$a + \mathbf{Z}^+ \cdot r$. Thus the properties of containing arbitrarily long proper arithmetic progressions, and infinitely long proper arithmetic progressions, are distinct.

10.0.6 Show that the Erdős–Turán conjecture is equivalent to the absolute convergence of the sum

$$\sum_{n=1}^{\infty} \frac{r_k([1, 2^n])}{2^n}.$$

10.0.7 Show that if A and B are additive sets which are Freiman isomorphic of order 2, then $r_k(A) = r_k(B)$ for all k.

10.0.8 If A and B are additive sets (possibly in different groups), show that $r_k(A \times B) \geq r_k(A) r_k(B)$.

10.0.9 If Z, Z' are two finite additive groups, show that $r_k(Z \times Z') \leq r_k(Z)|Z'|$.

10.0.10 Show that to prove Theorem 10.5 for arbitrary groups Z, it suffices to verify it for cyclic groups $\mathbf{Z}_N$ and for vector spaces $\mathbf{Z}_p^n$ over fields of prime order. (Hint: use Corollary 3.8 and the previous exercise.) A similar claim applies of course to Roth's theorem.

10.0.11 Let $n \geq 1$. Define a *capset* of order n to be any subset of the vector space F_3^n over the finite field F_3 which contains no (affine) lines. Show that the largest possible cardinality of a capset of order n is $r_3(F_3^n)$. Using Exercise 10.0.8, show that $r_3(F_3^n) \geq 2^n$.

10.0.12 If Z is a finite additive group whose order is coprime to $k!$, show that $r_k(Z) \leq (1 - \frac{1}{k})|Z|$. (Hint: if $A \subset Z$ has cardinality greater than $(1 - \frac{1}{k})|Z|$, choose $a \in Z$, $r \in Z \backslash \{0\}$ randomly and consider the probability of the events $a + jr \notin A$ for $j = 0, 1, \ldots, k - 1$.)

10.1 General strategy

In this section we make some general observations concerning progressions of length 3, and describe in high-level terms the various strategies one could employ to prove Roth-like theorems.

Let us work in a fixed finite additive group Z of odd order, and let A be a subset of Z. We shall think of A as being rather dense, so that the density $0 \leq \mathbf{P}_Z(A) \leq 1$ is moderately large. Roth's theorem is then an assertion that if $|Z|$ is sufficiently large, then A must contain progressions of length three.

To explain why this should be the case, it is convenient to introduce the trilinear form

$$\Lambda_3(f, g, h) := \mathbf{E}_{x,r \in Z} f(x) g(x + r) h(x + 2r) \tag{10.1}$$

for any $f, g, h : Z \to \mathbf{C}$. Note in particular that

$$\Lambda_3(1_A, 1_A, 1_A) = \mathbf{P}_{x,r\in Z}(x, x+r, x+2r \in A) \tag{10.2}$$

so the quantity $\Lambda_3(1_A, 1_A, 1_A)$ measures the proportion of arithmetic progressions $(x, x+r, x+2r)$ in Z which are completely contained in A. Intuitively, if A is "randomly" distributed, then the events $x \in A$, $x+r \in A$, $x+2r \in A$ should be "independent", and we then expect

$$\begin{aligned}\Lambda_3(1_A, 1_A, 1_A) &\approx \mathbf{P}_{x,r\in Z}(x \in A)\mathbf{P}_{x,r\in Z}(x+r \in A)\mathbf{P}_{x,r\in Z}(x+2r \in A)\\ &= \mathbf{P}_Z(A)^3. \end{aligned}\tag{10.3}$$

Thus if A is fairly dense in Z, we expect $\Lambda_3(1_A, 1_A, 1_A)$ to be large. On the other hand, if $|Z|$ is odd and A has no proper progressions of length 3, then the only progressions $(x, x+r, x+2r)$ which can lie in A are those for which $x \in A$ and $r = 0$, whence

$$\Lambda_3(1_A, 1_A, 1_A) = \mathbf{P}_Z(A)/|Z|. \tag{10.4}$$

If $|Z|$ is sufficiently large, this seems to be in conflict with the heuristic (10.3). Thus to prove Roth's theorem it will suffice to establish some rigorous analog of (10.3). In particular, Roth's theorem will be implied by the following result.

