9

Progressions in Sparse Pseudorandom Sets

Chapter Highlights

- The Green–Tao theorem: proof strategy
- A relative Szemerédi theorem and its proof: a central ingredient in the proof of the Green–Tao theorem
- Transference principle: applying Szemerédi's theorem as a black box to the sparse pseudorandom setting
- A graph theoretic approach
- Dense model theorem: modeling a sparse set by a dense set
- Sparse triangle counting lemma

In this chapter we discuss a celebrated theorem by Green and Tao (2008) that settled a folklore conjecture about primes.

Theorem 9.0.1 (Green–Tao theorem)
The primes contain arbitrarily long arithmetic progressions.

The proof of this stunning result uses sophisticated ideas from both combinatorics and number theory. As stated in the abstract of their paper:

[T]he main new ingredient of this paper . . . is a certain transference principle. This allows us to deduce from Szemerédi's theorem that any subset of a sufficiently pseudorandom set (or measure) of positive relative density contains progressions of arbitrary length.

The main goal of this chapter is to explain what the above paragraph means. As Green (2007b) writes (emphasis in original):

Our main advance, then, lies not in our understanding of the primes but rather in what we can say about *arithmetic progressions*.

We will abstract away ingredients related to prime numbers (see Further Reading at the end of the chapter) and instead focus on the central combinatorial result: a **relative Szemerédi theorem**. We follow the graph theoretic approach by Conlon, Fox, and Zhao (2014, 2015), which simplified both the hypotheses and the proof of the relative Szemerédi theorem.

9.1 Green–Tao Theorem

In this section, we give a high-level overview of the proof strategy of the Green–Tao theorem. Recall Szemerédi's theorem:

Theorem 9.1.1 (Szemerédi's theorem)
Fix $k \geq 3$. Every k-AP-free subset of $[N]$ has size $o(N)$.

By the prime number theorem,

$$\#\{\text{primes} \leq N\} = (1 + o(1))\frac{N}{\log N}.$$

So Szemerédi's theorem does not automatically imply the Green–Tao theorem.

***Remark* 9.1.2** (Quantitative bounds)**.** It is possible that better quantitative bounds on Szemerédi's theorem might eventually imply the Green–Tao theorem based on the density of primes alone. For example, Erdős famously conjectured that any $A \subseteq \mathbb{N}$ with divergent harmonic series (i.e., $\sum_{a \in A} 1/a = \infty$) contains arbitrarily long arithmetic progressions (Conjecture 0.2.5). The current best quantitative bounds on Szemerédi's theorem for k-APs is $|A| \leq N(\log \log N)^{-c_k}$ (Gowers 2001), which are insufficient for the primes, although better bounds are known for $k = 3, 4$. More recently, Bloom and Sisask (2020) proved that for $k = 3$, $|A| \leq N(\log N)^{-1-c}$ for some constant $c > 0$, thereby implying the Green–Tao theorem for 3-APs via the density of primes alone.

We will be quite informal here in order to highlight some key ideas of the proof of the Green–Tao theorem. Fix $k \geq 3$. The idea is to embed the primes in a slightly larger "pseudorandom host set":

$$\{\text{primes}\} \subseteq \{\text{“almost primes”}\}.$$

Very roughly speaking, "almost primes" are numbers with no small prime divisors. The "almost primes" are much easier to analyze compared to the primes. Using analytic number theory (involving techniques related to the problem of *small gaps between primes*), one can construct "almost primes" satisfying the following properties.

Properties of the "almost primes":

(1) The primes occupy at least a positive constant fraction of the "almost primes":

$$\frac{\#\{\text{primes} \leq N\}}{\#\{\text{“almost primes”} \leq N\}} \geq \delta_k.$$

(2) The "almost primes" behave pseudorandomly with respect to certain pattern counts.

The next key ingredient plays a central role in the proof of the Green–Tao theorem, as mentioned at the beginning of this chapter. It will be nicer to work in $\mathbb{Z}/N\mathbb{Z}$ rather than $[N]$.

Relative Szemerédi theorem (informal). *Fix $k \geq 3$. If $S \subseteq \mathbb{Z}/N\mathbb{Z}$ satisfies certain pseudorandomness hypotheses, then every k-AP-free subset of S has size $o(|S|)$.*

Here imagine a sequence $S = S_N \subseteq \mathbb{Z}/N\mathbb{Z}$ of size $o(N)$ (or else the relative Szemerédi theorem would already follow from Szemerédi's theorem), and $|S| \geq N^{1-c_k}$ for some small constant $c_k > 0$. In the proof of the Green–Tao theorem, the set S will be the "almost primes" (so that $|S| = \Theta(N/\log N)$), subject to various other technical modifications such as the W-trick discussed in Remark 9.1.4.

The relative Szemerédi theorem and the construction of the "almost primes" together tell us that the primes contain a k-AP. It also implies the following.

Theorem 9.1.3 (Green–Tao)
Fix $k \geq 3$. If A is a k-AP-free subset of the primes, then

$$\lim_{N\to\infty} \frac{|A \cap [N]|}{|\text{Primes} \cap [N]|} = 0.$$

In other words, every subset of primes with positive *relative density* contains arbitrarily long arithmetic progressions.

***Remark* 9.1.4** (Residue biases in the primes and the W-trick)**.** There are certain local biases that get in the way of pseudorandomness for primes. For example, all primes greater than 2 are odd, all primes greater than 3 are not divisible by 3, and so on. In this way, the primes look different from a subset of positive integers where each n is included with probability $1/\log n$ independently at random.

The ***W-trick*** corrects these residue class biases. Let $w = w(N)$ be a function with $w \to \infty$ slowly as $N \to \infty$. Let $W = \prod_{p\leq w} p$ be the product of primes up to w. The W-trick tells us to only consider primes that are congruent to 1 mod W. The resulting set of "W-tricked primes" $\{n : nW + 1 \text{ is prime}\}$ does not have any bias modulo a small fixed prime. The relative Szemerédi theorem should be applied to the W-tricked primes.

We shall not dwell on the analytic number theoretic arguments here. See Further Reading at the end of the chapter for references. For example, Conlon, Fox, and Zhao (2014, Sections 8 and 9) gives an exposition of the construction of the "almost primes" and the proofs of its properties. The goal of the rest of the chapter is to state and prove the relative Szemerédi theorem.

9.2 Relative Szemerédi Theorem

In this section, we formulate a relative Szemerédi theorem. For concreteness, we mostly discuss 3-APs, though everything generalizes to k-APs straightforwardly.

Recall Roth's theorem:

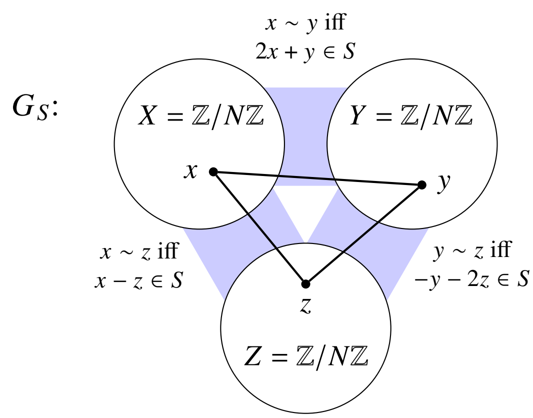
Theorem 9.2.1 (Roth's theorem)
Every 3-AP-free subset of $\mathbb{Z}/N\mathbb{Z}$ has size $o(N)$.

