this scheme work, one must borrow heavily from the tools developed from the other approaches to Szemerédi's theorem, notably the energy increment strategy used to prove the regularity lemma, the Gowers uniformity norms used in the Fourier-analytic approach, and the relativization to pseudo-random sets which first appears in the graph and hypergraph regularity approach; in addition, some number-theoretic arguments (involving some analysis of the Riemann zeta function) is necessary to ensure that the almost primes have the desired amount of pseudo-randomness. We will discuss this result in Section 11.7.

Unfortunately we will not have space to give a full proof of Szemerédi's theorem or the Green–Tao theorem in this chapter, as all of the known proofs are either quite lengthy or require a fair amount of supporting theory. We will however give several partial results and key lemmas, and describe the main steps of most of these proofs, referring the reader to the original references for the complete details. Thus this chapter should be viewed as an introduction to the very active current area of research surrounding Szemerédi's theorem, rather than a comprehensive treatment of the field. One theme we wish to emphasize here is that despite the extraordinary diversity of methods and techniques in all the various proofs of Szemerédi's theorem, that there is some very strong unifying themes among all these approaches, such as the exploitation of a dichotomy between randomness and structure, and this chapter intends to highlight such common themes between all the known proofs of Szemerédi's theorem and related results.

11.1 Gowers uniformity norms

As in the previous chapter, it is convenient to attack Szemerédi's theorem by studying the k-linear form

$$\Lambda_k(f_0, \ldots, f_{k-1}) := \mathbf{E}_{x,r \in Z} f_0(x) f_1(x+r) \cdots f_{k-1}(x+(k-1)r)$$

defined for any finite additive group Z and any functions $f_0, \ldots, f_{k-1} : Z \to \mathbf{C}$. Thus, for instance, $\Lambda_k(1_A, \ldots, 1_A)$ is at least as large as $\mathbf{P}(A)/|Z|$, and will be larger if and only if A contains an arithmetic progression of length k with non-zero step; note that such progressions are proper if $|Z|$ is coprime to $k!$. This form of course generalizes the form $\Lambda_3(f, g, h)$ which featured prominently in the previous chapter.

Just as Varnavides' theorem (Theorem 10.9) is equivalent to Roth's theorem (Theorem 10.8), Szemerédi's theorem is equivalent to the following.

Theorem 11.1 (Szemerédi's theorem again) *Let $k \geq 3$, let Z be a finite cyclic group of prime order $|Z| \geq k$, and let $f : Z \to \mathbf{R}^+$ be a non-negative function* *which is not identically zero, and obeys the bound* $0 \leq f(x) \leq 1$ *for all* $x \in Z$*; then*

$$\Lambda_k(f, \ldots, f) = \Omega_{k,\mathbf{E}_Z(f)}(1).$$

In fact, this theorem is valid for all finite abelian groups, not just the cyclic group, as we shall see in Section 11.6.

Thus one strategy to prove Szemerédi's theorem is to obtain good bounds for quantities of the form $\Lambda_k(f_0, \ldots, f_{k-1})$ for various choices of $f_0, \ldots, f_{k-1}$). This is the approach taken by both Gowers' Fourier-analytic proof and in the finitary ergodic proof (and variants of this strategy also are used in the infinitary ergodic proof and the hypergraph proof). In the previous chapter, the linear bias norm $\|f\|_{u^2(Z)}$ was used to control this quantity effectively when $k = 3$, but this norm turns out to not be appropriate for higher k (see exercises). There are higher-order generalizations $\|f\|_{u^{k-1}(Z)}$ of the linear bias norm which we will discuss later, but in the absence of any useful quadratic or higher-order analog of Plancherel's theorem, it is difficult (though not impossible, see below) to use this norm to control Λ_k. Instead, there is a related norm, the *Gowers uniformity norm* $\|f\|_{U^{k-1}(Z)}$, which is more combinatorial than Fourier-analytic in nature, but controls the form Λ_k very easily. It is defined as follows.

