CHAPTER 5

Longer progressions and higher Gowers norms

5.1. Introducing the higher Gowers norms

Roth's theorem, proved in Chapter **??**, showed that subsets of $\{1, \dots, N\}$ with positive density contain nontrivial 3-term APs. The same statement is true for progressions of any length, a famous theorem of Szemerédi. Roth in fact establish a quantitative version of this result, as we saw in the first chapter. The first decent quantitative bounds for Szemerédi's theorem were obtained by Gowers in 1998.

Theorem 5.1 (Gowers). *Let $k \geqslant 3$ be an integer. Suppose that $A \subseteq \{1, \dots, N\}$ is a set with cardinality at least $N(\log\log N)^{-c_k}$. Then A contains a nontrivial k-term arithmetic progression.*

This theorem is *much* harder when $k \geqslant 4$ than it was for $k = 3$. Our aim in this chapter is to give the proof of the case $k = 4$.

As a consequence of the way we proved Roth's theorem, much of this consists of routine generalization. Recall that we were working in a group $G = \mathbb{Z}/N'\mathbb{Z}$, where $N' = 2N + 1$. To deal with k-term progressions one may instead take $N' = (k-1)!N + 1$ (say). We introduced the Gowers U^2-norm on G, defined by

$$\|f\|_{U^2} := \big(\mathbb{E}_{x,h_1,h_2 \in G} f(x)\overline{f(x+h_1)f(x+h_2)}f(x+h_1+h_2)\big)^{1/4}.$$

This is a kind of average of f over parallelepipeds; the definition makes perfect sense in any abelian group G. The Gowers U^k-norm is a very similar average over k-dimensional parallelepipeds. In order to keep track of complex conjugates in the definition we introduce the notation $\mathcal{C}f := \overline{f}$.

Definition 5.1 (Gowers U^k-norm). Suppose that $f : G \to \mathbb{C}$ is a function. Let $k \geqslant 2$ be an integer. Then we define

$$\|f\|_{U^k} := \big(\mathbb{E}_{x,h_1,\dots,h_k \in G} \prod_{\omega \in \{0,1\}^k} \mathcal{C}^{|\omega|} f(x + \omega \cdot \mathbf{h})\big)^{1/2^k}.$$

Here $\omega = (\omega_1, \dots, \omega_k)$ ranges over the cube $\{0,1\}^k$, $\omega \cdot \mathbf{h}$ means $\omega_1 h_1 + \dots + \omega_k h_k$, and $|\omega|$ denotes $|\omega_1| + \dots + |\omega_k|$.

In the first chapter we scarcely developed the theory of Gowers norms, relying on the "accident" that $\|f\|_{U^2} = \|\widehat{f}\|_4$. There seems to be no such happy accident for the higher norms.

It is convenient to introduce, in addition to the Gowers norm, the *Gowers inner product* $\langle f_\omega \rangle_{U^k}$. This is defined not for one function but for a family $(f_\omega)_{\omega \in \{0,1\}^k}$, and the definition consists in nothing more than replacing f by f_ω in Definition5.1 and omitting to take $1/2^k$-th roots. For example,

$$\langle f_{00}, f_{10}, f_{01}, f_{11} \rangle_{U^2} := \mathbb{E}_{x,h_1,h_2 \in G} f_{00}(x) \overline{f_{10}(x+h_1) f_{01}(x+h_2)} f_{11}(x+h_1+h_2),$$

and

$$\langle f, f, \dots, f \rangle_{U^k} = \|f\|_{U^k}^{2^k}.$$

PROPOSITION 5.1 (Basic properties). *Let $k \geqslant 2$ be an integer, and let $(f_\omega)_{\omega \in \{0,1\}^k}$ be a family of functions mapping G to $\mathbb{C}$. Then*

(i) *We have the Gowers-Cauchy-Schwarz inequality, stating that*

$$\langle f_\omega \rangle_{U^k} \leqslant \prod_{\omega in \{0,1\}^k} \|f_\omega\|_{U^k};$$

(ii) *The Gowers norms are nested, that is to say*

$$\|f\|_{U^2} \leqslant \|f\|_{U^3} \leqslant \dots$$

(iii) *The Gowers norms really are norms, that is to say they take non-negative values, vanish only when $f = 0$, and satisfy the triangle inequality $\|f + g\|_{U^k} \leqslant \|f\|_{U^k} + \|g\|_{U^k}$.*

Part (i) is a repeated application of the Cauchy-Schwarz inequality. Perhaps the cleanest way to see this is to reparametrise the Gowers inner product, writing for example

$$\langle f_{00}, f_{10}, f_{01}, f_{11} \rangle_{U^2} = \mathbb{E}_{x,x',y,y'} f_{00}(x+y) \overline{f_{10}(x'+y) f_{01}(x+y')} f_{11}(x'+y').$$

By Cauchy-Schwarz "in the x's" this is at most

$$\left(\mathbb{E}_{x,x'} \big(\mathbb{E}_y f_{00}(x+y) \overline{f_{10}(x'+y)} \big)^2 \right)^{1/2} \left(\mathbb{E}_{x,x'} \big(\mathbb{E}_{y'} \overline{f_{01}(x+y')} f_{11}(x'+y') \big)^2 \right)^{1/2},$$

which is equal to $\langle f_{00}, f_{10}, f_{00}, f_{10} \rangle_{U^2}^{1/2} \langle f_{01}, f_{11}, f_{01}, f_{11} \rangle_{U^2}^{1/2}$. Now one may do a further Cauchy-Schwarz "in the y's" on each of the factors on the right hand side to bound this above by $\|f_{00}\|_{U^2} \|f_{10}\|_{U^2} \|f_{01}\|_{U^2} \|f_{11}\|_{U^2}$, as claimed.

The proof for $k > 2$ is identical, except that the notation is a lot worse and k applications of the Cauchy-Schwarz inequality are required rather than two.