Theorem 10.9 (Varnavides' theorem) *[372] Let Z be a finite additive group of odd order. Then for any non-empty set $A \subseteq Z$ we have*

$$\Lambda_3(1_A, 1_A, 1_A) = \Omega_{\mathbf{P}_Z(A)}(1).$$

In other words, we have $\Lambda_3(1_A, 1_A, 1_A) \geq c(\mathbf{P}_Z(A))$ where $c(\mathbf{P}_Z(A)) > 0$ depends only on the density $\mathbf{P}_Z(A)$ of A and not on the group Z. More generally, if $f : Z \to \mathbf{R}^+$ is a non-negative function which is not identically zero, and obeying the bound $0 \leq f(x) \leq 1$ for all $x \in Z$, then

$$\Lambda_3(f, f, f) = \Omega_{\mathbf{E}_Z(f)}(1).$$

Note that Varnavides' theorem is in fact a bit stronger than Roth's theorem, as it implies that any subset of Z of density δ will contain $\Omega_\delta(|Z|^2)$ proper arithmetic progressions of length 3, if Z is sufficiently large depending on δ. This is in contrast with Roth's theorem which would only provide a single proper arithmetic progression of length 3. Nevertheless, a simple averaging argument shows that the two theorems are equivalent: see exercises.

It is still not clear how to convert the heuristic (10.3) into a rigorous statement such as Theorem 10.9. Indeed (10.3) can fail for certain special A, with $\Lambda_3(1_A, 1_A, 1_A)$ ranging as high as $\mathbf{P}_Z(A)^2$ if A is a subgroup of Z, and as low as $\mathbf{P}_Z(A)^{\Omega(\log \frac{1}{\mathbf{P}_Z(A)})}$ if A is given by the Behrend example (see exercises). However, it

turns out that $\Lambda_3(1_A, 1_A, 1_A)$ will be very close to $\mathbf{P}_Z(A)^3$ (as predicted by (10.3)) as long as A has very little *linear bias*. Recall from Definition 4.12 that the linear bias (or Fourier bias) $\|A\|_u$ of an additive set A was defined as

$$\|A\|_u := \sup_{\xi \in Z \backslash 0} |\hat{1}_A(\xi)| = \sup_{\xi \in Z \backslash 0} |\mathbf{E}_{x \in Z} 1_A(x) e(-\xi \cdot x)|.$$

Proposition 10.10 (Lack of progressions implies non-uniformity) *[287] Let A be an additive set in a finite additive group Z of odd order. Then*

$$|\Lambda_3(1_A, 1_A, 1_A) - \mathbf{P}_Z(A)^3| \leq \|A\|_u \mathbf{P}_Z(A).$$

In particular, if A contains no proper arithmetic progressions of length 3, then we have the linear bias estimate

$$\|A\|_u \geq \mathbf{P}_Z(A)^2 - \frac{1}{|Z|}.$$

Proof From the identity $a - 2(a + r) + (a + 2r) = 0$, and the observation that the map $x \mapsto 2 \cdot x$ is bijective on Z when $|Z|$ is odd, we see that

$$\Lambda_3(1_A, 1_A, 1_A) = \frac{1}{|Z|^2} |\{(a_1, a_2, a_3) \in A \times (-2 \cdot A) \times A : 0 = a_1 + a_2 + a_3\}|.$$

Applying Lemma 4.13 we obtain the first inequality. The second claim then follows from (10.4). □

This shows that the only way the heuristic (10.3) can fail is if the function 1_A has a large correlation with a linear character $e(\xi \cdot x)$. This very important observation can be viewed as an *inverse theorem* for Λ_3; we will return to this perspective in the next chapter. There is an analog of the above proposition for functions. Define the *linear bias* $\|f\|_{u^2(Z)}$ of a function $f : Z \to \mathbf{C}$ to be the quantity

$$\|f\|_{u^2(Z)} := \sup_{\xi \in Z} |\hat{f}(\xi)|. \tag{10.5}$$

The reason for the notation $u^2(Z)$ will be made clearer in the next chapter. Note for instance that $\|A\|_u = \|1_A - \mathbf{P}_Z(A)\|_{u^2(Z)}$ for any $A \subseteq Z$.

Proposition 10.11 *Let Z have odd order. For any functions $f, g, h : Z \to \mathbf{C}$, we have the identity*

$$\Lambda_3(f, g, h) = \sum_{\xi \in Z} \hat{f}(\xi) \hat{g}(-2\xi) \hat{h}(\xi). \tag{10.6}$$

We can then conclude the estimate

$$|\Lambda_3(f, g, h)| \leq \|f\|_{u^2(Z)} \|g\|_{L^2(Z)} \|h\|_{L^2(Z)}$$

and similarly with f, g, h permuted on the right-hand side.