We would like to formulate a result of the following form, where $\mathbb{Z}/N\mathbb{Z}$ is replaced by a **sparse pseudorandom host set** S.

Relative Roth theorem (informal). *If $S \subseteq \mathbb{Z}/N\mathbb{Z}$ satisfies certain pseudorandomness conditions, then every 3-AP-free subset of S has size $o(|S|)$.*

In what sense should S behave pseudorandomly? It will be easiest to explain the pseudorandom hypothesis using a graph.

Consider the following construction of a graph G_S that we saw in Chapter 6 (in particular Sections 2.4 and 2.10).



Here G_S is a tripartite graph with vertex sets X, Y, Z, each being a copy of $\mathbb{Z}/N\mathbb{Z}$. Its edges are:

- $(x, y) \in X \times Y$ whenever $2x + y \in S$;
- $(x, z) \in X \times Z$ whenever $x - z \in S$;
- $(y, z) \in Y \times Z$ whenever $-y - 2z \in S$.

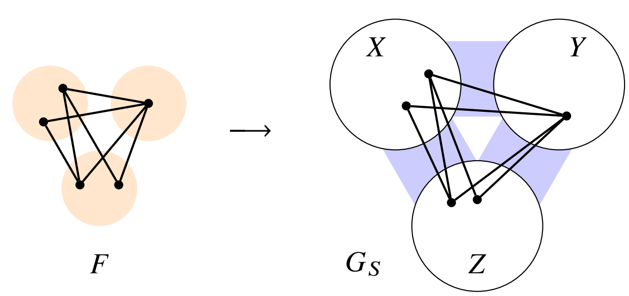
This graph G_S is designed so that $(x, y, z) \in X \times Y \times Z$ is a triangle if and only if

$$2x + y, \quad x - z, \quad -y - 2z \in S.$$

Note that these three terms form a 3-AP with common difference $-x - y - z$. So the triangles in G_S precisely correspond to 3-APs in S (it is an N-to-1 correspondence).

The following definition is a variation of homomorphism density from Section 4.3.

Definition 9.2.2 (F-density)
Let F and G be tripartite graphs with three labeled parts. Define ***F-density in G***, denoted $\boldsymbol{t(F, G)}$, to be the probability that a random map $V(F) \to V(G)$ is a graph homomorphism $F \to G$, where each vertex in the first vertex part of F is sent to a uniform vertex of the first vertex part of G, and likewise with the second and third parts, all independently.

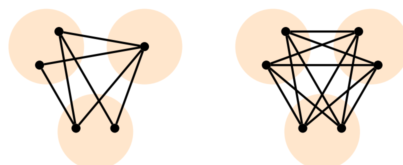


Now we define the desired pseudorandomness hypotheses on $S \subseteq \mathbb{Z}/N\mathbb{Z}$, which says that the associated graph G_S has certain subgraph counts close to random.

Definition 9.2.3 (3-linear forms condition)
We say that $S \subseteq \mathbb{Z}/N\mathbb{Z}$ satisfies the ***3-linear forms condition with tolerance ε*** if, setting $p = |S|/N$, one has

$$(1-\varepsilon)p^{e(F)} \le t(F, G_S) \le (1+\varepsilon)p^{e(F)} \qquad \text{whenever } F \subseteq K_{2,2,2}.$$

(Here $F \subseteq K_{2,2,2}$ means that is a subgraph of the labeled tripartite graph $K_{2,2,2}$; an example is illustrated below.)



$F \quad \subseteq \quad K_{2,2,2}$

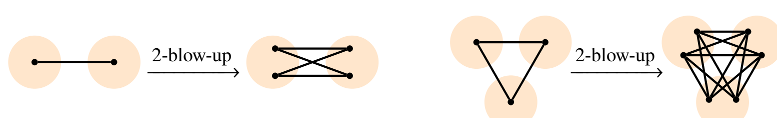
In other words, comparing the graph G_S to a random tripartite graph with the same edge density p, these two graphs have approximately the same F-density whenever $F \subseteq K_{2,2,2}$.

Alternatively, we can state the 3-linear forms condition explicitly without referring to graphs. This is done by expanding the definition of G_S. Let $x_0, x_1, y_0, y_1, z_0, z_1 \in \mathbb{Z}/N\mathbb{Z}$ be chosen independently and uniformly at random. Then $S \subseteq \mathbb{Z}/N\mathbb{Z}$ with $|S| = pN$ satisfies the 3-linear forms condition with tolerance ε if the probability that

$$\left\{\begin{array}{lll} -y_0 - 2z_0, & x_0 - z_0, & 2x_0 + y_0, \\ -y_1 - 2z_0, & x_1 - z_0, & 2x_1 + y_0, \\ -y_0 - 2z_1, & x_0 - z_1, & 2x_0 + y_1, \\ -y_1 - 2z_1, & x_1 - z_1, & 2x_1 + y_1 \end{array}\right\} \subseteq S$$

lies in the interval $(1 \pm \varepsilon)p^{12}$, and furthermore the same holds if we erase any subset of the above 12 linear forms and also change the "12" in p^{12} to the number of linear forms remaining.

***Remark* 9.2.4.** This $K_{2,2,2}$ condition is reminiscent of the C_4-count condition for the quasirandom graph in Theorem 3.1.1 by Chung, Graham, and Wilson (1989). Just as how $C_4 = K_{2,2}$ is a 2-blow-up of a single edge, $K_{2,2,2}$ is a 2-blow-up of a triangle.



The 3-linear forms condition can be viewed as a "second moment" condition with respect to triangles. It is needed in the proof of the sparse triangle counting lemma later.

We are now ready to state a precise formulation of the relative Roth theorem.

Theorem 9.2.5 (Relative Roth theorem)
For every $\delta > 0$, there exist $\varepsilon > 0$ and N_0 so that for all odd $N \geq N_0$, if $S \subseteq \mathbb{Z}/N\mathbb{Z}$ satisfies the 3-linear forms condition with tolerance ε, then every 3-AP-free subset of S has size less than $\delta |S|$.

To extend these definitions and results to k-APs, we set up a $(k-1)$-uniform hypergraph. We use a procedure similar to the deduction of Szemerédi's theorem from the hypergraph removal lemma in Section 2.10.

