph, we have $|\Gamma' + A| = |\Gamma'| |A|$. On the other hand, $\Gamma' + A$ is contained in $9\Gamma' - 8\Gamma'$. Applying (6.1), we conclude that $|A| \leq K^2$. Now let $m = \lceil \log_5 |A| \rceil$ and let $\Psi : (\mathbb{F}_5^n)^* \rightarrow \mathbb{F}_5^m$ be a randomly chosen linear transformation from $(\mathbb{F}_5^n)^*$ to \mathbb{F}_5^m . Observe that for each non-zero $\xi \in A$, we have $\Psi(\xi) \neq 0$ with probability at least $1 - 5^{-m}$. Thus we see that with non-zero probability Ψ is non-zero on all of $A \setminus \{0\}$.

Fix Ψ with the above properties, let c be a randomly selected point in \mathbb{F}_5^m and define $\Gamma'' := \{(h, \xi_h) \in \Gamma' : \Psi(\xi_h) = c\}$. Then the size of $|\Gamma''|$ is expected to be at least $|\Gamma'|/5^m \geq N/5K^2$. Also, observe that if $(0, \xi)$ lies in $8\Gamma'' - 8\Gamma''$, then ξ lies in A and (by linearity of Ψ) $\Psi(\xi) = 0$, hence $\xi = 0$. Since $8\Gamma'' - 8\Gamma''$ is the difference set of $4\Gamma'' - 4\Gamma''$, this implies that $4\Gamma'' - 4\Gamma''$ is a graph. The claim follows. \square

Define H'' by $\Gamma'' := \{(h, \xi_h) : h \in H''\}$. Now we use a result of Bogolyubov [7] (which we shall also use later in this paper) to obtain some control on the set $2H'' - 2H''$.

Lemma 6.4 (Bogolyubov lemma [7]). *Let A be a subset of a finite additive group G such that $|A| \geq \delta|G|$. Then there exists a set $S \subseteq \hat{G}$ with $|S| \leq 2\delta^{-2}$ such that $B(S, \frac{1}{4}) \subseteq 2A - 2A$.*

Remark 6.5. Somewhat sharper versions of this lemma are known (see [12]) but we will not need these here.

Proof. From Fourier inversion we have

$$1_A(x) = \sum_{\xi \in \hat{G}} \hat{1}_A(\xi) e(\xi \cdot x).$$

Convolving this with itself four times we obtain

$$1_A * 1_A * 1_{-A} * 1_{-A}(x) = \sum_{\xi \in G} |\hat{1}_A(\xi)|^4 e(\xi \cdot x),$$

where $f * g(x) := \mathbb{E}_y f(y)g(x - y)$ is the normalized convolution operation. Since the left-hand side is only non-zero on $2A - 2A$, we conclude that

$$\left\{ x \in G : \operatorname{Re} \sum_{\xi \in G} |\hat{1}_A(\xi)|^4 e(\xi \cdot x) > 0 \right\} \subseteq 2A - 2A.$$

Let us now study the sum on the left-hand side. Let $0 < \alpha < \delta$ be a parameter to be chosen later and let $S := \{\xi \in G : |\hat{1}_A(\xi)| \geq \alpha\}$ denote the large Fourier coefficients of

A. Since $\alpha < \mathbb{E}(1_A)$, we have $0 \in S$. Also, from the Plancherel identity

$$\sum_{\xi \in G} |\hat{1}_A(\xi)|^2 = \mathbb{E}(1_A) \geq \delta$$

and Chebyshev’s inequality we have an upper bound on the cardinality of S :

$$|S| \leq \delta/\alpha^2.$$

Now suppose that $x \in B(S, \frac{1}{4})$. Then we have $\cos(2\pi\xi \cdot x) \geq 0$ for all $\xi \in S$. Thus

$$\begin{aligned} \operatorname{Re} \sum_{\xi \in G} |\hat{1}_A(\xi)|^4 e(\xi \cdot x) &= \sum_{\xi \in S} |\hat{1}_A(\xi)|^4 \cos(2\pi\xi \cdot x) + \sum_{\xi \notin S} |\hat{1}_A(\xi)|^4 \cos(2\pi i\xi \cdot x) \\ &\geq |\hat{1}_A(0)|^4 - \sum_{\xi \notin S} |\hat{1}_A(\xi)|^4 \\ &\geq \delta^4 - \sum_{\xi \in G} \alpha^2 |\hat{1}_A(\xi)|^2 \\ &\geq \delta^4 - \alpha^2 \delta \end{aligned}$$

by another application of Plancherel’s theorem. Thus if we set $\alpha := \delta^{3/2}/\sqrt{2}$ (for instance), the claim follows. □

We apply Lemma 6.4 with $G = \mathbb{F}_5^n$ and $A = H''$, the set coming from Lemma 6.2. The set S that we produce satisfies $|S| \leq 50K^6$. Let $V \subseteq \mathbb{F}_5^n$ be the subspace $V := \{x \in \mathbb{F}_5^{n+m} : x \cdot \xi = 0 \text{ for all } \xi \in S\}$. Then by linear algebra we have

$$\dim(V) \geq n - |S| \geq n - 50K^6 \tag{6.4}$$

and, since clearly $V \subseteq B(S, \frac{1}{4})$, we have

$$V \subseteq 2H'' - 2H''.$$

Thus there exists a map $M : V \rightarrow \hat{G}$ such that

$$Z := \{(h, 2Mh) : h \in V\}$$

is a subset of $2\Gamma'' - 2\Gamma''$; since $2\Gamma'' - 2\Gamma''$ is a graph and contains the origin, we have $M0 = 0$. Also, since $4\Gamma'' - 4\Gamma''$ is a graph and V is closed under addition, we see that $M(h + h') = Mh + Mh'$ for all $h, h' \in V$, thus M is linear.

Consider the set $Z + \Gamma''$. On one hand, this set can be foliated into (say) L disjoint cosets of the linear space Z . On the other hand, it is contained in $3\Gamma'' - 2\Gamma''$, and

$$|3\Gamma'' - 2\Gamma''| \leq |9\Gamma' - 8\Gamma'| \leq KN.$$

This gives the bound $|Z| \leq KN/L$. Since $Z + \Gamma''$ also contains Γ'' , we thus see from the pigeonhole principle and (6.3) that there exists a coset $(x_0, \xi_0) + Z$ of Z such that

$$|\Gamma'' \cap ((x_0, \xi_0) + Z)| \geq |Z|/5K^4.$$

Since $\dim(Z) = \dim(V) \geq n - 50K^6$, the claim (6.2) follows. □

Combining Proposition 6.1 with Proposition 5.4, we quickly conclude the following proposition.

Proposition 6.6 (large $U^3(\mathbb{F}_5^n)$ gives linear phase derivative). *Let $G = \mathbb{F}_5^n$, and let $f : G \rightarrow \mathcal{D}$ be a bounded function such that $\|f\|_{U^3(G)} \geq \eta$ for some $\eta > 0$. Then there exists a linear subspace V of G with the codimension bound*

$$n - \dim(V) \leq 2^{C_1} \eta^{-C'_1} \tag{6.5}$$

and a translate $x_0 + V$ of this subspace, together with a linear transformation $M : V \rightarrow \hat{G}$ and an element $\xi_0 \in \hat{G}$ such that

$$\mathbb{E}_{h \in V} |\mathbb{E}_{x \in G} T^{x_0+h} f(x) \overline{f(x)} e(-(2Mh + \xi_0) \cdot x)| \geq 2^{-C_2} \eta^{C'_2}. \tag{6.6}$$

It is permissible to take $C_i, C'_i = 2^{16}$ for $i = 1, 2$.

6.2. Step 2: the symmetry argument

We now establish some symmetry properties on M , closely following the argument in Lemma 3.1. As we shall be focusing more on M than on f in this step, we shall suppress the terms involving f (and ξ_0 and x_0) using the \mathbf{b} notation. Indeed from (6.6) we have

$$|\mathbb{E}_{x \in G, h \in V} \mathbf{b}(x+h) \mathbf{b}(x) \mathbf{b}(h) e(-2Mh \cdot x)| \geq 2^{-C_2} \eta^{C'_2}. \tag{6.7}$$

Once again we use and similar tools to eliminate all the bounded functions. By Lemma 4.1 there exists $x_1 \in G$ such that

$$|\mathbb{E}_{x, h \in V} \mathbf{b}(x+x_1+h) \mathbf{b}(x+x_1) \mathbf{b}(h) e(-2Mh \cdot (x+x_1))| \geq 2^{-C_2} \eta^{C'_2}, \tag{6.8}$$

which after redefining the bounded functions to absorb the x_1 terms implies that

$$|\mathbb{E}_{x, h \in V} \mathbf{b}(x+h) \mathbf{b}(x) \mathbf{b}(h) e(-2Mh \cdot x)| \geq 2^{-C_2} \eta^{C'_2}. \tag{6.9}$$

Applying the Cauchy–Schwarz inequality (Lemma 4.3) to eliminate $\mathbf{b}(h)$, we see that

$$|\mathbb{E}_{x, y, h \in V} \mathbf{b}(y+h) \mathbf{b}(y+h) \mathbf{b}(x+h) \mathbf{b}(x) e(-2Mh \cdot (y-x))| \geq 2^{-2C_2} \eta^{2C'_2}. \tag{6.10}$$

Making the substitution $z = x + y + h$, this becomes

$$|\mathbb{E}_{x, y, z \in V} \mathbf{b}(z, y) \mathbf{b}(z, x) e(-2M(z-x-y) \cdot (y-x))| \geq 2^{-2C_2} \eta^{2C'_2}. \tag{6.11}$$

Absorbing as many phase terms into the functions \mathbf{b} as we can, we infer that

$$|\mathbb{E}_{x, y, z \in V} \mathbf{b}(z, y) \mathbf{b}(z, x) e(2\{x, y\})| \geq 2^{-2C_2} \eta^{2C'_2}, \tag{6.12}$$

where $\{x, y\}$ is the antisymmetric form defined in (3.13), that is to say, $\{x, y\} = Mx \cdot y - My \cdot x$. By the pigeonhole principle in z we conclude that

$$|\mathbb{E}_{x, y \in V} \mathbf{b}(y) \mathbf{b}(x) e(2\{x, y\})| \geq 2^{-2C_2} \eta^{2C'_2} \tag{6.13}$$

for some bounded functions $\mathbf{b}(y), \mathbf{b}(x)$. Applying the Cauchy–Schwarz inequality again to eliminate the $\mathbf{b}(x)$ factor, we deduce that

$$|\mathbb{E}_{x,y,y' \in V} \mathbf{b}(y)\mathbf{b}(y')e(2\{x,y'-y\})| \geq 2^{-4C_2}\eta^{4C'_2}. \tag{6.14}$$

By the triangle inequality we obtain

$$\mathbb{E}_{y,y' \in V} |\mathbb{E}_{x \in V} e(2\{x,y'-y\})| \geq 2^{-4C_2}\eta^{4C'_2}. \tag{6.15}$$

Making the substitution $h = y' - y$ we conclude that

$$\mathbb{E}_{h \in V} |\mathbb{E}_{x \in V} e(2\{x,h\})| \geq 2^{-4C_2}\eta^{4C'_2}. \tag{6.16}$$

The map $x \mapsto 2\{x,h\}$ is a homomorphism from V to \mathbb{R}/\mathbb{Z} . Thus if we write

$$W := \{h \in V : \{x,h\} = 0 \text{ for all } x \in V\},$$

then W is a linear subspace of V and

$$\mathbb{E}_{x \in V} e(2\{x,h\}) = 1_W(h).$$

Thus, in view of (6.16), we see that W is extremely large relative to V :

$$|W|/|V| = \mathbb{E}_{y,h \in V} 1_W(h) \geq 2^{-4C_2}\eta^{4C'_2}.$$

In particular, from (6.5) we have

$$n - \dim(W) \leq 2^{C_1}\eta^{-C'_1} + 4C'_2 \log_5(2/\eta) + 4C_2 \log_5 2 \leq 2^{C_1+1}\eta^{-C'_1}.$$

By construction of W we see that M is self-adjoint on W , that is to say,

$$Mw \cdot w' = Mw' \cdot w \quad \text{for all } w, w' \in W. \tag{6.17}$$

Let us remark that the analogy between this argument and that of §3 is exceptionally close. The seven equations (3.9), (3.10), (3.11), (3.12), (3.14), (3.15) and (3.16) are analogues of (6.9), (6.10), (6.11), (6.12), (6.13), (6.14) and (6.15), respectively.

6.3. Step 3: eliminating the quadratic phase component

We now give the final part of the proof of Theorem 2.3, which also closely follows the proof of Lemma 3.1. We return to (6.6), which we write as

$$|\mathbb{E}_{x \in G; h \in V} \mathbf{b}(h)\mathbf{b}(x+h)\overline{f(x)}e(-2Mh \cdot x)| \geq 2^{-C_2}\eta^{C'_2},$$

where we have distributed the phase factor $e(-\xi_0 \cdot x) = e(-\xi_0 \cdot (x+h))e(\xi_0 \cdot h)$ among the functions \mathbf{b} . Here the focus will be on the $\overline{f(x)}$ factor, the aim being to demonstrate that this function exhibits some quadratic bias. We first observe that the simple averaging argument of Lemma 4.1 allows us to find $h' \in V$ such that

$$|\mathbb{E}_{x \in G; h \in W} \mathbf{b}(h+h')\mathbf{b}(x+h+h')\overline{f(x)}e(-2M(h+h') \cdot x)| \geq 2^{-C_2}\eta^{C'_2}.$$

Once again we can absorb $e(-2Mh' \cdot x)$ into the functions \mathbf{b} , and conclude that

$$|\mathbb{E}_{x \in G; h \in W} \mathbf{b}(h) \mathbf{b}(x+h) \overline{f(x)} e(-2Mh \cdot x)| \geq 2^{C_2} \eta^{C'_2},$$

which implies that

$$|\mathbb{E}_{y \in G; x, h \in W} \mathbf{b}(h) \mathbf{b}(x+h+y) \overline{f(x+y)} e(-2Mh \cdot (x+y))| \geq 2^{C_2} \eta^{-C'_2}.$$

Using the triangle inequality we deduce that

$$\mathbb{E}_{y \in G} |\mathbb{E}_{x, h \in W} \mathbf{b}(h, y) \mathbf{b}(x+h, y) \overline{f(x+y)} e(-2Mh \cdot x)| \geq 2^{C_2} \eta^{-C'_2}. \tag{6.18}$$

Now we observe from (6.17) that we have the identity

$$2Mh \cdot x = M(x+h) \cdot (x+h) - Mx \cdot x - Mh \cdot h,$$

and hence

$$e(-2Mh \cdot x) = \mathbf{b}(x+h) \mathbf{b}(h) e(Mx \cdot x).$$

Therefore, (6.18) implies that

$$\mathbb{E}_{y \in G} |\mathbb{E}_{x, h \in W} \mathbf{b}(h, y) \mathbf{b}(x+h, y) \overline{f(x+y)} e(-Mx \cdot x)| \geq 2^{-C_2} \eta^{C'_2}.$$

Applying Lemma 4.4 for each $y \in G$ separately and with $B = B' = W$, we conclude that

$$\mathbb{E}_{y \in G} \|\overline{f(x+y)} e(Mx \cdot x)\|_{u^2(W)} \geq 2^{-C_2} \eta^{C'_2},$$

and hence of course

$$\mathbb{E}_{y \in G} \|\overline{f(x+y)} e(Mx \cdot x)\|_{u^3(W)} \geq 2^{-C_2} \eta^{C'_2}.$$

But the $u^3(W)$ norm is invariant under quadratic phase modulations, conjugation and translation, and so we have

$$\mathbb{E}_{y \in G} \|f\|_{u^3(y+W)} \geq 2^{-C_2} \eta^{C'_2}$$

which gives (2.6).

7. Application: Szemerédi’s theorem in finite-field geometries

As a sample application of Theorem 2.3, we can now prove a quantitative Szemerédi theorem for progressions of length four in \mathbb{F}_5^n . If G is any finite abelian group of order N , where $(6, N) = 1$, we define $r_4(G)$ to be the cardinality of the largest set $A \subseteq G$ which does not contain four distinct elements in arithmetic progression.

Theorem 7.1 (Szemerédi theorem for $r_4(\mathbb{F}_5^n)$). *Write $N = 5^n$. Then we have the bound*

$$r_4(\mathbb{F}_5^n) \ll N(\log \log N)^{-2^{-21}}.$$

Remark 7.2. In §11 we will prove a similar result for an arbitrary G . Although that result will supersede the present one, the proof is quite a bit more complicated so we give the finite-field argument separately now. In [35] we improve the bound to $r_4(\mathbb{F}_5^n) \ll N(\log N)^{-c}$ using substantially lengthier arguments.

As with all analytic arguments for proving Szemerédi-type theorems, the key step is the establishment of the following ‘density increment’ result.

Proposition 7.3. *Let $\delta > 0$, suppose that*

$$n > 6(2/\delta)^{2^{20}}, \tag{7.1}$$

and let $A \subseteq \mathbb{F}_5^n$ be a set with size at least δN . Suppose that A contains no four-term arithmetic progressions. Then we can find an affine subspace $x_0 + V$ of \mathbb{F}_5^n with dimension $\dim(V) \geq n/3$ such that we have the density increment

$$\mathbb{E}_{x \in x_0 + V} 1_A(x) \geq \mathbb{E}_{x \in \mathbb{F}_5^n} 1_A(x) + (\delta/2)^{2^{20}}.$$

Proof. Write

$$\alpha := \mathbb{E}_{x \in \mathbb{F}_5^n} 1_A(x) = |A|/N.$$

By Corollary 1.12 and the lower bound $N \geq 2/\delta^3$ (which is very much a consequence of (7.1)!), we conclude that

$$\|1_A - \alpha\|_{U^3(\mathbb{F}_5^n)} \geq \delta^3/8.$$

Applying Theorem 2.3, we can thus find a subspace $W \leq \mathbb{F}_5^n$ with codimension at most $(2/\delta)^{4C}$ and quadratic phase functions ϕ_y for each $y \in \mathbb{F}_5^n$ such that

$$\mathbb{E}_{y \in \mathbb{F}_5^n} |\mathbb{E}_{x \in y+W} (1_A(x) - \alpha)e(-\phi_y(x))| \geq (\delta/2)^{4C}. \tag{7.2}$$

On the other hand, we may assume that

$$\mathbb{E}_{x \in y+W} (1_A(x) - \alpha) < \frac{1}{4}(\delta/2)^{4C}$$

for all $y \in \mathbb{F}_5^n$, since the proposition is immediate otherwise. Now since $\mathbb{E}_{x \in y+W} (1_A(x) - \alpha)$ has mean zero, we thus conclude that

$$\mathbb{E}_{y \in \mathbb{F}_5^n} |\mathbb{E}_{x \in y+W} (1_A(x) - \alpha)| < \frac{1}{2}(\delta/2)^{4C}.$$

Subtracting this from (7.2) and using the pigeonhole principle, we infer that there exists $y \in \mathbb{F}_5^n$ such that

$$|\mathbb{E}_{x \in y+W} (1_A(x) - \alpha)e(-\phi_y(x))| > \frac{1}{2}(\delta/2)^{4C} + |\mathbb{E}_{x \in y+W} (1_A(x) - \alpha)|.$$

Now observe (using Lemma 3.1) that ϕ_y takes values in the set $T = \{0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}\}$. For each $t \in T$ let S_t be the quadratic surface $S_t := \{x \in y + W : \phi_y(x) = t\}$. Then by the triangle inequality we have

$$|\mathbb{E}_{x \in y+W} (1_A(x) - \alpha)e(-\phi_y(x))| \leq \sum_{t \in T} |\mathbb{E}_{x \in y+W} (1_A(x) - \alpha)1_{S_t}(x)|$$

while

$$|\mathbb{E}_{x \in y+W}(1_A(x) - \alpha)| \geq -\mathbb{E}_{x \in y+W}(1_A(x) - \alpha) = -\sum_{t \in T} \mathbb{E}_{x \in y+W}(1_A(x) - \alpha)1_{S_t}(x).$$

Combining these estimates, and using the pigeonhole principle, we deduce that there exists $t \in T$ such that

$$|\mathbb{E}_{x \in y+W}(1_A(x) - \alpha)1_{S_t}(x)| > \frac{1}{2}(\delta/2)^{4C} \mathbb{E}_{x \in y+W}1_{S_t}(x) - \mathbb{E}_{x \in y+W}(1_A(x) - \alpha)1_{S_t}(x),$$

and hence

$$\mathbb{E}_{x \in y+W}(1_A(x) - \alpha)1_{S_t}(x) \geq \frac{1}{4}(\delta/2)^{4C} \mathbb{E}_{x \in y+W}1_{S_t}(x). \tag{7.3}$$

This gives a density increment on a quite large quadratic hypersurface $S_t \cap (y + W)$. Our job is now to convert this into a density increment on a subspace. Observe from Lemma 3.1 that we can write $S_t \cap (y + W)$ in the form

$$S_t \cap (y + W) = y + \{x \in W : \frac{1}{2}x \cdot Mx + v \cdot x = c\}$$

for some self-adjoint linear transformation $M : W \rightarrow \hat{W}$, some $v \in W^*$ and some $c \in F$. We now locate a large subspace on which M is degenerate. To this end we need a simple lemma.* For future reference we shall phrase this lemma for more general finite fields than \mathbb{F}_5 .

Lemma 7.4 (Gauss sum lemma). *Let F be a finite field of odd characteristic, let W be a vector space over F and let $M : W \rightarrow \hat{W}$ be a self-adjoint linear transformation. If $\dim(W) \geq 3$, then there exists a non-zero $x \in W$ such that $x \cdot Mx = 0$.*

Proof. If M has a non-trivial kernel, then the claim is easy, so assume that M has no kernel. Then a standard Gauss sum computation (using (4.1), for instance) shows that $|\mathbb{E}_{x \in W} e(a(x \cdot Mx))|^2 = |F|^{-\dim(W)}$ for all $a \in |F| \setminus \{0\}$. But by Fourier inversion we have

$$\mathbb{E}_{x \in W} 1_{x \cdot Mx=0} = \frac{1}{|F|} \sum_{a \in F} \mathbb{E}_{x \in W} e(a(x \cdot Mx)) \geq \frac{1}{|F|} (1 - (|F| - 1) \cdot |F|^{-\dim(W)/2}).$$

Since $\dim(W) \geq 3$ we thus see that $\{x \in W : x \cdot Mx = 0\}$ must contain at least one non-zero element, and we are done. \square

In our situation, the space W has dimension substantially larger than 3—in fact, $\dim(W) \geq n - (2/\delta)^{4C}$. Let U be a subspace of W that is degenerate in the sense that $x \cdot My = 0$ for all $x, y \in U$ and that is maximal with respect to this property. We claim that $\dim(U) \geq n/2 - (2/\delta)^{4C}$. Indeed if this were not the case, the space $U^\perp := \{x \in W : x \cdot My = 0 \text{ for all } y \in U\}$ would have at least three more dimensions than U (in fact, vastly more than this), and one could apply the previous lemma to U^\perp/U to contradict the maximality of U .

* One could also proceed here using the theory of Witt groups, but that would be far more advanced technology than what is actually needed here.

Let U be as above. By splitting (7.3) into cosets of U and applying the pigeonhole principle, we may find a coset $z + U$ of U in $y + W$ such that

$$\mathbb{E}_{x \in z+U} (1_A(x) - \alpha) 1_{S_t}(x) \geq \frac{1}{4} (\delta/2)^{4C} \mathbb{E}_{x \in z+U} 1_{S_t}(x).$$

Note that on this space $z + U$, the form $x \cdot Mx$ becomes linear, which means that $S_t \cap (z + U)$ is an affine subspace of $z + U$ of codimension at most 1. Letting $x_0 + V$ denote this subspace, and recalling that the constant C in Theorem 2.3 could certainly be taken to be 2^{17} , Proposition 7.3 follows. Note that (7.1) suffices to guarantee the stated bound $\dim(V) \geq n/3$. \square

Proof of Theorem 7.1. Suppose that $A \subseteq \mathbb{F}_5^n$ has cardinality δN yet contains no four-term arithmetic progression. Then by repeated application of Proposition 7.3 we may construct a sequence

$$\mathbb{F}_5^n \supseteq x_1 + V_1 \supseteq x_2 + V_2 \supseteq \dots$$

of affine subspaces such that $\dim(V_j) \geq n/3^j$ and

$$\mathbb{E}_{x \in x_j + V_j} 1_A(x) \geq \delta + j(\delta/2)^{2^{20}}, \quad (7.4)$$

provided only that at all stages the condition (7.1) is satisfied, that is to say,

$$\dim(V_j) > 6(2/\delta)^{2^{20}}. \quad (7.5)$$

Equation (7.4) is impossible if $j \geq j_0 = (2/\delta)^{2^{20}}$, and so (7.5) must be violated by some $j \leq j_0$. This means that

$$n > 6 \times (2/\delta)^{2^{20}} \times 3^{(2/\delta)^{2^{20}}},$$

which certainly implies that $\delta \ll (\log \log N)^{-2^{21}}$. \square

Remark 7.5. A more-or-less identical argument shows that

$$r_4(\mathbb{F}_p^n) \ll N(\log \log N/p)^{-c},$$

uniformly for all primes $p \geq 5$, for some absolute constant $c > 0$. In combination with Gowers's result that $r_4(\mathbb{Z}/N\mathbb{Z}) \ll N(\log \log N)^{-c}$, this may be used to give a fairly cheap proof that $r_4(G) = o(N)$ for all finite abelian G , a result which was first obtained by Frankl and Rödl [17] by rather different means. The key to the argument is that G contains either a large subgroup of the form \mathbb{F}_p^n or else a large cyclic subgroup. The bound obtained is of the form

$$r_4(G) \ll N(\log \log \log \log N)^{-c};$$

we omit the details since a far superior bound will be obtained in § 11.