Definition 11.2 (Gowers uniformity norm) Let $f : Z \to \mathbf{C}$ and $d \geq 1$. Then the *Gowers uniformity norm* $\|f\|_{U^d(Z)}$ of order d is defined recursively by[1]

$$\|f\|_{U^1(Z)} := |\mathbf{E}_Z(f)|; \quad \|f\|_{U^{d+1}(Z)} := \left(\mathbf{E}_{h\in Z}\|T^h f \overline{f}\|_{U^d(Z)}^{2^d}\right)^{1/2^{d+1}}$$

for all $d \geq 1$, where $T^h f(x) := f(x+h)$ is the shift of f by h.

Thus for instance we have

$$\begin{aligned}\|f\|_{U^2(Z)} &= (\mathbf{E}_{h\in Z}|\mathbf{E}_Z(T^h f\overline{f})|^2)^{1/4}\\ &= (\mathbf{E}_{x,h_1,h_2\in Z} f(x+h_1+h_2)\overline{f(x+h_1)f(x+h_2)}f(x))^{1/4}\end{aligned}$$

and more generally

$$\|f\|_{U^d(Z)} = \left(\mathbf{E}_{x,h_1,\ldots,h_d\in Z}\prod_{\omega\in\{0,1\}^d}\mathcal{C}^{|\omega|}f(x+\omega\cdot h)\right)^{1/2^d} \tag{11.1}$$

where $\mathcal{C}f := \overline{f}$ is the conjugation operator, $\omega = (\omega_1, \ldots, \omega_d)$, $h := (h_1, \ldots, h_d)$, and $|\omega| := \omega_1 + \cdots + \omega_d$. In the case of an indicator function $f = 1_A$, we have

$$\|1_A\|_{U^d(Z)}^{2^d} = \mathbf{P}_{x,h_1,\ldots,h_d\in Z}(x + [0,1]^d\cdot(h_1,\ldots,h_d) \subset A); \tag{11.2}$$

[1] It would also be consistent to define $\|f\|_{U^0(Z)} = \mathbf{E}_Z(f)$, but this quantity is signed and thus is too pathological to be called a norm.

thus $\|1_A\|_{U^d(Z)}$ is a normalized measure of how many d-dimensional cubes are contained in A. In particular, we have the identity

$$\|1_A\|_{U^2(Z)}^4 = \mathbf{E}(A, A)/|Z|^3$$

which relates the U^2 norm of 1_A to the additive energy of A, defined in Definition 2.8.

At first glance, the Gowers uniformity norm $\|f\|_{U^{k-1}(Z)}$ norm looks even more complicated than the expression $\Lambda_k(f, \ldots, f)$ which it is meant to control, but as we shall see it has a significantly better structure which makes it more amenable to analysis.

In the $d = 2$ case the Gowers uniformity norm is also related to the Fourier transform by the simple formula

$$\|f\|_{U^2(Z)} = \|\hat{f}\|_{l^4(Z)}; \tag{11.3}$$

we leave the verification of this identity as an exercise. (Compare also with (4.18).) In particular this shows that the $U^2(Z)$ norm is indeed a norm. It turns out that the higher $U^d(Z)$ norms are also norms as well. To see this it is convenient to introduce the *Gowers inner product* $\langle (f_\omega)_{\omega\in\{0,1\}^d}\rangle_{U^d(Z)}$ of 2^d functions f_ω by the formula

$$\big\langle (f_\omega)_{\omega\in\{0,1\}^d}\big\rangle_{U^d(Z)} := \mathbf{E}_{x,h_1,\ldots,h_d\in Z} \prod_{\omega\in\{0,1\}^d} \mathcal{C}^{|\omega|} f_\omega(x + \omega\cdot h).$$