Let us illustrate it first for 4-APs. We say that $S \subseteq \mathbb{Z}/N\mathbb{Z}$ satisfies the ***4-linear forms condition with tolerance ε*** if given random $w_0, w_1, x_0, x_1, y_0, y_1, z_0, z_1 \in \mathbb{Z}/N\mathbb{Z}$ (independent and uniform as always), the probability that

$$\left\{\begin{array}{llll} 3w_0 + 2x_0 + y_0, & 2w_0 + x_0 - z_0, & w_0 - y_0 - 2z_0, & -x_0 - 2y_0 - 3z_0, \\ 3w_0 + 2x_0 + y_1, & 2w_0 + x_0 - z_1, & w_0 - y_0 - 2z_1, & -x_0 - 2y_0 - 3z_1, \\ 3w_0 + 2x_1 + y_0, & 2w_0 + x_1 - z_0, & w_0 - y_1 - 2z_0, & -x_0 - 2y_1 - 3z_0, \\ 3w_0 + 2x_1 + y_1, & 2w_0 + x_1 - z_1, & w_0 - y_1 - 2z_1, & -x_0 - 2y_1 - 3z_1, \\ 3w_1 + 2x_0 + y_0, & 2w_1 + x_0 - z_0, & w_1 - y_0 - 2z_0, & -x_1 - 2y_0 - 3z_0, \\ 3w_1 + 2x_0 + y_1, & 2w_1 + x_0 - z_1, & w_1 - y_0 - 2z_1, & -x_1 - 2y_0 - 3z_1, \\ 3w_1 + 2x_1 + y_0, & 2w_1 + x_1 - z_0, & w_1 - y_1 - 2z_0, & -x_1 - 2y_1 - 3z_0, \\ 3w_1 + 2x_1 + y_1, & 2w_1 + x_1 - z_1, & w_1 - y_1 - 2z_1, & -x_1 - 2y_1 - 3z_1 \end{array}\right\} \subseteq S$$

lies within the interval $(1 \pm \varepsilon)p^{32}$, and furthermore the same is true if we erase any subset of the above 32 factors and replace the "32" in p^{32} by the number of linear forms remaining.

Here is the statement for k-APs. (You may wish to skip it and simply imagine how it should generalize based on the above examples.)

Definition 9.2.6 (k-linear forms condition)
For each $1 \leq r \leq k$, let

$$L_r(x_1, \ldots, x_k) = kx_1 + (k-1)x_2 + \cdots + x_k - r(x_1 + \cdots + x_k).$$

We say that $S \subseteq \mathbb{Z}/N\mathbb{Z}$ satisfies the ***k-linear forms condition with tolerance ε*** if for every $R \subseteq [k] \times \{0,1\}^k$, with each variable $x_{i,j} \in \mathbb{Z}/N\mathbb{Z}$ chosen independently and uniformly at random, the probability that

$$L_r(x_{1,j_1}, \ldots, x_{k,j_k}) \in S \quad \text{for all } (r, j_1, \ldots, j_k) \in R$$

lies within the interval $(1 \pm \varepsilon)p^{|R|}$.

Theorem 9.2.7 (Relative Szemerédi theorem)
For every $k \geq 3$ and $\delta > 0$, there exist $\varepsilon > 0$ and N_0 so that for all $N \geq N_0$ coprime to $(k-1)!$, if $S \subseteq \mathbb{Z}/N\mathbb{Z}$ satisfies the k-linear forms condition with tolerance ε, then every k-AP-free subset of S has size less than $\delta |S|$.

***Remark* 9.2.8** (History)**.** The above formulations of relative Roth and Szemerédi theorems are due to Conlon, Fox, and Zhao (2015). The original approach by Green and Tao (2008) required in addition another technical hypothesis on S known as the "correlation condition," which is no longer needed.

***Remark* 9.2.9 (Szemerédi's theorem in a random set).** Instead of a pseudorandom host set S, what happens if S is a random subset of $\mathbb{Z}/N\mathbb{Z}$ obtained by keeping each element with probability $p = p_N \to 0$ as $N \to \infty$? A second moment argument shows that, provided that p_N tends to zero sufficiently slowly, the random set S indeed satisfies the k-linear forms condition (see Exercise 9.2.11 below). However, this argument is rather lossy. The following sharp result was proved independently by Conlon and Gowers (2016) and Schacht (2016). In the statement below, there is no substantive difference between $[N]$ and $\mathbb{Z}/N\mathbb{Z}$.

Theorem 9.2.10 (Szemerédi's theorem in a random set)

For every $k \geq 3$ and $\delta > 0$, there is some C such that as long as $p > CN^{-1/(k-1)}$, with probability approaching 1 as $N \to \infty$, given a random $S \subseteq [N]$ where every element is included independently with probability p, every k-AP-free subset of S has size at most $\delta\,|S|$.

The threshold $CN^{-1/(k-1)}$ is optimal up to the constant C. Indeed, the expected number of k-APs in S is $O(p^k N^2)$, which is less than half of $\mathbb{E}\,|S| = pN$ if $p < cN^{-1/(k-1)}$ for a sufficiently small constant $c > 0$. One can delete from S an element from each k-AP contained in S. So with high probability, this process deletes at most half of S, and the remaining subset of S is k-AP-free.

The **hypergraph container method** gives another proof of the above result, plus much more (Balogh, Morris, and Samotij 2015; Saxton and Thomason 2015). See the survey *The method of hypergraph containers* by Balogh, Morris, and Samotij (2018) for more on this topic.

Exercise 9.2.11 (Random sets and the linear forms condition). Let $S \subseteq \mathbb{Z}/N\mathbb{Z}$ be a random set where every element of $\mathbb{Z}/N\mathbb{Z}$ is included in S independently with probability p.

Prove that there is some $c > 0$ so that for every $\varepsilon > 0$ there is some $C > 0$ so that as long as $p > CN^{-c}$ and N is large enough, with probability at least $1 - \varepsilon$, S satisfies the 3-linear forms condition with tolerance ε . What is the optimal c?

Hint: Use the second moment method; see Alon and Spencer (2016, Chapter 4).

9.3 Transference Principle

To prove the relative Szemerédi theorem, we shall assume Szemerédi's theorem and apply it as a black box to the sparse pseudorandom setting. It may be surprising that we can apply Szemerédi's theorem this way. Green and Tao developed a method known as a ***transference principle*** for bringing Szemerédi's theorem to the sparse pseudorandom setting . The idea also appeared earlier in the work of Green (2005b) establishing Roth's theorem in the primes. The transference principle is an influential idea, and it can be applied to other extremal problems in combinatorics.

Let us sketch the outline of the proof of the relative Szemerédi theorem. We are given

$$A \subseteq S \qquad \text{with } |A| \geq \delta\,|S|\,.$$

Here $S \subseteq \mathbb{Z}/N\mathbb{Z}$ is a sparse pseudorandom set satisfying the k-linear forms condition.

Step 1. Approximate A by a dense model.

We will prove a **dense model theorem** that produces a "dense model" B of A. In particular, the density of B in $\mathbb{Z}/N\mathbb{Z}$ is similar to the relative density of A in S:

$$\frac{|B|}{N} \approx \frac{|A|}{|S|} \geq \delta.$$

And furthermore, B will be close to A with respect to a "cut norm" derived from the graphon cut norm (see Chapter 4 on graph limits). Recall that the graphon cut norm is closely linked to ε-regularity from the regularity lemma (Chapter 2) and the discrepancy condition **DISC** from quasirandom graphs (Chapter 3).

