that with large probability, the number of 3-edges joining any three large subsets A, B, C of $V \cup W$ is about the same for H and H', but that H and H' have very different numbers of 3-simplices. (Of course, one should quantify these vague statements precisely, for instance using Chernoff's inequality.) This shows that regularization based entirely on vertex partition will not be sufficient to easily conclude the simplex removal lemma.

11.6.5 Prove Proposition 11.33.

11.7 Arithmetic progressions in the primes

We now discuss the Green–Tao theorem, Theorem 10.7. We will not give a complete proof of this theorem here, referring the reader to the original paper [158] and to the survey articles [358], [217], [184], [153], [361] for further details. Instead we shall give a somewhat informal discussion, in particular focusing on the connections with the other arguments discussed in this chapter.

We begin by a very brief history of the problem. This result has been conjectured for some time; indeed, long progressions of primes were already studied by Lagrange and Waring in 1770. The Erdős–Turan conjecture (Conjecture 10.6), formulated in 1936, was certainly motivated in part by this problem; it implies Theorem 10.7 but is much stronger (and still open). The first significant progress on the problem was in 1939, when Van der Corput [370] used Fourier-analytic methods (but not the density increment or energy increment arguments) to establish that the primes contained infinitely many progressions of length three. A key step of the argument is to obtain good bounds for exponential sums such as $\mathbf{E}_{1 \leq n \leq N} \Lambda(n) e(\alpha n)$, where Λ is the von Mangoldt function and α is a real number (which may be close to a rational with small denominator, or far away from one). However, as discussed earlier, Fourier methods (also known as the *Hardy–Littlewood circle method* in analytic number theory) do not directly work for progressions of length 4 or higher. Progress on this problem thus became very slow. Szemerédi's theorem did not directly give any new results on the primes, as they had density zero, and even the powerful quantitative bounds of Bourgain (Theorem 10.30) for $k = 3$ and Gowers (11.23) were insufficient to attack the primes (which would require a bound roughly of the form $r_k(\mathbf{Z}_N) = o(N \log\log N / \log N)$).

Meanwhile, the methods of sieve theory were developed by analytic number theorists, in part to solve questions concerning the existence of patterns of primes such as arithmetic progressions. While these methods seem unable by themselves to count primes directly (due to the notorious *parity problem* in sieve theory, the discussion of which is beyond the scope of this book), they have proven to be enormously successful in counting *almost-primes* – products of very few primes.

For instance, it is not too hard to use sieve theory methods to show that for any given k, there are infinitely many progressions of length k, the elements of which are each the product of $O_k(1)$ prime factors. However to pass from the almost-primes to the primes remained difficult; one notable result is that of Heath-Brown [179] in 1981, who showed that there were infinitely many progressions of length 4 where three elements were prime and the fourth was the product of at most two primes. In another direction, Balog [15] in 1992 was able to find infinitely many k-tuples of primes $p_1, \ldots, p_k$ whose midpoints $(p_i + p_j)/2$ were also prime. Meanwhile, in 1996, Kohayakawa, Luczak, and Rödl [212] extended the Szemerédi regularity lemma to subgraphs of a certain type of random subgraph, and in so doing extended Roth's theorem to show that *relatively* dense subsets of a random set contained many progressions of length 3 (see Theorem 10.18). More recently, Green [147] used Fourier methods to obtain a Roth theorem for the primes, in other words showing that any subset of the primes of positive relative density contained infinitely many arithmetic progressions of length 3. This was then refined by Green and Tao [159], who showed (roughly speaking) that any dense subset of a set which was well controlled by a sieve would contain infinitely many progressions of length 3.

In [158] this type of result was extended to arbitrary k. The precise statement requires some notation.

Definition 11.34 (Pseudo-random measure) [158] A function $\nu : \mathbf{Z}_N \to \mathbf{R}^+$ is said to be *k-pseudo-random* if we have $\mathbf{E}_{\mathbf{Z}_N} \nu = 1 + o_{N\to\infty}(1)$, and more generally we have the *linear forms condition*

$$\mathbf{E}_{x_1,\ldots,x_t \in \mathbf{Z}_N} \prod_{i=1}^{m} \nu\left(\sum_{j=1}^{t} L_{ij} x_j + b_i\right) = 1 + o_{N\to\infty;k}(1)$$