Part (ii) is an immediate consequence of part (i), since

$$\|f\|_{U^{k-1}}^{2^{k-1}} = \langle f, f, \dots, f, 1, 1, \dots, 1 \rangle_{U^k} \leqslant \|f\|_{U^k}^{2^{k-1}} \|1\|_{U^k}^{2^{k-1}},$$

and of course $\|1\|_{U^k} = 1$.

Finally we turn to (iii). The nonnegativity follows immediately from (ii). To prove the triangle inequality, we expand $\|f+g\|_{U^k}^{2^k}$ as $\langle f+g, f+g, \dots, f+g \rangle_{U^k}$. Since the Gowers inner product is patently multilinear, this splits as a sum of 2^{2^k} terms, each of something like the form $\langle f, g, f, f, g, \dots \rangle_{U^k}$. The number of such terms with i

copies of f and $2^k - i$ copies of g is, of course, $\binom{2^k}{i}$. By the Gowers-Cauchy-Schwarz inequality each such term is bounded by $\|f\|_{U^k}^i \|g\|_{U^k}^{2^k - i}$, and so

$$\|f + g\|_{U^k}^{2^k} \leqslant \sum_{i=0}^{2^k} \binom{2^k}{i} \|f\|_{U^k}^i \|g\|_{U^k}^{2^k - i} = (\|f\|_{U^k} + \|g\|_{U^k})^{2^k}.$$

Taking 2^kth roots completes the proof. □

5.2. Progressions of length 4 – first steps

Let us turn our attention to the case $k = 4$ of Gowers' version of Szemerédi's theorem. Everything we say in this section generalises rather easily to arbitrary $k \geqslant 3$, but the same is not true of subsequent sections and so we restrict to the special case here too, for notational simplicity. The general strategy of the proof will be exactly the same as the proof of Roth's theorem as presented in the first chapter, and the reader is encouraged to look at that again now.

Here, as there, the proof is based on an iterative procedure. The density increment step for progressions of length four is the following slight variant of Proposition 1.1.

PROPOSITION 5.2. *Suppose that* $0 < \alpha < \frac{9}{10}$ *and that* $N > \alpha^{-C}$. *Suppose that* $P \subseteq \mathbb{Z}$ *is an arithmetic progression of length* N *and that* $A \subseteq P$ *is a set with cardinality at least* αN. *Then at least one of the following two alternatives holds:*

(i) *A contains at least $\frac{1}{20}\alpha^4 N^2$ nontrivial three-term progressions and in particular at least one;*
(ii) *There is an arithmetic progression P' of length at least $\exp(-1/\alpha^C)N^{\alpha^C}$ such that, writing $A' := A \cap P'$ and $\alpha' := |A'|/|P'|$, we have $\alpha' > \alpha + \alpha^C$.*

Remark. The assumption that $\alpha < \frac{9}{10}$ is just for notational convenience, since then any expression of the form C/α^C is bounded by $\alpha^{C'}$, for some other absolute constant C'. Note that Gowers' theorem is trivial for sets of density at least $9/10$.

The size of the progression P' is rather smaller than it was in Proposition 1.1, where we could take $|P'| \geqslant N^{1/3}$. This makes remarkably little difference to the final bound, and we leave it to the reader to check that Proposition 5.2 implies Gowers' result for progressions of length 4 by mimicking the argument given in Section 1.1.

It remains, of course, to prove Proposition 5.2. By rescaling we may assume that $P = \{1, \ldots, N\}$. Let N' be a prime between $6N$ and $12N$ and set $G = \mathbb{Z}/N'\mathbb{Z}$. (such a prime exists by Bertrand's postulate, and there is one small point later, in the proof of Proposition 5.4, where the primeness of N will be convenient. We will use the same notation as in the first chapter. In particular if $A \subseteq \{1, \ldots, N\}$ is a set of size αN then we regard A as a subset of G in the obvious way, and consider also the balanced function $f : G \to \mathbb{C}$ defined by $f := 1_A - \alpha 1_{[N]}$.

If $f_1, f_2, f_3, f_4 : G \to \mathbb{C}$ are functions we define

$$\mathrm{AP}_4(f_1, f_2, f_3, f_4) = \mathbb{E}_{x,d \in G} f_1(x) f_2(x + d) f_3(x + 2d) f_4(x + 3d).$$

It is a very easy matter indeed to establish the following analogue of Lemma 1.1.

Lemma 5.1. *Let $0 < \alpha < \frac{9}{10}$ be a parameter. Suppose that $N > \alpha^{-C}$ and that the set A contains fewer than $\frac{1}{20}\alpha^4 N^2$ nontrivial 4-term progressions. Then there are 1-bounded functions g_1, g_2, g_3, g_4, at least one of which is equal to the balanced function f, such that $|\mathrm{AP}_4(g_1, g_2, g_3, g_4)| \geqslant c\alpha^4$.*

The next lemma is, quite obviously, analogous to Lemma 1.3.

Lemma 5.2 (Generalised von Neumann theorem). *Suppose that $f_1, f_2, f_3, f_4 : G \to \mathbb{C}$ are 1-bounded functions. Then*

$$|\mathrm{AP}_4(f_1, f_2, f_3, f_4)| \leqslant \|f_i\|_{U^3}$$

for $i = 1, 2, 3, 4$.

Proof. The proof is very closely analogous to that of Lemma 1.3, and begins with a clever rewriting of the left-hand side. Once again we indicate the proof for $i = 1$, the proof of the other cases being similar. The rewriting we choose is

$$\mathrm{AP}_4(f_1, f_2, f_3, f_4) = \mathbb{E}_{x,y,z} f_1(x + y + z) f_2(\frac{y}{2} + \frac{2z}{3}) f_3(-x + \frac{z}{3}) f_4(-2x - \frac{y}{2}).$$

There is actually nothing particularly clever about this: we have simply parametrised in such a way that only f_1 involves all three of the variables x, y and z. One now applies Cauchy-Schwarz three times, eliminating f_2, f_3 and f_4 in turn. When this is done we have

$$|\mathrm{AP}_4(f_1, f_2, f_3, f_4)|^8 \leqslant \mathbb{E}_{x,x',y,y',z,z' \in G} f_1(x+y+z)\overline{f_1(x' + y + z)} \dots \overline{f_1(x' + y' + z')},$$

and the right hand side is just a reparametrisation of $\|f_1\|_{U^3}^8$. □

Combining this with Lemma 5.1 leads quickly to the following analogue of Corollary 1.1.