Step 2. Count k-APs in A and B.

We will prove a **sparse counting lemma** to show that the number of k-APs in A is similar to the number of k-APs in B, after an appropriate density normalization. In other words, setting $p = |S|/N$ for the normalizing density, we will show

$$|\{k\text{-APs in } A\}| \approx p^k \, |\{k\text{-APs in } B\}| \, .$$

Szemerédi's theorem says that every subset of $[N]$ with size $\geq \delta N$ contains a k-AP (provided that N is sufficiently large given the constant $\delta > 0$). In fact, we can bootstrap Szemerédi's theorem to show that a subset of $[N]$ with size $\geq \delta N$ must contain lots of k-APs. The deduction uses a sampling argument and is attributed to Varnavides (1959). (This was Exercise 1.3.7 from Section 1.3 on supersaturation.)

Theorem 9.3.1 (Szemerédi's theorem, counting version)
For every $\delta > 0$, there exists $c > 0$ and N_0 such that for every $N \geq N_0$, every subset of $\mathbb{Z}/N\mathbb{Z}$ with $\geq \delta N$ elements contains $\geq cN^2$ k-APs.

Since the "dense model" B has size $\geq \delta N/2$, by the counting version of Szemerédi's theorem, B has $\gtrsim_\delta N^2$ k-APs, and hence A has $\gtrsim_\delta p^k N^2$ k-APs by the sparse counting lemma. So in particular, A cannot be k-AP-free. This finishes the proof sketch of the relative Szemerédi theorem.

Now that we have seen the above outline, it remains to formulate and prove:

- a dense model theorem, and
- a sparse counting lemma.

We will focus on explaining the 3-AP case (i.e., relative Roth theorem) in the rest of this chapter. The 3-AP setting is notationally simpler than that of k-AP. It is straightforward to generalize the 3-AP proof to k-APs following the $(k-1)$-uniform hypergraph setup discussed in the previous section.

9.4 Dense Model Theorem

In this section, Γ is any finite abelian group. We will only need the case $\Gamma = \mathbb{Z}/N\mathbb{Z}$ in subsequent sections.

Given $f : \Gamma \to \mathbb{R}$, we define the following "cut norm" similar to the cut norm from graph limits (Chapter 4):

$$\|f\|_{\square} := \sup_{A,B\subseteq\Gamma} \left|\mathbb{E}_{x,y\in\Gamma}[f(x+y)1_A(x)1_B(y)]\right|.$$

This is essentially the graphon cut norm applied to the function $\Gamma \times \Gamma \to \mathbb{R}$ given by $(x, y) \mapsto f(x+y)$.

As should be expected from the equivalence of **DISC** and **EIG** for quasirandom Cayley graphs (Theorem 3.5.3), having small cut norm is equivalent to being Fourier uniform.

Exercise 9.4.1. Show that for all $f : \Gamma \to \mathbb{R}$,

$$c\|\widehat{f}\|_\infty \le \|f\|_\square \le \|\widehat{f}\|_\infty,$$

where c is some absolute constant (not depending on Γ or f).

***Remark* 9.4.2** (Generalizations to k-APs)**.** The above definition is tailored to 3-APs. For 4-APs, we should define the corresponding norm of f as

$$\sup_{A,B,C\subseteq\Gamma\times\Gamma} \left|\mathbb{E}_{x,y,z\in\Gamma}[f(x+y+z)1_A(x,y)1_B(x,z)1_C(y,z)]\right|.$$

(The more obvious guess of using $1_A(x)1_B(y)1_C(z)$ instead of the above turns out to be insufficient for proving the relative Szemerédi theorem. A related issue in the context of hypergraph regularity was discussed in Section 2.11.) The generalization to k-APs is straightforward. However, for $k \ge 4$, the above norm is no longer equivalent to Fourier uniformity. This is why we study $\|f\|_\square$ norm instead of $\|\widehat{f}\|_\infty$ in this section.

Informally, the main result of this section says that if a sparse set S is close to random in normalized cut norm, then every subset $A \subseteq S$ can be approximated by some dense $B \subseteq \mathbb{Z}/N\mathbb{Z}$ in normalized cut norm.

Theorem 9.4.3 (Dense model theorem)

For every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. For every finite abelian group Γ and sets $A \subseteq S \subseteq \Gamma$ such that, setting $p = |S|/|\Gamma|$,

$$\|1_S - p\|_\square \le \delta p,$$

there exists $g : \Gamma \to [0, 1]$ such that

$$\|1_A - pg\|_\square \le \varepsilon p.$$

***Remark* 9.4.4** (3-linear forms condition implies small cut norm)**.** The cut norm hypothesis is weaker than the 3-linear forms condition, as can be proved by two applications of the Cauchy–Schwarz inequality (for example, see the proof of Lemma 9.5.2 in the next section). In short, $\|\nu - 1\|_\square^4 \le t(K_{2,2}, \nu - 1)$.

***Remark* 9.4.5** (Set instead of function)**.** We can replace the function g by a random set $B \subseteq \Gamma$ where each $x \in \Gamma$ is included in B with probability $g(x)$. By standard concentration bounds, changing g to B induces a negligible effect on ε if Γ is large enough. It is important here that $g(x) \in [0, 1]$ for all $x \in \Gamma$.

So the above theorem says, given a sparse pseudorandom host set S, any subset of S can be modeled by a dense set B that is close to A with respect to the normalized cut norm.

It will be more natural to prove the above theorem a bit more generally where sets $A \subseteq S \subseteq \Gamma$ are replaced by functional analogs. Since these are sparse sets, we should scale indicator functions as follows:

$$f = p^{-1} 1_A \quad \text{and} \quad \nu = p^{-1} 1_S.$$

Then $f \le \nu$ pointwise. Note that f and ν take values in $[0, p^{-1}]$, unlike g, which takes values in $[0, 1]$. The normalization is such that $\mathbb{E}\nu = 1$. Here is the main result of this section.

Theorem 9.4.6 (Dense model theorem)

For every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. For every finite abelian group Γ and functions $f, \nu \colon \Gamma \to [0, \infty)$ satisfying

$$\|\nu - 1\|_\square \le \delta$$

and

$$f \le \nu \quad \text{pointwise},$$

there exists a function $g \colon \Gamma \to [0, 1]$ such that

$$\|f - g\|_\square \le \varepsilon.$$

The rest of this section is devoted to proving the above theorem. First, we reformulate the cut norm using convex geometry.

Let Φ denote the set of all functions $\Gamma \to \mathbb{R}$ that can be written as a convex combination of convolutions of the form $1_A * 1_B$ or $-1_A * 1_B$, where $A, B \subseteq \Gamma$. Equivalently,

$$\Phi = \text{ConvexHull}\left(\{1_A * 1_B : A, B \subseteq \Gamma\} \cup \{-1_A * 1_B : A, B \subseteq \Gamma\}\right).$$

Note that Φ is a centrally symmetric convex set of functions $\Gamma \to \mathbb{R}$.