8. Some results on Bohr sets

Let G be an arbitrary finite abelian group with $|G| = N$, let $S \subseteq \hat{G}$ be a set of d characters, and suppose that $\rho \in (0, 1)$ is a positive parameter. We will collect some basic facts about Bohr sets $B(S, \rho)$ which we will need to prove Theorem 2.7. These Bohr sets play the role that subspaces did in the finite geometry setting, with the quantity d corresponding, roughly speaking, to the codimension of the subspace.

Note that a Bohr set always contains 0 and is symmetric around the origin. In fact we have the following easy bounds on the size of Bohr sets.

Lemma 8.1 (bounds for the size of Bohr sets). *We have*

$$|B(S, \rho)| \geq \rho^d N$$

and

$$|B(S, 2\rho)| \leq 5^d |B(S, \rho)|.$$

Proof. By the triangle inequality, we see that for any $y = (y_\xi)_{\xi \in S}$ in the torus $(\mathbb{R}/\mathbb{Z})^S$ we have

$$\begin{aligned} \sum_{x \in G} \prod_{\xi \in S} 1_{\|\xi \cdot x - y_\xi\|_{\mathbb{R}/\mathbb{Z}} < \rho/2} &= \left| \left\{ x \in G : \sup_{\xi \in S} \|\xi \cdot x - y_\xi\|_{\mathbb{R}/\mathbb{Z}} < \rho/2 \right\} \right| \\ &\leq \left| \left\{ x \in G : \sup_{\xi \in S} \|\xi \cdot x\|_{\mathbb{R}/\mathbb{Z}} < \rho \right\} \right| \\ &= |B(S, \rho)|. \end{aligned}$$

Integrating this over all $y \in (\mathbb{R}/\mathbb{Z})^S$, we conclude that

$$\sum_{x \in G} \rho^d \leq |B(S, \rho)|,$$

which gives the first bound.

To establish the second bound we integrate the same expression but only over the cube $\{y : \sup_{\xi \in S} \|y_\xi\| \leq \frac{5}{2}\rho\}$. Note from the triangle inequality that this does not affect the components of the integral for which $x \in B(S, 2\rho)$. Thus we have

$$\sum_{x \in B(S, 2\rho)} \rho^d \leq |B(S, \rho)|(5\rho)^d,$$

which gives the second bound. □

Next we establish, following Bourgain, that regular Bohr sets (as defined in Definition 2.6) exist in abundance.

Lemma 8.2 (regular Bohr sets are ubiquitous [10]). *Let $0 < \varepsilon < 1$. Then there exists $\rho \in [\varepsilon, 2\varepsilon]$ such that $B(S, \rho)$ is regular.*

Proof. We may assume that S is non-empty since the claim is trivial otherwise. Let $f : [0, 1] \rightarrow \mathbb{R}$ be the function $f(a) := (1/d) \log_2 |B(S, 2^a \varepsilon)|$. Observe that f is non-decreasing in a , and from Lemma 8.1 we have $f(1) - f(0) \leq \log_2 5$.

Suppose we could find a , $0.1 \leq a \leq 0.9$, such that $|f(a') - f(a)| \leq 20|a - a'|$ for all $|a| \leq 0.1$. Then it is easy to see that the Bohr set $B(S, 2^a \varepsilon)$ is regular. Thus, it suffices to obtain an a with this property. This can be done directly from the Hardy–Littlewood maximal inequality (applied to the Lebesgue–Stieltjes measure df), or as follows. If no such a exists, then for every $a \in [0.1, 0.9]$ there exists an interval I of length at most 0.1 and with one endpoint equal to a , such that $\int_I df > \int_I 20 dx$. These intervals cover $[0.1, 0.9]$, which has measure 0.8. By the Vitali covering lemma,* one can thus find a finite subcollection of disjoint intervals I_1, \dots, I_n of total length $|I_1| + \dots + |I_n| \geq 0.8/5$ (say). But we then have

$$\log_2 5 \geq \int_0^1 df \geq \sum_{i=1}^n \int_{I_i} df \geq \sum_{i=1}^n \int_{I_i} 20 dx \geq \frac{0.8}{5} \times 20,$$

which is a contradiction. □

Next we show that given any small set of points in G , one can find a large Bohr set which avoids all of them except possibly zero. We will need this to develop the analogue of Lemma 6.2 (see Lemma 9.2 below).

Lemma 8.3 (separation lemma). *Let G be a finite additive group and let $A \subseteq G$ be a set of elements containing zero. Then there exists a set $S \subseteq \hat{G}$ with $|S| \leq 1 + \log_2 |A|$ such that $A \cap B(S, \frac{1}{4}) = \{0\}$.*

Proof. We induct on $|A|$. When $|A| = 1$ the claim is trivial (set $S = \emptyset$). Now suppose that $|A| \geq 2$ and the claim has already been proven for smaller sets A . Suppose that $2^n < |A| \leq 2^{n+1}$. Let $\xi \in \hat{G}$ be chosen randomly. Observe that for each $x \in A \setminus \{0\}$, the map $\xi \mapsto \xi \cdot x$ is a non-trivial group homomorphism from \hat{G} to \mathbb{R}/\mathbb{Z} , thus the random variable $\xi \cdot x$ is uniformly distributed over a cyclic subgroup of \mathbb{R}/\mathbb{Z} . In particular, we have

$$\mathbb{P}(x \in B(\xi, \frac{1}{4})) = \mathbb{P}(\|\xi \cdot x\|_{\mathbb{R}/\mathbb{Z}} < \frac{1}{4}) \leq \frac{1}{2}.$$

Summing this over all non-zero A , we conclude that

$$\mathbb{E}|(A \setminus \{0\}) \cap B(\xi, \frac{1}{4})| \leq (|A| - 1)/2.$$

In particular, we can find $\xi \in \hat{G}$ such that

$$|A \cap B(\xi, \frac{1}{4})| \leq \left\lfloor \frac{|A| + 1}{2} \right\rfloor \leq 2^n.$$

By the induction hypothesis we conclude that there exists a set S' of cardinality at most n such that

$$A \cap B(\xi, \frac{1}{4}) \cap B(S', \frac{1}{4}) = \{0\},$$

and the claim follows by setting $S := S' \cup \{\xi\}$. □

* One can also use the Besicovitch covering lemma at this point, which would in fact give slightly better bounds. Indeed, one can improve the constant $\frac{1}{5}$ to $\frac{1}{2}$ (see, for example, [15]).

Our next task is to investigate the Fourier-analytic behaviour of Bohr sets. In the first instance we deal with a concept somewhat more general than that of a Fourier coefficient, replacing a linear phase function by a locally linear phase function.

Lemma 8.4 (generalized Fourier decay). *Let $S \subseteq \hat{G}$, $|S| = d$, be a set of characters. Let $B := B(S, \rho)$ be a regular Bohr set and let $\phi : B(S, 2\rho) \rightarrow \mathbb{R}/\mathbb{Z}$ be a function which is locally linear in the sense that $\phi(x + y) = \phi(x) + \phi(y)$ whenever $x, y \in B(S, \rho)$. Suppose that*

$$|\mathbb{E}_{x \in B}(e(\phi(x)))| \geq \eta$$

for some $0 < \eta \leq 1$ (large generalized Fourier coefficient). Then ϕ is close to constant in the sense that for every $y \in B$ we have

$$\|\phi(y)\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{2^{12}d\|y\|_S}{\rho\eta^2}.$$

Proof. Let $y \in B$ and let $M \geq 0$ be the largest integer such that $M\|y\|_S \leq \eta\rho/400d$. If $M = 0$ then we have $\|y\|_S \geq \eta\rho/400d$, and the claim is trivial. By Lemma 4.2 (ii) we have

$$|\mathbb{E}_{x \in B}(e(\phi(x))) - \mathbb{E}_{x \in B} \mathbb{E}_{-M \leq n \leq M}(e(\phi(x + ny)))| \leq \eta/2$$

and hence, by the triangle inequality,

$$|\mathbb{E}_{x \in B} \mathbb{E}_{-M \leq n \leq M} e(\phi(x + ny))| \geq \eta/2.$$

On the other hand, by the local linearity of ϕ we have $\phi(x + ny) = \phi(x) + n\phi(y)$ for all $|n| \leq M$ and hence $e(\phi(x + ny)) = \mathbf{b}(x)e(n\phi(y))$. By the triangle inequality we conclude that

$$|\mathbb{E}_{-M \leq n \leq M} e(n\phi(y))| \geq \eta/2.$$

But by the geometric series formula, the left-hand side is bounded by $2/M\|\phi(y)\|_{\mathbb{R}/\mathbb{Z}}$. This implies that

$$\|\phi(y)\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{4}{M\eta} \leq \frac{3200d}{\rho\eta^2} \|y\|_S,$$

which implies the result. □

As a corollary we see that the normalized Fourier transform of a Bohr set decays away from the ‘polar body’ of that Bohr set.* If $\xi \in \hat{G}$ then define

$$\|\xi\|_{B(S, \rho)} := \sup_{y \in B(S, \rho)} \|\xi \cdot y\|_{\mathbb{R}/\mathbb{Z}}.$$

Note that if $\xi \in S$ then $\|\xi\|_{B(S, \rho)} \leq \rho$.

* One could obtain much better Fourier localization properties by replacing the Bohr sets by smoother weight functions; see, for example, [30, 61] for examples of this approach. This also conveys the slight advantage that all weight functions can automatically be made regular. However, these functions have the disadvantage of being spread out in physical space, and we found it more convenient to use Bourgain’s machinery of regular Bohr sets from [10] instead.

Corollary 8.5 (Fourier decay). *Let $S \subseteq \hat{G}$ be a set of d characters, let $B := B(S, \rho)$ be a regular Bohr set and let $0 < \theta \leq 1$. Then for any $\xi \in \hat{G}$, we have*

$$|\mathbb{E}_{x \in B} e(\xi \cdot x)| \leq 64 \left(\frac{\theta d}{\|\xi\|_{B(S, \theta \rho)}} \right)^{1/2}.$$

Proof. Apply Lemma 8.4 with $\phi(x) := \xi \cdot x$, with $\eta := |\mathbb{E}_{x \in B} e(\xi \cdot x)|$ and with y being an arbitrary element of $B(S, \theta r)$. □

We can exploit this decay via a Tomas–Stein almost-orthogonality type argument (also used by Bombieri [9] in the context of the large sieve; see also [47]) to conclude the following corollary.

Corollary 8.6 (local Bessel inequality). *Let $S \subseteq \hat{G}$ be a set of d characters, let $B := B(S, \rho)$ be a regular Bohr set, let $0 < \theta \leq 1$ and let $\xi_1, \dots, \xi_k \in \hat{G}$ be frequencies such that $\|\xi_i - \xi_j\|_{B(S, \theta \rho)} \geq \delta$ for all $1 \leq i < j \leq k$ and some $\delta > 0$. Then*

$$\mathbb{E}_{x \in B} \left| \sum_{j=1}^k \mathbf{b}(j) e(\xi_j(x)) \right|^2 \leq k + 2^7 k^2 (\theta d / \delta)^{1/2}$$

for any bounded complex numbers $\mathbf{b}(j)$.

Proof. Thanks to Corollary 8.5 we have

$$\begin{aligned} \mathbb{E}_{x \in B} \left| \sum_{j=1}^k \mathbf{b}(j) e(\xi_j(x)) \right|^2 &= \sum_{1 \leq i, j \leq k} \mathbf{b}(i, j) \mathbb{E}_{x \in B} e((\xi_i - \xi_j) \cdot x) \\ &\leq k + 2 \sum_{1 \leq i < j \leq k} |\mathbb{E}_{x \in B} e((\xi_i - \xi_j) \cdot x)| \\ &\leq k + 2^7 k^2 (\theta d / \delta)^{1/2}, \end{aligned}$$

The claim follows. □

We can dualize the above corollary to give the following result. This allows us to generalize, to the relative setting, a frequently used consequence of Parseval’s identity: a large set $A \subseteq G$ cannot have too many large Fourier coefficients.

Corollary 8.7 (local Bessel inequality, dual version). *Let $S \subseteq \hat{G}$ be a set of d characters, let $B := B(S, \rho)$ be a regular Bohr set, let $0 < \theta, \eta \leq 1$ and suppose that $A \subseteq B$. Let*

$$\Gamma := \{ \xi \in \hat{G} : |\hat{1}_A(\xi)| \geq \eta \mathbb{E}(1_B) \}.$$

Then there exist frequencies $\xi_1, \dots, \xi_k \in \hat{G}$ with $k \leq 2/\eta^2$ such that any $\xi \in \Gamma$ is close to some ξ_i in the $\|\cdot\|_{B(S, \theta \rho)}$ norm:

$$\Gamma \subseteq \{ \xi \in \hat{G} : \|\xi - \xi_j\|_{B(S, \theta \rho)} \leq 2^{16} \theta d / \eta^4 \text{ for some } 1 \leq j \leq k \}. \tag{8.1}$$

Proof. Let $\{\xi_1, \dots, \xi_k\}$ be a maximal set of frequencies in Γ such that $\|\xi_i - \xi_j\|_{B(S, \theta\rho)} \geq \delta$. Then it is clear that (8.1) holds. For each $1 \leq j \leq k$ we have $\xi_j \in \Gamma$. Hence there exists a bounded complex number $\mathbf{b}(j)$ such that

$$\operatorname{Re} \mathbb{E}_{x \in B} \mathbf{b}(j) e(\xi_j(x)) 1_A(x) \geq \eta.$$

Summing this in j and applying the Cauchy–Schwarz inequality, we conclude that

$$\mathbb{E}_{x \in B} \left| \sum_{j=1}^k \mathbf{b}(j) e(\xi_j(x)) \right|^2 \geq \eta^2 k^2.$$

Applying Corollary 8.6, we conclude that $\eta^2 k^2 \leq k + \frac{1}{2} \eta^2 k^2$, and hence $k \leq 2/\eta^2$. The claim follows. \square

As a consequence, we can now generalize Bogolyubov’s argument (Lemma 6.4) to subsets of Bohr sets.

Lemma 8.8 (local Bogolyubov lemma). *Let $S \subseteq \hat{G}$ be a set of d characters and let $B := B(S, \rho)$. Let $A \subseteq B$ be a set with $|A| = \delta|B|$. Then there exists a set $S' \subseteq \hat{G}$ with $|S'| \leq 2^7 \delta^{-3}$ such that $B(S \cup S', 2^{-33} \delta^6 \rho/d) \subseteq 2A - 2A$.*

Proof. It is convenient to replace $2A - 2A$ by the slightly smaller set $A + A' - A - A'$. Let $\varepsilon = \delta/400d$. By Lemma 8.2, there exists $\rho' \in [\varepsilon\rho, 2\varepsilon, \rho]$ such that the Bohr set $B' := B(S, \rho')$ is regular. By Lemma 4.2 (iii) we can find $x \in B$ such that

$$\mathbb{E}_{y \in x+B'} 1_A(y) \geq \mathbb{E}_{y \in B} 1_A(y) - 200d\varepsilon \geq \delta/2.$$

Let $A' := A \cap (x + B')$. From the Fourier inversion formulae

$$1_A(x) = \sum_{\xi \in \hat{G}} \hat{1}_A(\xi) e(\xi \cdot x) \quad \text{and} \quad 1_{A'}(x) = \sum_{\xi \in \hat{G}} \hat{1}_{A'}(\xi) e(\xi \cdot x),$$

we conclude that

$$1_A * 1_{A'} * 1_{-A} * 1_{-A'}(x) = \sum_{\xi \in \hat{G}} |\hat{1}_A(\xi)|^2 |\hat{1}_{A'}(\xi)|^2 e(\xi \cdot x). \tag{8.2}$$

In particular, applying (8.2) with $x = 0$ we conclude that

$$\mathbb{E}_{x \in G} |1_A * 1_{A'}(x)|^2 = \sum_{\xi \in \hat{G}} |\hat{1}_A(\xi)|^2 |\hat{1}_{A'}(\xi)|^2.$$

The function $1_A * 1_{A'}$ is supported on $B(S, \rho + \rho')$, which has cardinality at most $2|B|$ since B is regular and $\varepsilon < 1/200d$. Thus by the Cauchy–Schwarz inequality

$$\begin{aligned} \mathbb{E}_{x \in G} |1_A * 1_{A'}(x)|^2 &\geq (\mathbb{E}_{x \in G} 1_A * 1_{A'}(x))^2 / 2\mathbb{E}(1_B) \\ &= \mathbb{E}(1_A)^2 \mathbb{E}(1_{A'})^2 / 2\mathbb{E}(1_B) \\ &= \frac{1}{2} (\mathbb{E}_{y \in B} 1_A(y))^2 (\mathbb{E}_{y \in x+B'} 1_A(y))^2 \mathbb{E}(1_B) \mathbb{E}(1_{B'})^2 \\ &\geq \frac{1}{8} \delta^4 K^{-4} \mathbb{E}(1_B) \mathbb{E}(1_{B'})^2. \end{aligned}$$

Thus we have

$$\sum_{\xi \in \hat{G}} |\hat{1}_A(\xi)|^2 |\hat{1}_{A'}(\xi)|^2 \geq \frac{1}{8} \delta^4 \mathbb{E}(1_B) \mathbb{E}(1_{B'})^2. \tag{8.3}$$

Now let

$$R := \{ \xi \in G : |\hat{1}_{A'}(\xi)| \geq \frac{1}{8} \delta^{3/2} \mathbb{E}(1_{B'}) \}$$

and let $x \in B(R, \frac{1}{10})$. Then, by taking real parts of both sides of (8.2), we conclude that

$$\begin{aligned} 1_A * 1_{A'} * 1_{-A} * 1_{-A'}(x) &= \sum_{\xi \in \hat{G}} |\hat{1}_A(\xi)|^2 |\hat{1}_{A'}(\xi)|^2 \cos(2\pi \xi \cdot x) \\ &\geq \sum_{\xi \in R} |\hat{1}_A(\xi)|^2 |\hat{1}_{A'}(\xi)|^2 \cos(2\pi/10) - \sum_{\xi \notin R} |\hat{1}_A(\xi)|^2 |\hat{1}_{A'}(\xi)|^2 \\ &= \sum_{\xi \in \hat{G}} |\hat{1}_A(\xi)|^2 |\hat{1}_{A'}(\xi)|^2 \cos(2\pi/10) \\ &\quad - (\cos(2\pi/10) + 1) \sum_{\xi \notin R} |\hat{1}_A(\xi)|^2 |\hat{1}_{A'}(\xi)|^2 \\ &\geq \frac{1}{2} \left(\sum_{\xi \in \hat{G}} |\hat{1}_A(\xi)|^2 |\hat{1}_{A'}(\xi)|^2 \right) - 2 \sum_{\xi \notin R} |\hat{1}_A(\xi)|^2 |\hat{1}_{A'}(\xi)|^2 \\ &\geq \frac{1}{16} \delta^4 \mathbb{E}(1_B) \mathbb{E}(1_{B'})^2 - \frac{1}{32} \delta^3 \mathbb{E}(1_{B'})^2 \sum_{\xi \in G} |\hat{1}_A(\xi)|^2 \end{aligned}$$

using (8.3) and the definition of R . On the other hand, from Plancherel's identity we have

$$\sum_{\xi \in G} |\hat{1}_A(\xi)|^2 = \mathbb{E}(1_A) \leq \delta \mathbb{E}(1_B),$$

and hence

$$1_A * 1_{A'} * 1_{-A} * 1_{-A'}(x) \geq \left(\frac{\delta^4}{16} - \frac{\delta^4}{32} \right) \mathbb{E}(1_B) \mathbb{E}(1_{B'})^2 > 0,$$

which implies that x is contained in $A + A' - A - A'$ and hence in $2A - 2A$. Hence we have

$$B(R, \frac{1}{10}) \subseteq 2A - 2A.$$

We are not done yet, because we do not have good bounds for $|R|$. Let $\theta > 0$ be a small parameter to be chosen later. Invoking Corollary 8.7, we conclude the existence of frequencies $\xi_1, \dots, \xi_k \in \hat{G}$ with $k \leq 128\delta^3$ such that

$$R \subseteq \{ \xi \in \hat{G} : \|\xi - \xi_j\|_{B(S, \theta\rho)} \leq 2^{28} \delta^{-6} \theta d \text{ for some } 1 \leq j \leq k \}.$$

Let $S' := \{ \xi_1, \dots, \xi_k \}$. If $x \in B(S \cup S', \theta\rho)$, then, in particular, $x \in B(S, \theta\rho)$. Thus if $\xi \in R$, then by the preceding inclusion we have

$$\|\xi \cdot x - \xi_j \cdot x\|_{\mathbb{R}/\mathbb{Z}} \leq 2^{28} \delta^{-6} \theta d \text{ for some } j.$$

Also, since $x \in B(S', \theta\rho)$, we get

$$\|\xi_j \cdot x\|_{\mathbb{R}/\mathbb{Z}} \leq \theta\rho \leq \theta;$$

by the triangle inequality we then obtain

$$\|\xi \cdot x\|_{\mathbb{R}/\mathbb{Z}} \leq 2^{29}\delta^{-6}\theta d.$$

We thus conclude that

$$B(S \cup S', \theta\rho) \subseteq B(R, 2^{29}\delta^{-6}\theta d).$$

Thus if we choose $\theta := 2^{-33}\delta^6/d$, we have $B(S \cup S', \theta\rho) \subseteq B(R, \frac{1}{10})$, and since $B(R, \frac{1}{10}) \subseteq 2A - 2A$ the claim follows. \square

9. The general group case

We now prove Theorem 2.7, which generalizes Theorem 2.3 to the case of arbitrary finite additive groups G . We begin by disposing of part (ii) of the theorem, which is rather easier to establish than (i).

Proof. Recall that in order to establish Theorem 2.7 (ii) we are to prove that if $S \subseteq \hat{G}$ is a set of d characters, if $B = B(S, \rho)$ is a regular Bohr set, if $f : G \rightarrow \mathcal{D}$ is a bounded function and if $\|f\|_{u^3(y+B)} \geq \eta$, then we have

$$\|f\|_{U^3(G)} \geq (\eta^3 \rho^2 / C d^3)^d$$

for some absolute constant C . By translation invariance we may take $y = 0$. Let $\phi : B \rightarrow \mathbb{R}/\mathbb{Z}$ be a locally quadratic phase function on B and suppose that

$$|\mathbb{E}_{x \in B}(f(x)e(-\phi(x)))| = \eta.$$

It turns out to be convenient to have ϕ defined, and to be a quadratic form, on a slightly larger Bohr set than B . In general this is not possible, but the same effect can be achieved by first passing to a smaller Bohr set. In fact, in the argument which follows we will have two smaller Bohr sets $B' = B(S, \rho')$ and $B'' = B(S, \rho'')$. Set $\varepsilon = c\eta/d$, where c is a small constant to be specified later. We will take $\rho' \in [\varepsilon\rho/2, \varepsilon\rho]$ so that B' is regular (this is possible by Lemma 8.2) and $\rho'' = \varepsilon\rho'$ (we will not require B'' to be regular). It will be convenient to write $\beta' := \mathbb{E}1_{B'}$ and $\beta'' := \mathbb{E}1_{B''}$. By Lemma 4.2 we have

$$\eta = \mathbb{E}_{z \in B} \mathbb{E}_{x \in z+B'}(f(x)e(-\phi(x))) + O(\varepsilon d).$$

Observe that the contribution from $z \in B \setminus B(S, (1 - 10\varepsilon)\rho)$ is at most $O(\varepsilon d)$, thanks to the regularity of B . Thus we in fact have

$$\mathbb{E}_{x \in B}(f(x)e(-\phi(x))) = \mathbb{E}_{z \in B} 1_{B(S, (1-10\varepsilon)\rho)}(z) \mathbb{E}_{x \in z+B'}(f(x)e(-\phi(x))) + O(\varepsilon d),$$

and hence by the pigeonhole principle there exists $z \in B(S, (1 - 10\varepsilon)\rho)$ such that

$$|\mathbb{E}_{x \in z+B'}(f(x)e(-\phi(x)))| \geq 2\eta/3 \tag{9.1}$$

provided that c is chosen sufficiently small.