whenever $0 \leq m \leq k2^{k-1}$, $t \leq 3k-4$, and $b_1, \ldots, b_m \in \mathbf{Z}_N$ are arbitrary, and L_{ij} are rational numbers with numerator and denominator of magnitude at most k, such that none of the m t-tuples $(L_{ij})_{j=1}^t$ are rational multiples of any other. Furthermore we assume the *correlation condition*

$$\mathbf{E}_{x \in \mathbf{Z}_N} \prod_{i=1}^{m} \nu(x + h_i) \leq \sum_{1 \leq i < j \leq m} \tau(h_i - h_j)$$

for all $1 \leq m \leq 2^{k-1}$ and all $h_1, \ldots, h_m \in \mathbf{Z}_N$, where $\tau : \mathbf{Z}_N \to \mathbf{R}^+$ is a function obeying the moment conditions $\mathbf{E}\tau^q = O_{q,k}(1)$ for all $1 \leq q < \infty$.

The above definition is rather complicated, but one should view these conditions as an assertion that the weight function (or "measure") ν is very randomly distributed. If we have $\nu = \frac{1}{\mathbf{P}(A)} 1_A$ for some set $A \subset \mathbf{Z}_N$, these conditions are

essentially asserting that the events $\sum_{i=1}^m L_{ij} x_j + b_i \in A$ are essentially independent of each other if the $(L_{ij})_{j=1}^t$ are not commensurate, and the events $x + h_i \in A$ are only mildly correlated to each other for generic choices of $h_1, \ldots, h_m$.

The key result in [158] then takes the Szemerédi theorem, in the form of Theorem 11.1, and generalizes it to pseudo-random measures.

Theorem 11.35 (Relative Szemerédi theorem) *Let $k \geq 3$, let $\mathbf{Z}_N$ be a finite cyclic group of large prime order N, and let $f : Z \to \mathbf{R}^+$ is a non-negative function which is not identically zero, and obeys the bounds $0 \leq f(x) \leq \nu(x)$ and $\mathbf{E}_{\mathbf{Z}_N}(f) \geq \delta > 0$ for all $x \in \mathbf{Z}_N$ and some k-pseudo-random measure ν, then*

$$\Lambda_k(f, \ldots, f) = \Omega_{k,\delta}(1) - o_{N\to\infty;k,\delta}(1).$$

This strengthening of Szemerédi's theorem allows one to detect arithmetic progressions not just in sets of positive density, but now also in sets of positive *relative* density with respect to sufficiently "pseudo-random" sets, even if the latter sets have density zero. For instance, given any set $B \subset \mathbf{Z}_N$ for which $\frac{1}{\mathbf{P}(B)} 1_B$ is k-pseudo-random, the above theorem will guarantee that $r_k(B) = o_{N\to\infty;k}(|B|)$, provided one has a mild condition such as $\mathbf{P}(B) \geq N^{-1/k}$ in order to neglect the diagonal $r = 0$ term in $\Lambda_k(f, \ldots, f)$. In particular, any subset A of B of large relative density $|A|/|B| \geq \delta$ will contain a proper arithmetic progression of length k as soon as N is sufficiently large depending on δ and k.

As it turns out, the primes P do not quite fall into the above framework, because they are unevenly distributed with respect to small residue classes (e.g. they are almost all odd), and any set B containing P for which P has positive relative density will also necessarily have some uneven distribution in small residue classes (this is ultimately due to the divergence of the Euler product $\prod_p (1 - \frac{1}{p})^{-1}$). On the other hand, pseudo-random measures are necessarily evenly distributed among such classes (see exercises). However, this can be easily fixed, by the simple trick of using the pigeonhole principle to pass to a single residue class among small divisors. More precisely, one defines $W := \prod_{p<w} p$ for some small w (e.g. $w = \log\log N$ will suffice), and replaces the primes P by the set $P_{W,b,N} = \{q \in [\varepsilon_k N, 2\varepsilon_k N] : Wp + b \in P\}$ for some b coprime to W (in fact one can use Dirichlet's theorem on distribution of primes in residue classes to take $b = 1$). Here $\varepsilon_k := 1/2^k(k+4)!$ is a small number needed for some minor technical reasons (related to the denominators of the L_{ij} in the k-pseudo-random condition). See [158], [361] for more details of this "W-trick".