Lemma 9.4.7 (Multiplicative closure)

The set Φ is closed under pointwise multiplication; that is, if $\varphi, \varphi' \in \Phi$, then $\varphi\varphi' \in \Phi$.

Proof. Given $A, B, C, D \subseteq \Gamma$, we have

$$\begin{aligned}
(1_A * 1_B)(x)(1_C * 1_D)(x) &= \mathbb{E}_{a,b,c,d:a+b=c+d=x} 1_A(a) 1_B(b) 1_C(c) 1_D(d) \\
&= \mathbb{E}_{a,b,s:a+b=x} 1_A(a) 1_B(b) 1_C(a+s) 1_D(b-s) \\
&= \mathbb{E}_s \mathbb{E}_{a,b:a+b=x} 1_{A\cap(C-s)}(a) 1_{B\cap(D+s)}(b). \\
&= \mathbb{E}_s (1_{A\cap(C-s)} * 1_{B\cap(D+s)})(x).
\end{aligned}$$

Thus the pointwise product of $1_A * 1_B$ and $1_C * 1_D$ lies in Φ since it is an average of various functions of the form $1_S * 1_T$. Since Φ is the convex hull of functions of the form $1_A * 1_B$ and $-1_A * 1_B$, Φ is closed under pointwise multiplication. $\square$

Given $f, g \colon \Gamma \to \mathbb{R}$, define their inner product by

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle := \mathbb{E}_{x\in\Gamma} f(x) g(x).$$

Since

$$\mathbb{E}_{x,y\in\Gamma} f(x+y) 1_A(x) 1_B(y) = \langle f, 1_A * 1_B \rangle,$$

we have

$$\|f\|_\square = \sup_{A,B\subseteq\Gamma} |\langle f, 1_A * 1_B\rangle| = \sup_{\varphi\in\Phi} \langle f,\varphi\rangle .$$

Since Φ is a centrally symmetric convex body, $\|\ \|_\square$ is indeed a norm. Its dual norm is thus given by, for any nonzero $\psi\colon \Gamma\to\mathbb{R}$,

$$\|\psi\|_\square^* = \sup_{\substack{f\colon\Gamma\to\mathbb{R}\\ \|f\|_\square\le 1}} \langle f,\psi\rangle = \sup\left\{r\in\mathbb{R} : r^{-1}\psi\in\Phi\right\} .$$

In other words, Φ is the unit ball for $\|\ \|_\square^*$ norm. The following inequality holds for all $f,\psi\colon\Gamma\to\mathbb{R}$:

$$\langle f,\psi\rangle \le \|f\|_\square \|\psi\|_\square^* .$$

Lemma 9.4.8 (Submultiplicativity of the dual cut norm)
The norm $\|\cdot\|_\square^*$ is submultiplicative; that is, for all $\psi,\psi'\colon\Gamma\to\mathbb{R}$,

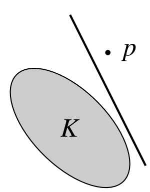
$$\|\psi\psi'\|_\square^* \le \|\psi\|_\square^* \|\psi'\|_\square^* .$$

Proof. The inequality is not affected if we multiply ψ and ψ' each by a constant. So we can assume that $\|\psi\|_\square^* = \|\psi'\|_\square^* = 1$. Then $\psi,\psi'\in\Phi$. Hence $\psi\psi'\in\Phi$ by Lemma 9.4.7. This implies that $\|\psi\psi'\|_\square' \le 1$. □

We need two classical results from analysis and convex geometry.

Theorem 9.4.9 (Weierstrass polynomial approximation theorem)
Let $a,b\in\mathbb{R}$ and $\varepsilon>0$. Let $F\colon[a,b]\to\mathbb{R}$ be a continuous function. Then there exists a polynomial P such that $|F(t)-P(t)|\le\varepsilon$ for all $t\in[a,b]$.

Theorem 9.4.10 (Separating hyperplane theorem)
Given a closed convex set $K\subseteq\mathbb{R}^n$ and a point $p\notin K$, there exists a hyperplane separating K and p.



Proof idea of the dense model theorem. If no $g\colon\Gamma\to[0,1]$ satisfies $\|f-g\|_\square\le\varepsilon$, then f does not lie in the convex set containing all functions of the form $g+g'$ where $g\colon\Gamma\to[0,1]$ and $\|g'\|_\square\le\varepsilon$. The separating hyperplane theorem then gives us a function ψ so that $\langle f,\psi\rangle>1$ and $\langle g+g',\psi\rangle\le 1$ for all such g,g'. (It helps to pretend a bit of extra slack here, say $\langle f,\psi\rangle>1+\varepsilon$.) Using the Weierstrass polynomial approximation theorem, choose a polynomial $P(t)$ so that $P(t)\approx\max\{0,t\}$ pointwise for all $|t|\le\|\psi\|_\square^* = O_\varepsilon(1)$. Writing $\psi_+(x)=\max\{0,\psi(x)\}$ for the positive part of ψ, we have

$$\langle f,\psi\rangle \le \langle f,\psi_+\rangle \le \langle \nu,\psi_+\rangle \approx \langle \nu, P\psi\rangle = \langle \nu-1, P\psi\rangle + \langle 1, P\psi\rangle .$$

We can show that $\|\psi\|_\square^* = O_\varepsilon(1)$. As P is a polynomial, by the triangle inequality and the submultiplicativity of $\| \ \|_\square^*$, we find that $\|P\psi\|_\square^* = O_\varepsilon(1)$. And so

$$\langle \nu - 1, P\psi \rangle \le \|\nu - 1\|_\square \|P\psi\|_\square^* \le \delta \|P\psi\|_\square^*$$

can be made arbitrarily small by making δ small. We also have $\langle 1, P\psi\rangle \approx \langle 1, \psi_+\rangle$, which is at most around 1. Together, we see that $\langle f, \psi\rangle$ is at most around 1, which would contradict $\langle f, \psi\rangle > 1$ from earlier (assuming enough slack). ∎

Proof of the dense model theorem (Theorem 9.4.6). We will show that the conclusion holds with $\delta > 0$ chosen to be sufficiently small as a function of ε. We may assume that $0 < \varepsilon < 1/2$. We will prove the existence of a function $g\colon \Gamma \to [0, 1 + \varepsilon/2]$ such that $\|f - g\|_\square \le \varepsilon/2$. (To obtain the function $\Gamma \to [0,1]$ in the theorem, we can replace g by $\min\{g, 1\}$.)

We are trying to prove that one can write f as $g + g'$ with

$$g \in K := \left\{\text{functions } \Gamma \to [0, 1 + \tfrac{\varepsilon}{2}]\right\}$$

and

$$g' \in K' := \left\{\text{functions } \Gamma \to \mathbb{R} \text{ with } \|\cdot\|_\square \le \tfrac{\varepsilon}{2}\right\}.$$

We can view the sets K and K' as convex bodies (both containing the origin) in the space of all functions $\Gamma \to \mathbb{R}$. Our goal is to show that $f \in K + K'$.