We are going to compare $1_{z+B'}$, which is relevant to (9.1), with the function

$$F(x) := \frac{1}{\beta''} \mathbb{E}_h 1_{z+B''}(x+h) 1_{z+B'}(x+2h).$$

Write $B'_- = B(S, (1 - 2\epsilon)\rho')$ and $B'_+ = B(S, (1 + 2\epsilon)\rho')$. Note that if $x \in z + B'_-$ and $x + h \in z + B''$, then $x + 2h = 2(x + h) - x$ is contained in $z + B'$. For such x , then, we have $F(x) = 1$. Also, if $F(x) \neq 0$ then there is some h such that $x + h \in z + B''$ and $x + 2h \in z + B'$, which means that $x = 2(x + h) - (x + 2h)$ lies in $z + B'_+$. We have, then,

$$|F(x) - 1_{z+B'}(x)| \leq 21_{B'_+ \setminus B'_-}(x). \tag{9.2}$$

Now note further that if $x + h \in z + B''$ and $x + 2h \in z + B'$, then

$$x \in z + B' + 2B'' \subseteq B,$$

and also

$$x + 3h = 2(x + 2h) - (x + h) \in z + 2B' + B'' \subseteq B.$$

Both of these are consequences of the fact that $z \in B(S, (1 - 10\epsilon)\rho)$. In such an eventuality, then, all four of the elements $x, x + h, x + 2h, x + 3h$ lie in B and, since ϕ is quadratic, we have $(h \cdot \nabla_x)^3 \phi(x) = 0$, or, in other words,

$$\phi(x) - 3\phi(x + h) + 3\phi(x + 2h) - \phi(x + 3h) = 0. \tag{9.3}$$

Therefore,

$$\begin{aligned} & F(x)e(-\phi(x)) \\ &= \frac{1}{\beta''} \mathbb{E}_h 1_{z+B''}(x+h)e(3\phi(x+h))1_{z+B'}(x+2h) \times e(-3\phi(x+2h))e(\phi(x+3h)) \\ &=: \frac{1}{\beta''} \mathbb{E}_h g_1(x+h)g_2(x+2h)g_3(x+3h), \end{aligned}$$

say, where the functions g_1, g_2, g_3 are all bounded by 1. It is immediate from (9.2) that

$$|F(x)e(-\phi(x)) - 1_{z+B'}(x)e(-\phi(x))| \leq 21_{B'_+ \setminus B'_-}(x).$$

From (9.1) we infer, then, that

$$\begin{aligned} \frac{1}{\beta''} \mathbb{E}_{x,h} f(x)g_1(x+h)g_2(x+2h)g_3(x+3h) &= \mathbb{E}_x f(x)F(x)e(-\phi(x)) \\ &= \mathbb{E}_x f(x)1_{z+B'}(x)e(-\phi(x)) + O(\mathbb{E}1_{B'_+ \setminus B'_-}) \\ &= \mathbb{E}_x f(x)1_{z+B'}(x)e(-\phi(x)) + O(\epsilon d\beta'), \end{aligned}$$

the penultimate step being a consequence of the regularity of B' . If c is chosen small enough, this means in view of (9.1) that

$$\mathbb{E}_{x,h} f(x)g_1(x+h)g_2(x+2h)g_3(x+3h) \geq \beta'\beta''\eta/3. \tag{9.4}$$

However, Proposition 1.11 implies that we have

$$|\mathbb{E}_{x,h} f(x)g_1(x+h)g_2(x+2h)g_3(x+3h)| \leq \|f\|_{U^3(G)},$$

and hence from (9.4) we have

$$\|f\|_{U^3(G)} \geq \beta' \beta'' \eta / 3 \geq (\varepsilon^3 \rho^2 / 4)^d \eta / 3 \geq (\eta^3 \rho^2 / C d^3)^d$$

for some absolute constant C . □

Now we turn to the proof of Theorem 2.7 (i). As in §6 our starting point is Proposition 5.4, which the reader may care to recall now. The argument is closely analogous to that in §6, and hence in turn to that in §3.

9.1. Step 1: linearization of the phase derivative

We begin by carrying out the first major step, which is to show that the function $h \mapsto \xi_h$, which roughly speaking captures the derivative of the phase of f , matches up with a locally linear function.

Proposition 9.1. *Let $H' \subseteq G$ and suppose that $\xi : H' \rightarrow \hat{G}$ is a function whose graph*

$$\Gamma' := \{(h, \xi_h) : h \in H'\} \subseteq G \times \hat{G}$$

obeys the estimates

$$K^{-1}N \leq |\Gamma'| \leq |9\Gamma' - 8\Gamma| \leq KN$$

for some $K \geq 1$. Then there is a set $S \subseteq \hat{G}$,

$$d_1 := |S| \leq 2^{13} K^{26},$$

*a regular Bohr set $B_1 := B(S, \rho)$, where $\rho \in [\frac{1}{16}, \frac{1}{8}]$, elements $x_0 \in G$, $\xi \in \hat{G}$ and a function $M : B(S, \frac{1}{4}) \rightarrow \hat{G}$ such that the following two properties hold. Firstly, we have the local linearity condition**

$$M(h \pm h') = Mh \pm Mh' \quad \text{whenever } \|h\|_S, \|h'\|_S \leq \frac{1}{8}. \tag{9.5}$$

Secondly, we have

$$\mathbb{E}(1_{H'}(x_0 + h)1_{\xi_{x_0+h}=2Mh+\xi_0} | h \in B_1) \geq 2^{-6} K^{-13}.$$

Proof. As in the finite-field case, the first step is to refine the graph Γ' so that certain of the iterated sum-difference sets $k\Gamma' - l\Gamma'$ are also graphs. To do this we use the following generalization of Lemma 6.2.

Lemma 9.2. *There exists a subset $\Gamma'' = \{(h, \xi_h) : h \in H''\}$ of Γ' with*

$$|\Gamma''| \geq 2^{-6} K^{-13} N$$

such that $4\Gamma'' - 4\Gamma''$ is a graph.

* Recall that $\|h\|_S := \sup_{\xi \in S} \|\xi_h\|_{\mathbb{R}/\mathbb{Z}}$.

Proof. Let $A \subseteq \hat{G}$ be the set of all ξ such that $(0, \xi) \in 8\Gamma' - 8\Gamma''$. Arguing as in the proof of Lemma 6.2 we conclude that $|A| \leq K^2$. Applying Lemma 8.3, we can find a set $S \in \hat{G}$ with $|S| \leq 1 + 2 \log_2 K$ such that $A \cap B(S, \frac{1}{4}) = \{0\}$.

Let $\Psi : \hat{G} \rightarrow (\mathbb{R}/\mathbb{Z})^S$ be the homomorphism $\Psi(\xi) := (s(\xi))_{s \in S}$. Now let us cover the torus $(\mathbb{R}/\mathbb{Z})^S$ by $2^{6|S|} \leq 2^6 K^{12}$ cubes of side-length $\frac{1}{64}$. Since $|\Gamma'| \geq N/K$, the pigeonhole principle implies that there exists one of these cubes Q for which the set

$$\Gamma'' := \{(h, \xi_h) \in \Gamma' : \Psi(\xi_h) \in Q\}$$

has cardinality at least $2^{-6} K^{-13} N$. Now observe from the linearity of s that if $(0, \xi) \in 8\Gamma'' - 8\Gamma''$ then $\|s(\xi)\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{16}{64}$ for all $s \in S$. In other words, $\xi \in B(S, \frac{1}{4})$. But ξ also lies in A , and hence $\xi = 0$ by construction. Since $8\Gamma'' - 8\Gamma''$ is the difference set of $4\Gamma'' - 4\Gamma''$, we conclude that $4\Gamma'' - 4\Gamma''$ is a graph as desired. \square

Define H'' so that $\Gamma'' = \{(h, \xi_h) : h \in H''\}$. Applying Lemma 6.4 with $A := H''$, we obtain a set $S \subseteq \hat{G}$ with $|S| \leq 2^{13} K^{26}$ such that the Bohr set $B_0 := B(S, \frac{1}{4})$ is completely contained inside $2H'' - 2H''$. We will now work inside this Bohr set B_0 and pass to progressively narrower Bohr sets B_1, B_2, \dots when necessary. We will eventually end up at B_5 ; the dimension of B_j will be denoted d_j , and so in particular $d_0 = |S|$. It will turn out that $d_0 = d_1 = d_2 < d_3 = d_4 = d_5$, that is to say, it is only in passing from B_2 to B_3 that we shall increment the dimension of B_j . This is because that passage will involve Lemma 8.8.

Since $2\Gamma'' - 2\Gamma''$ is a graph, we can find a (unique) function $M : B_0 \rightarrow \hat{G}$ such that

$$\{(h, 2M(h)) : h \in B_0\} \subseteq 2H'' - 2H''.$$

Since $2\Gamma'' - 2\Gamma''$ contains 0, we conclude that $\phi(0) = 0$. Also, since $8\Gamma'' - 8\Gamma''$ is a graph, we see that

$$M(h_1) + M(h_2) = M(h'_1) + M(h'_2) \tag{9.6}$$

whenever $h_1, h_2, h'_1, h'_2 \in B_0$ is such that $h_1 + h_2 = h'_1 + h'_2$; in other words, M is a *Freiman homomorphism* of order 2. In particular, since $M0 = 0$, we have the local linearity relationship (9.5).

By Lemma 8.2 there is $\rho \in [\frac{1}{16}, \frac{1}{8}]$ such that the Bohr set $B_1 := B(S, \rho_1)$ is regular. By Lemma 4.1 there exists $x_0 \in G$ such that

$$\mathbb{E}_{h \in B_1} 1_{H''}(x_0 + h) \geq 2^{-6} K^{-13}.$$

Let us fix this x_0 , and set $A := \{h \in B_1 : x_0 + h \in H''\}$, so that we have

$$|A| \geq 2^{-6} K^{-13} |B_1|.$$

Observe that if $h, h' \in A$, then $(h - h', \xi_{x_0+h} - \xi_{x_0+h'})$ lies in $\Gamma'' - \Gamma''$, which is a subgraph of $2\Gamma'' - 2\Gamma''$. Thus we have $\xi_{x_0+h} - \xi_{x_0+h'} = 2M(h - h')$. Combining this with (9.5), we conclude that there exists $\xi_0 \in \hat{G}$ such that

$$\xi_{x_0+h} = \xi_0 + 2Mh \quad \text{for all } h \in A.$$

This concludes the proof of Proposition 9.1. \square

Combining Proposition 9.1 with Proposition 5.4 leads immediately to the following, which generalizes Proposition 6.6 to arbitrary G .

Proposition 9.3 (large $U^3(G)$ norm implies locally linear phase derivative). *Let G be an arbitrary finite additive group and let $f : G \rightarrow \mathcal{D}$ be a bounded function such that $\|f\|_{U^3(G)} \geq \eta$ for some $\eta > 0$. Then there exists a set $S \subseteq \hat{G}$ with*

$$d_1 := |S| \leq 2^{C_3} \eta^{-C'_3},$$

a regular Bohr set $B_1 := B(S, \rho) \subseteq B(S, \frac{1}{4}) = B_0$ with $\rho \in [\frac{1}{16}, \frac{1}{8}]$, elements $x_0 \in G$ and $\xi_0 \in \hat{G}$, and a function $M : B_0 \rightarrow \hat{G}$ obeying the local linearity property (9.5), such that

$$\mathbb{E}_{h \in B_1} |\mathbb{E}_{x \in G} T^{x_0+h} f(x) \overline{f(x)} e(-(\xi_0 + 2Mh) \cdot x)| \geq 2^{-C_4} \eta^{C'_4}. \tag{9.7}$$

We could take $C_i, C'_i = 2^{18}$, $i = 3, 4$.

9.2. Step 2: the symmetry argument

Let S, B_1, x_0, ξ_0, M be as in Proposition 9.3. Using the proof of Theorem 2.3 as a model, the next step would be to establish some symmetry property on $M : B_0 \rightarrow \hat{G}$, in the sense that the form $\{x, y\} := M(x) \cdot y - M(y) \cdot x$ is small. More precisely, we shall establish the following lemma.

Lemma 9.4 (symmetry of the derivative). *Let the notation be as in Proposition 9.3. For any $x, y \in B_0$, let $\{x, y\}$ denote the antisymmetric form*

$$\{x, y\} := M(x) \cdot y - M(y) \cdot x.$$

Then there exists a set S_3 of frequencies with $S_3 \supseteq \frac{1}{2} \cdot S$ and $d_3 := |S_3| \leq 2^{C_5} \eta^{-C'_5}$, and a Bohr set $B_3 = B(S_3, 2^{-C_6} \eta^{C'_6}) \subseteq B_1$, such that

$$\|\{x, z\}\|_{\mathbb{R}/\mathbb{Z}} \leq 2^{C_7} \eta^{-C'_7} \|x\|_{S_3} \quad \text{for all } x, z \in B_3. \tag{9.8}$$

It is permissible to take all of the $C_i, C'_i, i = 5, 6, 7$, equal to 2^{23} .

Proof. Let $\varepsilon_2 = 2^{-C_3-2C_4-10} \eta^{C'_3+2C'_4}$. By Lemma 8.2, we can find $\rho_2 \in [\varepsilon_2, 2\varepsilon_2]$ such that $B_2 := B(S, \rho_2) \subseteq B_1$ is a regular Bohr set. Of course we have

$$d_2 = d_1 \leq 2^{C_3} \eta^{-C_3}.$$

We write (9.7) as

$$|\mathbb{E}_{h \in B_1; x \in G} \mathbf{b}(x+h) \mathbf{b}(x) \mathbf{b}(h) e(-2Mh \cdot x)| \geq 2^{-C_4} \eta^{C'_4}, \tag{9.9}$$

absorbing all the phase terms into the functions \mathbf{b} . Applying Lemma 4.1 we can find $x_1 \in G$ such that

$$|\mathbb{E}_{h \in B_1; x \in B_2} \mathbf{b}(x+x_1+h) \mathbf{b}(x+x_1) \mathbf{b}(h) e(-2Mh \cdot (x+x_1))| \geq 2^{-C_4} \eta^{C'_4}. \tag{9.10}$$

Absorbing the x_1 terms into the functions \mathbf{b} we conclude that

$$|\mathbb{E}_{h \in B_1; x \in B_2} \mathbf{b}(h) \mathbf{b}(x+h) \mathbf{b}(x) e(-2Mh \cdot x)| \geq 2^{-C_4} \eta^{C'_4}. \tag{9.11}$$

Applying the Cauchy–Schwarz inequality (Lemma 4.3) to eliminate $\mathbf{b}(h)$, we then deduce

$$\mathbb{E}_{h \in B_1; x, y \in B_2} \mathbf{b}(x+h) \mathbf{b}(x) \mathbf{b}(y+h) \mathbf{b}(y) e(-2Mh \cdot (y-x)) \geq 2^{-2C_4} \eta^{2C'_4}. \tag{9.12}$$

Making the substitution $z := x + y + h$, this becomes

$$\mathbb{E}_{x, y \in B_2} \mathbb{E}_{z \in x+y+B_2} \mathbf{b}(z, x) \mathbf{b}(z, y) e(-2M(z-x-y) \cdot (y-x)) \geq 2^{-2C_4} \eta^{2C'_4}. \tag{9.13}$$

Absorbing as many phase terms into the functions $\mathbf{b}(z, x)$ and $\mathbf{b}(z, y)$ as we can, we conclude that

$$\mathbb{E}_{x, y \in B_2} \mathbb{E}_{z \in x+y+B_1} \mathbf{b}(z, x) \mathbf{b}(z, y) e(2\{x, y\}) \geq 2^{-2C_4} \eta^{2C'_4}.$$

Next, by the regularity of B_1 and Lemma 4.2 (i), we observe that

$$\begin{aligned} &|\mathbb{E}_{x, y \in B_2} \mathbb{E}_{z \in x+y+B_1} \mathbf{b}(z, x) \mathbf{b}(z, y) e(2\{x, y\}) \\ &\quad - \mathbb{E}_{x, y \in B_2} \mathbb{E}_{z \in B_1} \mathbf{b}(z, x) \mathbf{b}(z, y) e(2\{x, y\})| \leq 2^9 \varepsilon_2 d, \end{aligned}$$

which, due to the choice of ε_2 and the bound $d \leq 2^{C_3} \eta^{-C'_3}$, implies that

$$\mathbb{E}_{x, y \in B_2} \mathbb{E}_{z \in B_1} \mathbf{b}(z, x) \mathbf{b}(z, y) e(2\{x, y\}) \geq 2^{-2C_4-1} \eta^{2C'_4}. \tag{9.14}$$

In particular, by the pigeonhole principle in z we have

$$|\mathbb{E}_{x, y \in B_2} \mathbf{b}(x) \mathbf{b}(y) e(2\{x, y\})| \geq 2^{-2C_4-1} \eta^{2C'_4} \tag{9.15}$$

for some bounded functions $\mathbf{b}(x)$, $\mathbf{b}(y)$. At this point we observe the local bilinearity relationships

$$\{x + x', y\} = \{x, y\} + \{x', y\} \quad \text{and} \quad \{x, y + y'\} = \{x, y\} + \{x, y'\}, \tag{9.16}$$

which hold whenever all four of $\|x\|_S, \|x'\|_S, \|y\|_S, \|y'\|_S$ are at most $\frac{1}{8}$. We can then apply the Cauchy–Schwarz inequality (Lemma 4.3) to eliminate $\mathbf{b}(x)$ and conclude that

$$|\mathbb{E}_{x, y, y' \in B_2} \mathbf{b}(y, y') e(2\{x, y' - y\})| \geq 2^{-4C_4-2} \eta^{4C'_4},$$

and hence, by the triangle inequality,

$$\mathbb{E}_{y, y' \in B_2} |\mathbb{E}_{x \in B_2} e(2\{x, y' - y\})| \geq 2^{-4C_4-2} \eta^{4C'_4}.$$

By the pigeonhole principle, there exists $y' \in B_2$ such that

$$\mathbb{E}_{y \in B_2} |\mathbb{E}_{x \in B_2} e(2\{x, y' - y\})| \geq 2^{-4C_4-2} \eta^{4C'_4}. \tag{9.17}$$

Fix this y' . Since $|\mathbb{E}_{x \in B_2} e(2\{x, y' - y\})|$ is bounded above by 1, we conclude that there exists a set $A \subseteq B_2$ with $|A| \geq 2^{-4C_4-3} \eta^{4C'_4} |B_2|$ such that

$$|\mathbb{E}_{x \in B_2} e(2\{x, y' - y\})| \geq 2^{-4C_4-3} \eta^{4C'_4} \quad \text{for all } y \in A.$$

Applying Lemma 8.4 (and recalling that $\rho_2 \geq \varepsilon_2 = 2^{-C_3-2C_4-10} \eta^{C'_3+2C'_4}$ and $d_2 \leq 2^{C_3} \eta^{-C'_3}$) we conclude that

$$\|2\{x, y' - y\}\|_{\mathbb{R}/\mathbb{Z}} \leq 2^{24+2C_3+8C'_4} \eta^{-2C'_3-10C'_4} \|x\|_S \quad \text{for all } x \in B_2, y \in A.$$

Applying (9.16) (and recalling that $\rho_2 \leq 2\varepsilon \leq \frac{1}{32}$), we conclude that

$$\|2\{x, z\}\|_{\mathbb{R}/\mathbb{Z}} \leq 2^{26+2C_3+8C'_4} \eta^{-2C'_3-10C'_4} \|x\|_S \quad \text{for all } x \in B_2, z \in 2A - 2A.$$

On the other hand, by applying Lemma 8.8, we can find $S' \subseteq \hat{G}$ with $|S \cup S'| \leq 2^{12C_4+16} \eta^{-12C'_4}$ and a Bohr set $B'_3 := B(S \cup S', 2^{-61-2C_3-26C_4} \eta^{2C'_3+26C'_4})$ which is completely contained in $2A - 2A$ and inside B_2 . Thus we have

$$\|2\{x, z\}\|_{\mathbb{R}/\mathbb{Z}} \leq 2^{26+2C_3+8C'_4} \eta^{-2C'_3-10C'_4} \|x\|_S \quad \text{for all } x, z \in B'_3.$$

Let us now eliminate the factor 2. Observe that $B_3 := \{2x : x \in B'_3\}$ is also a Bohr set (with the frequency set $S \cup S'$ replaced by $S_3 := \frac{1}{2} \cdot (S \cup S')$). Since $\|x\|_S \leq 2\|x\|_{S_3}$, we also observe that $B_3 \subseteq B(S, 2\rho_2) \subseteq B_1$. By (9.16), which implies that $\{2x, z\} = 2\{x, z\}$, we conclude (9.8) as desired. \square

There are extremely close analogies between the above argument and that of § 6. Equations (9.9), (9.10), (9.11), (9.12), (9.13), (9.14), (9.15) and (9.17) are analogous to (6.7), (6.8), (6.9), (6.10), (6.11), (6.12), (6.13) and (6.16), respectively.