Corollary 5.1 (Gowers norm dichotomy). *Let α, $0 < \alpha < \frac{9}{10}$, be a real number. Suppose that $N > \alpha^{-C}$ and that A is a subset of $\{1, \dots, N\}$ with cardinality αN and fewer than $\frac{1}{20}\alpha^4 N^2$ nontrivial 4-term progressions. Let $f : G \to \mathbb{R}$ be the balanced function of A. Then $\|f\|_{U^3} \geqslant c\alpha^4$.*

Putting all this together, we see that the task of proving Proposition 5.2 reduces to the following result.

Proposition 5.3 (Density increment for the U^3-norm). *Suppose that $f : \{1, \dots, N\} \to \mathbb{R}$ is a 1-bounded function with $\sum_x f(x) = 0$ and that $\|f\|_{U^3} \geqslant \delta$, where $\delta < \frac{1}{2}$. Then there is a progression $P \subseteq \{1, \dots, N\}$ with $|P| \geqslant N^{\delta^C}$ such that $\mathbb{E}_{x \in P} f(x) \geqslant \delta^C$.*

We shall deduce this result from a "local inverse theorem for the U^3-norm" and some tools from diophantine approximation. The (slightly complicated) statement of this result is given in Proposition 5.5 below. We note that Gowers obtained corresponding results for the U^k-norm, $k \geqslant 4$, as well: this is a remarkable achievement and has yet to be supplanted by corresponding "global" results. There *is* a global inverse theorem for the U^3-norm and we shall give its statement later on.

5.3. Gowers' local inverse theorem for the U^3-norm, I

We now turn to the rather long process of proving Proposition 5.3. To motivate the proof let us mention an example of a complex-valued 1-bounded function $F : G \to \mathbb{C}$ with *very* large U^3-norm, namely $f(x) = e(x^2/N')$. For this function we have $\|f\|_{U^3} = 1$, because

$$x^2 - (x+h_1)^2 - (x+h_2)^2 - (x+h_3)^2 + (x+h_1+h_2)^2 + (x+h_1+h_3)^2 + (x+h_2+h_3)^2 - (x+h_1+h_2+h_3)^2 = 0$$

(the familiar principle that third differences of the sequence of squares vanish).

With the benefit of hindsight this example is extremely natural: for functions f of the form $f(x) = e(\psi(x))$ the U^k-norm takes the phase ψ and differences or "differentiates it" k times. Thus if the U^k-norm of a function f is large then this detects some kind of bias in the kth derivative of f, and to find examples of such functions it is certainly natural to look at polynomials of degree $k-1$ or functions which are biased with respect to them.

In reality the situation is more complicated, as we shall explain later. For now let us return to the U^3-norm and have in mind the example $f(x) = e(x^2/N')$ just discussed. Suppose then that $f : G \to \mathbb{R}$ has $\|f\|_{U^3} \geqslant \delta$. We introduce the *difference functions* $\Delta(f,h)(x) := f(x)\overline{f(x+h)}$, and observe that

$$\|f\|_{U^3}^8 = \mathbb{E}_{h \in G} \|\Delta(f,h)\|_{U^2}^4.$$

Indeed both sides are counting the number of 3-dimensional parallelepipeds weighted by f, the left-hand side directly and the right hand side by first fixing one of the directions h and then looking at the possible 2-dimensional parallelograms that may complete the parallelepiped.

Now since f is 1-bounded we obviously have $\|\Delta(f,h)\|_{U^2} \leqslant 1$ for all h, and so it follows from the above observation and the assumption $\|f\|_{U^3} \geqslant \delta$ that there are at least $\delta^8|G|/2$ values of $h \in G$ for which $\|\Delta(f,h)\|_{U^2} \geqslant \delta^2/2^{1/4}$.

Write S for the set of such h. For each $h \in S$ we may apply the inverse theorem for the U^2-norm, Theorem 1.2, to conclude that there is some $\phi(h) \in G$ for which

$$|\Delta(f;h)^\wedge(\phi(h))| \geqslant \delta^4/2.$$

Recall once again the example $f(x) = e(x^2/N')$. For this function $\Delta(f,h)(x) = f(x)\overline{f(x+h)} = e(-2hx/N')e(-h^2/N')$, a function which manifestly has a large Fourier coefficient at $\phi(h) = -2h$. Thus in this example the function ϕ (which is unique in this example, though in general it need not be) is very far from arbitrary: it is in fact *linear*.

Gowers' argument hinges on the rather miraculous observation that this linearity is not only a feature of *this* example.

LEMMA 5.3 (Additive structure of large Fourier coefficients). *Suppose that $S \subseteq G$ is a set of size $\sigma|G|$, and that $f : G \to \mathbb{C}$ is a 1-bounded function with the property that $|\Delta(f,h)^\wedge(\phi(h))| \geqslant \eta$ for al $h \in S$ and for some function $\phi : S \to G$. Write $\Gamma = \{(h, \phi(h)) : h \in S\}$ for the graph of ϕ considered as a subset of $G \times G$. Then the additive energy $\omega_+(\Gamma)$ is at least $\sigma\eta^8$.*

Remark. The conclusion is a slightly fancy way of saying that there are many solutions to $\phi(h_1) + \phi(h_2) = \phi(h_3) + \phi(h_4)$ with $h_1 + h_2 = h_3 + h_4$, a condition which makes it rather clearer that some kind of approximate linearity is being encoded.