Thus for instance

$$\begin{aligned}
\langle (f_0, f_1)\rangle_{U^1(Z)} &= \mathbf{E}_{x,h\in Z}\overline{f_1(x+h)}f_0(x)\\
&= \mathbf{E}_Z(f_0)\overline{\mathbf{E}_Z(f_1)}\\
\langle (f_{00}, f_{01}, f_{10}, f_{11})\rangle_{U^2(Z)} &= \mathbf{E}_{x,h_1,h_2\in Z} f_{11}(x+h_1+h_2)\overline{f_{10}(x+h_1)}\\
&\quad \overline{f_{01}(x+h_2)}f_{00}(x)\\
&= \sum_{\xi\in Z} \hat{f}_{11}(\xi)\overline{\hat{f}_{10}(\xi)\hat{f}_{01}(\xi)\hat{f}_{00}(\xi)}.
\end{aligned}$$

Furthermore we see that

$$\|f\|_{U^d(Z)} = \big\langle (f)_{\omega\in\{0,1\}^d}\big\rangle_{U^d(Z)}^{1/2^d}. \tag{11.4}$$

An application of the Cauchy–Schwarz inequality in the h_d variable gives the bound

$$\left|\big\langle (f_\omega)_{\omega\in\{0,1\}^d}\big\rangle_{U^d(Z)}\right| \leq \big\langle (f_{\omega',0})_{\omega\in\{0,1\}^d}\big\rangle_{U^d(Z)}^{1/2}\big\langle (f_{\omega',1})_{\omega\in\{0,1\}^d}\big\rangle_{U^d(Z)}^{1/2} \tag{11.5}$$

where $\omega' := (\omega_1, \ldots, \omega_{d-1}) \in \{0, 1\}^{d-1}$ is the first $d-1$ components of ω. Similarly for permutations. Applying this inequality d times one obtains

$$\left|\big\langle (f_\omega)_{\omega\in\{0,1\}^d}\big\rangle_{U^d(Z)}\right| \leq \prod_{\tilde{\omega}\in\{0,1\}^d} \left|\big\langle (f_{\tilde{\omega}})_{\omega\in\{0,1\}^d}\big\rangle_{U^d(Z)}\right|^{1/2^d}.$$

Applying (11.4), we conclude the *Gowers–Cauchy–Schwarz inequality*

$$\left|\left\langle (f_\omega)_{\omega\in\{0,1\}^d}\right\rangle_{U^d(Z)}\right| \le \prod_{\omega\in\{0,1\}^d} \|f_\omega\|_{U^d(Z)}. \tag{11.6}$$

Applying (11.6) to the special case when $f_\omega = f$ for $\omega_d = 0$ and $f_\omega = 1$ otherwise, one easily verifies the useful *monotonicity formula*

$$\|f\|_{U^{d-1}(Z)} \le \|f\|_{U^d(Z)} \tag{11.7}$$

for all $d \ge 1$. Next, from (11.4), multilinearity, and the Gowers–Cauchy–Schwarz inequality we have

$$\begin{aligned}\|f_0 + f_1\|_{U^d(Z)}^{2^d} &= \left\langle (f_0 + f_1)_{\omega\in\{0,1\}^d}\right\rangle_{U^d(Z)} \\ &= \sum_{I\subseteq\{0,1\}^d} \left\langle (f_{\mathbf{I}(\omega\in I)})_{\omega\in\{0,1\}^d}\right\rangle_{U^d(Z)} \\ &\le \sum_{I\subseteq\{0,1\}^d} \prod_{\omega\in\{0,1\}^d} \left\| f_{\mathbf{I}(\omega\in I)}\right\|_{U^d(Z)} \\ &= \prod_{\omega\in\{0,1\}^d} \left(\|f_0\|_{U^d(Z)} + \|f_1\|_{U^d(Z)}\right)\end{aligned}$$

from which we deduce the *Gowers triangle inequality*

$$\|f_0 + f_1\|_{U^d(Z)} \le \|f_0\|_{U^d(Z)} + \|f_1\|_{U^d(Z)}.$$

This argument should be compared with the standard derivation of the Hilbert space triangle inequality from the Hilbert space Cauchy–Schwarz inequality. Since the Gowers norm $\|f\|_{U^d(Z)}$ is clearly non-negative and homogeneous, it is at least a semi-norm. When $d = 1$ it is not necessarily a norm (because $\|f\|_{U^1(Z)} = |\mathbf{E}_Z(f)|$ can vanish without f being identically zero), but from (11.3) and the injectivity of the Fourier transform we see that the U^2 norm, at least, is a norm, and then (11.7) implies that the higher U^d are also norms.