9.3. Step 3: eliminating the quadratic phase component

We now return to the conclusion of Proposition 9.3 and localize the x and h variables to a small Bohr set. Let

$$\varepsilon_4 = \min(2^{-C_3-C_4-10} \eta^{C'_3+C'_4}, 2^{-5-C_4-C_7} \eta^{C'_4}),$$

let S_3 be the set of characters coming from the previous subsection and let $B_4 := B(S_3, \rho_4) \subseteq B_3$ be a regular Bohr set such that $\rho_4 \in [\varepsilon_4, 2\varepsilon_4]$. By the previous estimate we have

$$\|\{x, z\}\|_{\mathbb{R}/\mathbb{Z}} \leq 2^{C_7} \eta^{-C'_7 \varepsilon_4} \quad \text{for all } x, z \in B_4. \tag{9.18}$$

Let us write (9.7) as

$$|\mathbb{E}_{h \in B_1; x \in G} \mathbf{b}(h) \mathbf{b}(x+h) \overline{f(x)} e(-2Mh \cdot x)| \geq 2^{-C_4} \eta^{C'_4},$$

where we have absorbed some phase terms into the functions \mathbf{b} as before. Since $B_1 = B(S, \rho)$ for some $\rho \geq \frac{1}{16}$, we conclude from Lemma 4.2 (ii) that

$$\begin{aligned} & |\mathbb{E}_{h \in B_1; x \in G} \mathbf{b}(h) \mathbf{b}(x+h) \overline{f(x)} e(-2Mh \cdot x) \\ & - \mathbb{E}_{h' \in B_1; h \in B_4; x \in G} \mathbf{b}(h+h') \mathbf{b}(x+h+h') \overline{f(x)} e(-2M(h+h') \cdot x)| \leq 2^9 \varepsilon_4 d_1. \end{aligned}$$

This is at most $2^{-C_4-1}\eta^{C'_4}$, and therefore

$$|\mathbb{E}_{h' \in B_1; h \in B_4; x \in G} \mathbf{b}(h+h') \mathbf{b}(x+h+h') \overline{f(x)} e(-2M(h+h') \cdot x)| \geq 2^{-C_4-1}\eta^{C'_4}.$$

Hence, by the pigeonhole principle, there exists $h' \in B_r$ such that

$$|\mathbb{E}_{h \in B_4; x \in G} \mathbf{b}(h+h') \mathbf{b}(x+h+h') \overline{f(x)} e(-2M(h+h') \cdot x)| \geq 2^{-C_4-1}\eta^{C'_4}.$$

Since $e(-Mh' \cdot x) = e(-2Mh' \cdot (x+h))e(2Mh' \cdot h)$, we can absorb all the h' terms into the functions \mathbf{b} to conclude that

$$|\mathbb{E}_{h \in B_4; x \in G} \mathbf{b}(h) \mathbf{b}(x+h) \overline{f(x)} e(-2Mh \cdot x)| \geq 2^{-C_4-1}\eta^{C'_4}.$$

By Lemma 4.1 we then have

$$|\mathbb{E}_{y \in G; x, h \in B_4} \mathbf{b}(h) \mathbf{b}(x+y+h) \overline{f(x+y)} e(-2Mh \cdot (x+y))| \geq 2^{-C_4-1}\eta^{C'_4},$$

and hence, by the triangle inequality,

$$\mathbb{E}_{y \in G} |\mathbb{E}_{x, h \in B_4} \mathbf{b}(h, y) \mathbf{b}(x+h, y) \overline{f(x+y)} e(-2Mh \cdot x)| \geq 2^{-C_4-1}\eta^{C'_4}.$$

Now we observe from (9.5) that

$$2Mh \cdot x = M(x+h) \cdot (x+h) - Mx \cdot x - Mh \cdot h - \{x, h\},$$

and hence, by (9.18),

$$|e(-2Mh \cdot x) - \mathbf{b}(x+h) \mathbf{b}(h) e(Mx \cdot x)| \leq 2\pi \cdot 2^{C_7} \eta^{-C'_7 \varepsilon_4} \leq 2^{-C_4-2}\eta^{C'_4}.$$

Thus we have

$$\mathbb{E}_{y \in G} |\mathbb{E}_{x, h \in B_4} \mathbf{b}(h, y) \mathbf{b}(x+h, y) \overline{f(x+y)} e(Mx \cdot x)| \geq 2^{-C_4-2}\eta^{C'_4}.$$

It is convenient to localize x further. Let $\varepsilon_5 = 2^{-C_4-C_5-13}\eta^{C'_4+C'_5}\varepsilon_4$. By Lemma 8.2 we can find a regular Bohr set $B_5 = B(S_3, \rho_5)$ with $\rho_5 \in [\varepsilon_5, 2\varepsilon_5]$. By Lemma 4.2 (ii) we have that

$$\begin{aligned} & |\mathbb{E}_{y \in G} |\mathbb{E}_{x, h \in B_4} \mathbf{b}(h, y) \mathbf{b}(x+h, y) \overline{f(x+y)} e(Mx \cdot x)| \\ & - \mathbb{E}_{y \in G} |\mathbb{E}_{w, h \in B_4; x \in B_5} \mathbf{b}(h, y) \mathbf{b}(x+w+h, y) \overline{f(x+w+y)} e(M(x+w) \cdot (x+w))| | \end{aligned}$$

is at most $2^{10}\varepsilon_5 d_3/\varepsilon_4$, which on account of the choice of ε_5 implies that

$$\mathbb{E}_{y \in G} |\mathbb{E}_{w, h \in B_4; x \in B_5} \mathbf{b}(h, y) \mathbf{b}(x+w+h, y) \overline{f(x+w+y)} e(M(x+w) \cdot (x+w))| \geq 2^{-C_4-3}\eta^{C'_4}.$$

By the pigeonhole principle there exists $w \in B_4$ such that

$$\mathbb{E}_{y \in G} |\mathbb{E}_{h \in B_4; x \in B_5} \mathbf{b}(h, y) \mathbf{b}(x+w+h, y) \overline{f(x+w+y)} e(M(x+w) \cdot (x+w))| \geq 2^{-C_4-3}\eta^{C'_4}.$$

Let us now apply Lemma 4.4. Since B_4 is regular and ε_5 is so small, we certainly have $\mathbb{E}(1_{B_4})/\mathbb{E}(1_{B_4+B_5}) \geq \frac{1}{2}$, and so lemma allows us to conclude that

$$\mathbb{E}_{y \in G} \|\overline{f(x+w+y)} e(M(x+w) \cdot (x+w))\|_{u^2(B_5)} \geq 2^{-C_4-4}\eta^{C'_4}.$$

This, of course, implies that

$$\mathbb{E}_{y \in G} \overline{\|f(x+w+y)e(M(x+w) \cdot (x+w))\|_{u^3(B_5)}} \geq 2^{-C_4-4}\eta^{C'_4}.$$

The u^3 norm being invariant under translation, conjugation and quadratic phase modulation, we conclude that

$$\mathbb{E}_{y \in G} \|f\|_{u^3(y+w+B_5)} \geq 2^{-C_4-4}\eta^{C'_4}.$$

Making the change of variables $y \mapsto y + w$, and completing a small computation, we obtain (2.8) as desired. \square

We remark that we have proved slightly more than (2.8), in that the quadratic phase functions used to demonstrate the largeness of the $\|f\|_{u^3(y+B)}$ norm all agree up to lower-order (i.e. linear and constant) terms. However, we were unable to find any way to exploit this additional fact.

10. Bohr sets and generalized arithmetic progressions

Our focus from this point on is largely on the group $G = \mathbb{Z}/N\mathbb{Z}$, as we are working towards connections with themes in ergodic theory, in particular involving \mathbb{Z} -actions. As we have stressed, $\mathbb{Z}/N\mathbb{Z}$ is an appropriate group to consider if one is interested in discrete questions concerning the integers. However, some of what we have to say, particularly in the present section, can be generalized without undue pain to arbitrary additive groups.

We have obtained an inverse theorem, Theorem 2.7, for the U^3 norm which relates that norm to the quadratic bias norm $u^3(y+B)$ on Bohr sets B (or on subspaces W , in finite-field cases such as $G = \mathbb{F}_5^n$). This is a fairly satisfactory state of affairs, except for the presence of the Bohr set B ; in particular, it is not clear at present what exactly the locally quadratic phase functions are on B . In this section we show how the Bohr set can, if desired, be replaced with a *generalized arithmetic progression*, and how to characterize the locally quadratic phase functions on such progressions.

We begin by recalling what a generalized arithmetic progression is.

Definition 10.1 (generalized arithmetic progression). A *generalized arithmetic progression* P in an additive group G is any set of the form

$$P := \{a + l_1v_1 + \dots + l_dv_d : 0 \leq l_j < L_j \text{ for all } 1 \leq j \leq d\},$$

where $d \geq 0$, $a, v_1, \dots, v_d \in G$ and $L_1, \dots, L_d \geq 1$. We shall abbreviate the right-hand side as $P = a + [0, L] \cdot v$, where $L := (L_1, \dots, L_d)$ and $v := (v_1, \dots, v_d)$. We call a the *base point* of the progression, d the *rank*, v_1, \dots, v_d the *generators* and L_1, \dots, L_d the *lengths* of the progression. If all the sums in P are distinct, so that $|P| = L_1 \cdots L_d$, we say that P is *proper*. A *coset progression* is any set of the form $P + H$ where H is a subgroup of G . We say that the coset progression $P + H$ is *proper* if P is proper and $|P + H| = |P| |H|$ (i.e. all the sums in $P + H$ are distinct); we define the rank, base point, etc., of the coset progression $P + H$ to be the same as that of its component P .

The need to generalize from generalized arithmetic progressions to coset progressions in the setting of a general group G was first noted in [33].*

We now use standard facts from the geometry of numbers to show that every Bohr set contains a large proper coset progression. The first lemma follows from a result of Mahler [11, Chapter VIII, Corollary to Theorem VII] together with Minkowski's second theorem [11, Chapter VIII, Theorem V]. This result (in fact, a rather stronger one) was used in an additive-combinatorial context in Bilu's work on Freiman's theorem [6, Lemma 2.1].

Lemma 10.2. *Let Γ be a lattice of full rank in \mathbb{R}^d . Then there exist linearly independent vectors $w_1, \dots, w_d \in \Gamma$ which generate Γ , and such that*

$$|w_1| \cdots |w_d| \leq 2 \cdot d! \cdot \text{mes}(\mathbb{R}^d/\Gamma), \quad (10.1)$$

where $\text{mes}(\mathbb{R}^d/\Gamma)$ is the volume of a fundamental domain of Γ .

Next, we give a 'discrete John's theorem' that shows that the intersection of a convex symmetric body and a lattice of full rank is essentially equivalent to a progression.

Lemma 10.3 (discrete John's theorem). *Let B be a convex symmetric body in \mathbb{R}^d , and let Γ be a lattice in \mathbb{R}^d of full rank. Then there exists a d -tuple*

$$w = (w_1, \dots, w_d) \in \Gamma^d$$

of linearly independent vectors in Γ and a d -tuple $L = (L_1, \dots, L_d)$ of positive integers such that

$$(d^{-2d} \cdot B) \cap \Gamma \subseteq (-L, L) \cdot w \subseteq B \cap \Gamma \subseteq (-d^{2d}L, d^{2d}L) \cdot w.$$

Here, of course,

$$(-L, L) \cdot w := \{l_1 w_1 + \cdots + l_d w_d : -L_j < l_j < L_j \text{ for all } 1 \leq j \leq d\}.$$

Proof. We first observe using John's theorem [44] (see also [8, 51]) and an invertible linear transformation that we may assume without loss of generality that $B_d \subseteq B \subseteq d \cdot B_d$, where B_d is the unit ball in \mathbb{R}^d . We may also assume that $d \geq 2$, since the claim is easy otherwise.

Now let $w = (w_1, \dots, w_d)$ be as in Lemma 10.2. For each j , let L_j be the least integer greater than $1/d|w_j|$. Then from the triangle inequality we see that $|l_1 w_1 + \cdots + l_d w_d| < 1$ whenever $|l_j| < L_j$, and hence $(-L, L) \cdot w$ is contained in B_d and hence in B .

Now let $x \in B \cap \Gamma$. Since w generates Γ , we have $x = l_1 w_1 + \cdots + l_d w_d$ for some integers l_1, \dots, l_d ; since $B \subseteq d \cdot B_d$, we have $|x| \leq d$. Applying Cramer's rule to solve for l_1, \dots, l_d and (10.1), we have

$$|l_j| = \frac{|x \wedge w_1 \cdots w_{j-1} \wedge w_{j+1} \wedge w_d|}{|w_1 \wedge \cdots \wedge w_d|} \leq \frac{|x| |w_1| \cdots |w_d|}{|w_j| |w_1 \wedge \cdots \wedge w_d|} = \frac{|x| \text{mes}(\mathbb{R}^d/\Gamma)}{|w_j|} \leq \frac{2d \cdot d!}{|w_j|},$$

* Of course, the classification of finite abelian groups tells us that every coset progression is also a generalized arithmetic progression (by expanding H as the direct sum of cyclic groups, which can each be interpreted as an arithmetic progression), but in representing the coset progression in this manner one may cause its rank of the coset progression to increase enormously (by the number of generators needed to span H).

which is certainly no greater than $d^{2d}L_j$. It follows that $x \in (-d^{2d}L, d^{2d}L) \cdot w$, which is what we wanted to prove. A more-or-less identical argument gives the inclusion $(d^{-2d} \cdot B) \cap \Gamma \subseteq (-L, L) \cdot w$. \square

Let $x \mapsto \{x\}$ denote the fractional part map from \mathbb{R}/\mathbb{Z} to the fundamental domain $(-\frac{1}{2}, \frac{1}{2}]$.

Lemma 10.4 (Bohr sets contain large coset progressions). *Let $S \subseteq \hat{G}$ be a set of d characters, let $\rho < \frac{1}{4}$ be a real number and let $B(S, \rho) \subseteq G$ be a Bohr set. Then there exists a proper coset progression $P + H$ of rank d' , $0 \leq d' \leq d$, where $P = (-L, L) \cdot v$ for some $L_1, \dots, L_{d'} \geq 1$ and $v_1, \dots, v_{d'} \in G$, and we have the inclusions*

$$B(S, d'^{-2d'} \rho) \subseteq P + H \subseteq B(S, \rho). \tag{10.2}$$

In particular, from Lemma 8.1 we have

$$|P + H| \geq \rho^d d^{-2d^2} N. \tag{10.3}$$

Furthermore, the vectors $(\{\xi \cdot v_j\})_{\xi \in S} \in \mathbb{R}^S$, $1 \leq j \leq d'$, can be chosen to be linearly independent, and H can be taken to be the orthogonal complement of S , that is to say, the group

$$H := \{x \in G : \xi \cdot x = 0 \text{ for all } S\}. \tag{10.4}$$

Remark 10.5. The lemma is at the same time a refinement and a weakening of a lemma from [33]. The refinement, corresponding to the fact that we use Lemma 10.2 rather than Minkowski’s second theorem, is that we obtain the left-hand inclusion in (10.2) and not just the right-hand one. The weakening is that using just Minkowski’s second theorem (and thus sacrificing the left-hand inclusion in (10.2)) gives a stronger bound than (10.3).

Proof. Let $\phi : G \rightarrow (\mathbb{R}/\mathbb{Z})^S$ be the group homomorphism $\phi(x) := (\xi \cdot x)_{\xi \in S}$. Observe that $\phi(G)$ is a finite subgroup of the torus $(\mathbb{R}/\mathbb{Z})^S$ and that $B(S, \rho)$ is the inverse image of the cube $Q := \{(y_\xi)_{\xi \in S} : |y_\xi| \leq \rho\}$ under ϕ .

Let $\Gamma \subseteq \mathbb{R}^S$ be the lattice $\phi(G) + \mathbb{Z}^S$. Though it is a slight abuse of notation, we consider $\phi(G) \cap Q$ to be the same as $\Gamma \cap Q$. Applying Lemma 10.3, we can find a progression $\tilde{P} := (-L, L) \cdot w$ for some linearly independent $w_1, \dots, w_{d'} \subseteq \Gamma$ with $0 \leq d' \leq d$ such that

$$\Gamma \cap d'^{-2d'} \cdot Q \subseteq \tilde{P} \subseteq \Gamma \cap Q.$$

Since the w_j are independent, \tilde{P} is necessarily proper. The claim now follows by setting v_j to be an arbitrary element of $\phi^{-1}(w_j)$ for each $1 \leq j \leq d'$ and setting H equal to the kernel of ϕ , which is of course just (10.4). \square

When $G = \mathbb{Z}/N\mathbb{Z}$ is a cyclic group of prime order, the subgroup H has no role to play* and we conclude the following corollary.

* At the other extreme, in the finite-field geometry setting $G = \mathbb{F}_5^n$ it is the progression component P which is irrelevant, because the properness of P forces all the lengths to be less than 5. Since the rank of P is also under control, we see that H is a substantial portion of $P + H$. Indeed, the fact that Bohr sets in finite-field geometries contain large subspaces was already exploited in the proof of Theorem 2.3.

Corollary 10.6. *Let $G = \mathbb{Z}/N\mathbb{Z}$ be a cyclic group of prime order, let $S \subseteq \hat{G}$ be a set of d characters and let $\rho < \frac{1}{4}$ be a parameter. Then there is a proper generalized arithmetic progression $P = (-L, L) \cdot v$ of rank at most d and size at least $\rho^d d^{-2d^2} N$ such that $B(S, d^{-2d}\rho) \subseteq P \subseteq B(S, \rho)$. Furthermore, the vectors $(\{\xi \cdot v_j\})_{\xi \in S}$ are linearly independent in \mathbb{R}^S .*

In this paper, it is in general more convenient technically to work with Bohr sets than progressions or coset progressions. However, there is one task which is much easier to achieve on progressions than on Bohr sets, and that is to classify quadratic phase functions:

Lemma 10.7 (inverse theorem for locally quadratic functions). *Let G be a finite additive group of odd order. Let $P + H$ be a coset progression in G , let a be the base point of P and let v_1, \dots, v_d be the generators. Let $\phi : P + H \rightarrow \mathbb{R}/\mathbb{Z}$ be a locally quadratic phase function on $P + H$. Then there exists a self-adjoint homomorphism $M : H \mapsto \hat{H}$, elements $\xi_0, \xi_1, \dots, \xi_d \in \hat{H}$, elements $\eta_i, \lambda_{ij} \in \mathbb{R}/\mathbb{Z}$ for $1 \leq i, j \leq d$, and $c \in \mathbb{R}/\mathbb{Z}$ such that $\lambda_{ij} = \lambda_{ji}$ and*

$$\phi(a + l_1 v_1 + \dots + l_d v_d + h) = Mh \cdot h + 2 \sum_{i=1}^d l_i \xi_i \cdot h + \sum_{1 \leq i, j \leq d} l_i l_j \lambda_{ij} + \xi_0 \cdot h + \sum_{i=1}^d l_i \eta_i + c \tag{10.5}$$

for all l_1, \dots, l_d, h with $0 \leq l_j < L_j$ for all $1 \leq j \leq d$ and $h \in H$.

Remark 10.8. In the converse direction, it is easy to show that (10.5) is indeed well defined and gives a locally quadratic function if $3P + H$ is a proper coset progression, but we will not need that fact here.

Proof. We may assume that $L_j \geq 2$ for all j , and we may translate so that $a = 0$. Let $\phi|_H$ be the restriction of ϕ to H . By Theorem 3.2, the quadratic extension theorem, we can extend $\phi|_H$ to a globally quadratic phase function ψ on G . Using Lemma 3.1, it is easy to see that ψ , when restricted to $P + H$, has the form (10.5). Thus we may subtract ψ from ϕ , which means that ϕ now vanishes on H . We now claim that under this reduction, ϕ takes the simpler form

$$\phi(l_1 v_1 + \dots + l_d v_d + h) = \sum_{i=1}^d l_i \xi_i \cdot h + \sum_{1 \leq i, j \leq d} l_i l_j \lambda_{ij} + \sum_{i=1}^d l_i \eta_i.$$

Observe that for any $h \in H$, the function $(h \cdot \nabla)\phi$ is locally linear on $P + H$ and vanishes on H , and hence takes the form

$$(h \cdot \nabla)\phi(l_1 v_1 + \dots + l_d v_d + h') = \sum_{i=1}^d l_i (h \cdot \nabla)\phi(v_i)$$

for all l_1, \dots, l_d, k with $0 \leq l_j < L_j$ and $h' \in H$. It is then easy to see that $h \mapsto (h \cdot \nabla)\phi(v_i)$ is a group homomorphism from H to \mathbb{R}/\mathbb{Z} and thus there exists $\xi_i \in \hat{H}$ such that

$(h \cdot \nabla)\phi(v_i) = 2\xi_i \cdot h$ for all $h \in H$. Using this, we thus reduce matters to showing that

$$\phi(l_1v_1 + \dots + l_dv_d) = \sum_{1 \leq i, j \leq d} l_i l_j \lambda_{ij} + \sum_{i=1}^d l_i \eta_i,$$

or equivalently that

$$\phi(l_1v_1 + \dots + l_dv_d) = \sum_{1 \leq i < j \leq d} 2l_i l_j \lambda_{ij} + \sum_{i=1}^d l_i \eta_i + l_i^2 \lambda_i.$$

We induct on d . When $d = 0$ there is nothing to prove. Now suppose that the claim is already proven for $d - 1$. We observe that the derivative $(v_d \cdot \nabla)\phi$ is linear, and hence

$$(v_d \cdot \nabla)\phi(l_1v_1 + \dots + l_dv_d) = \sum_{1 \leq i \leq d} 2l_i \lambda_{id} + \eta_d$$

for some $\lambda_{1d}, \dots, \lambda_{dd}, \eta_d \in \mathbb{R}/\mathbb{Z}$, with the caveat that l_d must now be less than $L_d - 1$ rather than L_d . (Note that it is always possible to divide by two in \mathbb{R}/\mathbb{Z} , though the value obtained need not be unique). The claim then follows from the induction hypothesis and a simple ‘integration’ argument, which we omit. \square

Now, let us specialize to the setting of cyclic groups $\mathbb{Z}/N\mathbb{Z}$ of prime order. We begin by defining some special functions on this set.

Definition 10.9 (bracket polynomials). Let $\mathbb{Z}/N\mathbb{Z}$ be a cyclic group of prime order. If $k \geq 0$, we define a *bracket monomial of degree k* on $\mathbb{Z}/N\mathbb{Z}$ to be any function $\phi : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ of the form

$$\phi(x) = a\{\xi_1 \cdot x\} \cdots \{\xi_k \cdot x\} \pmod{1},$$

where $\xi_1, \dots, \xi_k \in \widehat{\mathbb{Z}/N\mathbb{Z}}$ and $a \in \mathbb{R}$; we refer to ξ_1, \dots, ξ_k as the *frequencies* of the monomial. If $L \geq 0$ and $S \subseteq \widehat{\mathbb{Z}/N\mathbb{Z}}$, we define a *bracket polynomial of degree at most k , length at most L and frequency set S* to be any function $\phi : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ which can be expressed as the sum of L or fewer bracket monomials of degree at most k and frequencies inside S . We write $\text{Freq}(\phi) \subseteq S$. Note that if $|S| = d$ then we may always assume that $L \leq kd^k$; for this reason there will be little subsequent discussion of length.

These bracket polynomials are special cases of the *generalized polynomials* considered in various papers of Håland, Håland and Knuth and Bergelson and Leibman [2, 4, 38–40], though with the (minor) caveat that our fractional parts take values from $-\frac{1}{2}$ to $\frac{1}{2}$, while the ones in those papers take values from 0 to 1.

Define a *bracket quadratic* to be a bracket polynomial of degree at most 2. We can now link locally quadratic phase functions with bracket quadratics.

Proposition 10.10. *Let $G = \mathbb{Z}/N\mathbb{Z}$ be a cyclic group of prime order, let $S \subseteq \widehat{G}$ be a set of d characters and suppose that $\rho \in (0, \frac{1}{4}]$. Let P be the proper progression contained in $B(S, \rho)$, which was constructed in Corollary 10.6, and let $\phi : B(S, \rho) \rightarrow \mathbb{R}/\mathbb{Z}$ be a locally quadratic phase function on $B(S, \rho)$. Then there exists a bracket quadratic $\tilde{\phi} : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ with $\text{Freq}(\tilde{\phi}) \subseteq S$ such that $\tilde{\phi} = \phi$ on P .*

Proof. Let v_1, \dots, v_d and L_1, \dots, L_d be the generators and lengths of P . By Lemma 10.7 we have

$$\phi(l_1 v_1 + \dots + l_d v_d) = \sum_{1 \leq i \leq j \leq d} l_i l_j \lambda_{ij} + \sum_{i=1}^d l_i \eta_i + c \pmod{1} \tag{10.6}$$

for some real numbers λ_{ij}, η_i .

Next, let $\Phi : P \rightarrow \mathbb{R}^S$ be the map

$$\Phi(x) := (\{\xi \cdot x\})_{\xi \in S}.$$

Since $P \subseteq B(S, \rho)$, we see that $\Phi(P)$ lies inside the cube $Q := \{(y_\xi)_{\xi \in S} : |y_\xi| \leq \rho \text{ for all } \xi \in S\}$. Since $\rho < \frac{1}{4}$, it is also easy to verify that

$$\Phi(l_1 v_1 + \dots + l_d v_d) = l_1 \Phi(v_1) + \dots + l_d \Phi(v_d).$$

From Corollary 10.6 we know that the $\Phi(v_j)$ are linearly independent. Thus there exists a vector $u_i \in \mathbb{R}^S$ such that

$$\Phi(l_1 v_1 + \dots + l_d v_d) \cdot u_i = l_i.$$

Writing $u_i = (u_{i\xi})_{\xi \in S}$ and $x = l_1 v_1 + \dots + l_d v_d$ we conclude that

$$l_i = \sum_{\xi \in S} u_{i\xi} \{\xi \cdot x\}.$$

Inserting this formula into (10.6) we obtain the claim. □

We can now give a version of Theorem 2.7, the $U^3(G)$ inverse theorem, in the case $G = \mathbb{Z}/N\mathbb{Z}$, which involves bracket quadratic functions. In this theorem, C_0, C_1, \dots and c denote absolute constants which *do not* vary from line to line.