Proof From the Fourier inversion formula (4.4) we have

$$f = \sum_{\xi_1} \hat{f}(\xi_1) e_{\xi_1}; \quad g = \sum_{\xi_2} \hat{g}(\xi_2) e_{\xi_2}; \quad h = \sum_{\xi_3} \hat{h}(\xi_3) e_{\xi_3}$$

and hence

$$\Lambda_3(f, g, h) = \sum_{\xi_1, \xi_2, \xi_3 \in Z} \hat{f}(\xi_1)\hat{g}(\xi_2)\hat{h}(\xi_3)\Lambda(e_{\xi_1}, e_{\xi_2}, e_{\xi_3}).$$

On the other hand, a direct computation using Lemma 4.5 shows

$$\Lambda(e_{\xi_1}, e_{\xi_2}, e_{\xi_3}) = \mathbf{I}(\xi_2 = -2\xi_1; \xi_3 = \xi_1)$$

which gives (10.6). From Parseval's identity (4.3) and the hypothesis that Z has odd order, we have

$$\sum_{\xi \in Z} |\hat{g}(-2\xi)|^2 = \|g\|^2_{L^2(Z)}; \quad \sum_{\xi \in Z} |\hat{h}(\xi)|^2 = \|h\|^2_{L^2(Z)}$$

and the claim then follows from Hölder's inequality. Similarly if the roles of f, g, h are permuted. □

To exploit inverse results such as Proposition 10.10 or Proposition 10.11, there are two arguments available: the *density increment argument* of Roth, and the *energy increment argument* developed separately by Furstenberg and Szemerédi (in very different contexts). The density increment argument proceeds informally as follows. To prove Roth's theorem, suppose for contradiction that one can find a dense set A in a large group Z (or interval $[1, N]$) which contains no progressions of length three. Proposition 10.10 then implies that A has large linear bias, thus 1_A correlates with some linear phase function $e(\xi \cdot x)$. It then turns out that this linear bias can be converted into a *density increment*, or more precisely some structured subset Z' (such as a subgroup, a sub-progression, or a Bohr set) of the original space Z on which A has larger density, thus $\mathbf{P}_{Z'}(A) > \mathbf{P}_Z(A)$. (Recall that $\mathbf{P}_{Z'}(A) = |A \cap Z'|/|Z'|$ and $\mathbf{P}_Z(A) = |A|/|Z|$.) One then passes to this structured subset and repeats the argument. If the original space Z was large enough, we can run this argument for so many steps that the relative density of A eventually exceeds 1, a contradiction.

The energy increment argument proceeds differently, aiming to prove Varnavides' theorem instead of Roth's theorem (i.e. one seeks non-trivial lower bounds on $\Lambda_3(f, f, f)$). Instead of continually changing the ambient space Z, we now hold Z fixed, but instead construct certain *low complexity approximations* $f_{U^\perp}$ to the original function f. Initially, our approximation will just be the density, $f_{U^\perp} = \mathbf{P}_Z(A)$. We now consider the error $f_U := f - f_{U^\perp}$ between the indicator function and the approximation. If this error is very linearly uniform (in the sense

that the Fourier bias $\|f_U\|_{u^2(Z)}$ is small), then Proposition 10.11 can be used to approximate $\Lambda_3(f, f, f)$ by $\Lambda_3(f_{U^\perp}, f_{U^\perp}, f_{U^\perp})$, and one can exploit the low complexity of $f_{U^\perp}$ to obtain a non-trivial lower bound on the latter quantity. If instead the error exhibits linear bias, one can exploit this by refining the approximation $f_{U^\perp}$ to absorb this bias; this will increase the *energy* $\|f_{U^\perp}\|^2_{L^2(Z)}$ of $f_{U^\perp}$ by a significant amount. One then repeats the argument until the error f_U contains no further bias; a key point will be that that f (and hence $f_{U^\perp}$) remain bounded throughout the iteration and so the energy of $f_{U^\perp}$ cannot increase indefinitely.

Exercises

10.1.1 Let Z be a finite additive group of odd order, let $0 < \delta < 1$, and let A be a random subset of Z such that the events $x \in A$ are independent with probability $\mathbf{P}(x \in A) = \delta$. Show that with probability $1 - o_{|Z|\to\infty;\delta}(1)$, we have $\mathbf{P}_Z(A) = \delta + o_{|Z|\to\infty;\delta}(1)$ and $\Lambda_3(1_A, 1_A, 1_A) = \delta^3 + o_{|Z|\to\infty;\delta}(1)$, thus confirming (10.3) in the random case. (Hint: use Corollary 1.9.)

10.1.2 Let Z be a finite additive group of odd order. Show that $\Lambda_3(1_A, 1_A, 1_A) \leq \mathbf{P}_Z(A)^2$, with equality attained if and only if A is the translate of a subgroup of Z.