We now relate the Gowers uniformity norms to the forms Λ_k which are relevant to Szemerédi's theorem. It is convenient to introduce the following notation: we use $\mathbf{b}(x_1, \ldots, x_n)$ to denote any function of n variables $x_1, \ldots, x_n$ that is bounded in magnitude by 1. As with the $O()$ notation, the exact function used in the $\mathbf{b}()$ notation will vary from case to case. This notation is useful whenever dealing with complicated multilinear expressions involving an interesting function f, and several other less interesting functions whose only important features are their boundedness and the precise set of variables that they depend on; the $\mathbf{b}()$ notation can then be used to conceal the uninteresting functions and focus attention on the important terms in the expression.

We begin with a simple but very useful lemma, which controls the correlations of several functions f_a with an arbitrary bounded function $\mathbf{b}(a)$, in terms of the correlations of f_a with (shifts of) itself.

Lemma 11.3 (Van der Corput lemma) *Let Z be a finite additive group, and let A be a non-empty set. For each $a \in A$ let $f_a : Z \to \mathbf{C}$ be a function. Then we have*

$$|\mathbf{E}_Z(\mathbf{E}_{a\in A}\mathbf{b}(a) f_a)| \leq |\mathbf{E}_{a\in A, h\in Z}\mathbf{E}_Z(T^h f_a \overline{f_a})|^{1/2}.$$

Proof From the triangle inequality followed by Cauchy–Schwarz we have

$$\begin{aligned} |\mathbf{E}_Z(\mathbf{E}_{a\in A}\mathbf{b}(a) f_a)| &\leq \mathbf{E}_{a\in A}|\mathbf{E}_Z(f_a)| \\ &\leq (\mathbf{E}_{a\in A}|\mathbf{E}_Z(f_a)|^2)^{1/2} \\ &= (\mathbf{E}_{a\in A, x, x'\in Z} f_a(x')\overline{f_a(x)})^{1/2}. \end{aligned}$$

The claim then follows by making the substitution $x' = x + h$. □

As a consequence we have

Lemma 11.4 (Generalized von Neumann theorem) *Let Z be a finite additive group, let $k \geq 2$, and let $c_0, \ldots, c_{k-1}$ be distinct integers such that $c_i - c_j$ is coprime to $|Z|$ for all distinct i, j. Then for any function $f : Z \to \mathbf{C}$ we have*

$$|\mathbf{E}_{x,r\in Z} f(x + c_0 r)\mathbf{b}(x + c_1 r)\cdots\mathbf{b}(x + c_{k-1}r)| \leq \|f\|_{U^{k-1}(Z)}.$$

As a particular corollary, we see that if $|Z|$ is coprime to $(k-1)!$ then we have

$$|\Lambda_k(f_0, \ldots, f_{k-1})| \leq \min_{0\leq j\leq k-1} \|f_j\|_{U^{k-1}(Z)} \tag{11.8}$$

whenever $f_0, \ldots, f_{k-1} : Z \to \mathbf{C}$ are bounded in magnitude by 1. This result should be compared with Proposition 10.11.