Theorem 10.11 (inverse theorem for $U^3(\mathbb{Z}/N\mathbb{Z})$, bracket quadratic functions). *Let $\eta \in [0, c)$ and let $\mathbb{Z}/N\mathbb{Z}$ be a cyclic group of prime order. If $f : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathcal{D}$ is a bounded function such that $\|f\|_{U^3(\mathbb{Z}/N\mathbb{Z})} \geq \eta$, then there exists a set $S \subseteq \widehat{\mathbb{Z}/N\mathbb{Z}}$ of size $d \leq \eta^{-C}$ and a proper progression P of rank at most d and size $|P| \geq \exp(-\eta^{-C})N$ with the inclusions*

$$B(S, \exp(-\eta^{-C_1})) \subseteq P \subseteq B(S, \frac{1}{8}), \tag{10.7}$$

and there exists a generalized quadratic $\phi : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ with $\text{Freq}(\phi) \subseteq S$ such that we have the local quadratic bias estimate

$$|\mathbb{E}_{x \in P}(T^h f(x)e(-\phi(x)))| \geq \eta^{C_1} \tag{10.8}$$

for some $h \in \mathbb{Z}/N\mathbb{Z}$. More generally, for any non-empty set $A \subseteq P$ there exists an $h_A \in \mathbb{Z}/N\mathbb{Z}$ such that

$$|\mathbb{E}_{x \in A}(T^{h_A} f(x)e(-\phi(x)))| \geq \eta^{C_1}. \tag{10.9}$$

Furthermore, there exists another generalized quadratic $\tilde{\phi} : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ with $\text{Freq}(\phi) \subseteq S$ such that we have the global quadratic bias estimate

$$|\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(f(x)e(-\tilde{\phi}(x)))| \geq \exp(-\eta^{-C_2}). \tag{10.10}$$

Conversely, suppose that S is a set of d frequencies and that $\tilde{\phi} : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ is a bracket quadratic with $\text{Freq}(\phi) \subseteq S$. Suppose that $f : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathcal{D}$ is a function such that $|\mathbb{E}(f(x)e(-\tilde{\phi}(x)))| \geq \eta$. Then we have

$$\|f\|_{U^3} \geq (\eta^3 \rho^2 / C_3 d^3)^d. \tag{10.11}$$

Proof. Applying Theorem 2.7, we can find a regular Bohr set $B := B(S, \rho)$ in $\mathbb{Z}/N\mathbb{Z}$ with $d = |S| \leq \eta^{-C_0}$ and $\rho \geq \eta^{C_0}$, a $y \in \mathbb{Z}/N\mathbb{Z}$, and a locally quadratic phase function $\phi_0 : B \rightarrow \mathbb{Z}/N\mathbb{Z}$ on B such that

$$|\mathbb{E}_{x \in B}(T^y f(x)e(-\phi_0(x)))| \geq \eta^{C_0}.$$

We could take $C_0 = 2^{25}$. Next, let $\varepsilon = \eta^{2C_0+11}$ and apply Corollary 10.6 to find a proper progression P of rank at most d such that

$$B(S, \varepsilon \exp(-\eta^{-C_0-1})\rho) \subseteq P \subseteq B(S, \varepsilon\rho).$$

By Lemma 4.2 (iii) we can find $w \in B(S, (1 - \varepsilon)\rho)$ such that

$$|\mathbb{E}_{x \in w+P}(T^y f(x)e(-\phi_0(x)))| \geq \eta^{C_0+1},$$

and (10.8) follows after translating by w and applying Proposition 10.10. A very similar argument gives (10.9). Note that the inclusions on P follow by choice of η , and the lower bound on $|P|$ follows from Lemma 8.1.

Now we prove (10.10). From (10.7) and Lemma 8.2 we can find a regular Bohr set $B' := B(S, \rho')$ with $\rho' \geq \exp(-\eta^{-C_1-3})$ which is contained in P . Applying (10.9) we can find a shift $h' \in \mathbb{Z}/N\mathbb{Z}$ such that

$$|\mathbb{E}(T^{h'} f(x)e(-\phi(x))1_{B'}(x))| \geq \eta^{C_1+1}\mathbb{E}1_{B'}.$$

Let $\varepsilon' = \eta^{2C_1+9}$ and let $\chi : \mathbb{R} \rightarrow [0, 1]$ be a smooth cutoff such that $\chi(s) = 1$ when $|s| \leq \rho'(1 - \varepsilon')$, $\chi(s) = 0$ when $|s| \geq \rho'(1 + \varepsilon')$ and such that the derivative estimate $\|\chi''\|_\infty \leq 100/\varepsilon'^2 \rho'^2$ holds true. By the regularity of B' we see that

$$\left| \mathbb{E}_x T^{h'} f(x)e(-\phi(x)) \prod_{\xi \in S} \chi(\{\xi \cdot x\}) \right| \geq (\eta^{C_1+1} - 200\varepsilon'd)\mathbb{E}1_{B'} \geq \eta^{C_1+2}\mathbb{E}1_{B'},$$

and this is at most $\exp(-\eta^{-2C_1-4})$ by Lemma 8.1. Next, we use Fourier expansion on \mathbb{R} to write

$$\chi(s) = \int_{\mathbb{R}} \hat{\chi}(t)e(ts) dt, \quad \text{where } \hat{\chi}(t) := \int_{\mathbb{R}} \chi(s)e(-ts) ds.$$

This allows us to conclude that

$$\left| \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \left[\mathbb{E}_x T^{h'} f(x) e\left(-\phi(x) + \sum_{\xi \in S} t_{\xi} \{\xi \cdot x\}\right) \right] \prod_{\xi \in S} \hat{\chi}(t_{\xi}) dt_{\xi} \right| \geq \exp(-\eta^{-2C_1-4}). \tag{10.12}$$

Now set $\lambda := \varepsilon'^{-1/2} \rho'^{-1}$. Then by integration by parts, applied twice, we have

$$\int_{\lambda}^{\infty} |\hat{\chi}(t)| dt = \frac{1}{4\pi^2} \int_{\lambda}^{\infty} \frac{dt}{t^2} \left| \int_{\mathbb{R}} \chi''(s) e(-ts) ds \right| \leq \frac{100}{\varepsilon' \rho'} \int_{\lambda}^{\infty} \frac{dt}{t^2} \leq \frac{100}{\varepsilon^{1/2}}$$

and hence

$$\int_{\mathbb{R}} |\hat{\chi}(t)| dt \leq 2\lambda \|\hat{\chi}\|_{\infty} + \frac{200}{\varepsilon^{1/2}} \leq 8\lambda \rho' + \frac{200}{\varepsilon^{1/2}} \leq 2^8 \varepsilon^{-1/2}.$$

This implies that

$$\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{\xi \in S} |\hat{\chi}(t_{\xi})| dt_{\xi} \leq \exp(-\eta^{-C_1-1}).$$

Comparing this with (10.12), we conclude the existence of real numbers t_{ξ} for $\xi \in S$ such that

$$\left| \mathbb{E}_x T^{h'} f(x) e\left(-\phi(x) + \sum_{\xi \in S} t_{\xi} \{\xi \cdot x\}\right) \right| \geq \exp(-\eta^{-3C_1-5}),$$

and (10.10) follows.

Finally, we prove (10.11). Assume then that $\tilde{\phi}$ is a bracket quadratic with $\text{Freq}(\phi) \subseteq S$, $|S| = d$, and that $f : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathcal{D}$ is a function with $|\mathbb{E}(f(x)e(-\tilde{\phi}(x)))| \geq \eta$. Set $\varepsilon := \eta/160d$ and select a $\rho \in [\varepsilon, 2\varepsilon]$ such that $B := B(S, \rho)$ is regular. Introducing an averaging over translates of B , we see that

$$|\mathbb{E}_y \mathbb{E}_{x \in y+B} (f(x)e(-\tilde{\phi}(x)))| \geq \eta. \tag{10.13}$$

Now write $U := B(S, \frac{1}{2} - 10\rho)$. We wish to exclude from (10.13) those y which lie in the complement U^c of U , since the bracket functions $\{\xi \cdot x\}$, $\xi \in S$, fail to be linear when $\{\xi \cdot x\} \approx \pm \frac{1}{2}$. To this end, we estimate

$$\begin{aligned} |\mathbb{E}_y \mathbb{E}_{x \in y+B} (f(x)e(-\tilde{\phi}(x))) 1_{U^c}(y)| &\leq \mathbb{E}_y 1_{U^c}(y) \\ &\leq \mathbb{E}_y \sum_{\xi \in S} 1_{\|y \cdot \xi\| \geq \frac{1}{2} - 10\rho}(y) \\ &\leq 40d\rho \leq 80d\varepsilon, \end{aligned}$$

the penultimate estimate following from the fact that N is prime, so $\xi \cdot x$ takes on the values r/N , $r \in \mathbb{Z}/N\mathbb{Z}$, precisely once each as x varies. Combining this with (10.13) we see that

$$|\mathbb{E}_y \mathbb{E}_{x \in y+B} (f(x)e(-\tilde{\phi}(x))) 1_U(y)| \geq \eta/2,$$

and hence there is $y \in U$ such that

$$|\mathbb{E}_{x \in y+B} (f(x)e(-\tilde{\phi}(x)))| \geq \eta/2. \tag{10.14}$$

We claim that $\tilde{\phi}$ is a quadratic phase function on $y+B$. To check this, we must show that if the cube $(x + \omega_1 h_1 + \omega_2 h_2 + \omega_3 h_3)_{\omega \in \{0,1\}^3}$ is contained in $y+B$, then $(h_1 \cdot \nabla_x)(h_2 \cdot \nabla_x)(h_3 \cdot \nabla_x)\phi(x) = 0$. This is easy to prove once one appreciates that (for example) if $x, x+h_1 \in y+B$ and $\xi \in S$, then $\{\xi \cdot (x+h_1)\} = \{\xi \cdot x\} + \{\xi \cdot h_1\}$. Indeed, this identity is patently true (mod 1), and furthermore one has the bounds $|\{\xi \cdot x\}| \leq \frac{1}{2} - 9\rho$ and $|\{\xi \cdot h_1\}| \leq 2\rho$, whence $|\{\xi \cdot (x+h_1)\}| \leq \frac{1}{2} - 7\rho$.

Equation (10.14), then, implies that

$$\|f\|_{u^3(y+B)} \geq \eta/2.$$

The result is now an immediate consequence of Theorem 2.7 (ii). □

11. An application: a bound for $r_4(G)$

As an application of Theorem 2.7 we obtain a bound for $r_4(G)$, the size of the largest set $A \subseteq G$ with no four-term arithmetic progressions.

Theorem 11.1 (Szemerédi’s theorem for G). *Let G be a finite additive group of order N , where $(N, 6) = 1$. Then we have the bound*

$$r_4(G) \ll N(\log \log N)^{-c}$$

for some absolute constant $c > 0$.

The reader may find it helpful to recall Gowers’s argument [25] in the case $G = \mathbb{Z}/N\mathbb{Z}$, which is explained, for example, in [26]. Our argument here will be similar, though we must handle torsion in G . Gowers did not use the so-called ‘symmetry argument’ of §9, since he was able to apply the weak inverse theorem, Theorem 1.14, where we shall apply Theorem 2.7. It is likely that if our only interest was in proving Theorem 11.1, then we could do likewise. However, as remarked in the introduction, our work is ultimately directed towards a study of 4-tuples $p_1 < p_2 < p_3 < p_4 \leq N$ of primes in arithmetic progression, and potentially towards the bound $r_4(\mathbb{Z}/N\mathbb{Z}) \ll N(\log N)^{-c}$. For these applications one does need the full strength of Theorem 2.7 (when $G = \mathbb{Z}/N\mathbb{Z}$).

The key to the proof of Theorem 11.1 is the following density increment result, which is in the spirit of Proposition 7.3 but is rather more complicated.

Proposition 11.2. *Let G be an abelian group of size N and suppose that all elements of G have order at most $e^{\sqrt{\log N}}$. Suppose that $(6, N) = 1$, that N is sufficiently large, that $\delta \geq 1/\log \log N$ is smaller than some absolute constant and that $A \subseteq G$ has size at least δN . Suppose that A contains no four-term arithmetic progression. Then there is some subgroup $G' \leq G$, $|G'| \geq N^{1/8}$, together with a coset $t + G'$ such that*

$$\mathbb{E}_{x \in t+G'} 1_A(x) \geq \mathbb{E}_{x \in G} 1_A(x) + \delta^C,$$

where C is some absolute constant.

Proof. Write $\alpha := \mathbb{E}1_A$ and $f := 1_A - \alpha$. Applying Corollary 1.12 we have $\|f\|_{U^3(G)} \geq \delta^3/8$; applying Theorem 2.7 we can then find a regular Bohr set $B(S, \rho)$ in G with $|S| \leq \delta^{-C}$ and $\rho \geq \delta^C$ such that

$$\mathbb{E}_{y \in G} \|f\|_{u^3(y+B)} \geq \delta^C.$$

Applying Lemma 10.4 we see that B contains a proper coset progression $P + H$ of rank at most δ^{-C} such that

$$|P + H| \geq \exp(-\delta^{-C})N.$$

Since every element in G has order at most $e^{\sqrt{\log N}}$, we see that the lengths of the proper progression P are also at most $e^{\sqrt{\log N}}$. Thus $|P| \leq N^{1/2}$, which implies that

$$|H| \geq N^{1/3}. \quad (11.1)$$

By the definition (10.2) of H , we see that B can be partitioned into cosets of H , and hence

$$\mathbb{E}_{y \in G} \|f\|_{u^3(y+H)} \geq \delta^C.$$

Set $\alpha'(y) := \mathbb{E}_{x \in y+H} 1_A(x)$. From the triangle inequality we have

$$\begin{aligned} \|f\|_{u^3(y+H)} &\leq \|1_A - \alpha'(y)\|_{u^3(y+H)} + |\alpha'(y) - \alpha| \\ &= \|1_A - \alpha'(y)\|_{u^3(y+H)} - 8|\alpha'(y) - \alpha| + 9|\alpha'(y) - \alpha|, \end{aligned}$$

and so either

$$\mathbb{E}_{y \in G} \|1_A - \alpha'(y)\|_{u^3(y+H)} - 8|\alpha'(y) - \alpha| \geq \delta^C/2 \quad (11.2)$$

or

$$\mathbb{E}_{y \in G} |\alpha'(y) - \alpha| \geq \delta^C/18.$$

Suppose that the latter inequality holds. From Lemma 4.1 we have

$$\mathbb{E}_{y \in G} (\alpha'(y) - \alpha) = 0;$$

adding this to the preceding estimate and applying the pigeonhole principle, we conclude that there exists y such that

$$\mathbb{E}_{x \in y+H} 1_A(x) = \alpha'(y) \geq \alpha + \delta^C/36 = \mathbb{E}_{x \in G} 1_A(x) + \delta^C/36, \quad (11.3)$$

which implies the proposition (with a change to the absolute constant C). Suppose, then, that (11.2) holds. By the pigeonhole principle, we can find $y \in G$ such that

$$\|1_A(x) - \alpha'(y)\|_{u^3(y+H)} \geq 8|\alpha'(y) - \alpha| + \delta^C/2.$$

By translating A we may take $y = 0$. Writing $\alpha' := \mathbb{E}_{x \in H} 1_A(x)$ we conclude the existence of a quadratic phase function $\phi : H \rightarrow \mathbb{R}/\mathbb{Z}$ such that

$$|\mathbb{E}_{x \in H} (f_H(x) e(-\phi(x)))| \geq 8|\alpha' - \alpha| + \delta^C/2,$$

where $f_H(x) := 1_A(x) - \alpha'$. Applying Lemma 3.1 (or Lemma 10.7), we may thus find a self-adjoint homomorphism $M : H \rightarrow \hat{H}$ and $\xi \in \hat{H}$ such that

$$|\mathbb{E}_{x \in H} f_H(x) e(-Mx \cdot x) e(-\xi \cdot x)| \geq 8|\alpha' - \alpha| + \delta^C/2. \tag{11.4}$$

As in the proof of Theorem 7.1, the next step is to locate a large subgroup of H on which M vanishes. To achieve this we need some preliminary algebraic (and Fourier-analytic) lemmas, of similar flavour to Lemma 7.4.

Lemma 11.3 (orthogonal complements). *Let K be any subgroup of H and let $K^\perp \subseteq H$ be the subgroup*

$$K^\perp := \{y \in H : Mx \cdot y = 0 \text{ for all } x \in K\}.$$

Then $|K^\perp| \geq |H|/|K|$.

Proof. Let $\phi : H \rightarrow \hat{K}$ be the homomorphism $\phi(y)(x) := Mx \cdot y$. Then K^\perp is precisely the kernel of ϕ . But since ϕ is a homomorphism with a domain of size $|H|$ and a range of size at most $|\hat{K}| = |K|$, the claim follows. \square

Lemma 11.4 (Gauss sum lemma). *Let K be any finite group with $(6, K) = 1$ and suppose that every non-zero element of K has order at most t for some $t > 2$. Let $M : K \rightarrow \hat{K}$ be a self-adjoint homomorphism. Then, if $|K| \geq 100t^4$, there exists a non-zero element $x \in K \setminus \{0\}$ such that $Mx \cdot x = 0$.*

Proof. We can assume that M is injective (and hence bijective), since otherwise we can just set x to equal a non-zero element in the kernel of M . Using the classification of finite abelian groups, we can write K as the direct sum of cyclic groups of odd prime power order. Note that if p is a prime such that at least three cyclic groups of order equal to a power of p appear in this direct sum, then K contains a subgroup isomorphic to \mathbb{F}_p^3 . Restricting M to \mathbb{F}_p^3 (note that M will still be self-adjoint) and applying Lemma 7.4, we can then conclude the existence of a non-zero $x \in K$ such that $Mx \cdot x = 0$. Thus we may assume that for each $p \geq 5$ there are at most two cyclic groups of order equal to a power of p in the direct sum decomposition of K , in which case we may write

$$K = \otimes_{j=1}^k (\mathbb{Z}/p_j^{u_j} \mathbb{Z}) \times (\mathbb{Z}/p_j^{u'_j} \mathbb{Z})$$

for some distinct primes p_1, \dots, p_k and exponents u_j, u'_j . Let $n \neq 0$ be any integer, and define

$$n^\perp := \{x \in K : nx = 0\}.$$

Writing

$$n = (-1)^v p_1^{v_1} \cdots p_k^{v_k},$$

one confirms the estimate

$$|n^\perp| = \prod_{j=1}^k \min(p_j^{u_j}, p_j^{v_j}) \min(p_j^{u'_j}, p_j^{v_j}) \leq |n|^2. \tag{11.5}$$

Let $\chi : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}$ be a smooth bump function such that $\chi(s) = 1$ when $\|s\|_{\mathbb{R}/\mathbb{Z}} < \frac{1}{2}t$, $\chi(s) = 0$ when $\|s\|_{\mathbb{R}/\mathbb{Z}} \geq 1/t$, and for which the derivative estimate $\|\chi'''\|_{\infty} \leq 100/t^3$ holds true. Observe that if $Mx \cdot x$ is non-zero, then $\|Mx \cdot x\|_{\mathbb{R}/\mathbb{Z}} \geq 1/t$ since x (and hence $Mx \cdot x$) has order at most t . Thus

$$\mathbb{E}_{x \in K} \mathbf{1}_{Mx \cdot x=0} = \mathbb{E}_{x \in K} \chi(Mx \cdot x).$$

Expanding in a Fourier series gives

$$\mathbb{E}_{x \in K} \mathbf{1}_{Mx \cdot x=0} = \sum_{n \in \mathbb{Z}} \hat{\chi}(n) \mathbb{E}_x e(nMx \cdot x), \quad \text{where } \hat{\chi}(n) = \int_{\mathbb{R}/\mathbb{Z}} \chi(s) e(-ns) ds.$$

Isolating the term $n = 0$ we obtain the inequality

$$|\mathbb{E}_{x \in K} \mathbf{1}_{Mx \cdot x=0} - \hat{\chi}(0)| \leq \sum_{n \in \mathbb{Z} \setminus \{0\}} |\hat{\chi}(n)| |\mathbb{E}_x e(nMx \cdot x)|.$$

Now by (4.1), Fourier inversion, the injectivity and self-adjointness of M and the hypothesis that $|K|$ is odd, we have

$$\begin{aligned} |\mathbb{E}_x e(nMx \cdot x)| &\leq |\mathbb{E}_h \mathbb{E}_x e(nM(x+h) \cdot (x+h) - nMx \cdot x)|^{1/2} \\ &\leq (\mathbb{E}_h |\mathbb{E}_x e(2nMx \cdot h)|)^{1/2} \\ &= (\mathbb{E}_h \mathbf{1}_{M2nh=0})^{1/2} \\ &= (\mathbb{E}_h \mathbf{1}_{nh=0})^{1/2} \\ &= \left(\frac{|n^\perp|}{|K|} \right)^{1/2} \\ &\leq \frac{n}{\sqrt{|K|}}, \end{aligned}$$

the last estimate following from (11.5). It follows that

$$|\mathbb{E}_{x \in K} \mathbf{1}_{Mx \cdot x=0} - \hat{\chi}(0)| \leq \sum_{n \in \mathbb{Z} \setminus \{0\}} |\hat{\chi}(n)| |n| |K|^{-1/2}. \quad (11.6)$$

Now by integrating by parts three times and using the bound on $\|\chi'''\|_{\infty}$ one sees that $|\hat{\chi}(n)| \leq t^2/|n|^3$. In combination with the trivial bound $\|\hat{\chi}\|_{\infty} \leq 2/t$, we obtain

$$\sum_{n \neq 0} |\hat{\chi}(n)| |n| \leq \sum_{|n| < t} \frac{|n|}{t} + \sum_{|n| \geq t} \frac{t^2}{n^2} \leq 6t.$$

Thus, since $|K| \geq 100t^4$, (11.6) implies that

$$\mathbb{E}_{x \in K} \mathbf{1}_{Mx \cdot x=0} \geq \frac{1}{3t} > \frac{1}{|K|},$$

which immediately implies the result. \square

Corollary 11.5. *Let H be a finite additive group, $(|H|, 6) = 1$, such that every element has order at most t , $t \geq 2$, and let $M : H \rightarrow \hat{H}$ be a self-adjoint homomorphism. Then there exists a subgroup K of H such that*

$$|K| \geq |H|^{1/2}/10t^2 \tag{11.7}$$

and $Mx \cdot y = 0$ for all $x, y \in K$.

Proof. Let K be a subgroup of H on which the quadratic form $Mx \cdot y$ vanishes (i.e. $Mx \cdot y = 0$ for all $x, y \in K$), and which is maximal with respect to set inclusion. Observe that the orthogonal complement K^\perp of K contains K , and hence by Lemma 11.3 the quotient group K^\perp/K has cardinality at least $|H|/|K|^2$. Also, every element in this group has order at most t . Since $Mx \cdot y = My \cdot x = 0$ whenever $x \in K$ and $y \in K^\perp$, we see that the bilinear form $Mx \cdot y$ descends to a bilinear form on K^\perp/K . If the associated quadratic form vanished for at least one non-zero element of K^\perp/K , then by adjoining this element to K we could contradict the maximality of K . Thus we may assume that there is no such form. But then by Lemma 11.4 we have $|K^\perp/K| \leq 100t^4$. Combining this with our lower bound for $|K^\perp/K|$ we obtain the result. \square

Let us return now to the situation (11.4), and let K be the subgroup obtained by the above Corollary. By Lemma 4.1 we have

$$|\mathbb{E}_{y \in H} \mathbb{E}_{x \in y+K} f_H(x) e(-Mx \cdot x) e(-\xi \cdot x)| \geq 8|\alpha' - \alpha| + \delta^C/2.$$

Setting $\alpha''(y) := \mathbb{E}_{x \in y+K} 1_A(x)$, the triangle inequality implies that either

$$\mathbb{E}_{y \in H} |\alpha''(y) - \alpha'| \geq 2|\alpha' - \alpha| + \delta^C/12 \tag{11.8}$$

or

$$\mathbb{E}_{y \in H} (|\mathbb{E}_{x \in y+K} f_H(x) e(-Mx \cdot x) e(-\xi \cdot x)| - 3|\alpha''(y) - \alpha'|) \geq 2|\alpha' - \alpha| + \delta^C/4. \tag{11.9}$$

Suppose that (11.8) holds. From Lemma 4.1 we have

$$\mathbb{E}_{y \in H} (\alpha''(y) - \alpha) = 0$$

and hence

$$2\mathbb{E}_{y \in H} \max(\alpha''(y) - \alpha', 0) = \mathbb{E}_y |\alpha''(y) - \alpha'| \geq 2|\alpha' - \alpha| + \delta^C/12.$$

By the pigeonhole principle we thus conclude that there exists $y \in H$ such that

$$2(\alpha''(y) - \alpha') \geq 2|\alpha' - \alpha| + \delta^C/12,$$

and hence, by the triangle inequality,

$$\mathbb{E}_{x \in y+K} 1_A(x) = \alpha''(y) \geq \alpha' + |\alpha' - \alpha| + \delta^C/24 \geq \alpha + \delta^C/24 = \mathbb{E}_{x \in G} 1_A(x) + \delta^C/24. \tag{11.10}$$

This, together with the lower bound

$$|K| \geq |H|^{1/2}/10e^{2\sqrt{\log N}} > N^{1/8}$$

(cf. (11.7)) implies the proposition under the assumption that (11.8) holds.