It turns out that $P_{W,b,N}$ can be contained effectively in a k-pseudo-random measure. More precisely, there exists a k-pseudo-random measure $\nu : \mathbf{Z}_N \to \mathbf{R}^+$ such that $\mathbf{E}_{\mathbf{Z}_N} 1_{P_{W,b,N}} \nu = \Theta_k(1)$, and also one has the mild upper bound $\|\nu\|_{L^\infty(\mathbf{Z}_N)} = O(N^{1/k})$ (again needed to order to neglect the $r = 0$ diagonal term).

This fact, combined with Theorem 11.1, is enough to establish arithmetic progressions of length k in the primes, and even to establish the stronger result that $r_k(P \cap [1, N]) = o_{N\to\infty;k}(|P \cap [1, N]|) = o_{N\to\infty;k}(N/\log N)$. The construction of this measure relies on a version of the Selberg sieve used by Goldston and Yıldırım [134], [132], [133] (see also [363], [184], [361]); it is purely number-theoretical in nature and we do not reproduce it here. However, we do remark that ν can be thought of as being a (smoothed out) version of the normalized indicator function on the *almost-primes* $P_k = \{n : n$ is the product of $O_k(1)$ primes$\}$, or more precisely of the portion of P_k in the residue class $b \pmod W$. As mentioned earlier, modern sieve theory techniques such as the Selberg sieve are very accurate at counting correlations of almost-primes, and thus can verify the k-pseudo-randomness of ν by fairly standard arguments. In contrast, verifying the k-pseudo-randomness of a normalized counting function of the primes themselves (or of a related object such as $P_{W,b,N}$) is still beyond the reach of current technology, being roughly equivalent to the notorious *Hardy–Littlewood prime tuples conjecture*, which would imply not just the Green–Tao theorem but also the twin prime conjecture, Goldbach's conjecture, and many other difficult and unsolved problems in additive number theory. Thus one crucially needs a tool such as the relative Szemerédi theorem to bridge the gap between the almost-primes (which we understand quite well) and the primes (which are still very mysterious).

We briefly discuss the proof of Theorem 11.35. It turns out that this theorem is proven by a means very similar to that to the proof of Szemerédi's theorem outlined in Section 11.4, but now the functions involved are not bounded by 1, but are instead bounded by some k-pseudo-random measure ν. Nevertheless, it is still possible to adapt most of the arguments in that section (with the exception of the useful UAP^{k-2} norms, which do not seem to have a suitable analog in this setting). First of all one can generalize the generalized von Neumann theorem (11.8) to obtain the bound

$$|\Lambda_k(f_0, \ldots, f_{k-1})| = O_k\left(\min_{0\le j\le k-1} \|f_j\|_{U^{k-1}(\mathbf{Z}_N)}\right) + o_{N\to\infty;k}(1) \qquad (11.34)$$

whenever $f_0, \ldots, f_{k-1} : \mathbf{Z}_N \to \mathbf{R}^+$ are bounded in magnitude by $\nu + 1$. The original bound (11.8) was proven using multiple applications of the van der Corput lemma, which in turn is essentially just the Cauchy–Schwartz inequality; similarly, the bound (11.34) is also proven using several applications of the Cauchy–Schwarz inequality, the main task being to keep track of all the weights involving ν and to use the linear forms condition to ensure that after a certain point these weights can be replaced by 1 with only a negligible error. See [158] for full details.

The bound (11.34) tells us that even in the pseudo-random setting, functions which are Gowers uniform of order $k - 2$ can still be safely ignored. This opens

the way to prove Theorem 11.35 by using a Koopman–von Neumann theorem. Here, the relevant theorem is as follows.

Proposition 11.36 (Generalized Koopman–von Neumann structure theorem) *[158] Let ν be a k-pseudo-random measure, and let $f : \mathbf{Z}_N \to \mathbf{R}^+$ be such that $0 \leq f(x) \leq \nu(x)$ for all $x \in \mathbf{Z}_N$. Let $0 < \varepsilon \ll 1$ be a small parameter, and assume $N > N_0(\varepsilon)$ is sufficiently large. Then there exists a σ-algebra $\mathcal{B}$ and an exceptional set $\Omega \in \mathcal{B}$ such that:*