Proof We induce on k. When $k = 2$ we observe that the map $(x, r) \mapsto (x + c_0 r, x + c_1 r)$ is bijective on $Z \times Z$ (since $|Z|$ is coprime to $c_0 - c_1$) and hence

$$|\mathbf{E}_{x,r\in Z} f(x + c_0 r)\mathbf{b}(x + c_1 r)| = |\mathbf{E}_Z(f)\mathbf{E}_Z(\mathbf{b})| \leq |\mathbf{E}_Z(f)| = \|f\|_{U^1(Z)}$$

as desired. Now suppose that $k \geq 3$ and the claim has already been proven for $k - 1$. By shifting x by $c_{k-1}r$ if necessary (and replacing c_j by $c_j - c_{k-1}$) we may take $c_{k-1} = 0$, so we can write the left-hand side as

$$|\mathbf{E}_{x\in Z}\mathbf{b}(x)\mathbf{E}_{r\in Z} f(x + c_0 r)\mathbf{b}(x + c_1 r)\cdots\mathbf{b}(x + c_{k-2}r)|.$$

Applying Lemma 11.3, we can bound this by

$$|\mathbf{E}_{x,h\in Z}\mathbf{E}_{r\in Z}(T^{c_0 h} f\overline{f})(x + c_0 r)\mathbf{b}(x + c_1 r, h)\cdots\mathbf{b}(x + c_{k-2}r, h)|^{1/2}.$$

Applying the induction hypothesis, we can bound this by

$$\left|\mathbf{E}_{h\in Z}\|T^{c_0 h} f\overline{f}\|_{U^{k-2}(Z)}\right|^{1/2}.$$

Since $c_0 = c_0 - c_{k-1}$ is coprime to $|Z|$ we can change variables and replace $c_0 h$ by h. Applying Hölder we can the bound the previous expression by

$$\left(\mathbf{E}_{h\in Z}\|T^h f\overline{f}\|_{U^{k-2}(Z)}^{2^{k-2}}\right)^{1/2^{k-1}},$$

and the claim now follows from the recusive definition of the $U^{k-1}(Z)$ norm. □

Let us informally refer to a function f as *Gowers uniform of order* $k-2$ if the quantity $\|f\|_{U^{k-1}(Z)}$ is small. It is easy to verify the bounds

$$\|f\|_{u^2(Z)} \leq \|f\|_{U^2(Z)} \leq \|f\|_{u^2(Z)}^{1/2} \tag{11.9}$$

whenever f is bounded in magnitude by 1 (see exercises), thus Gowers uniformity of order 1 is the same as linear (or Fourier) uniformity. In analogy with this, we shall refer to Gowers uniformity of order 2 as *quadratic uniformity*, Gowers uniformity of order 3 as *cubic uniformity*, and so forth. A partial explanation for this terminology can be found in Exercise 11.1.12; see also the next section.

The estimate (11.8) shows that functions which are Gowers uniform of order $k-2$ are negligible for the purposes of counting progressions of length k. One is then naturally led to the strategy of approximating an arbitrary function f by a much more structured function f, up to errors which are Gowers uniform. For instance, if one is lucky enough that $f - \mathbf{E}(f)$ is Gowers uniform of order $k-2$, then one can use (11.8) to approximate $\Lambda_k(f, \ldots, f)$ by $\Lambda_k(\mathbf{E}(f), \ldots, \mathbf{E}(f)) = \mathbf{E}(f)^k$. Of course, it is not always the case that $f - \mathbf{E}(f)$ is Gowers uniform. In such an event, it is important to understand which functions are *not* Gowers uniform, and more precisely what the *obstructions* to Gowers uniformity are. This will be the focus of the next section.

Exercises

11.1.1 Show that Theorem 11.1 is equivalent to Theorem 10.1. Also show that to prove Theorem 11.1 it suffices to do so in the special case when f is an indicator function, $f = 1_A$.

11.1.2 Let $Z = \mathbf{Z}_N$ for some prime $N > 3$, let ξ be a non-zero element of Z, and define the functions

$$\begin{aligned} f_0(x) &:= e(\xi x^2/N); \\ f_1(x) &:= e(-3\xi x^2/N); \\ f_2(x) &:= e(3\xi x^2/N); \\ f_3(x) &:= e(-\xi x^2/N). \end{aligned}$$

Show that $\Lambda_4(f_0, f_1, f_2, f_3) = 1$, but that $\|f_j\|_{u^2(Z)} = N^{-1/2}$ for $j = 0, 1, 2, 3$. This shows that there is no direct analog of Proposition 10.11.