Suppose, then, that (11.9) holds instead. By the pigeonhole principle, we can find $y \in H$ such that

$$|\mathbb{E}_{x \in y+K} f_H(x) e(-Mx \cdot x) e(-\xi \cdot x)| \geq 3|\alpha''(y) - \alpha'| + 2|\alpha' - \alpha| + \delta^C/4.$$

Splitting f_H as $(1_A - \alpha''(y)) + (\alpha''(y) - \alpha')$ and using the triangle inequality, we conclude that

$$|\mathbb{E}_{x \in y+K} (1_A(x) - \alpha''(y)) e(-Mx \cdot x) e(-\xi \cdot x)| \geq 2|\alpha''(y) - \alpha| + \delta^C/4.$$

We write $x = y + z$ and use the fact that the bilinear form $My \cdot z$ is symmetric and vanishes on K to conclude that

$$|\mathbb{E}_{z \in K} (1_A(y+z) - \alpha''(y)) e(-(2My + \xi) \cdot z)| \geq 2|\alpha''(y) - \alpha| + \delta^C/4.$$

Consider now the homomorphism $\phi : K \rightarrow \mathbb{R}/\mathbb{Z}$ defined by $\phi(x) := (2My + \xi) \cdot x$. A simple pigeonhole argument shows that there exist at least $|K|e^{-\sqrt{\log N}}$ elements x of K for which $\|\phi(x)\|_{\mathbb{R}/\mathbb{Z}} < e^{\sqrt{\log N}}$. But since every element of K has order at most $e^{\sqrt{\log N}}$, we conclude that $\phi(x) = 0$. Thus if we set K' to be the kernel of ϕ , then K' is a subgroup of K with

$$|K'| \geq |K|e^{-\sqrt{\log N}} > N^{1/8}. \quad (11.11)$$

We then apply Lemma 4.1 again to conclude that

$$|\mathbb{E}_{w \in K} \mathbb{E}_{z \in w+K'} ((1_A(y+z) - \alpha''(y)) e(-(2My + \xi) \cdot z))| \geq 2|\alpha''(y) - \alpha| + \delta^C/4.$$

Since the phase $e(-(2My + \xi) \cdot z)$ is constant for $z \in w + K'$, for fixed w , we conclude that

$$\mathbb{E}_{w \in K} |\mathbb{E}_{z \in w+K'} (1_A(y+z) - \alpha''(y))| \geq 2|\alpha''(y) - \alpha| + \delta^C/4$$

and thus, writing $\alpha'''(t) := \mathbb{E}_{x \in t+K'} 1_A(x)$, that

$$\mathbb{E}_{w \in K} |\alpha'''(w+y) - \alpha''(y)| \geq 2|\alpha''(y) - \alpha| + \delta^C/4. \quad (11.12)$$

Now from Lemma 4.1 we have

$$\mathbb{E}_{w \in K} (\alpha'''(w+y) - \alpha''(y)) = 0.$$

Together with (11.12) this implies that

$$2\mathbb{E}_{w \in K} \max(\alpha'''(w+y) - \alpha''(y), 0) \geq 2|\alpha''(y) - \alpha| + \delta^C/4,$$

and so there exists $w \in K$ such that

$$\alpha'''(w+y) - \alpha''(y) \geq |\alpha''(y) - \alpha| + \delta^C/8.$$

One final application of the triangle inequality at last gives that

$$\mathbb{E}_{x \in w+y+K'} 1_A(x) = \alpha'''(w+y) \geq \alpha + \delta^C/8 = \mathbb{E}_{x \in G} 1_A(x) + \delta^C/8.$$

Together with the lower bound (11.11), this concludes the proof of Proposition 11.2. \square

Proof of Theorem 11.1. Let G be an abelian group and let $A \subseteq G$ be a set with cardinality at least δN which contains no four distinct elements in arithmetic progression. We wish to show that $\delta \ll (\log \log N)^{-c}$ for some absolute constant c ; thus we may certainly suppose that $\delta \geq 2/\log \log N$.

Suppose that G has an element g of order greater than $\exp((\log N)^{1/3})$. Writing $H = \langle g \rangle$, we see that there is some coset $y + H$ such that $\mathbb{E}_{x \in y+H} 1_A(x) \geq \delta$. The result of Gowers [25] then immediately implies that

$$\delta \ll (\log \log |H|)^{-c} \ll (\log \log N)^{-c}.$$

Suppose, then, that all elements of G have order at most $\exp((\log N)^{1/3})$. We will define a sequence $G = G_0 \supseteq G_1 \supseteq \dots$ of subgroups of G with cardinalities $N = N_0 \supseteq N_1 \supseteq \dots$. The sequence will be defined in such a way that

$$N_j \geq \exp((\log N)^{2/3}), \tag{11.13}$$

which means that no element in G_j has order greater than $\exp((\log N_j)^{1/2})$, and also that $\delta \geq 1/\log \log N_j$. This is to enable us to apply Proposition 11.2.

Suppose that we have defined G_j . For any x_j , the set $(x_j + A) \cap G_j$ does not contain a four-term arithmetic progression. Suppose that x_j is such that $\mathbb{E}_{x \in x_j+G_j} 1_A(x) \geq \delta$. Then, applying Proposition 11.2, we see that there is some G_{j+1} , $N_{j+1} := |G_{j+1}| \geq N_j^{1/8}$, together with some x_{j+1} such that

$$\mathbb{E}_{x \in x_{j+1}+G_{j+1}} 1_A(x) \geq \mathbb{E}_{x \in x_j+G_j} 1_A(x) + \delta^C.$$

Iterating this construction leads to a contradiction for some $j \leq \delta^{-C}$ unless (11.13) is violated. Since $N_j \geq N^{(1/8)^j}$, we must therefore have

$$N^{(1/8)^{\delta^{-C}}} < \exp((\log N)^{2/3}),$$

which implies the required bound $\delta \ll (\log \log N)^{-c}$. □

Remarks 11.6. We hope to prove a bound of the form $r_4(G) \ll N(\log N)^{-c}$ in a future paper by combining the ideas of [35] with nested Bohr set technology in the spirit of that used in §§ 8 and 9 of the present paper. These methods were first introduced by Bourgain [10], who obtained the bound $r_3(G) \ll N(\log N)^{-(1/2)+\epsilon}$, which is still the best currently known when $G = \mathbb{Z}/N\mathbb{Z}$. A feature of this approach is that, unlike in the present section, it is no easier to deal with $\mathbb{Z}/N\mathbb{Z}$ than it is with an arbitrary abelian G . We note that the celebrated Erdős–Turán conjecture [16] is roughly equivalent to a bound of the form $r_k(\mathbb{Z}/N\mathbb{Z}) \ll_k N(\log N)^{-1+\epsilon}$, and so even in the case $k = 3$ there is an awful lot left to be done.

12. An ergodic-theory interpretation

We now connect the U^3 inverse theorems discussed earlier to ergodic theory, and in particular to the recent work of Host and Kra [43] and Ziegler [64].

Define a *measure-preserving system* $(X, \mathcal{B}, T, \mathbb{P})$ to be a probability space $(X, \mathcal{B}, \mathbb{P})$ with an invertible measure-preserving (i.e. probability-preserving) shift operator $T : X \rightarrow X$. This induces a shift operator T on random variables $f : X \rightarrow \mathbb{R}$ by the formula $Tf(x) := f(T^{-1}x)$, and more generally $T^n f(x) := f(T^{-n}x)$ for any $n \in \mathbb{Z}$. We use $\mathbb{E}_X(f)$ to denote the expectation of f .

The *Furstenberg correspondence principle* (see, for example, [19]) equates combinatorial theorems such as Szemerédi’s theorem to recurrence results in ergodic theory. In particular, Theorem 1.1 is logically equivalent (using the axiom of choice) to the following theorem.

Theorem 12.1 (Furstenberg recurrence theorem [18, 24]). *Let $(X, \mathcal{B}, T, \mathbb{P})$ be a measure-preserving system and let $f \in L^\infty(X, \mathcal{B})$ be any non-negative random variable with $\mathbb{E}_X f > 0$. Then for every $k \geq 1$ we have*

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{-N \leq n \leq N} \mathbb{E}_X f T^n f \cdots T^{(k-1)n} f > 0.$$

In particular, if $A \in \mathcal{B}$ is any event with positive probability $\mathbb{P}(A) > 0$, then

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{-N \leq n \leq N} \mathbb{P}(A \cap T^n A \cap \cdots \cap T^{(k-1)n} A) > 0.$$

Recently, it was shown in [43] and [64] that this limit inferior can in fact be replaced by a limit; earlier work related to the $k = 4$ case can be found in [14, 23, 41, 42]. The two approaches are slightly different; the argument in [43] proceeds by establishing the ergodic-theory analogue of an inverse theorem for the Gowers uniformity norm $U^d(\mathbb{Z}/N\mathbb{Z})$. Indeed, if $f \in L^\infty(X, \mathcal{B})$ is any complex-valued random variable, define the quantity $\|f\|_{U^d(T)}$ for $d \geq 0$ by the formula

$$\|f\|_{U^d(T)} := \lim_{N \rightarrow \infty} \left(\mathbb{E}_{-N \leq h_1, \dots, h_d \leq N} \mathbb{E}_X \prod_{\omega \in \{0,1\}^d} \mathcal{C}^{|\omega|} T^{\omega \cdot h} f(x) \right)^{1/2^d}.$$

It can be shown [42, 43] that this limit actually exists, and it can also be shown that $\|f\|_{U^d(T)}$ is in fact a seminorm on bounded random variables for any $d \geq 1$ (see [43]). This seminorm is clearly related to the $U^d(G)$ norms defined in Definition 1.8. For instance, if T is periodic of order N , then it is easy to see that the $U^d(T)$ norm of $f(x)$ is the L^{2^d} average of the $U^d(\mathbb{Z}/N\mathbb{Z})$ norm of f restricted to the orbits of T :

$$\|f\|_{U^d(T)}^{2^d} = \mathbb{E}_x \|(T^h f(x))_{h \in \mathbb{Z}/N\mathbb{Z}}\|_{U^d(\mathbb{Z}/N\mathbb{Z})}^{2^d}.$$

Here of course we take advantage of the periodicity of T to define T^h for $h \in \mathbb{Z}/N\mathbb{Z}$ in the obvious manner.

In [43] it was observed that this seminorm controls expressions such as those appearing in the Furstenberg recurrence theorem. Indeed, there is an analogue of Proposition 1.11 which asserts that if f_0, \dots, f_{k-1} are bounded random variables and at least one of them has vanishing $U^{k-2}(T)$ norm, then

$$\lim_{N \rightarrow \infty} \mathbb{E}_{-N \leq n \leq N} \mathbb{E}_X f_0 T^n f_1 \cdots T^{(k-1)n} f_n = 0;$$

in fact one can make the slightly stronger claim that $\mathbb{E}_{-N \leq n \leq N} T^n f_1 \cdots T^{(k-1)n} f_n$ converges to zero in (for instance) the $L^2(X)$ sense. Informally, this fact shows that functions with vanishing $U^{k-2}(T)$ norm are irrelevant for understanding k -fold recurrence.

It is thus of interest to determine when the $U^{k-2}(T)$ norm is positive. This question is answered in [43] using the language of nilsystems. We first recall some notation. If G is a (not necessarily abelian) group written multiplicatively and if $g, h \in G$, we let $[g, h] := g^{-1}h^{-1}gh$ denote the commutator of g and h . If G' and G'' are subgroups of G , we let $[G', G''] = [G'', G']$ be the subgroup generated by the commutators $\{[g', g''] : g' \in G', g'' \in G''\}$. We then define the *lower central series*

$$G = G_1 \supseteq G_2 \supseteq G_3 \supseteq \cdots$$

of subgroups of G by the recursive definition $G_1 := G$; $G_{k+1} := [G, G_k]$. We say that G is $(k-2)$ -step nilpotent for some $k \geq 2$ if G_{k-1} is trivial. Thus, for instance, a group is 1-step nilpotent if and only if it is abelian.

A $(k-2)$ -step nilmanifold is defined to be a manifold of the form $G/\Gamma := \{x\Gamma : x \in G\}$, where G is a finite-dimensional nilpotent Lie group and Γ is a discrete subgroup of G which is co-compact (i.e. the nilmanifold G/Γ is compact). Note that we do not assume that Γ is normal, and hence a nilmanifold need not have a group structure. It is, however, a compact symmetric space, with a left-action of the group G . Thus there is a unique invariant Haar measure \mathbb{P} on a nilmanifold, which we normalize to be a probability measure. Thus every nilmanifold is a probability space, taking the σ -algebra to be the Borel σ -algebra.

If $g \in G$ then we write T_g for the *shift operator* from G to itself defined by $T_g(x) = gx$, and also (by abuse of notation) for the map from G/Γ to itself defined by $T_g(x\Gamma) := gx\Gamma$. This latter map is measure preserving and invertible. Let us call a $(k-2)$ -step nilmanifold with one of these shift operators a $(k-2)$ -step nilflow. A $(k-2)$ -step nilfunction is defined to be any continuous function $F : G/\Gamma \rightarrow \mathbb{C}$ on a $(k-2)$ -step nilmanifold; given such a nilfunction, a point $x_0 \in G/\Gamma$ and a group element $g \in G$, we define the associated *basic $(k-2)$ -step nilsequence* $F_{g,x_0} : \mathbb{Z} \rightarrow \mathbb{C}$ by the formula*

$$F_{g,x_0}(n) := F(T_g^n x_0) \quad \text{for all } n \in \mathbb{Z}.$$

We can truncate this to $\mathbb{Z}/N\mathbb{Z}$ and define the truncated nilsequence $F_{N,g,x_0} : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{C}$ by the formula

$$F_{N,g,x_0}(n) := F_{g,x_0}(n) = F(T_g^n x_0) \quad \text{for all } -N/2 < n \leq N/2,$$

where we identify the integers from $-N/2$ to $N/2$ with $\mathbb{Z}/N\mathbb{Z}$ in the usual manner.

We now give three key examples of nilflows and nilsequences.

Example 12.2 (the circle nilflow). Let G be the one-dimensional matrix group

$$G := \begin{pmatrix} 1 & \mathbb{R} \\ 0 & 1 \end{pmatrix} := \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{R} \right\}$$

* A general $(k-2)$ -step nilsequence is defined as the uniform limit of basic $(k-2)$ -step sequences; see [5] for further analysis of these nilsequences.

and let Γ be the discrete subgroup

$$\Gamma := \begin{pmatrix} 1 & \mathbb{Z} \\ 0 & 1 \end{pmatrix} := \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} : n \in \mathbb{Z} \right\}.$$

Then G/Γ is a 1-step nilmanifold (and hence also a 2-step nilmanifold)—indeed we can easily identify it with the unit circle \mathbb{R}/\mathbb{Z} . A shift T_g on this nilmanifold then corresponds to a simple translation $x \mapsto x + \alpha$, where $\alpha \in \mathbb{R}$ is the upper-right matrix entry of g . In particular, we observe that if $F : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{C}$ is any function, we see that the sequence

$$n \mapsto F(T_g^n x) = F(x + n\alpha)$$

is a basic 1-step nilsequence. Thus, for instance, the linear phase function $n \mapsto e(n\alpha)$ is a basic 1-step nilsequence (and hence also a 2-step nilsequence). More generally, any quasiperiodic sequence is a basic 1-step nilsequence, and any almost periodic sequence can be expressed as the uniform limit of basic 1-step nilsequences.

Example 12.3 (the skew shift nilflow). Now we consider the example

$$G := \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}, \quad \Gamma := \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}.$$

Then G/Γ is a 2-step nilmanifold, and one can identify it topologically with the 2-torus $(\mathbb{R}/\mathbb{Z})^2$ by the identification

$$(x, y) \equiv \begin{pmatrix} 1 & 0 & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} \Gamma.$$

If we let

$$g := \begin{pmatrix} 1 & m & \beta \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix}$$

be a typical element of G (thus $m \in \mathbb{Z}$ and $\alpha, \beta \in \mathbb{R}$), then the shift T_g is given by $(x, y) \mapsto (x + \alpha, y + \beta + mx)$, and thus if $F : (\mathbb{R}/\mathbb{Z})^2 \rightarrow \mathbb{C}$ is any function, then the sequence

$$n \mapsto F(T_g^n(x, y)) = F(x + n\alpha, y + n\beta + \frac{1}{2}mn(n+1)\alpha)$$

is a basic 2-step nilsequence. Thus, for instance, the quadratic phase function $n \mapsto e(\frac{1}{2}\alpha n(n+1))$ is a basic 2-step nilsequence. More generally, any quadratic phase $n \mapsto e(\alpha n^2 + \beta n + \gamma)$, or finite linear combination of such phases, is a basic 2-step nilsequence.

Example 12.4 (the Heisenberg nilflow). Now we consider the example

$$G := \begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}, \quad \Gamma := \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}.$$

Then G/Γ is a 2-step nilmanifold. By using the identification

$$(x, y, z) \equiv \begin{pmatrix} 1 & z & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} \Gamma,$$

we can identify G/Γ (as a set) with \mathbb{R}^3 , quotiented out by the equivalence relations

$$(x, y, z) \sim (x + a, y + b + az, z + c) \quad \text{for all } a, b, c \in \mathbb{Z}.$$

This can in turn be coordinatized by the cylinder $[-\frac{1}{2}, \frac{1}{2}] \times (\mathbb{R}/\mathbb{Z})^2$ with the identification $(-\frac{1}{2}, y, z) \sim (\frac{1}{2}, y + z, z)$.

Let $F : G/\Gamma \rightarrow \mathbb{C}$ be a function. We may lift this to a function $\tilde{F} : G \rightarrow \mathbb{C}$, defined by $\tilde{F}(g) := F(g\Gamma)$. In coordinates, this lift takes the form

$$\tilde{F}(x, y, z) = F(\{x\}, y - [x]z \pmod{1}, z \pmod{1}),$$

where $[x] = x - \{x\}$ is the nearest integer to x (we round half-integers up). If we let

$$g := \begin{pmatrix} 1 & \gamma & \beta \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix}$$

be an element of G , then the shift $T_g : G \rightarrow G$ is given by

$$T_g(x, y, z) = (x + \alpha, y + \beta + \gamma x, z + \gamma),$$

from which a short induction confirms that

$$T_g^n(x, y, z) = (x + n\alpha, y + n\beta + \frac{1}{2}n(n + 1)\alpha, z + n\gamma).$$

Therefore, if $F : G/\Gamma \rightarrow \mathbb{C}$ is any function, written as a function $F : [-\frac{1}{2}, \frac{1}{2}] \times (\mathbb{R}/\mathbb{Z})^2 \rightarrow \mathbb{C}$ with $F(-\frac{1}{2}, y, z) = F(\frac{1}{2}, y + z, z)$, then we have

$$F(T_g^n(x, y, z)) = F(\{x + n\alpha\}, y + n\beta + \frac{1}{2}n(n + 1)\alpha\gamma - [x + n\alpha](z + n\gamma), z + n\gamma).$$

This, of course, is a basic 2-step nilsequence. We see, for instance, that the generalized quadratic phase function $n \mapsto e(\frac{1}{2}n(n + 1)\alpha\gamma - [n\alpha]n\gamma)$ is a basic 2-step nilsequence.

We call the three basic examples just discussed the *fundamental* 2-step nilsequences. They may be used in a straightforward product construction to construct further nilsequences, as we now describe.

If $(G/\Gamma, T_g)$ and $(G'/\Gamma', T_{g'})$ are 2-step nilflows, then so is the direct sum $((G \oplus G')/(\Gamma \oplus \Gamma'), T_{(g, g')})$. Also, if $F : G/\Gamma \rightarrow \mathbb{C}$ and $F' : G'/\Gamma' \rightarrow \mathbb{C}$ are functions, and we define the tensor product $F \otimes F' : (G \oplus G')/(\Gamma \oplus \Gamma')$ in the usual manner as

$$F \otimes F'(x, x') := F(x)F'(x'),$$

then we see that the function $F \otimes F'(T_{(g,g')}^n(x, x'))$ factors as

$$F \otimes F'(T_{(g,g')}^n(x, x')) = F(T_g^n x) F'(T_{g'}^n x').$$

Now suppose we take n_1 nilsequences coming from circle nilflows, n_2 nilsequences coming from skew shift nilflows and n_3 nilsequences coming from Heisenberg nilflows, and tensor them all together. What results is a nilsequence on a 2-step nilmanifold G/Γ that is topologically the cube $[-\frac{1}{2}, \frac{1}{2}]^{n_1+2n_2+3n_3}$ with faces identified. Two-step nilsequences of this type—that is to say, tensor products of fundamental nilsequences—are in a sense the only important ones if one is interested in the Gowers U^3 norm. We call them the *elementary 2-step nilsequences* (we also refer to *elementary 2-step nilmanifolds* and *elementary 2-step nilflows*).

Given an elementary 2-step nilsequence, it is natural to refer to $n_1 + 2n_2 + 3n_3$ as its *dimension*. It is also of interest to have a notion of how continuous the underlying function $F : G/\Gamma \rightarrow \mathbb{C}$ is. We adopt a rather low-brow approach to this concept which is sufficient for our purposes. Let $\delta = 1/m$ be the reciprocal of an integer. Then we may divide $(-\frac{1}{2}, \frac{1}{2})^d$ into m^d cubes of side-length δ , for any d , and in fact these subdivisions respect the quotienting of Examples 12.2, 12.3 and 12.4 (for $d = 1, 2, 3$, respectively), giving what we refer to as δ -nets on the three fundamental 2-step nilmanifolds G/Γ . We refer to the building blocks of these nets as the δ -atoms. Taking products, we may obtain δ -atoms and a δ -net on any elementary 2-step nilmanifold G/Γ . Finally, if $F : G/\Gamma \rightarrow \mathbb{C}$ is a function and if $K > 0$ is a constant, we say that F is K -Lipschitz if, for all $\delta = 1/m$ and for all δ -atoms A , we have

$$|F(x) - F(x')| \leq K\delta$$

whenever $x, x' \in A$. The following lemma is straightforward.

Lemma 12.5. *Suppose that $F_i : G_i/\Gamma_i \rightarrow \mathbb{C}$, $i = 1, \dots, k$, are K -Lipschitz functions on elementary 2-step nilmanifolds G_i/Γ_i . Then the tensor product $F_1 \otimes \dots \otimes F_k$ is Kk -Lipschitz.*

Now, it has been known since the work of Furstenberg and Weiss [23] that random variables on a $(k-2)$ -step nilflow $(G/\Gamma, T_g)$ have a non-trivial behaviour with respect to k -term recurrence. Indeed, given any bounded random variable $f_0 : G/\Gamma \rightarrow \mathbb{C}$ which is not identically zero, one can find f_1, \dots, f_{k-1} such that the averages

$$\mathbb{E}_{-N \leq n \leq N} \mathbb{E}_X f_0 T^n f_1 \dots T^{(k-1)n} f_n$$

do not converge to zero; this is basically due to non-trivial algebraic relations between the k points $x\Gamma, g^n x\Gamma, \dots, g^{(k-1)n} x\Gamma$. In particular, the $U^{k-1}(T)$ norm is non-degenerate on this nilmanifold (and is thus a genuine norm); see [43, 64] for some further discussion of this fact. We will prove a variant of this statement.

Proposition 12.6 (nilsequences obstruct uniformity). *Let $k \geq 3$, and let $(G/\Gamma, T)$ be a $(k-2)$ -step nilsystem. Let $F : G/\Gamma \rightarrow \mathbb{C}$ be a continuous function on G/Γ*

which is not identically zero. Suppose that $N > k - 1$ is a prime and that $f : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathcal{D}$ is a function such that

$$|\mathbb{E}_{-N/2 \leq n \leq N/2} f(n) \overline{F(T_g^n x)}| \geq \eta.$$

Then we have

$$\|f\|_{U^3} \geq c_{F,G/\Gamma}(\eta) > 0$$

uniformly in $x \in G/\Gamma$, $g \in G$ and N .

Proof. In proving this proposition we will use the following lemma to the effect that the point $T_g^{n+(k-1)r}x$ is completely constrained by the cosets $T_g^n x\Gamma, T_g^{n+r}x\Gamma, \dots, T_g^{n+(k-2)r}x\Gamma$.

Lemma 12.7. *Let $(G/\Gamma, T)$ be a $(k - 2)$ -step nilsystem. Then there is a compact set $\Sigma \subseteq (G/\Gamma)^{k-1}$ and a continuous function $P : \Sigma \rightarrow G$ such that for all $n, r \in \mathbb{Z}$, $g \in G$ and $x \in G/\Gamma$ we have*

$$(T_g^n x, T_g^{n+r} x, \dots, T_g^{n+(k-2)r} x) \in \Sigma$$

and

$$P(T_g^n x, T_g^{n+r} x, \dots, T_g^{n+(k-2)r} x) = T_g^{n+(k-1)r} x\Gamma.$$

The existence of a constraint of this type is discussed in several places in the ergodic-theory literature [5, 20, 21, 63]. For the convenience of the reader we supply a self-contained proof in Appendix A.

The proof of Proposition 12.6 is not dissimilar to that of Theorem 2.7 (ii), which was given at the start of §9, but here we use Lemma 12.7 in place of (9.3), and the technical details are rather different.