Proof. The statement certainly implies that

$$\mathbb{E}_h 1_S(h)|\Delta(f,h)^\wedge(\phi(h))|^2 \geqslant \sigma\eta^2.$$

Expanding out and making an obvious substitution, this implies that

$$\mathbb{E}_{x,h,k} 1_S(h) f(x)\overline{f(x+h)f(x+k)} f(x+h+k) e(\phi(h)k/N') \geqslant \sigma\eta^2.$$

Using the inequality $|f(x)\overline{f(x+k)}| \leqslant 1$, we get

$$\mathbb{E}_{x,k}|\mathbb{E}_h 1_S(h)\overline{f(x+h)} f(x+h+k) e(\phi(h)k/N')| \geqslant \sigma\eta^2$$

and hence, by Cauchy-Schwarz, that

$$\mathbb{E}_{x,k}|\mathbb{E}_h 1_S(h)\overline{f(x+h)} f(x+h+k) e(\phi(h)k/N')|^2 \geqslant \sigma^2\eta^4.$$

Writing

$$F_k(t) := 1_S(t) e(\phi(t)k/N') \quad \text{and} \quad G_k(t) := \Delta(f,k)(t),$$

this last inequality may be rewritten as

$$\mathbb{E}_k \|F_k^\circ * G_k\|_2^2 \geqslant \sigma^2\eta^4,$$

where $F_k^\circ(t) = F_k(-t)$. By Parseval's identity this may be written in terms of Fourier coefficients as

$$\mathbb{E}_k \sum_r |\widehat{F_k}(-r)|^2 |\widehat{G_k}(r)|^2 \geqslant \sigma^2\eta^4.$$

Now since G_k is bounded by 1 so is $G_k * G_k$, and hence from Parseval's identity we have $\sum_r |\widehat{G_k}(r)|^4 \leqslant 1$. Applying Hölder's inequality to the preceding line, we thus obtain

$$\mathbb{E}_k \big(\sum_r |\widehat{F_k}(r)|^4\big)^{1/2} \geqslant \sigma^2\eta^4.$$

By one further application of Cauchy-Schwarz this implies that

$$\mathbb{E}_k \sum_r |\widehat{F_k}(r)|^4 \geqslant \sigma^4\eta^8.$$

Substituting the definition of F_k this means that

$$\mathbb{E}_k \sum_r |\mathbb{E}_t 1_S(t) e((\phi(t)k - rt)/N')|^4 \geqslant \sigma^4\eta^8,$$

that is to say

$$\mathbb{E}_{t_1,t_2,t_3,t_4} 1_S(t_1)1_S(t_2)1_S(t_3)1_S(t_4)\mathbb{E}_k e\Big(\frac{k}{N'}(\phi(t_1)+\phi(t_2)-\phi(t_3)-\phi(t_4))\Big) \times \sum_r e\Big(\frac{-r}{N'}(t_1+t_2-t_3-t_4)\Big).$$

But this is precisely $1/N'^3$ times the number of quadruples $t_1, t_2, t_3, t_4 \in S$ with $t_1 + t_2 = t_3 + t_4$ and $\phi(t_1)+\phi(t_2) = \phi(t_3)+\phi(t_4)$. The additive energy of the graph Γ is equal to the number of such quadruples divided by $|S|^3$, and so we obtain the stated bound. □

If the function $\phi : G \to G$ were actually linear then the additive energy $\omega_+(\Gamma)$ would equal 1. The conclusion of the last lemma is a weak version of this statement. It is closely analogous to the weak version of "being a group" that is encoded by a set A having large additive energy $\omega_+(A)$. We now know how to deal with such a situation by applying Balog-Szemerédi-Gowers followed by the Freĭman-Ruzsa theorem.

This is precisely what we shall do, but we need to prepare in two ways. Firstly, the Freĭman-Ruzsa theorem has been stated for subsets of $\mathbb{Z}$ and not for subsets of $G^2 = (\mathbb{Z}/N'\mathbb{Z})^2$. It turns out that we can easily adapt the theorem to cover this case. Secondly, it turns out to be rather wasteful to apply the full strength of the theorem: ducking out at an intermediate point, prior to the use of Chang's covering lemma, is more sensible.

Proposition 5.4. *Let $S \subseteq G$ and suppose that $\phi : S \to G$ is a function. Consider the graph $\Gamma \subseteq G^2$ and suppose that the doubling $\sigma[\Gamma]$ is at most K, where $K \geqslant 2$. Suppose also that $|S| > \exp(K^C)$, for a suitably large absolute constant C. Then there is an arithmetic progression $Q \subseteq G$ of cardinality at least N^{1/K^C} together with a linear function $\psi(x) = ax + b$ such that $\phi(x) = \psi(x)$ for at least a proportion K^{-C} of the elements $x \in Q$.*

Proof. Consider first a set $A \subseteq \mathbb{Z}$ with $\sigma[A] \leqslant K$. By following the proof of the Freĭman-Ruzsa theorem but ignoring the covering argument at the end we obtain a proper GAP, P, with dimension at most K^C and size at least $\exp(-K^C)|A|$, contained in $2A - 2A$. Now we have $\sum_x |A \cap (P + x)| = |A||P|$, but the support of this sum is contained in $A - P \subseteq 3A - 2A$, a set which has size at most $K^C|A|$ by Ruzsa calculus. It follows that there is some x such that $|A \cap (P + x)| \geqslant K^{-C}|P|$.

Suppose that $P + x = \{x_0 + l_1x_1 + \cdots + l_dx_d : 0 \leqslant l_i < L_i, i = 1, \ldots, d\}$, where L_d is, without loss of generality, at least as large as the other L_i. By foliating P into (one-dimensional) arithmetic progressions of the form $\{x_0 + l_1x_1 + \cdots + l_dx_d : 0 \leqslant l_d < L_d\}$, we conclude that there is some arithmetic progression Q of length at least $\exp(-K^C)|A|^{1/K^C}$ such that $|A \cap Q| \geqslant K^{-C}|Q|$.

