that for any distinct $x, x' \in V$, we have $f_i(x) = f_i(x')$ for at most $3n/4$ values of i, or in other words $|\mathbf{E}_{i\in[1,n]} f_i(x) f_i(x')| \leq 1/2$. (Hint: use the probabilistic method. Alternatively, identify V with an error-correcting code in $\{-1, +1\}^n$, constructed for instance using the greedy algorithm.) If $\lambda : V \to \mathbf{R}^+$ is any function such that $\|\lambda\|_{l^1(V)} = 1$ and $\|\lambda\|_{l^\infty(V)} \leq 1 - \varepsilon$ for some $\varepsilon > 0$, show that

$$\mathbf{E}_{i\in[1,n]} \left| \sum_{x\in V} \lambda(x) f_i(x) \right|^2 \leq 1 - \Omega(\varepsilon).$$

Conclude in particular that $|\sum_{x\in V} \lambda(x) f_i(x)| \leq 1 - \Omega(\varepsilon)$ for at least $\Omega(\varepsilon n)$ values of i.

10.6.11 [136] Let V be a large finite set, and let $f_1, \ldots, f_n : V \to \{-1, +1\}$ be as in the preceding exercise. Let W be another large finite set, let G be the graph with vertex set $[1, n] \times V \times W$, with any two distinct vertices $(i, x, w), (j, y, z)$ being connected by an edge if and only if $f_i(y) = f_j(x)$. Let $\varepsilon > 0$, and suppose that $[1, n] \times V$ is partitioned into $[1, n] \times V \times W = V_1 \cup \cdots \cup V_k$ as in the regularity lemma. Suppose further that for all but $O(\varepsilon k)$ of the sets V_s, there exists an $i_s \in [1, n]$ such that $|V_s \cap (\{i_s\} \times V \times W)| \geq (1 - O(\varepsilon))|V_s|$; thus up to errors of $O(\varepsilon)$, most of the cells V_s of the partition are essentially contained in one of the $\{i\} \times V$. Conclude that for all but $O(\varepsilon k)$ of the sets V_s, there exists $i_s \in [1, n]$ and $x_s \in V$ such that $|V_s \cap (\{i_s\} \times \{x_s\} \times W)| \geq (1 - O(\varepsilon))|V_s|$; thus any regular partition which essentially refines the partition $\{\{i\} \times V \times W\}$, must automatically essentially refine the finer partition $\{\{i\} \times \{x\} \times W\}$. (This is a more complicated version of Exercise 10.6.9, and requires use of the previous exercise, with $\lambda(x)$ being equal to the relative density of $V_s \cap (\{i_s\} \times \{x_s\} \times W)$ in $V_s \cap (\{i_s\} \times V \times W)$.) An iteration of this fact can be used to establish a lower bound of tower type for the Szemerédi regularity lemma; see [136].

10.7 Szemerédi's argument

In this section we give another proof of Roth's theorem due to Szemerédi (see e.g. [143]). This argument gives slightly better bounds than that obtained from the regularity lemma, but still worse than that given from the Fourier-analytic argument. However, it has the advantage of being completely elementary and rather short. A more complex version of this argument was also used in [343] to establish Szemerédi's theorem for progressions of length 4, but the general k case requires a rather different (and even more complex) combinatorial argument which we will not discuss here; see [345].

Intuitively, the idea is as follows. If A is a dense set of an interval $[1, N]$, then it should contain a large cube $a + [0, 1]^d \cdot v$. If A also has no proper progressions of length three, then this implies that A must be disjoint from a sumset $2a + [0, 1]^d \cdot 2v - A_0$ of a large set and a cube. This disjointness "squeezes" A into a collection of moderately long progressions, and the density of A must increase on at least one of them. This creates a density increment that one can then iterate as in the Fourier-analytic proof of Roth's theorem. Thus one is using the disjointness from a sumset as a substitute for Fourier bias (cf. Exercise 4.3.12).

We now give the main steps of the argument, leaving the proofs as exercises. First we need to show that dense sets contain cubes.