Assume without loss of generality that $\|F\|_\infty \leq 1$, and let $\varepsilon \leq 1/100k$ be a small constant (it will be chosen to be a small multiple of η). Observe that the function $F \circ P : \Sigma \rightarrow \mathbb{C}$ is continuous, hence uniformly continuous, on the compact set Σ . In particular we can find a neighbourhood $V \subseteq G$ of the identity 1 which depends on $\varepsilon, F, G/\Gamma$ such that

$$F(P(x_0, \dots, x_{k-2})) = F(P(y_0, \dots, y_{k-2})) + O(\varepsilon) \tag{12.1}$$

whenever $(x_0, \dots, x_{k-2}), (y_0, \dots, y_{k-2}) \in \Sigma$ and $x_j \in Vy_j$ for all $0 \leq j \leq k - 2$. Applying Lemma 12.7 and exploiting compactness again, we conclude that there exists another neighbourhood $V' \subseteq V$ of the identity 1 such that, given any $z_1, \dots, z_{k-2} \in G/\Gamma$, there exists a bounded function $Q_{z_1, \dots, z_{k-2}} : G/\Gamma \rightarrow \mathcal{D}$ such that

$$F(T_g^{n+(k-1)r}x) = Q_{z_1, \dots, z_{k-2}}(T_g^n x) + O(\varepsilon) \quad \text{whenever } T_g^{n+jr}x \in V'z_j \text{ for all } 1 \leq j \leq k-2.$$

It is not hard to ensure that the function $Q_{z_1, \dots, z_{k-2}}$ depends in a measurable manner on z_1, \dots, z_{k-2} . In particular we see that

$$\begin{aligned}
 & F(T_g^{n+(k-1)r}x)f(n+(k-1)r) \prod_{j=1}^{k-2} 1_{V'_{z_j}}(T_g^{n+jr}x) \\
 &= Q_{z_1, \dots, z_{k-2}}(T_g^n x)f(n+(k-1)r) \prod_{j=1}^{k-2} 1_{V'_{z_j}}(T_g^{n+jr}x) \\
 &+ O\left(\varepsilon F(T_g^{n+(k-1)r}x)f(n+(k-1)r) \prod_{j=1}^{k-2} 1_{V'_{z_j}}(T_g^{n+jr}x)\right). \tag{12.2}
 \end{aligned}$$

for all $n, r \in \mathbb{Z}$, $g \in G$ and $x \in G/\Gamma$. We are going to average this over n and r , but for technical reasons related to the difference between $\{n : |n| \leq N/2\}$ and $\mathbb{Z}/N\mathbb{Z}$ as additive objects, we shall restrict the range of r . To this end, take a function $\chi : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{R}^+$ such that $\mathbb{E}_r \chi(r) = 1$, $\text{supp}(\chi) \subseteq [-\varepsilon N, \varepsilon N]$ and for which we have the Fourier estimate

$$\sum_{m \in \mathbb{Z}/N\mathbb{Z}} |\hat{\chi}(m)| \leq 100/\varepsilon. \tag{12.3}$$

Such a function can easily be constructed, for example by convolving an interval with itself. Averaging (12.2) over $|n| \leq N/2 - \varepsilon kN$ and over r weighted by χ , one obtains

$$\begin{aligned}
 & \mathbb{E}_{|n| \leq N/2 - \varepsilon kN} \mathbb{E}_{r \in \mathbb{Z}/N\mathbb{Z}} \chi(r) F(T_g^{n+(k-1)r}x)f(n+(k-1)r) \prod_{j=1}^{k-2} 1_{V'_{z_j}}(T_g^{n+jr}x) \\
 &= \mathbb{E}_{|n| \leq N/2 - \varepsilon kN} \mathbb{E}_{r \in \mathbb{Z}/N\mathbb{Z}} \chi(r) h(n)f(n+(k-1)r) \prod_{j=1}^{k-2} h_j(n+jr) \\
 &+ O\left(\varepsilon \prod_{j=1}^{k-2} 1_{V'_{z_j}}(T_g^{n+jr}x)\right), \tag{12.4}
 \end{aligned}$$

where

$$h(n) := Q_{z_1, \dots, z_{k-2}}(T_g^n x) \quad \text{and} \quad h_j(n) := 1_{V'_{z_j}}(T_g^n x). \tag{12.5}$$

Note that as a consequence of the restrictions we have made on the support of n and of r , the expressions $n+jr$, $j = 0, 1, \dots, k-1$, are the same whether we regard the addition as taking place in \mathbb{Z} or in $\mathbb{Z}/N\mathbb{Z}$. In particular, (12.4) remains valid if one imagines that these additions are made in $\mathbb{Z}/N\mathbb{Z}$.

Now the second line of (12.4) can be written as

$$\mathbb{E}_{|n| \leq N/2} \mathbb{E}_{r \in \mathbb{Z}/N\mathbb{Z}} \chi(r) \tilde{h}(n)f(n+(k-1)r) \prod_{j=1}^{k-2} h_j(n+jr),$$

where

$$\tilde{h}(n) := (1 - 2\varepsilon k)^{-1} h(n) 1_{|n| \leq N/2 - \varepsilon k N},$$

and in particular $\|\tilde{h}\|_\infty \leq 2$. Writing $\chi(r)$ in terms of its Fourier transform on $\mathbb{Z}/N\mathbb{Z}$ and using Proposition 1.11 we can bound this expression above as follows, where $e_N(r) := e(r/N)$:

$$\begin{aligned} & \left| \mathbb{E}_{|n| \leq N/2} \mathbb{E}_{r \in \mathbb{Z}/N\mathbb{Z}} \chi(r) \tilde{h}(n) f(n + (k-1)r) \prod_{j=1}^{k-2} h_j(n + jr) \right| \\ &= \left| \mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{r \in \mathbb{Z}/N\mathbb{Z}} \sum_{m \in \mathbb{Z}/N\mathbb{Z}} \hat{\chi}(m) e_N(-mr) \tilde{h}(n) f(n + (k-1)r) \prod_{j=1}^{k-2} h_j(n + jr) \right| \\ &= \left| \mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{r \in \mathbb{Z}/N\mathbb{Z}} \sum_{m \in \mathbb{Z}/N\mathbb{Z}} \hat{\chi}(m) \tilde{h}(n) e_N(mn) h_1(n+r) e_N(-m(n+r)) \right. \\ & \qquad \qquad \qquad \left. \times \prod_{j=2}^{k-2} h_j(n + jr) \cdot f(n + (k-1)r) \right| \\ &\leq 2 \sum_{m \in \mathbb{Z}/N\mathbb{Z}} |\hat{\chi}(m)| \|f\|_{U^{k-1}(\mathbb{Z}/N\mathbb{Z})} \\ &\leq \frac{200}{\varepsilon} \|f\|_{U^{k-1}(\mathbb{Z}/N\mathbb{Z})}. \end{aligned}$$

Now we substitute this into (12.5) and average over z_1, \dots, z_{k-2} (picking these $k-2$ elements uniformly according to the Haar measure \mathbb{P} on G/Γ). This yields

$$\begin{aligned} & |\mathbb{E}_{|n| \leq N/2 - \varepsilon k N} \mathbb{E}_r \chi(r) F(T_g^{n+(k-1)r} x) f(n + (k-1)r)| \\ & \qquad \qquad \qquad \leq \frac{200}{\varepsilon \mathbb{P}(\pi(V'))^{k-2}} \|f\|_{U^{k-1}(\mathbb{Z}/N\mathbb{Z})} + O(\varepsilon). \end{aligned} \tag{12.6}$$

Write $G(n) := F(T_g^n x) f(n)$. Then $\|G\|_\infty \leq 1$, and so we have

$$\begin{aligned} & \mathbb{E}_{|n| \leq N/2 - \varepsilon k N} \mathbb{E}_r \chi(r) G(n + (k-1)r) \\ & \qquad \qquad \qquad = \mathbb{E}_r \chi(r) \mathbb{E}_{-N/2 + \varepsilon k N - (k-1)r \leq n' \leq N/2 - \varepsilon k N - (k-1)r} G(n') \\ & \qquad \qquad \qquad = \mathbb{E}_r \chi(r) (\mathbb{E}_{|n| \leq N/2} G(n') + O(\varepsilon)) \\ & \qquad \qquad \qquad = \mathbb{E}_{|n| \leq N/2} G(n) + O(\varepsilon). \end{aligned}$$

Comparing this with (12.6) gives

$$|\mathbb{E}_{|n| \leq N/2} f(n) F(T_g^n x)| \leq \frac{200}{\varepsilon \mathbb{P}(\pi(V'))^{k-2}} \|f\|_{U^{k-1}(\mathbb{Z}/N\mathbb{Z})} + O(\varepsilon),$$

which implies the result if $\varepsilon = c\eta$ for $c = c_k$ sufficiently small. □

Remark 12.8. An alternative way to obtain this lemma is to establish that the $(k-2)$ -step nilsequence $n \mapsto F(T_g^n x)$ can be approximated to high accuracy by a function

that is uniformly almost periodic of order $k - 2$ in the sense of [62]; this approach has the advantage of not requiring an explicit algebraic constraint such as that given in Lemma 12.7, but we do not pursue it here. This approach corresponds closely to the observation that a $(k - 2)$ -step nilflow can be constructed as a tower of $k - 2$ compact extensions of the trivial measure-preserving system; see [23, 43, 63] for further discussion.

Proposition 12.6 shows (essentially) that the basic $(k - 2)$ -step nilsequences form ‘obstructions to quadratic uniformity’, in the sense that functions $f : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathcal{D}$ which have a large inner product with such functions cannot have small $U^{k-1}(\mathbb{Z}/N\mathbb{Z})$ norm. The remarkable result of Host and Kra [43] asserts, roughly speaking, that these are in fact the *only* obstructions to having small U^{k-1} norm. More precisely, they work in the infinitary setting of arbitrary measure-preserving systems (as opposed to the shift on $\mathbb{Z}/N\mathbb{Z}$) and show that this system contains as an invariant factor an inverse limit of $(k - 2)$ -step nilflows, such that the $U^{k-1}(T)$ norm vanishes on the orthogonal complement of this inverse limit. In particular, this inverse limit is a characteristic factor for the $U^{k-1}(T)$ norm and for all quantities controlled by this norm, including the k -term recurrence expressions appearing, for instance, in the Furstenberg recurrence theorem; this fact is crucial in establishing the convergence of these recurrence expressions. We remark that the work of Ziegler [63, 64] achieves a very similar result, but avoids use of the $U^{k-1}(T)$ norm and obtains a characteristic factor (and convergence results) for the recurrence expressions directly. Also, the subsequent work of Bergelson, Host and Kra [5] gives a further discussion of the connection between the $U^{k-1}(T)$ norm and $(k - 2)$ -step nilsequences.

We now use Theorem 10.11 to obtain a finitary (and reasonably quantitative) version of the Host–Kra theorem in the case $k = 4$, with very explicit nilsequences; in fact, they will be none other than the elementary 2-step nilsequences defined earlier.

Theorem 12.9 (inverse theorem for $U^3(\mathbb{Z}/N\mathbb{Z})$, elementary nilsequence version). *Let $N > 2$ be a prime, let $0 < \eta \leq 1$ be sufficiently small and suppose that $f : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathcal{D}$ is a function with $\|f\|_{U^3(\mathbb{Z}/N\mathbb{Z})} \geq \eta$. Then there exists an elementary 2-step nilsystem G/Γ of dimension less than or equal to η^{-C} , an η^{-C} -Lipschitz function $F : G/\Gamma \rightarrow \mathcal{D}$ and elements $g \in G$, $x_0 \in G/\Gamma$ and $h \in \mathbb{Z}/N\mathbb{Z}$ such that*

$$|\mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}}(T^h f \overline{F_{N,g,x_0}})| \geq \exp(-\eta^{-C}).$$

Here we define $F_{N,g,x_0} : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{C}$ by

$$F_{N,g,x_0}(n) := F(T_g^n x_0) \quad \text{for all } -N/2 < n < N/2.$$

Remarks 12.10. The function F_{N,g,x_0} is just a 2-step nilsequence, adapted to $\mathbb{Z}/N\mathbb{Z}$. The analogous theorem for $U^2(\mathbb{Z}/N\mathbb{Z})$ is a trivial consequence of Proposition 2.2; the only linear nilfunction that needs to be considered is the function $F(x) := e(x)$ on the unit circle \mathbb{R}/\mathbb{Z} from Example 12.2, with $h = 0$ and $x_0 = 0$. A modification of Example 2.4 can be used to show that in formulating this theorem we must take into account Example 12.4; the other two fundamental 2-step nilsystems are in fact embedded inside this one and one could have dispensed with them altogether, but we have kept them for expository

purposes. One could also easily eliminate the role of x_0 (which is harmless anyway, since it ranges over a compact set) and of the shift h , but the parameter g ranges over a genuinely non-compact set and cannot be eliminated from this theorem (this can be seen even in the linear case; the frequency ξ in Proposition 2.2 is not restricted to a bounded set of values independently of N).

Proof. Applying Theorem 10.11 (and Lemma 8.2), we obtain a set $S \subseteq \mathbb{Z}/N\mathbb{Z}$ with $d := |S| \leq \eta^{-C}$, a regular Bohr set $B = B(S, \rho)$ with $\rho \geq \eta^C$, and a bracket quadratic

$$\phi(n) := \sum_{\xi, \xi' \in S} a_{\xi, \xi'} \{\xi \cdot n\} \{\xi' \cdot n\} + \sum_{\xi \in S} a_{\xi} \{\xi \cdot n\} \tag{12.7}$$

with $\text{Freq}(\phi) \subseteq S$ such that

$$|\mathbb{E}_{n \in B}(T^h f(n) e(-\phi(n)))| \geq \eta^C$$

and thus

$$|\mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}}(T^h f(n) e(-\phi(n)) 1_B(n))| \geq \eta^C \mathbb{E} 1_B. \tag{12.8}$$

Now let $\varepsilon := \eta^C/400d$ and let $\chi : \mathbb{R}/\mathbb{Z} \rightarrow [0, 1]$ be a continuous function such that $\chi(x) = 1$ when $|x| < \rho(1 - \varepsilon)$ and $\chi(x) = 0$ when $\frac{1}{2} \geq |x| > \rho(1 + \varepsilon)$. Consider the function*

$$\prod_{\xi \in S} \chi^{1+4d}(\xi \cdot n) - 1_B(n).$$

It is supported on $B(S, \rho(1 + \varepsilon)) \setminus B(S, \rho(1 - \varepsilon))$, which by the regularity of B has cardinality no more than $200d\varepsilon|B|$. Thus (12.8) implies that

$$\left| \mathbb{E}_n T^h f(n) e(-\phi(n)) \prod_{\xi \in S} \chi^{1+4d}(\xi \cdot n) \right| \geq (\eta^C - 200d\varepsilon) \mathbb{E} 1_B \geq \frac{1}{2} \eta^C \mathbb{E} 1_B \geq \exp(-\eta^{-C}),$$

the latter inequality being a consequence of Lemma 8.1.

Expanding ϕ as in (12.7), we see that our task is to show that the function

$$n \mapsto \left(\prod_{\xi \in S} \chi(\xi \cdot n) e(a_{\xi} \{\xi \cdot n\}) \right) \left(\prod_{\xi, \xi' \in S} \chi^2(\xi \cdot n) \chi^2(\xi' \cdot n) e(a_{\xi, \xi'} \{\xi \cdot n\} \{\xi' \cdot n\}) \right) \tag{12.9}$$

is an elementary 2-step nilsequence, for which it is enough to handle each of the functions in the product separately, as in the following lemma.

Lemma 12.11. *Each of the individual functions*

$$n \mapsto \chi(\xi \cdot n) e(a_{\xi} \{\xi \cdot n\}) \tag{12.10}$$

and

$$n \mapsto \chi^2(\xi \cdot n) \chi^2(\xi' \cdot n) e(a_{\xi, \xi'} \{\xi \cdot n\} \{\xi' \cdot n\}) \tag{12.11}$$

can be written as an elementary 2-step nilsequence with dimension no more than 9 and Lipschitz constant at most 50.

* The large power of χ here is so that we can distribute the cutoff χ among various factors later.

Proof. We begin by considering the functions (12.10), which are easier (corresponding to linear nilcharacters rather than quadratic ones). Split $a_\xi = q + s$, where q is an integer and $|s| \leq \frac{1}{2}$, and observe that $e(q\{\xi \cdot n\}) = e(q\xi n)$ if we identify $\mathbb{Z}/N\mathbb{Z}$ and $\widehat{\mathbb{Z}/N\mathbb{Z}}$ with the integers from $-N/2$ to $N/2$. Thus the function (12.10) takes the form

$$n \mapsto \chi(\xi n/N)e(s\{\xi n/N\})e(q\xi n/N).$$

This function may be identified as the elementary nilsequence $F_{N,(\xi/N,q\xi/N),0}$, where the underlying nilmanifold G/Γ is the direct sum of two copies of the unit circle shift (i.e. it is the torus $(\mathbb{R}/\mathbb{Z})^2$) and $F : (\mathbb{R}/\mathbb{Z})^2 \rightarrow \mathbb{C}$ is the function

$$F(x, y) := \chi(x)e(sx)e(y),$$

where we identify $x \in \mathbb{R}/\mathbb{Z}$ with a real number from $-\frac{1}{2}$ to $\frac{1}{2}$ in the usual manner. It is not hard to check that F is 50-Lipschitz.

Now consider the functions (12.11). We split $a_{\xi,\xi'}$ as $q + s$, much as before, so that (12.11) becomes

$$n \mapsto \chi(\alpha n)^2 \chi(\gamma n)^2 e(s\{\alpha n\}\{\gamma n\})e(q\{\alpha n\}\{\gamma n\}),$$

where $\gamma := \xi/N$ and $\gamma := \xi'/N$. Observe that since $\{\alpha n\} = \alpha n - [\alpha n]$ and $\{\gamma n\} = \gamma n - [\gamma n]$, we have the identity

$$q\{\alpha n\}\{\gamma n\} = q\alpha\gamma n^2 - q\alpha n[\gamma n] - q\gamma n[\alpha n] + q[\alpha n][\gamma n].$$

The last term is an integer, and hence

$$e(q\{\alpha n\}\{\gamma n\}) = e(q\alpha\gamma n^2)e(-q\alpha n[\gamma n])e(-q\gamma n[\alpha n]).$$

Thus it suffices to exhibit the three functions

$$\begin{aligned} n &\mapsto \chi(\alpha n)\chi(\gamma n)e(s\{\alpha n\}\{\gamma n\}), \\ n &\mapsto e(q\alpha\gamma n^2), \\ n &\mapsto \chi(\gamma n)e(-q\alpha n[\gamma n]), n \mapsto \chi(\alpha n)e(-q\gamma n[\alpha n]) \end{aligned}$$

as elementary 2-step nilsequences (note that the last two functions are essentially the same).

The first function can be obtained from a direct sum of two copies of the unit circle shift (Example 12.2) by repeating the analysis of (12.10), with F now defined by $F(x, y) := \chi(x)\chi(y)e(sxy)$ when $-\frac{1}{2} < x, y \leq \frac{1}{2}$.

The second function can easily be obtained from the skew shift (Example 12.3), by writing

$$q\alpha\gamma n^2 = -q\alpha\gamma n + 2q\alpha\gamma \frac{1}{2}(n(n+1))$$

and then taking $F(x, y) := e(y)$, $x_0 = (0, 0)$ and

$$g := \begin{pmatrix} 1 & 1 & -q\alpha\gamma \\ 0 & 1 & 2q\alpha\gamma \\ 0 & 0 & 1 \end{pmatrix}.$$

Finally, let us consider the third function. We write

$$e(-q\gamma n[\alpha n]) = e(\frac{1}{2}n(n+1)\alpha q\gamma - [n\alpha]nq\gamma)e(-\frac{1}{2}n(n+1)\alpha q\gamma).$$

The second factor can be generated using the skew shift as before.* We are thus left with

$$\chi(\alpha n)e(\frac{1}{2}(n(n+1))\alpha q\gamma - [n\alpha]nq\gamma).$$

But this can be generated from the Heisenberg shift (Example 12.4) with $x_0 = (0, 0, 0)$, $F(x, y, z) := \chi(x)e(y)$ on $[-\frac{1}{2}, \frac{1}{2}] \times (\mathbb{R}/\mathbb{Z})^2$ and

$$g := \begin{pmatrix} 1 & 0 & q\gamma \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix}.$$

It is easy to check that all of the functions F used in these constructions are 2π -Lipschitz, a bound which together with Lemma 12.5 completes the proof of the lemma. \square

Lemma 12.11, together with another application of Lemma 12.5, confirms that the function (12.9) is an elementary 2-step nilfunction with dimension at most $18d^2$ and Lipschitz constant no more than $100d^2$. This completes the proof of Theorem 12.9. \square

Remark 12.12. A pleasant reformulation of Theorem 12.9 may be obtained by considering the δ -atoms of G/Γ . Suppose that $F : G/\Gamma \rightarrow \mathcal{D}$ is K -Lipschitz. Let $(A_\omega)_{\omega \in \Omega}$ be the δ -atoms of G/Γ , pick arbitrary points $x_\omega \in A_\omega$ for each ω and write

$$\tilde{F}(n) := \sum_{\omega \in \Omega} F(x_\omega)1_{A_\omega}(n).$$

Since F is K -Lipschitz we clearly have the bound $\|F - \tilde{F}\| \leq K\delta$. Taking $\delta \leq \frac{1}{2}K^{-1} \exp(-\eta^{-C})$ we may replace the conclusion of Theorem 12.9 by

$$|\mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}}(T^h \overline{f \tilde{F}_{N,g,x_0}})| \geq \frac{1}{2} \exp(-\eta^{-C}).$$

Removing the sum over the atoms (of which there are at most $\delta^{-\eta^{-C}}$) by the pigeonhole principle, this implies that

$$|\mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}}(T^h \overline{f(1_{A_\omega})_{N,g,x_0}})| \geq \exp(-2\eta^{-2C}).$$

That is, if $\|f\|_{U^3}$ is large, then f correlates with the set of return times of a 2-step nilsequence to an atom.

* Indeed we could simply have rewritten our factorization of $e(q\{\alpha n\}\{\gamma n\})$ to incorporate these factors (and a linear phase correction), so as to then dispense with the second factor and the skew shift altogether. However, we have left this example in here to emphasize that purely quadratic phase functions such as $e(q\alpha\gamma n^2)$ are indeed examples of nilcharacters.

Remark 12.13. The space of quadratic nilsequences forms an algebra, being closed under multiplication, addition, subtraction and conjugation. This allows one to employ an ‘energy incrementation’ argument of the type used in [34, § 7] in order to decompose an arbitrary bounded function on $\mathbb{Z}/N\mathbb{Z}$ as the sum of a bounded function with U^3 norm smaller than some specified η , plus a 2-step nilsequence with dimension and Lipschitz constant controlled by functions of η . In the ergodic-theory setting an extremely similar decomposition was obtained in [5]. Informally speaking, the quadratic nilsequences form a *characteristic factor* for the U^3 norm, and hence for any expression controlled by that norm, and thus many questions involving such expressions can be reduced to questions concerning 2-step nilsequences.

It is perhaps of interest to briefly discuss a decomposition of this type for the U^2 norm, where the availability of harmonic analysis allows one to proceed more directly. If $f : \mathbb{Z}/N\mathbb{Z} \rightarrow \mathcal{D}$ is a function, then we write $R := \{r : |\hat{f}(r)| \geq \varepsilon_1\}$ for some suitable ε_1 , and define $\beta(x)$, $\mathbb{E}\beta = 1$, to be a normalized and suitably smoothed version of $1_B(x)$, where $B := B(R, \varepsilon_2)$. We then decompose

$$f(x) = f * \beta + (f - f * \beta). \quad (12.12)$$

It is easy to check that $\|f - f * \beta\|_{U^2}$ is small, but to write $f * \beta$ as a 1-step nilsequence it must be modified slightly. To do this, write

$$f * \beta(x) = \sum_{r \in \bar{R}} \hat{f}(r) \hat{\beta}(r) e(-rx/N) + \sum_{r \notin \bar{R}} \hat{f}(r) \hat{\beta}(r) e(-rx/N), \quad (12.13)$$

where $\bar{R} := \{m_1 r_1 + \dots + m_d r_d \mid r_j \in R, |m_j| \leq M\}$ for some M . Now if β is sufficiently smoothed and if M is sufficiently large, then

$$\sum_{r \notin \bar{R}} |\hat{\beta}(r)| \leq \varepsilon_3,$$

and so the second term in (12.13) is bounded by ε_3 , and in particular has small U^2 norm. The first term can be written as a 1-step nilfunction, the underlying nilmanifold being $(\mathbb{R}/\mathbb{Z})^d$ and the rotation T being $(x_1, \dots, x_d) \mapsto (x_1 + r_1/N, \dots, x_d + r_d/N)$. See [29] for an application of such a decomposition (the language of nilsystems and ergodic theory did not feature there, and the simpler decomposition (12.12) was used).