- (*smallness condition*)

$$\mathbf{E}(\nu 1_\Omega) = o_{N\to\infty;\varepsilon,k}(1); \tag{11.35}$$

- (ν *is uniformly distributed outside of* Ω)

$$\|(1 - 1_\Omega)\mathbf{E}(\nu - 1|\mathcal{B})\|_{L^\infty(\mathbf{Z}_N)} = o_{N\to\infty;\varepsilon,k}(1); \tag{11.36}$$

and

- (*Gowers uniformity estimate*)

$$\|(1 - 1_\Omega)(f - \mathbf{E}(f|\mathcal{B}))\|_{U^{k-1}(\mathbf{Z}_N)} \leq \varepsilon^{1/2^k}. \tag{11.37}$$

Assuming this proposition, one can now write $(1 - 1_\Omega)f = f_U + f_{U^\perp}$, where $f_U := (1 - 1_\Omega)(f - \mathbf{E}(f|\mathcal{B}))$ is Gowers uniform of order $k - 2$, and $f_{U^\perp} := (1 - 1_\Omega)\mathbf{E}(f|\mathcal{B})$ is bounded by $1 + o_{N\to\infty;\varepsilon,k}(1)$ (since $\mathbf{E}(f|\mathcal{B}) \leq 1 + \mathbf{E}(\nu - 1|\mathcal{B})$) and non-negative. Furthermore by using (11.35) one can show that $f_{U^\perp}$ almost has the same mean as f: $\mathbf{E}_{\mathbf{Z}_N} f_{U^\perp} = \mathbf{E}_{\mathbf{Z}_N} f - o_{N\to\infty;\varepsilon,k}(1)$. From the latter two facts one can use the ordinary Szemerédi theorem (Theorem 11.1) to establish that

$$\Lambda_k(f_{U^\perp}, \ldots, f_{U^\perp}) = \Omega_{k,\delta}(1) - o_{N\to\infty;k,\delta}(1).$$

Since f_U is Gowers uniform, we can easily use (11.34) to then conclude

$$\Lambda_k(f_{U^\perp} + f_U, \ldots, f_{U^\perp} + f_U) = \Omega_{k,\delta}(1) - o_{N\to\infty;k,\delta}(1)$$

and Theorem 11.35 then follows since $0 \leq f_{U^\perp} + f_U \leq f$.

It thus only remains to prove Proposition 11.36. Here we follow the energy increment strategy already used to prove Propositions 10.36, 11.18, and 11.29. The first step is the following generalization of Lemma 11.14:

Lemma 11.37 (Soft inverse theorem) *[158] Let $f : \mathbf{Z}_N \to \mathbf{C}$ be a function bounded in magnitude by $\nu + 1$, and let $F = \mathcal{D}_{k-1}(f)$ be the dual function. Then $\|F\|_{L^\infty(\mathbf{Z}_N)} \leq 2^{2^{k-1}-1} + o_{N\to\infty;k}(1)$. Furthermore, if $\|f\|_{U^{k-1}(Z)} \geq \eta$, then $|\langle f, F\rangle| \geq \eta^{2^d}$.*

The key feature here is that even though f may be unbounded (or at least very large), the dual function F is bounded quite concretely. This is a consequence of

the linear forms condition, which among other things provides a uniform bound for $\mathcal{D}_{k-1}(\nu + 1)$ and hence for $\mathcal{D}_{k-1}(f)$.

One can then run the same energy increment algorithm used in Propositions 10.36, 11.18, 11.29, to convert any lack of uniformity in the f_U term into a dual function which is then added to a σ-algebra in order to increase the energy of the $f_{U^\perp}$ term. The only difficulty with executing this strategy is to ensure that $f_{U^\perp}$ stays bounded. This is accomplished by the following somewhat technical result.

Proposition 11.38 *[158] Let ν be a k-pseudo-random measure. Let $0 < \varepsilon < 1$ and $0 < \eta < 1/2$ be parameters. Then to every function $F : \mathbf{Z}_N \to \mathbf{R}$ bounded in magnitude by $\nu + 1$, one can construct a σ-algebra $\mathcal{B}_{\varepsilon,\eta}(\mathcal{D}_{k-1}F)$ with the following property: for any $K \geq 1$ and any $F_1, \ldots, F_K : \mathbf{Z}_N \to \mathbf{R}$ functions bounded in magnitude by $\nu + 1$, if we set $\mathcal{B} := \mathcal{B}_{\varepsilon,\eta}(\mathcal{D}_{k-1}F_1) \vee \cdots \vee \mathcal{B}_{\varepsilon,\eta}(\mathcal{D}_{k-1}F_K)$, then if $\eta < \eta_0(\epsilon, K)$ is sufficiently small and $N > N_0(\epsilon, K, \eta)$ is sufficiently large we have*

$$\|\mathcal{D}_{k-1}F_j - \mathbf{E}(\mathcal{D}_{k-1}F_j|\mathcal{B})\|_{L^\infty(\mathbf{Z}_N)} \leq \varepsilon \text{ for all } 1 \leq j \leq K. \tag{11.38}$$