Let us assume the contrary. By the separating hyperplane theorem applied to $f \notin K + K'$, there exists a function $\psi\colon \Gamma \to \mathbb{R}$ (which is a normal vector to the separating hyperplane) such that

(a) $\langle f, \psi\rangle > 1$, and
(b) $\langle g + g', \psi\rangle \le 1$ for all $g \in K$ and $g' \in K'$

Taking $g = (1 + \frac{\varepsilon}{2})1_{\psi \ge 0}$ and $g' = 0$ in (b), we have

$$\langle 1, \psi_+\rangle \le \frac{1}{1 + \varepsilon/2}. \tag{9.1}$$

Here we write ψ_+ for the function $\psi_+(x) := \max\{\psi(x), 0\}$.

On the other hand, setting $g = 0$, we have

$$1 \ge \sup_{g' \in K'} \langle g', \psi\rangle = \sup_{\|g'\|_\square \le \varepsilon/2} \langle g', \psi\rangle = \frac{\varepsilon}{2}\|\psi\|_\square^*.$$

So

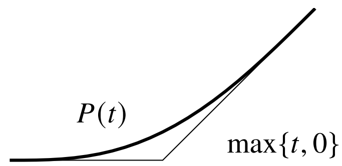
$$\|\psi\|_\square^* \le \frac{2}{\varepsilon}.$$

Setting $g = 0$ and $g' = \pm\frac{\varepsilon}{2}N1_x$ for a single $x \in \Gamma$ (i.e, g' is supported on a single element of Γ), we have $\|g'\|_\square \le \varepsilon/2$ and $1 \ge \langle g', \psi\rangle = \pm\frac{\varepsilon}{2}\psi(x)$. So $|\psi(x)| \le 2/\varepsilon$. This holds for every $x \in \Gamma$. Thus

$$\|\psi\|_\infty \le \frac{2}{\varepsilon}.$$

By the Weierstrass polynomial approximation theorem, there exists some real polynomial $P(x) = p_d x^d + \cdots + p_1 x + p_0$ such that

$$|P(t) - \max\{t, 0\}| \le \frac{\varepsilon}{20} \quad \text{whenever } |t| \le \frac{2}{\varepsilon}.$$



Set

$$R = \sum_{i=0}^{d} |p_i| \left(\frac{2}{\varepsilon}\right)^i,$$

which is a constant that depends only on ε. (A more careful analysis gives $R = \exp(\varepsilon^{-O(1)})$.)

Write $P\psi \colon \Gamma \to \mathbb{R}$ to mean the function given by $P\psi(x) = P(\psi(x))$. By the triangle inequality and the submulticativity of $\|\cdot\|_\square^*$ (Lemma 9.4.8) ,

$$\|P\psi\|_\square^* \le \sum_{i=0}^{d} |p_i| \, \|\psi^i\|_\square^* \le \sum_{i=0}^{d} |p_i| \, (\|\psi\|_\square^*)^i \le \sum_{i=0}^{d} |p_i| \left(\frac{2}{\varepsilon}\right)^i = R.$$

Let us choose

$$\delta = \min\left\{\frac{\varepsilon}{20R}, 1\right\}.$$

Then $\|\nu - 1\|_\square \le \delta$ implies that

$$|\langle \nu - 1, P\psi\rangle| \le \|\nu - 1\|_\square \, \|P\psi\|_\square^* \le \delta R \le \frac{\varepsilon}{20}. \tag{9.2}$$

Earlier we showed that $\|\psi\|_\infty \le 2/\varepsilon$, and also $|P(t) - \max\{t, 0\}| \le \varepsilon/20$ whenever $|t| \le 2/\varepsilon$. Thus

$$\|P\psi - \psi_+\|_\infty \le \frac{\varepsilon}{20}. \tag{9.3}$$

Hence,

$$\begin{aligned}
\langle \nu, P\psi\rangle &= \langle 1, P\psi\rangle + \langle \nu - 1, P\psi\rangle \\
&\le \langle 1, P\psi\rangle + \frac{\varepsilon}{20} && \text{[by (9.2)]} \\
&\le \langle 1, \psi_+\rangle + \frac{\varepsilon}{10} && \text{[by (9.3)]} \\
&\le \frac{1}{1 + \varepsilon/2} + \frac{\varepsilon}{10}. && \text{[by (9.1)].}
\end{aligned}$$

Also,

$$\langle \nu - 1, 1\rangle \le \|\nu - 1\|_\square \le \delta.$$

Thus

$$\|\nu\|_1 \le 1 + \|\nu - 1\|_1 \le 1 + \delta \le 2.$$

So by (9.3),

$$\langle \nu, \psi_+ - P\psi\rangle \le \|\nu\|_1 \, \|\psi_+ - P\psi\|_\infty \le 2 \cdot \frac{\varepsilon}{20} \le \frac{\varepsilon}{10}. \tag{9.4}$$

Thus, using that $0 \le f \le \nu$,

$$\begin{aligned}\langle f, \psi \rangle \le \langle f, \psi_+ \rangle &\le \langle \nu, \psi_+ \rangle \\ &\le \langle \nu, P\psi \rangle + \langle \nu, \psi_+ - P\psi \rangle \\ &\le \frac{1}{1+\varepsilon/2} + \frac{\varepsilon}{10} + \frac{\varepsilon}{10} \le 1 - \frac{\varepsilon}{10}.\end{aligned}$$

This contradicts (a) from earlier. This concludes the proof of the theorem. □

***Remark* 9.4.11 (History).** An early version of the density model theorem was used by Green and Tao (2008), where it was proved using a regularity-type energy increment argument. The above significantly simpler proof is due to Gowers (2010) and Reingold, Trevisan, Tulsiani, and Vadhan (2008) independently. Before the work of Conlon, Fox, and Zhao (2015), one needed to consider the Gowers uniformity norm rather than the simpler cut norm as we did above. The use of the cut norm further simplifies the proof of the corresponding dense model theorem, as noted by Zhao (2014).

Exercise 9.4.12. State and prove a dense model theorem for k-APs.

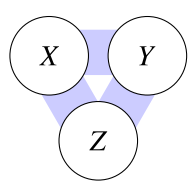
9.5 Sparse Counting Lemma

Let us prove an extension of the triangle counting lemma from Section 4.5. Here we work with a sparse graph (represented by f) that is a subgraph of a sparse pseudorandom host graph (represented by ν) satisfying a 3-linear forms condition (involving $K_{2,2,2}$ densities). The conclusion is that if f is close in cut norm to another dense graph g, then f and g have similar triangle densities (we normalize f for density).

Setup for this section. Throughout this section, we have three finite sets X, Y, Z (which can also be probability spaces) representing the vertex sets of a tripartite graph. The following functions represent edge-weighted tripartite graphs:

$$f, g, \nu \colon (X \times Y) \cup (X \times Z) \cup (Y \times Z) \to \mathbb{R}.$$

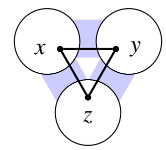
- ν represents the normalized edge-indicator function of a possibly sparse pseudorandom host graph (arising from $S \subseteq \mathbb{Z}/N\mathbb{Z}$ in the statement of the relative Roth theorem).
- f represents the normalized edge-indicator function of a relatively dense subset $A \subseteq S$.
- g represents the dense model of f.