13. Future prospects and open questions

It is natural to ask whether there are inverse theorems for the higher $U^k(G)$ norms, $k \geq 4$, that generalize Theorems 2.7, 10.11 and 12.9. We are certain that there are. It is easy to guess at the correct generalization of Theorem 12.9, which should simply involve replacing 2-step nilmanifolds by $(k-2)$ -step ones. Guessing at the generalization of Theorem 2.7 is a bit harder. We suspect that the correct objects to consider for the $U^4(G)$ norm are of the form $1_Q(x)e(-\lambda(x))$, where $Q := \{x : \psi_1(x), \dots, \psi_d(x) \approx 1\}$ is now a *quadratic Bohr set*, each ψ_j being of the form $\psi_j(x) = 1_{B_j}(x)e(-\phi_j(x))$ appearing

in Theorem 2.7. The phase $\lambda : Q \rightarrow \mathbb{R}/\mathbb{Z}$ is now cubic. It is easy to guess how functions appropriate for the $U^k(G)$ norm may be constructed inductively.

We think it likely that most of the ingredients necessary to prove such inverse theorems may be found in [27], and we intend to pursue this direction. The same major difficulty that Gowers encountered in dealing with the $U^k(G)$ norm for $k \geq 4$ is also present here. Suppose that f has large $U^4(G)$ norm. This means that $T^h f \bar{f}$ has large $U^3(G)$ norm for many values of h . Applying Theorem 2.7 we obtain a Bohr set B_h for each of these h , such that $T^h \bar{h}$ has large quadratic bias on several shifts of this Bohr set B_h . The problem is that we do not, *a priori*, have any control on how the Bohr set B_h depends on h .

Another interesting issue is that of obtaining better bounds in Theorems 2.3 and 2.7. It is quite possible that the codimension of W in Theorem 2.3 can be taken to be $O(\log(1/\eta))$ rather than $O(\eta^{-C})$. This would give bounds of the form

$$\|f\|_{U^3(\mathbb{F}_5^n)} \leq \|f\|_{u^3(\mathbb{F}_5^n)} \leq \|f\|_{U^3(\mathbb{F}_5^n)}^c \tag{13.1}$$

for some absolute constant c . Such a bound would be a consequence of the *polynomial Freiman–Ruzsa conjecture* (PFR), which is discussed in detail in [31], together with some mild adjustments to the arguments of § 6. We refer to the statement (13.1) as the *polynomial Gowers inverse conjecture* (PGI) for \mathbb{F}_5^n .

It would be nice to have a version of (13.1) in a general G . What we mean by this is a statement of the form

$$\|f\|_{U^3(G)} \geq \eta \implies |\mathbb{E}f(x)1_B(x)e(-\phi(x))| \geq c(\eta),$$

$B = B(S, \rho)$, which may be reversed with only polynomial losses in the constants—that is to say,

$$|\mathbb{E}f(x)1_B(x)e(-\phi(x))| \geq c(\eta) \implies \|f\|_{U^3(G)} \geq \eta^C.$$

This would seem to require that we can take $|S| = O(\log(1/\eta))$ and ρ greater than some absolute constant.* The methods of this paper seem to fall a long way short of proving such a statement. Even if one had an appropriate analogue of PFR (which might take the form of a stronger version of Lemma 6.4 in which the size of S is logarithmic in δ), we would have to find a way to avoid repeatedly passing to smaller Bohr sets as in § 9. Each such passage causes too much degradation in ρ .

Let us conclude this section by remarking that working out how to drop the restriction $(|G|, 6) = 1$ in Theorem 2.7 would at least be a diverting exercise, though we cannot think of any applications. The case $G = \mathbb{F}_2^n$ probably captures the essence of the problem.†

Appendix A. Algebraic constraints on nilmanifolds

In this section‡ we prove Lemma 12.7. For some further discussion of issues related to such constraints, see [5, 20–22, 63].

* Even then one would need to replace 1_B with something smoother, to avoid the losses in the argument at the beginning of § 9.

† The authors have recently learnt that Samorodnitsky [59] has resolved this issue.

‡ The authors are indebted to Sasha Liebman and Tamar Ziegler for conversations which were very helpful in preparing this appendix.

By replacing $g^r x$ by x and g^n by g , respectively, our task is to demonstrate, given any $(k - 2)$ -step nilmanifold G/Γ , the existence of a compact set $\Sigma \subseteq (G/\Gamma)^{k-1}$ and a continuous map $P : \Sigma \rightarrow G/\Gamma$ such that

$$\left. \begin{aligned} &(x, T_g x, \dots, T_g^{k-2} x) \in \Sigma, \\ &P(x, T_g x, \dots, T_g^{k-2} x) = T_g^{k-1} x \quad \text{for all } g \in G, x \in G/\Gamma. \end{aligned} \right\} \tag{A 1}$$

Remark A 1. One can prove (A 1) by direct algebraic computation in the cases $k = 3, 4$. Indeed, when $k = 3$ the 1-step nilpotent group G is abelian, as is the subgroup Γ , so G/Γ is also a group (indeed it is a torus). One can then take $\Sigma = (G/\Gamma)^2$ and $P(a, b) := a(a^{-1}b)^2$. When $k = 4$, so that G is a 2-step nilpotent group, things are a little more complicated. One needs to take

$$\Sigma := \{(x_0\Gamma, x_1\Gamma, x_2\Gamma) : x_2 \in x_0(x_0^{-1}x_1)^2G_2\};$$

this reflects the fact that G/G_2 is a 1-step nilpotent group and thus obeys the $k = 3$ constraints. Note that G_2 commutes with all elements of G and is thus easy to quotient out. One can then define $P : \Sigma \rightarrow G/\Gamma$ by setting

$$P(a\Gamma, b\Gamma, c\Gamma) := a(a^{-1}b)^3((a^{-1}b)^{-2}a^{-1}c)^3\Gamma \quad \text{whenever } (a\Gamma, b\Gamma, c\Gamma) \in \Sigma;$$

one can verify with some effort (using of course the fact that all commutators lie in G_2 , which commute with all elements of G) that this function is well defined, continuous on Σ and obeys (A 1). Unfortunately, in the $k > 4$ case it seems that the function P is significantly messier, and in particular requires choosing a partial inverse for projection maps $G_j \mapsto G_j/\Gamma$ for all $j < k - 2$, which in general cannot be done canonically when $j \geq 2$.

To prove (A 1) in the general case, we first need some notation.

Definition A 2 (continuous right invertibility). Let M, N be compact spaces, let $\pi : M \rightarrow N$ be a continuous map and let $\Sigma \subseteq M$. We say that π is *continuously right invertible on Σ* if for every $w \in \pi(\Sigma)$ there exists a neighbourhood $V_w \subseteq N$ of w and a continuous map $\pi_w^{-1} : V_w \rightarrow M$ such that $\pi_w^{-1} \circ \pi$ is the identity on $\Sigma \cap \pi^{-1}(V_w)$.

Lemma A 3. Let G/Γ be a $(k - 2)$ -step nilmanifold and let $\pi : (x_0, \dots, x_k) \rightarrow (x_0, \dots, x_{k-1})$ be the canonical projection from $(G/\Gamma)^k$ to $(G/\Gamma)^{k-1}$. Then π is continuously right invertible on the set

$$A := \{(x, T_g x, \dots, T_g^{k-1} x, T_g^k x) : x \in G/\Gamma, g \in G\}.$$

The existence of a constraint (A 1) then follows by taking Σ to be the closure of $\pi(A)$ and using a compactness argument (exploiting the fact that $\pi(A)$ is dense in the compact set Σ) to glue the various local right inverses π_w^{-1} together.

Proof. For any $0 \leq i \leq k - 1$, we define the *Hall–Petresco groups* $\text{HP}_{k,i} \subseteq G^k$ to be the sets

$$\text{HP}_{k,i} := \left\{ \left(x_i^{\binom{n}{i}} \cdots x_{k-2}^{\binom{n}{k-2}} \right)_{0 \leq n \leq k-1} : x_i \in G_i, \dots, x_{k-2} \in G_{k-2} \right\},$$

where we have the conventions that $G_0 := G_1 = G$ and that $\binom{n}{j} = 0$ if $j > n$. Thus when $k = 4$, for instance, we have

$$\begin{aligned} \text{HP}_{4,3} &= \{(1, 1, 1, 1)\}, \\ \text{HP}_{4,2} &= \{(1, 1, x_2, x_2^3) : x_2 \in G_2\}, \\ \text{HP}_{4,1} &= \{(1, x_1, x_1^2 x_2, x_1^3 x_2^3) : x_1 \in G_1, x_2 \in G_2\}, \\ \text{HP}_{4,0} &= \{(x_0, x_0 x_1, x_0 x_1^2 x_2, x_0 x_1^3 x_2^3) : x_0 \in G_0, x_1 \in G_1, x_2 \in G_2\}. \end{aligned}$$

It is known (see [45]) that the $\text{HP}_{k,i}$ are all subgroups of G^k , and we also have the nesting $\text{HP}_{k,i+1} \subseteq \text{HP}_{k,i}$ for all $0 \leq i < k - 1$.

Observe that Γ^k is a subgroup of G^k , so we may form the quotient space G^k/Γ^k , which we identify with the compact manifold $(G/\Gamma)^k$. Inside this space we have the submanifolds $\text{HP}_{k,i}/\Gamma^k$ for all $0 \leq i < k - 1$. Observe that if $x \in G/\Gamma$ and $g \in G$, and $y \in x\Gamma \subseteq G$ is any representative of x in G , then

$$\begin{aligned} (x, T_g x, \dots, T_g^k x)\Gamma^k &= (y, gy, \dots, g^k y)\Gamma^k \\ &= (y, \dots, y)(1, g, \dots, g^k)\Gamma^k \subseteq \text{HP}_{k,0}\text{HP}_{k,0}\Gamma^k \\ &\subseteq \text{HP}_{k,0}\Gamma^k \end{aligned}$$

and hence

$$(x, T_g x, \dots, T_g^k x) \in \text{HP}_{k,0}/\Gamma^k.$$

It thus suffices to show that π is right invertible on $\text{HP}_{k,0}/\Gamma^k$.

We shall show inductively, by backwards induction on i , that π is continuously right invertible on $\text{HP}_{k,i}/\Gamma^k$ for all $0 \leq i < k - 1$. The case $i = k - 1$ is trivial since $\text{HP}_{k,i}/\Gamma^k$ is just a point. Now suppose inductively that $0 \leq i < k - 1$ and that π was already shown to be continuously right invertible over $\text{HP}_{k,i+1}/\Gamma^k$.

Let $(z_0, \dots, z_{k-1}) \in \overline{\pi(\text{HP}_{k,i}/\Gamma^k)}$. Observe that the first i coefficients of z must be the origin $O \in G/\Gamma$, defined as the image of the identity $1 \in G$, and the coefficient z_i lies in the closed manifold G_i/Γ .

The projection map $\pi_i : G_i \rightarrow G_i/\Gamma$ is continuous and surjective from the manifold G_i to the manifold G_i/Γ , which is a submanifold of G/Γ . Thus we may find a continuous function $f : V_{z_i} \rightarrow G_i$ defined on a neighbourhood $V_{z_i} \subseteq G/\Gamma$ of z_i such that $\pi_i \circ f$ is the identity on $V_{z_i} \cap (G_i/\Gamma)$.

We need to right invert π on $\text{HP}_{k,i}/\Gamma^k$ in a neighbourhood of $\pi(z)$. To this end, let $x := (x_1, \dots, x_k) \in \text{HP}_{k,i}/\Gamma^k$ be such that $\pi(x)$ is close to $\pi(z)$; in particular, we may take $x_i \in V_{z_i}$. As before we have $x_n = 0$ for $n < i$, and $x_n \in G_i/\Gamma$ for all $n \geq i$. Thus $x_i \in V_{z_i} \cap (G_i/\Gamma)$ and hence $\pi_i \circ f(x_i) = x_i$, or, in other words, $x_i = f(x_i)\Gamma$. Now let $F(x_i) \in \text{HP}_{k,i}$ be the group element

$$F(x_i) := \left(f(x_i)^{\binom{n}{i}} \right)_{0 \leq n \leq k-1},$$

and observe that this depends continuously on x_i , and hence on $\pi(x)$, if $\pi(x)$ lies in a neighbourhood of $\pi(z)$.

On the other hand, since $x \in \text{HP}_{k,i}/\Gamma^k$, there exists $g = (1, \dots, 1, g_i, \dots, g_k) \in \text{HP}_{k,i}$ such that $g\Gamma^k = x$; in particular, $g_i \in G_i$ and $g_i \in x_i\Gamma$. Since $f(x_i) \in G_i$ and $f(x_i) \in x_i\Gamma$, we conclude that $f(x_i)^{-1}g_i$ lies in both G_i and Γ . Thus if we let $\tilde{g} \in \text{HP}_{k,i}$ be the group element

$$\tilde{g} = \left((f(x_i)^{-1}g_i)^{\binom{n}{i}} \right)_{0 \leq n \leq k-1},$$

then \tilde{g} lies in both $\text{HP}_{k,i}$ and Γ^k . Thus we can factorize

$$g = F(x_i)h\tilde{g},$$

where h lies in $\text{HP}_{k,i}$, and also has i th component equal to the identity. Thus h in fact lies in $\text{HP}_{k,i+1}$. Multiplying on the right by Γ^k , we conclude that

$$x = g\Gamma^k = F(x_i)h\tilde{g}\Gamma^k = F(x_i)h\Gamma^k$$

and hence

$$F(x_i)x = h\Gamma^k \in \text{HP}_{k,i+1}/\Gamma^k.$$

Since $\pi(x)$ is close to $\pi(z)$ and $F(x_i)$ depends continuously on $\pi(x)$, we see that $\pi(F(x_i)x)$ is close to $\pi(F(z_i)z)$. In particular, by the induction hypothesis we can find a continuous map $\pi_{\pi(F(z_i)z)}^{-1}$ mapping a neighbourhood $V_{F(z_i)z} \subseteq (G/\Gamma)^k$ of $\pi(F(z_i)z)$ to $\text{HP}_{k,i+1}/\Gamma^k$ that is a local right inverse of π on $\text{HP}_{k,i+1}/\Gamma^k \cap \pi^{-1}(V_{F(z_i)z})$. Thus we have

$$F(x_i)x = h\Gamma^k = \pi_{F(z_i)z}^{-1}(\pi(F(x_i)x))$$

and hence

$$x = F(x_i)^{-1}\pi_{F(z_i)z}^{-1}(\pi(F(x_i)x)).$$

Observe that $\pi(F(x_i)x) = \tilde{\pi}(F(x_i))\pi(x)$, where $\tilde{\pi} : G^k \rightarrow G^{k-1}$ is the canonical projection. Since x_i of course depends continuously on $\pi(x)$, the right-hand side then depends continuously on $\pi(x)$ when $\pi(x)$ lies in a sufficiently small neighbourhood of $\pi(z)$. We have achieved a right inverse for π on $\text{HP}_{k,i}/\Gamma^k$ in a neighbourhood of $\pi(z)$, thus closing the induction. \square

Acknowledgements. T.T. is supported by a grant from the Packard Foundation.

References

1. A. BALOG AND E. SZEMERÉDI, A statistical theorem of set addition, *Combinatorica* **14** (1994), 263–268.
2. V. BERGELSON AND I. J. HÅLAND, Sets of recurrence and generalized polynomials, in *Convergence in ergodic theory and probability*, Ohio State University Mathematical Research Institute Publications, Volume 5, pp. 91–110 (de Gruyter, Berlin, 1996).
3. V. BERGELSON AND A. LEIBMAN, Polynomial extensions of van der Waerden’s and Szemerédi’s theorems, *J. Am. Math. Soc.* **9** (1996), 725–753.
4. V. BERGELSON AND A. LEIBMAN, Distribution of values of bounded generalized polynomials, *Acta Mathematica* **198** (2007), 155–230.
5. V. BERGELSON, B. HOST AND B. KRA, Multiple recurrence and nilsequences, *Invent. Math.* **160** (2005), 261–303.

6. Y. BILU, Structure of sets with small sumset, *Astérisque* **258** (1999), 77–108.
7. N. N. BOGOLYUBOV, Sur quelques propriétés arithmétiques des presque-périodes, *Ann. Chaire Math. Phys. Kiev* **4** (1939), 185–194.
8. B. BOLLOBÁS, *Linear analysis, an introductory course*, 2nd edn (Cambridge University Press, 1999).
9. E. BOMBIERI, On the large sieve, *Mathematika* **12** (1965), 201–225.
10. J. BOURGAIN, On triples in arithmetic progression, *Geom. Funct. Analysis* **9** (1999), 968–984.
11. J. W. S. CASSELS, *An introduction to the geometry of numbers* (Springer, 1959).
12. M. C. CHANG, A polynomial bound in Freiman’s theorem, *Duke Math. J.* **113** (2002), 399–419.
13. M. C. CHANG, On problems of Erdős and Rudin, *J. Funct. Analysis* **207** (2004), 444–460.
14. J.-P. CONZE AND E. LESIGNE, Sur un théorème ergodique pour des mesures diagonales, *C. R. Acad. Sci. Paris Sér. I* **306** (1988), 491–493.
15. H. T. CROFT, Some problems, *Eureka* **31** (1968), 18–19.
16. P. ERDŐS AND P. TURÁN, On some sequences of integers, *J. Lond. Math. Soc.* **11** (1936), 261–264.
17. P. FRANKL AND V. RÖDL, Extremal problems on set systems, *Random Struct. Algor.* **20** (2002), 131–164.
18. H. FURSTENBERG, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.* **31** (1977), 204–256.
19. H. FURSTENBERG, *Recurrence in ergodic theory and combinatorial number theory* (Princeton University Press, 1981).
20. H. FURSTENBERG, Non-conventional ergodic averages, in *The legacy of John von Neumann*, Proceedings of Symposia in Pure Mathematics, Volume 50, pp. 43–56 (American Mathematical Society, Providence RI, 1990).
21. H. FURSTENBERG, A polynomial Szemerédi theorem, in *Combinatorics, Paul Erdős is eighty, II*, Bolyai Society Mathematical Studies, Volume 2, pp. 253–269 (Janos Bolyai Mathematical Society, Budapest, 1996).
22. H. FURSTENBERG, From the Erdős–Turán conjecture to ergodic theory: the contribution of combinatorial number theory to dynamics, in *Paul Erdős and his mathematics, I*, Bolyai Society Mathematical Studies, Volume 11, pp. 261–277 (Janos Bolyai Mathematical Society, Budapest, 2002).
23. H. FURSTENBERG AND B. WEISS, A mean ergodic theorem for $N^{-1} \sum_{n=1}^N f(T^n x)g(T^{n^2} x)$, in *Convergence in ergodic theory and probability*, Ohio State University Mathematical Research Institute Publications, Volume 5, pp. 193–227 (de Gruyter, Berlin, 1996).
24. H. FURSTENBERG, Y. KATZNELSON AND D. ORNSTEIN, The ergodic theoretical proof of Szemerédi’s theorem, *Bull. Am. Math. Soc.* **7** (1982), 527–552.
25. W. T. GOWERS, A new proof of Szemerédi’s theorem for progressions of length four, *Geom. Funct. Analysis* **8** (1998), 529–551.
26. W. T. GOWERS, Fourier analysis and Szemerédi’s theorem, in *Proceedings of the International Congress of Mathematicians*, Volume I, *Document. Math.* (1998), Extra Volume I, pp. 617–629.
27. W. T. GOWERS, A new proof of Szemerédi’s theorem, *Geom. Funct. Analysis* **11** (2001), 465–588.
28. W. T. GOWERS, Hypergraph regularity and the multidimensional Szemerédi theorem, preprint.
29. B. J. GREEN, Roth’s theorem in the primes, *Ann. Math.* **161** (2005), 1609–1636.
30. B. J. GREEN, A Szemerédi-type regularity lemma in abelian groups, *Geom. Funct. Analysis* **15** (2005), 340–376.

31. B. J. GREEN, Finite field models in additive combinatorics, in *Surveys in Combinatorics 2005*, London Mathematical Society Lecture Note Series, Volume 327, pp. 1–27 (Cambridge University Press, 2005).
32. B. J. GREEN AND S. KONYAGIN, On the Littlewood problem modulo a prime, *Can. J. Math.*, in press.
33. B. J. GREEN AND I. Z. RUZSA, Freiman’s theorem in an arbitrary abelian group, *J. Lond. Math. Soc.* **75** (2007), 163–175.
34. B. J. GREEN AND T. C. TAO, The primes contain arbitrarily long arithmetic progressions, *Ann. Math.*, in press.
35. B. J. GREEN AND T. C. TAO, New bounds for Szemerédi’s theorem, I, Progressions of length 4 in finite field geometries, preprint.
36. B. J. GREEN AND T. C. TAO, Quadratic uniformity of the Mobius function, *Annales Inst. Fourier*, in press.
37. B. J. GREEN AND T. C. TAO, Linear equations in primes, *Annals Math.*, in press.
38. I. J. HÅLAND, Uniform distribution of generalized polynomials, *J. Num. Theory* **45** (1993), 327–366.
39. I. J. HÅLAND, Uniform distribution of generalized polynomials of the product type, *Acta Arith.* **67** (1994), 13–27.
40. I. J. HÅLAND AND D. E. KNUTH, Polynomials involving the floor function, *Math. Scand.* **76** (1995), 194–200.
41. B. HOST AND B. KRA, Convergence of Conze–Lesigne averages, *Ergod. Theory Dynam. Syst.* **21** (2001), 493–509.
42. B. HOST AND B. KRA, Averaging along cubes, in *Modern dynamical systems and applications* (ed. M. Brin, B. Hasselblatt and Y. Pesin) (Cambridge University Press, 2004).
43. B. HOST AND B. KRA, Non-conventional ergodic averages and nilmanifolds, *Ann. Math.* **161** (2005), 397–488.
44. F. JOHN, Extremum problems with inequalities as subsidiary conditions, in *Studies and essays presented to R. Courant on his 60th birthday, January 8, 1948*, pp. 187–204 (Interscience, New York, 1948).
45. A. LEIBMAN, Polynomial sequences in groups, *J. Alg.* **201** (1998), 189–206.
46. R. MESHULAM, On subsets of finite abelian groups with no 3-term arithmetic progressions, *J. Combin. Theory A* **71** (1995), 168–172.
47. H. L. MONTGOMERY, *Topics in multiplicative number theory*, Lecture Notes in Mathematics, Volume 227 (Springer, 1971).
48. B. NAGLE AND V. RÖDL, Regularity properties for triple systems, *Random Struct. Algor.* **23** (2003), 264–332.
49. B. NAGLE, V. RÖDL AND M. SCHACHT, The counting lemma for regular k -uniform hypergraphs, *Random Struct. Algor.* **28** (2006), 113–179.
50. B. NAGLE, V. RÖDL AND M. SCHACHT, Note on the 3-graph counting lemma, *Discr. Math.*, in press.
51. G. PISIER, *The volume of convex bodies and Banach space geometry*, Cambridge Tracts in Mathematics, Volume 94 (Cambridge University Press, 1989).
52. H. PLÜNNECKE, *Eigenschaften und Abschätzungen von Wirkingsfunktionen*, BMwF-GMD-22 (Gesellschaft für Mathematik und Datenverarbeitung, Bonn 1969).
53. V. RÖDL AND J. SKOKAN, Regularity lemma for k -uniform hypergraphs, *Random Struct. Algor.* **25** (2004), 1–42.
54. V. RÖDL AND J. SKOKAN, Counting subgraphs in quasi-random 4-uniform hypergraphs, *Random Struct. Algor.* **26** (2005), 160–203.
55. V. RÖDL AND J. SKOKAN, Applications of the regularity lemma for uniform hypergraphs, *Random Struct. Algor.* **28** (2006), 180–194.
56. K. F. ROTH, On certain sets of integers, *J. Lond. Math. Soc.* **28** (1953), 245–252.

57. I. Z. RUZSA, An application of graph theory to additive number theory, *Scientia A* **3** (1989), 97–109.
58. I. Z. RUZSA, An analog of Freiman’s theorem in groups: structure theory of set addition, *Astérisque* **258** (1999), 323–326.
59. A. SAMORODNITSKY, Low degree tests at large distances, *Proc. 39th Annual ACM Symp. on Theory of Computing*, pp. 506–515 (ACM Press New York, 2007).
60. E. SZEMERÉDI, On sets of integers containing no four elements in arithmetic progression, *Acta Math. Hungar.* **20** (1969), 89–104.
61. T. C. TAO, Lecture notes 5 from Math 254A (available at www.math.ucla.edu/~tao/254a.1.03w/notes5.dvi).
62. T. C. TAO, A quantitative ergodic theory proof of Szemerédi’s theorem, *Electron. J. Combinat.* **13** (2006), 1–49.
63. T. ZIEGLER, A non-conventional ergodic theorem for a nilsystem, *Ergod. Theory Dynam. Syst.* **25** (2005), 1357–1370.
64. T. ZIEGLER, Universal characteristic factors and Furstenberg averages, *J. Am. Math. Soc.* **20** (2007), 53–97.