Our next task is to become convinced that the same statement is true for subsets of $\mathbb{Z}^2$. The definition of a GAP in $\mathbb{Z}^2$ makes perfect sense (simply take the "base elements" $x_0, \ldots, x_d$ to be elements of $\mathbb{Z}^2$), and after the point at which a large GAP was located inside $2A-2A$ we did not use any specific property of $\mathbb{Z}$ other than that it is an abelian group. Thus it suffices to show that if $A \subseteq \mathbb{Z}^2$ has small doubling then $2A-2A$ contains a large GAP. This follows from the corresponding statement over $\mathbb{Z}$ from the observation that A is Freiman 8-isomorphic to a subset of $\mathbb{Z}$. Indeed if $A \subseteq [-M, M]^2$ then an 8-isomorphism is given by the map $(x, y) \mapsto x + 8My$, as may be easily checked.

Finally, we must say something about the case of interest, namely subsets of G^2. Suppose that Γ is the graph of a function, as described, and that $\sigma[\Gamma] \leqslant K$. Consider the "unwrapped" version of Γ, $\overline{\Gamma}$, obtained by applying the unwrapping map $G^2 \mapsto \{0, 1 \dots, N'-1\}^2$ to Γ. Clearly $|\Gamma| = |\overline{\Gamma}|$, and each sum in $\Gamma+\Gamma$ gives rise to at most four sums in $\overline{\Gamma} + \overline{\Gamma}$. Therefore $\sigma[\overline{\Gamma}] \leqslant 4K$, and hence by the preceding discussion there is some arithmetic progression $Q \subseteq \mathbb{Z}^2$ with $|Q| \geqslant \exp(-K^C)|\Gamma|^{1/K^C}$ such that $|\overline{\Gamma} \cap Q| \geqslant K^{-C}|Q|$. Since $|S| = |\Gamma|$ is so large this implies that $|\overline{\Gamma} \cap Q| > 1$; the sole purpose of this observation is to note that the common difference of the projection of Q back down to G^2 is of the form (d_1, d_2) with $d_1 \neq 0$ (note that Γ, being a graph, has only one point (x, y) above each x). Writing Q' for this projection, we have $|\Gamma \cap Q'| \geqslant K^{-C}|Q'|$.

To conclude, note that a progression such as Q' is precisely the same thing as the graph of a linear function $x \mapsto ax + b$. □

Let us conclude the section by combining Lemmas 5.3 and the remarks preliminary to it with Proposition 5.4. Recall that $G = \mathbb{Z}/N'\mathbb{Z}$.

COROLLARY 5.2. *Let $\delta < 1/2$ be a real parameter, and suppose that $f : G \to \mathbb{C}$ is a 1-bounded function with $\|f\|_{U^3} \geqslant \delta$. Then there is an arithmetic progression $Q \subseteq G$ of length at least* $\exp(-1/\delta^C)N^{\delta^C}$ *as well as elements $a, b \in G$ such that*

$$\mathbb{E}_{h \in Q}|\Delta(f, h)^\wedge(ah + b)|^2 \geqslant \delta^C.$$

5.4. Gowers' local inverse theorem for the U^3-norm, II

In the last section we showed that if a 1-bounded function $f : G \to \mathbb{C}$ has large U^3-norm then a little piece of its derivative is linear. If we knew that its derivative was globally linear, say for example that

$$\mathbb{E}_{h \in G}|\Delta(f, h)^\wedge(2h)|^2 \geqslant \delta,$$

then it would be quite easy to proceed. Indeed upon expanding out this becomes

$$\mathbb{E}_{h \in G}\mathbb{E}_{x,y \in G}f(x)\overline{f(x+h)f(y)}f(y+h)e(-2h(x-y)/N') \geqslant \delta.$$

Making the obvious substitution $y - x = k$, this implies that

$$\mathbb{E}_{x,h,k}f(x)\overline{f(x+h)f(x+k)}f(x+h+k)e(2hk/N') \geqslant \delta.$$

This may be written as $\|\tilde{f}\|_{U^2} \geqslant \delta$, where $\tilde{f}(x) = f(x)e(x^2/N')$. Applying the inverse theorem for the U^2-norm, this immediately implies that there is some $r \in G$ such that

$$|\mathbb{E}_{x \in G}f(x)e((x^2 - rx)/N')| \geqslant \delta^2,$$

that is to say f correlates *globally* with a quadratic.

In actual fact we only have linearity on a small progression Q. The idea for handling this is quite similar but a little more technically involved and not so clean. Furthermore the conclusion is (necessarily) weaker, in that we only get the following kind of local correlation with quadratic behaviour.

PROPOSITION 5.5 (Local inverse theorem for the U^3-norm). *Let $0 < \delta \leqslant 1/2$ and suppose that $f : G \to \mathbb{C}$ is a 1-bounded function with $\|f\|_{U^3} \geqslant \delta$. Then there is an arithmetic progression Q of length at least $\exp(-1/\delta^C)N^{\delta^C}$ together with quadratics $\psi_1, \dots, \psi_{N'} : G \to G$ such that*

$$\mathbb{E}_j|\mathbb{E}_{x\in Q+j}f(x)e(-\psi_j(x))| \geqslant \delta^C.$$

Proof. By the analysis of the last section there is some progression $Q \subseteq G$ of length $\exp(-1/\delta^C)N^{\delta^C}$ together with a, b such that

$$\mathbb{E}_{h\in Q}|\Delta(f,h)^\wedge(ah+b)|^2 \geqslant \delta^C.$$

Expanding this out we obtain

$$\mathbb{E}_{h\in Q}E_{x,y\in G}f(x)\overline{f(x+h)f(y)}f(y+h)e(-(ah+b)(x-y)/N') \geqslant \delta^C$$

which, after an obvious substitution, becomes

$$\mathbb{E}_{h\in Q}\mathbb{E}_{x,k\in G}f(x)\overline{f(x+h)f(x+k)}f(x+h+k)e((ah+b)k/N') \geqslant \delta^C.$$

This may be written as

$$\mathbb{E}_{h\in Q}\mathbb{E}_{x,k\in G}f_1(x)\overline{f_2(x+h)f_3(x+k)}f_4(x+h+k) \geqslant \delta^C, \tag{5.1}$$

where $f_1(x) = f_3(x) = e(ax^2/2N' - bx/N')$ and $f_2(x) = f_4(x) = e(ax^2/2N')$.