10.1.3 Let $N, d, r \geq 1$ be integers, and consider the set

$$A = \big\{(n_1, \ldots, n_d) \in [0, N/2)^d : n_1^2 + \cdots + n_d^2 = r\big\},$$

viewed as a subset of $\mathbf{Z}_N^d$. Show that this set has no proper arithmetic progressions of length 3, and can have cardinality as large as $(N/2)^d/(d^2N^2)$ for a suitable choice of r. Conclude in particular that $r_3(\mathbf{Z}_N^d) \geq N^d/(2^d d^2 N^2)$.

10.1.4 (Behrend's example) [21] Using the preceding exercise and a Freiman isomorphism, establish the bounds

$$r_3(\mathbf{Z}_N), r_3([1, N]) = \Omega\big(Ne^{-O(\sqrt{\log N})}\big)$$

for all large N. In particular, it is not the case that $r_3([1, N])$, $r_3(\mathbf{Z}_N) = O(N^{1-\varepsilon})$ for any fixed $\varepsilon > 0$. This rules out a number of elementary approaches to proving Roth's theorem or Szemerédi's theorem (e.g arguments based entirely on Cauchy–Schwarz and pigeonhole principle type arguments) as these tend to only give polynomial type bounds. We remark that the more general estimate

$$r_k(\mathbf{Z}_N), r_k([1, N]) = \Omega_k\big(N \exp\big(-O_k(\log N)^{1/(1+\lfloor \log_2(k-1)\rfloor)}\big)\big)$$

for all $k \geq 3$ has been established in [277], [221] by a similar argument.

10.1.5 Given any $0 < \delta < 1$, give an example of an additive set A in a cyclic group $\mathbf{Z}_N$ such that $\mathbf{P}_Z(A) \geq \delta$ but

$$\Lambda(1_A, 1_A, 1_A) = O\big(\delta^{\Omega(\log \frac{1}{\delta})}\big).$$

(Hint: use the Behrend example.) Thus it is not possible to establish any lower bound of the form $\Lambda(1_A, 1_A, 1_A) = \Omega(\mathbf{P}_Z(A)^C)$ for any absolute constant $C > 0$.

10.1.6 [253] Let N be a large number. Show that one can color $\mathbf{Z}_N$ into $\exp(O(\sqrt{\log N}))$ color classes, such that none of the color classes contains aproper arithmetic progression of length three. Hint: modify the Behrend example.

10.1.7 Show that Varnavides' theorem for sets A implies Varnavides' theorem for functions f. (Hint: either bound f from below by a constant multiple of an indicator function, or construct a set A probabilistically using $f(x)$ as the probability that $x \in A$ and use the first moment method.)

10.1.8 Show that the special case $r_3([1, N]) = o_{N\to\infty}(N)$ of Roth's theorem implies Varnavides' theorem for $\mathbf{Z}_N$. (Hint: take a set A in $\mathbf{Z}_N$ and intersect it with a randomly chosen progression $a + [1, M] \cdot r$ for some moderately large M, and apply Roth's theorem to the progression $a + [1, M] \cdot r$. Then use the first moment method.)

10.1.9 Let F be a finite field. Show that the special case $r_3(F^n) = o_{n\to\infty;F}(N)$ of Roth's theorem implies Varnavides' theorem for F^n. (Hint: take a set A in F^n and intersect it with a randomly chosen m-dimensional affine subspace of F^n for some moderately large m. Then argue as in the preceding exercise.)

10.1.10 Show that Roth's theorem for arbitrary Z implies Varnavides' theorem for arbitrary Z.

10.1.11 Use Proposition 10.11 and the decomposition $1_A = (1_A - \mathbf{P}_Z(A)) + \mathbf{P}_Z(A)$ to provide an alternative proof of Proposition 10.10.

10.1.12 Assume Theorem 10.9. Let $(X, \mathcal{B}, d\mu)$ be any probability space (so $\mu(X) = 1$), and let $T : X \to X$ be any measure-preserving bijection on X, so $\mu(T^n(E)) = \mu(E)$ for all $E \in \mathcal{B}$ and $n \in \mathbf{Z}$. Show that if $f : X \to \mathbf{R}^+$ is any function with $0 \leq f(x) \leq 1$ almost everywhere and $\int_X f = \delta > 0$, then

$$\liminf_{N\to\infty} \mathbf{E}_{n\in[-N,N]} \int_X f(x) T^n f(x) T^{2n} f(x)\, d\mu(x) = \Omega_\delta(1).$$