Modify this example to show that there is no direct analog of Proposition 10.10 either. (Hint: it is simpler to construct an example in a vector space such as F_5^n, based on a quadratic hypersurface, than in a cyclic group such as $\mathbf{Z}_N$, which would require some sort of "quadratic Bohr set".)

11.1.3 Modify the proof of the Gowers triangle inequality to provide a proof of the triangle inequality for $l^{2^d}(Z)$ for $d = 1, 2, 3, \ldots$ based purely on the Cauchy–Schwarz inequality.

11.1.4 Prove (11.1) and (11.2).

11.1.5 Prove (11.3). Use (11.3) and Plancherel's theorem to prove (11.9).

11.1.6 Prove (11.5).

11.1.7 Prove (11.7).

11.1.8 For any finite additive group Z, any $f : Z \to \mathbf{C}$, and any $d \geq 1$, show that $\|\overline{f}\|_{U^d(Z)} = \|f\|_{U^d(Z)}$ and $\|\mathrm{Re}(f)\|_{U^d(Z)}, \|\mathrm{Im}(f)\|_{U^d(Z)} \leq \|f\|_{U^d(Z)}$.

11.1.9 Let $\phi : Z \to Z'$ be a Freiman isomorphism of order 2 from Z to Z'. Show that $\|f \circ \phi\|_{U^d(Z)} = \|f\|_{U^d(Z')}$ for any $d \geq 1$ and any $f : Z' \to \mathbf{C}$. In particular we have the translation invariance $\|T^h f\|_{U^d(Z)} = \|f\|_{U^d(Z)}$ for any $h \in Z$.

11.1.10 If $f : Z \to \mathbf{C}$ and $f' : \mathbf{Z}' \to \mathbf{C}$ are functions on two finite additive groups Z, Z', show that $\|f \otimes f'\|_{U^d(Z \otimes Z')} = \|f\|_{U^d(Z)} \|f'\|_{U^d(Z')}$.

11.1.11 Use (11.8) to give another proof that the $\|f\|_{U^k(Z)}$ norms are non-degenerate for $k \geq 2$, at least in the case when $|Z|$ is coprime to $(k-1)!$.

11.1.12 Let $d \geq 1$, let $F = F_p$ be a field of prime order $p > d$, and let $P : F \to F$ be a polynomial of degree exactly d with coefficients in F. Let $f : F \to \mathbf{C}$ be the function $f(x) := e(P(x)/p)$, where the map $x \mapsto x/p$ is defined from F to $\mathbf{R}/\mathbf{Z}$ in the obvious manner. Show that $\|f\|_{U^{d'}(F)} = 1$ for all $d' > d$, but that $\|f\|_{U^{d'}(F)} \leq ((d-1)/p)^{1/2^d}$ for all $1 \leq d' \leq d$; this shows that the $U^d(F)$ norms are genuinely different for each $1 \leq d < p$. Informally, we see that f is Gowers uniform of order $d-1$ or less, but is not Gowers uniform of order d or more. In particular establish the *Weyl exponential sum estimate*

$$\sum_{x \in F} e(P(x)/p) = O\big(p^{1-2^{-d}}\big).$$

Compare this with Lemma 4.14.

11.1.13 For any finite additive group Z and any $f : Z \to \mathbf{C}$, show that

$$\|f\|_{U^d(Z)} \leq \|f\|_{L^{2^d/(d+1)}(Z)}.$$

for all $d \geq 1$, and that $\lim_{d \to \infty} \|f\|_{U^d(Z)} = \|f\|_{L^\infty(Z)}$. Show that the exponent $2^d/(d+1)$ cannot be replaced by any smaller quantity. (Hint: consider a Dirac mass, or the characteristic function of a subgroup.)