The purpose of this rewriting, as in the simpler model case discussed above, is that we now have something that looks remarkably like a U^2-norm, or at least a U^2-inner product. The main barrier to saying anything about this last expression is the presence of Q. Suppose to begin with (for notational simplicity) that $Q = \{0, 1 \dots, L-1\}$. Introducing two extra averagings we obtain

$$\mathbb{E}_{i,j\in G}\mathbb{E}_{x,h,k\in Q}f_{1,i}(x)\overline{f_{2,i}(x+h)f_{3,i,j}(x+k)}f_{4,i,j}(x+h+k) \geqslant \delta^C, \tag{5.2}$$

where $f_{1,i}(x) = f_1(x+i)$, $f_{2,i}(x) = f_2(x+i)$, $f_{3,i,j}(x) = f_3(x+i+j)$ and $f_{4,i,j}(x) = f_4(x+i+j)$. Applying the pigeonhole principle we may remove the averaging over j, taking one specific j for which the remaining average is at least δ.

To handle expressions of this type we need the following lemma, a slight technical variant of the inverse theorem for the U^2-norm. I think the proof is quite instructive in that it demonstrates some of the power of harmonic analysis.

LEMMA 5.4. *Suppose that $g_1, g_2, g_3, g_4 : \mathbb{Z} \to \mathbb{C}$ are 1-bounded functions with*

$$\mathbb{E}_{x,h,k\in Q}g_1(x)\overline{g_2(x+h)g_3(x+k)}g_4(x+h+k) \geqslant \eta,$$

where $0 < \eta < 1/2$. Then there is some θ such that $|\mathbb{E}_{x\in Q}g_1(x)e(-\theta x)| \geqslant \eta^C$.

Proof. Let $Q' = 3Q$ (say). The assumption implies that

$$\mathbb{E}_{x,h,k\in\mathbb{Z}/Q'\mathbb{Z}} g_1(x)\overline{g_2(x+h)g_3(x+k)}g_4(x+h+k)1_Q(x)1_Q(h)1_Q(k) \geqslant \eta/9. \tag{5.3}$$

Our aim is to make this into a Gowers inner product and then apply the inverse theorem for the U^2-norm on $\mathbb{Z}/Q'\mathbb{Z}$. To do this we need to find a way of eliminating the cutoffs $1_Q(h)$ and $1_Q(k)$, and the rough idea is to expand them using the Fourier transform. This is generally a very useful way of handling irritating cutoffs: as Terry Tao once remarked to me, "all cutoffs are essentially the same".

This idea does not quite work in its most naïve form, since the cutoffs 1_Q are not "smooth" enough. We replace them by the function $\tilde{1}_Q$ defined by convolving 1_Q with $\frac{1}{2\varepsilon L}1_{\{-\varepsilon L,\dots,\varepsilon L\}}$, where $\varepsilon = \eta/100$ (say). Note that

$$\mathbb{E}_{x\in\mathbb{Z}/Q'\mathbb{Z}}|1_Q(x) - \tilde{1}_Q(x)| \leqslant \eta/50,$$

and so (5.3) may be replaced by

$$\mathbb{E}_{x,h,k\in\mathbb{Z}/Q'\mathbb{Z}} g_1(x)\overline{g_2(x+h)g_3(x+k)}g_4(x+h+k)1_Q(x)\tilde{1}_Q(h)\tilde{1}_Q(k) \geqslant \eta/18. \tag{5.4}$$

Suppose that $I \subseteq \mathbb{Z}/Q'\mathbb{Z}$ is an interval and that $r \in \mathbb{Z}/Q'\mathbb{Z}$ satisfies $|r| \leqslant Q'/2$. By the geometric series formula and the lower bound $|\sin t| \geqslant 2|t|/\pi$ for $|t| \leqslant \pi/2$, the Fourier transform $\widehat{1}_I(r)$ is bounded by $C\min(1,|r|^{-1})$. It follows easily that $\widetilde{1}^{\wedge}_Q(r)$ is bounded by $C\eta^{-1}\min(1,|r|^{-2})$, and hence by the inversion formula we have

$$\tilde{1}_Q(x) = \sum_r a_r e(rx/Q'),$$

where

$$\sum_r |a_r| = \sum_r |\widetilde{1}^{\wedge}_Q(r)| \leqslant C\eta^{-1}.$$

It was to obtain this estimate that we applied the smoothing device: if we had just used the unsmoothed cutoff 1_Q, the right hand side here would have had a fatal factor of $\log Q$. Substituting into (5.4) it follows that

$$\sum_{r,s} a_r a_s \mathbb{E}_{x,h,k} g_1(x)1_Q(x)\overline{g'_{2,r}(x+h)g'_{3,s}(x+k)}g'_{4,r,s}(x+h+k) \geqslant \eta/18,$$

where $g'_{2,r}(x) = g_2(x)e(rx/Q')$, $g'_{3,s}(x) = g_3(x)e(sx/Q')$ and $g'_{4,r,s}(x) = g_4(x)e((r+s)x/Q')$, and hence there is a choice of r,s such that

$$\mathbb{E}_{x,h,k} g_1(x)1_Q(x)\overline{g'_{2.r}(x+h)g'_{3,s}(x+k)}g'_{4,r,s}(x+h+k) \geqslant c\eta^C.$$