Lemma 10.49 *Let $0 < \delta < 1$, let P be a large proper arithmetic progression, and let A be a subset of P with $|A| \geq \delta|P|$. Then A contains a proper cube $a + [0, 1]^d \cdot v$ with $d = \Omega_\delta(\log \log |P|)$. In particular all the steps $v_1, \ldots, v_d$ are non-zero.*

The main point here is that the quantity d goes to infinity (somewhat slowly) as $|P| \to \infty$. As a corollary of this lemma, we have

Corollary 10.50 *Let $0 < \delta < 1$, let N be a sufficiently large integer depending on δ, and let A be a subset of $[1, N]$ with $|A| \geq \delta N$ which contains no proper progressions of length three. Then at least one of the following statements is true:*

- *(Density increment) There exists a progression $P \subset [1, N]$ of length $|P| \geq N/4 + O(1)$ such that $|A \cap P| \geq 1.1\delta|P|$.*
- *(Disjointness from sumset) There exists a set $A_0 \subset [1, N/4]$ and a cube $a + [0, 1]^d \cdot v \subset (N/4, N/2]$ with $d = \Omega_\delta(\log \log N)$ and all steps $v_1, \ldots, v_d$ non-zero such that $|A_0| = \Omega(\delta N)$ the sumset $2a + [0, 1]^d \cdot 2v - A_0 \subset [1, N]$ is disjoint from A.*

Proof Without loss of generality we may assume that $|A \cap (iN/4, (i + 1)N/4]| \leq 1.1\delta N/4 + O(1)$ for all $i = 0, 1, 2, 3$ otherwise we have a density increment. In particular this implies that $|A \cap (iN/4, (i + 1)N/4]| = \Omega(\delta N)$ for all i. Applying Lemma 10.49 to the set $A \cap (N/4, N/2]$ we see that this set contains a cube $a + [0, 1]^d \cdot v$ with the desired properties. If we then set $A_0 := A \cap [1, N/4)$, the claim then follows by observing that whenever $x \in A_0$ and $y \in a + [0, 1]^d \cdot v$, the sequence $x, y, 2y - x$ is a proper arithmetic progression and hence $2y - x$ cannot lie in A. □

Suppose we are in the situation in the above corollary, and the "disjointness from sumset" statement holds. Let $E_0 \subset E_1 \subset \cdots \subset E_d \subset [1, N]$ be the sets

$$E_i := 2a + [0, 1]^i \cdot (2v_1, \ldots, 2v_i) - A_0.$$

By the pigeonhole principle we can find $1 \leq i \leq d$ such that

$$|E_i| \leq |E_{i-1}| + O\left(\frac{N}{d}\right).$$

On the other hand, we have $E_i = E_{i-1} + \{0, 2v_i\}$. This shows that we can partition the set $[1, N]\backslash E_i$ into $O(\frac{N}{d})$ proper arithmetic progressions $P_1, \ldots, P_k$ of step $2v_i$ (see Exercises). Observe that

$$\sum_{j=1}^{k} |P_j| = N - |E_i| \leq N - |A_0| = (1 - \Omega(\delta))N.$$

On the other hand, since A is disjoint from E_i, we have

$$\sum_{j=1}^{k} |A \cap P_j| = |A| \geq \delta N.$$

We thus have

$$\sum_{j=1}^{k} |A \cap P_j| - (\delta + c\delta^2)|P_j| - cd > 0$$

for some small absolute constant $c > 0$. Thus by the pigeonhole principle, there exists a progression P_j such that $|P_j| \geq cd = \Omega_\delta(\log\log N)$ and $|A \cap P_j| \geq (\delta + c\delta^2)|P_j|$. This establishes a density increment of A on a progression whose length goes to infinity as $N \to \infty$. This is essentially Corollary 10.26 (but with somewhat worse explicit constants) and one can now iterate this Corollary as before to establish Roth's theorem.

A careful accounting of the bounds here yields the bound $r_3([1, N]) = O(N/\log_* N)$, which is marginally better than the bounds obtained by the regularity lemma.

Exercises

10.7.1 Prove Lemma 10.49. (Hint: first prove the following preliminary statement: if $|A| \geq \delta|P|$, and $|P|$ is large depending on δ, then there are $\Omega(\delta^2|P|)$ values of v such that $|A \cap (A + v)| = \Omega(\delta^2|P|)$.)

10.7.2 Let $A \subset [1, N]$ and $r \neq 0$ be such that $A + r \subset [1, N]$ and $|A + \{0, r\}| \leq |A| + k$. Show that $[1, N]\backslash(A + \{0, r\})$ can be partitioned into $O(k)$ disjoint arithmetic progressions of step r.