Furthermore there exists a set Ω which lies in $\mathcal{B}$ such that

$$\mathbf{E}_{\mathbf{Z}_N}((\nu + 1)1_\Omega) = O_{K,\varepsilon}\big(\eta^{1/2}\big) \tag{11.39}$$

and such that

$$\|(1 - 1_\Omega)\mathbf{E}(\nu - 1|\mathcal{B})\|_{L^\infty(\mathbf{Z}_N)} = O_{K,\varepsilon}\big(\eta^{1/2}\big). \tag{11.40}$$

The σ-algebras $\mathcal{B}_{\varepsilon,\eta}(\mathcal{D}_{k-1}F)$ are constructed very similarly to those in Proposition 10.38, the only real difference being that certain small atoms cause some difficulty and need to be placed in the exceptional set Ω. However these problems can be dealt with by taking η suitably small depending on K, ε, and then N suitably large depending on K, ε, η. The trickiest task is to establish (11.40). This ultimately comes down (using the Weierstrass approximation theorem as in the proof of Proposition 10.38) to establishing estimates of the form

$$\mathbf{E}((\nu - 1)\mathcal{D}_{k-1}F_1 \cdots \mathcal{D}_{k-1}F_K) = o_{N\to\infty;k,K}(1)$$

whenever $F_1, \ldots, F_K : \mathbf{Z}_N \to \mathbf{R}$ are functions bounded in magnitude by $\nu + 1$. This estimate turns out to be achievable by application of the Gowers–Cauchy–Schwarz inequality, Hölder's inequality, and both the linear forms and correlation conditions; see [158].

Finally, we apply the energy increment argument and combine Lemma 11.37 and Proposition 11.38 as in the proof of Proposition 10.36 to obtain Proposition 11.36. Actually the energy increment argument here is slightly simpler than that in Proposition 10.36 as there is no arbitrary growth function F to deal with. As such

one can use just a single loop iterative procedure rather than a double loop, which simplifies things slightly. On the other hand, the presence of the exceptional sets, and the unboundedness of several of the functions being manipulated, requires some additional care, in particular to ensure that one really does get a substantial energy increment at each stage in order to make the algorithm terminate in finite time (and to keep the quantity K appearing in Proposition 11.38 bounded by $O_\varepsilon(1)$).

Exercises

11.7.1 Suppose that one knew that $r_k(\mathbf{Z}_N) = o_{N\to\infty;k}(N \log\log N / \log N)$ for all $k \geq 3$. Derive the Green–Tao theorem as a consequence of this. (Hint: divide the primes from 1 to N into residue classes mod $P = \prod_{p<c\log N} P$ for some small absolute constant c, and use the pigeonhole principle (and Proposition 1.51) to conclude that the primes in one of these classes has density roughly $\log\log N / \log N$.)

11.7.2 Use Theorem 11.35 to prove a version of Theorem 10.18 for large cyclic groups $\mathbf{Z}_N$ and arbitrary k. (Hint: if B is a random subset of $\mathbf{Z}_N$ with expected density $\tau \geq N^{-\varepsilon}$ for some small $\varepsilon = \varepsilon_k > 0$, show using Chernoff's inequality that $\frac{1}{\tau}1_B$ is very likely to be k-pseudo-random.)

11.7.3 [158] Let $\nu : \mathbf{Z}_N \to \mathbf{R}$ be k-pseudo-random. Show that $\|\nu - 1\|_{U^{k-1}(\mathbf{Z}_N)} = o_{N\to\infty;k}(1)$. Conclude in particular that if $k \geq 3$, then one has the uniform distribution property

$$\mathbf{E}_{x\in\mathbf{Z}_N} 1_P(x)\nu(x) = \mathbf{P}_{\mathbf{Z}_N}(P) + o_{N\to\infty;k}(1)$$

for any arithmetic progression P. Thus pseudo-random measures must be evenly distributed in arithmetic progressions.

11.7.4 [158] Prove Lemma 11.37.