By the Gowers-Cauchy-Schwarz inequality we have $\|g_1 1_Q\|_{U^2} \geqslant c\eta^C$, and finally the inverse theorem for the U^2-norm lets us conclude. □

With this lemma in hand, let us return to (5.2). It is clear that for at least $\delta^C|G|$ values of i we have

$$\mathbb{E}_{x,h,k\in Q} f_{1,i}(x)\overline{f_{2,i}(x+h)f_{3,i,j}(x+k)}f_{4,i,j}(x+h+k) \geqslant \delta^C.$$

By the lemma it follows that for each such i there is some $\theta_i \in \mathbb{R}$ such that

$$|\mathbb{E}_{x\in Q} f_{i,1}(x)e(-\theta_i x)| \geqslant \delta^C.$$

Summing over i and writing out the definition of $f_{1,i}$ in full gives

$$\mathbb{E}_{i\in G}|\mathbb{E}_{x\in Q} f(x+i)e(a(x+i)^2/2N' - b(x+i)/N')e(-\theta_i x)| \geqslant \delta^C,$$

a conclusion which may be written much more succinctly as

$$\mathbb{E}_{i \in G} |\mathbb{E}_{x \in i+Q} f(x) e(-\psi_i(x))| \geqslant \delta^C,$$

where the ψ_i are quadratics (in fact, each has the same lead coefficient, namely $a/2N'$).

Now at an earlier point we made the assumption that $Q = \{0, \dots, L-1\}$. By making affine linear changes of variables in (5.1), we may reduce to this situation at the expense of making affine linear substitutions in f_1, f_2, f_3 and f_4. One may then apply the above argument, remembering at the end to undo those affine transformations. This, of course, still leaves us with quadratic phases ψ_i, and thus the proof of Proposition 5.5 is complete. □

5.5. A little diophantine approximation

We are now in a somewhat familiar position. Recall that in Chapter 1 our last task in the proof of Roth's theorem was to go from (1.1), which asserted that our balanced function f correlated with a linear phase $e(\theta x)$, to (1.2), asserting that f has large mass on a subprogression. Our aim here is the same, but our starting point is not quite so nice. We now have only *local* correlations, and now with *quadratic* phases and not linear ones.

It turns out to be the quadraticness which presents the most serious obstacle. To deal with it we need the following analogue of Dirichlet's lemma 1.5. The proof of that result was almost trivial, but now we need some reasonably serious machinery with the flavour of analytic number theory and the Hardy-Littlewood method.

Here, then, is the main result of this subsection.

LEMMA 5.5 (Diophantine approxiation by squares). *Let $\theta \in \mathbb{R}$, and suppose that $N \geqslant 1$. Then there is some $n \leqslant N$ such that $\|n^2\theta\|_{\mathbb{R}/\mathbb{Z}} \leqslant N^{-c}$.*

The proof of this depends on the follow estimate of Weyl concerning "exponential sums" over squares (exponential sum being an old-fashioned name for what is, essentially, a type of Fourier transform).

LEMMA 5.6 (Weyl's inequality). *Let $\delta \leqslant 1/2$ and suppose that $N > \delta^{-C}$. Suppose that $\frac{1}{N}|\sum_{n \leqslant N} e(\alpha n^2 + \beta n + \gamma)| \geqslant \delta$. Then there is some $q \leqslant \delta^{-C}$ such that $\|q\alpha\|_{\mathbb{R}/\mathbb{Z}} \leqslant \delta^C/N^2$.*

Remark. The following thoughts underpin this statement. The phase $\alpha n^2 + \beta n + \gamma$ can be expected, if α is "highly irrational", to be very uniformly distributed (mod 1). If that is so, the exponential sum will be small because the quantities $e(\alpha n^2 + \beta n + \gamma)$ will cancel out in the complex plane.

Proof. Squaring the assumption we of course obtain

$$\sum_{n,m\leqslant N} e(\alpha(m^2-n^2))+\beta(m-n))\geqslant \delta^2N^2.$$

Substituting $m=n+h$ it follows that

$$\sum_{|h|\leqslant N}|\sum_{n\in I_h} e(2\alpha hn)|\geqslant \delta^2N^2,$$

where, for each h, $I_h\subseteq\{1,\dots,N\}$ is some interval. By the geometric series formula the inner sum may be calculated quite explicitly, and it is certainly at most $2/|1-e(2\alpha h)|$, which may be bounded above by $C\|2\alpha h\|_{\mathbb{R}/\mathbb{Z}}^{-1}$. It is also, of course, at most N. It follows that there are at least $\delta^C N$ values of h, $|h|\leqslant N$, such that $\|2\alpha h\|_{\mathbb{R}/\mathbb{Z}}\leqslant \delta^{-C}/N$. It follows trivially from this that there are at least $\delta^{C_0}N$ values of $h\in\{0,1,\dots,2N\}$ such that $\|\alpha h\|_{\mathbb{R}/\mathbb{Z}}\leqslant\delta^{-C_0}/N$. For the remainder of this argument the constant C_0 will be fixed.