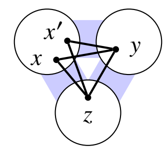
For any tripartite graph F, we write $t(F, f)$ for the F-density in f (and likewise with g

and ν). Some examples:

$$t(K_3, f) = \mathbb{E}_{x,y,z} f(x,y) f(x,z) f(y,z) \quad \text{and}$$



$$t(K_{2,1,1}, F) = \mathbb{E}_{x,x',y,z} f(x,y) f(x',y) f(x,z) f(x',z) f(y,z)$$

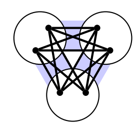


We maintain the convention that x, x' range uniformly over X, and so on.

The functions f, g, ν are assumed to satisfy:

- $0 \le f \le \nu$ pointwise;
- $0 \le g \le 1$ pointwise;
- The **3-linear forms condition**:

$$|t(F, \nu) - 1| \le \varepsilon \quad \text{whenever } F \subseteq K_{2,2,2};$$

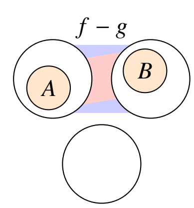


- When restricted to each of $X \times Y$, $X \times Z$, and $Y \times Z$, we have

$$\|f - g\|_\square \le \varepsilon.$$

For example, when restricted to $X \times Y$, the left-hand side quantity denotes

$$\sup_{A \subseteq X, B \subseteq Y} \left| \mathbb{E}_{x,y} (f-g)(x,y) 1_A(x) 1_B(y) \right|.$$



Throughout we assume that $\varepsilon > 0$ is sufficiently small, so that $\le \varepsilon^{\Omega(1)}$ means $\le C\varepsilon^c$ for some absolute constants $c, C > 0$ (which could change from line to line).

Here is the main result of this section, due to Conlon, Fox, and Zhao (2015).

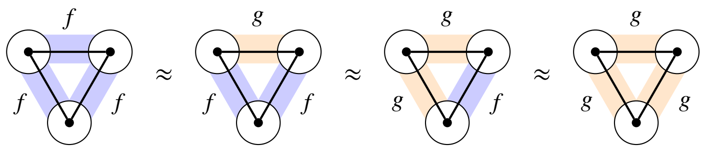
Theorem 9.5.1 (Sparse triangle counting lemma)
Assume the setup at the beginning of this section. Then

$$|t(K_3, f) - t(K_3, g)| \le \varepsilon^{\Omega(1)}.$$

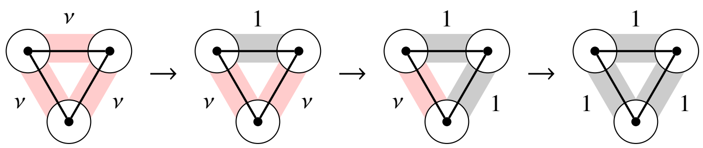
You should now pause and review the proof of the "dense" triangle counting lemma from Proposition 4.5.4, which says that if in addition we assume $0 \le f \le 1$ (that is, assuming $\nu = 1$ identically), then

$$|t(K_3, f) - t(K_3, g)| \le 3 \|f - g\|_\square \le 3\varepsilon.$$

Roughly speaking, the proof of the dense triangle counting lemma proceeds by replacing f by g one edge at a time, each time incurring at most an $\|f - g\|_\square$ loss.



Having $\nu = 1$ should be thought of as the “dense” case. Indeed, $\nu = 1$ correponds to $S = \mathbb{Z}/N\mathbb{Z}$ rather than having a sparse pseudorandom set S. In general, starting with a general “sparse” ν, our strategy is to reduce the problem to another triangle counting problem where ν is replaced by 1 on one of the edges of the triangle.



This **densification** strategy reduces a sparse triangle counting problem to a progressively easier triangle counting problem where some of the sparse bipartite graphs among X, Y, Z become dense.

Let **Sparsity(ν)** be the number of elements of $\{X \times Y, X \times Z, Y \times Z\}$ on which ν differs from 1. We will prove the statement:

SparseTCL(k): the sparse triangle counting lemma is true whenever Sparsity$(\nu) \leq k$. (The hidden constants may depend on k.)

We already proved the base case **SparseTCL**(0), which is the “dense” case corresponding to $\nu = 1$, as discussed earlier. We will prove the implications

$$\textbf{SparseTCL}(0) \implies \textbf{SparseTCL}(1) \implies \textbf{SparseTCL}(2) \implies \textbf{SparseTCL}(3).$$

We phrase our argument as an induction (a slightly unusual induction setup, as $0 \leq k \leq 3$). For the induction step, it suffices to prove the conclusion of the sparse triangle counting lemmas under the following hypothesis.

Induction hypothesis: **SparseTCL**$(k-1)$ holds with $k = \text{Sparsity}(\nu)$, and ν is not identically 1 on $X \times Y$.

The next lemma shows that ν is close to 1 in a strong sense, provided that ν satisfies the 3-linear forms condition.

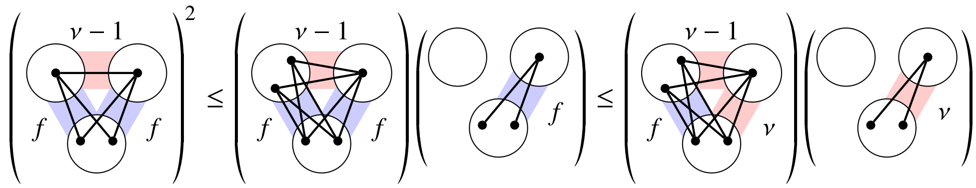
Lemma 9.5.2 (Strong linear forms)
Assume the setup at the beginning of this section. We have

$$\left|\mathbb{E}_{x,y,z,z'}(\nu(x,y)-1)f(x,z)f(x,z')f(y,z)f(y,z')\right| \leq \varepsilon^{\Omega(1)}.$$

The same statement holds if any subset of the four f factors are replaced by g.

Proof. The proof uses two applications of the Cauchy–Schwarz inequality. Let us write down the proof in the case when none of the four fs are replaced by gs. The other cases are similar (basically apply $g \leq 1$ instead of $f \leq \nu$ wherever appropriate).

Here is a figure illustrating the first application of the Cauchy–Schwarz inequality.