Now let $Q=\delta^{C_1}N^2$, where C_1 is a constant to be specified later (depending on C_0). By applying Dirichlet's pigeonhole argument there is some $q\leqslant Q$ such that

$$\|\alpha q\|_{\mathbb{R}/\mathbb{Z}}\leqslant 1/Q. \tag{5.5}$$

If $q\leqslant\delta^{-C_1}$ then the conclusion of the lemma is satisfied, so suppose this is not the case. Now (5.5) trivially implies that there is some integer a such that

$$|\alpha-\frac{a}{q}|\leqslant\frac{1}{qQ}.$$

We may assume without loss of generality that $\text{hcf}(a,q)=1$, since dividing through by any common factor can only decrease q. Using this, we may examine how often $h\alpha$, $h\in\{1,2,3,\dots\}$, can be near to an integer. Suppose that h and h' both lie in some interval of length $q/2$. Then $ah\neq ah'(\bmod q)$ and so ah/q and ah'/q differ by at least $1/q(\bmod 1)$. It follows that

$$\|\alpha(h-h')\|_{\mathbb{R}/\mathbb{Z}}\geqslant\frac{1}{q}-\frac{|h-h'|}{qQ}\geqslant\frac{1}{q}-\frac{1}{2Q}\geqslant\frac{1}{2q}.$$

Thus the numbers $\|\alpha h\|_{\mathbb{R}/\mathbb{Z}}$ are $(1/2q)$-spaced, as h ranges over any interval of length at most $q/2$. In particular the number of h in any such interval which satisfy $\|\alpha h\|_{\mathbb{R}/\mathbb{Z}}\leqslant\delta^{-C_0}/N$ is $O(1+q\delta^{-C_0}/N)$. The whole range $\{0,1,\dots,2N\}$ may be divided up into $O(1+N/q)$ such intervals, and so we may bound the number of $h\in\{0,1,\dots,2N\}$ such that $\|\alpha h\|_{\mathbb{R}/\mathbb{Z}}\leqslant\delta^{-C_0}/N$ by

$$O\big((1+N/q)(1+q\delta^{-C_0}/N)\big)=O(1+\frac{N}{q}+\frac{q\delta^{-C_0}}{N}+\delta^{-C_0}).$$

Since $q>\delta^{-C_1}$, $q\leqslant Q=\delta^{C_1}N^2$ and N is bigger than a large power of δ^{-1} this much less than $\delta^{C_0}N$, provided that $C_1\geqslant 2C_0+1$. This is contrary to our earlier reasoning. □

Remarks. This is not quite the usual way of stating Weyl's inequality but, except for the casual treatment of exponents C, it is equivalent to it. For the more traditional formulation see [**?**]. Much the same statement holds for exponential sums involving polynomials of degree k rather than simply quadratics, but instead of simply squaring one must apply Cauchy-Schwarz several times to linearise the phase in order that the geometric series formula may be applied.

Proof of Lemma 5.5. Suppose that θ is a real number for which the conclusion is invalid. This means that for every C_0 there are arbitrarily small ε such that $\|n^2\theta\|_{\mathbb{R}/\mathbb{Z}} \geqslant \varepsilon$ for all $n = 1, \dots, N$, where $N := \lfloor \varepsilon^{-C_0} \rfloor$. Let $\psi : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$ be a function with $\|\psi\|_1 = 1$, $\psi(x) = 0$ if $\|x\|_{\mathbb{R}/\mathbb{Z}} \geqslant \varepsilon$ and a Fourier expansion $\psi(x) = \sum_r a_r e(rx)$ with $a_0 = 1$ and $\sum_{r \leqslant 1/\varepsilon^C} |a_r| \leqslant \frac{1}{2}$ (for example, a "tent function" of width ε, for which the Fourier expansion may be verified directly without any general theory). Then we have

$$\sum_{n \leqslant N} \psi(\theta n^2) = 0.$$

Expanding ψ in Fourier gives

$$|\sum_{r \neq 0} a_r \sum_{n \leqslant N} e(r\theta n^2)| \geqslant a_0 N = N,$$

and so there is some $r \leqslant 1/\varepsilon^C$ such that

$$|\sum_{n \leqslant N} e(r\theta n^2)| \geqslant \varepsilon^C.$$

By Weyl's inequality this implies that there is some $q \leqslant \varepsilon^{-C}$ such that $\|qr\theta\|_{\mathbb{R}/\mathbb{Z}} \leqslant \varepsilon^{-C}/N^2$. But then $\|n^2\theta\|_{\mathbb{R}/\mathbb{Z}} \leqslant \varepsilon^{-3C}/N^2$, where $n = qr \leqslant \varepsilon^{-2C}$. This is contrary to assumption if C_0 is chosen large enough. □

The following lemma is the one we shall actually require in the next section.

LEMMA 5.7. *We have the following.*

(i) *Suppose that $P \subseteq \mathbb{Z}$ is an arithmetic progression of length L, and that $\phi(x) = \alpha x + \beta$ is a linear polynomial. Then P may be subdivided into $O(L^{3/4})$ subprogressions, on each of which the total variation of $\phi(x)$ is at most $L^{-1/2}$.*
(ii) *Suppose that $P \subseteq \mathbb{Z}$ is an arithmetic progression of length L, and that $\psi(x) = \alpha x^2 + \beta x + \gamma$ is a quadratic polynomial. Then P may be subdivided into $O(L^{1-c})$ subprogressions, on each of which the total variation of $\psi(x)$ is at most L^{-c}.*

Proof. For both parts we may apply an affine transformation to P, which does not change the form of the problem, so suppose $P = \{1, \dots, L\}$.

For the part (i), use Dirichlet's theorem to find a $d \leqslant L^{1/2}$ such that $\|\alpha d\|_{\mathbb{R}/\mathbb{Z}} \leqslant L^{-1/2}$; then any progression with common difference d and length at most $L^{1/4}$ will have the property. It is easy to see that $\{1, \dots, L\}$ may be divided into $O(L^{3/4})$ such subprogressions.

(ii) Using Lemma 5.5 select $d \leqslant L^{1/4}$ such that $\|d^2\alpha\|_{\mathbb{R}/\mathbb{Z}} \leqslant L^{-c}$. Consider a progression $Q = \{a, a+d, \dots, a+(L'-1)d\}$ where $L' := L^{c/3}$. We have

$$\alpha(a+td)^2 + \beta(a+td) + \gamma = t^2(\alpha d^2) + td(\beta + 2\alpha a) + \alpha a^2 + \beta a + \gamma.$$

The total variation of $t^2(\alpha d^2)$ is at most $L^{-c/3}$. Now subdivide Q into further subprogressions on which the total variation of $td(\beta + 2\alpha a)$ is small using the first part. □