Here are the inequalities written out:

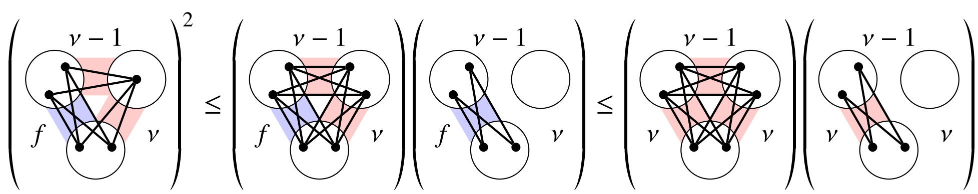
$$
\begin{aligned}
&\left|\mathbb{E}_{x,y,z,z'}(\nu(x,y)-1)f(x,z)f(x,z')f(y,z)f(y,z')\right|^2\\
&=\left|\mathbb{E}_{y,z,z'}\mathbb{E}_x\left[(\nu(x,y)-1)f(x,z)f(x,z')\right]f(y,z)f(y,z')\right|^2\\
&\le\left(\mathbb{E}_{y,z,z'}\left(\mathbb{E}_x(\nu(x,y)-1)f(x,z)f(x,z')\right)^2 f(y,z)f(y,z')\right)\mathbb{E}_{y,z,z'}f(y,z)f(y,z')\\
&\le\left(\mathbb{E}_{y,z,z'}\left(\mathbb{E}_x(\nu(x,y)-1)f(x,z)f(x,z')\right)^2 \nu(y,z)\nu(y,z')\right)\mathbb{E}_{y,z,z'}\nu(y,z)\nu(y,z').
\end{aligned}
$$

Note that we are able to apply $f \le \nu$ in the final step above due to the nonnegativity of the square, which arose from the Cauchy–Schwarz inequality. We could not have applied $f \le \nu$ at the very beginning.

The second factor above is at most $1+\varepsilon$ due to the 3-linear forms condition. It remains to show that the first factor is $\le \varepsilon^{\Omega(1)}$. The first factor expands to

$$\mathbb{E}_{x,x',y,z,z'}(\nu(x,y)-1)(\nu(x',y)-1)f(x,z)f(x,z')f(x',z)f(x',z')\nu(y,z)\nu(y,z').$$

We can upper bound the above quantity as illustrated below, using a second application of the Cauchy–Schwarz inequality.

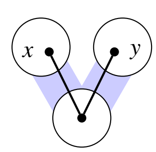


On the right-hand side, the first factor is $\le \varepsilon^{\Omega(1)}$ by the 3-linear forms condition. Indeed, $|t(F,\nu)-1| \le \varepsilon$ for any $F \subseteq K_{2,2,2}$. If we expand all the $\nu - 1$ in the first factor above, then it becomes an alternating sum of various $t(F,\nu) \in [1-\varepsilon, 1+\varepsilon]$ with $F \subseteq K_{2,2,2}$, with the main contribution 1 from each term canceling each other out. The second factor is $\le 1+\varepsilon$ again by the 3-linear forms condition.

Putting everything together, this completes the proof of the lemma. □

Define $\nu_\wedge, f_\wedge, g_\wedge \colon X \times Y \to [0,\infty)$ by

$$
\begin{aligned}
\nu_\wedge(x,y) &:= \mathbb{E}_z \nu(x,z)\nu(y,z),\\
f_\wedge(x,y) &:= \mathbb{E}_z f(x,z)f(y,z),\\
g_\wedge(x,y) &:= \mathbb{E}_z g(x,z)g(y,z).
\end{aligned}
$$

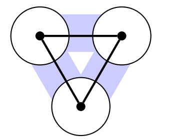


They represent codegrees. Even though ν and f are possibly unbounded, the new weighted graphs $\nu_\wedge$ and $f_\wedge$ behave like dense graphs because the sparseness is somehow smoothed out (this is a key observation). On a first reading of the proof, you may wish to pretend that $\nu_\wedge$

and $f_\wedge$ are uniformly bounded above by 1 (in reality, we need to control the negligible bit of ν exceeding 1).

We have

$$t(K_3, f) = \langle f, f_\wedge \rangle, \quad \text{and} \quad t(K_3, g) = \langle g, g_\wedge \rangle.$$

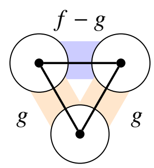


So

$$\begin{aligned} t(K_3, f) - t(K_3, g) &= \langle f, f_\wedge \rangle - \langle g, g_\wedge \rangle \\ &= \langle f, f_\wedge - g_\wedge \rangle + \langle f - g, g_\wedge \rangle. \end{aligned}$$

We have

$$|\langle f - g, g_\wedge \rangle| \le \|f - g\|_\square \le \varepsilon.$$



by the same argument as in the dense triangle counting lemma (Proposition 4.5.4), as $0 \le g \le 1$. So it remains to show $|\langle f, f_\wedge - g_\wedge \rangle| \le \varepsilon^{\Omega(1)}$.

By the Cauchy–Schwarz inequality, we have

$$\langle f, f_\wedge - g_\wedge \rangle^2 = \mathbb{E}[f(f_\wedge - g_\wedge)]^2 \le \mathbb{E}[f(f_\wedge - g_\wedge)^2]\, \mathbb{E} f \le \mathbb{E}[\nu(f_\wedge - g_\wedge)^2]\, \mathbb{E}\nu.$$

The second factor is $\mathbb{E}\nu \le 1 + \varepsilon$ by the 3-linear forms condition. So it remains to show that

$$\mathbb{E}[\nu(f_\wedge - g_\wedge)^2] = \langle \nu, (f_\wedge - g_\wedge)^2 \rangle \le \varepsilon^{\Omega(1)}.$$

By Lemma 9.5.2

$$\left|\langle \nu - 1, (f_\wedge - g_\wedge)^2 \rangle\right| \le \varepsilon^{\Omega(1)}$$

(to see this inequality, first expand $(f_\wedge - g_\wedge)^2$ and then apply Lemma 9.5.2 term by term). Thus

$$\mathbb{E}[\nu(f_\wedge - g_\wedge)^2] \le \mathbb{E}[(f_\wedge - g_\wedge)^2] + \varepsilon^{\Omega(1)}.$$

Thus, to prove the induction step (as stated earlier) for the sparse triangle counting lemma, it remains to prove the following.

Lemma 9.5.3 (Densified triangle counting)

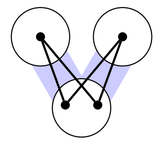
Assuming the setup at the beginning of the section as well as the induction hypothesis, we have

$$\mathbb{E}[(f_\wedge - g_\wedge)^2] \le \varepsilon^{\Omega(1)}. \tag{9.5}$$

Let us first sketch the idea of the proof of Lemma 9.5.3. Expanding, we have

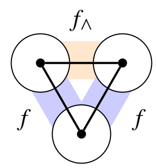
$$\text{LHS of (9.5)} = \langle f_\wedge, f_\wedge \rangle - \langle f_\wedge, g_\wedge \rangle - \langle g_\wedge, f_\wedge \rangle + \langle g_\wedge, g_\wedge \rangle. \tag{9.6}$$

Each term represents some 4-cycle density.



So it suffices to show that each of the four terms above differs from $\langle g_\wedge, g_\wedge\rangle$ by $\le \varepsilon^{\Omega(1)}$. We are trying to show that $\langle f_\wedge, f_\wedge\rangle \approx \langle g_\wedge, g_\wedge\rangle$. Expanding the second factor in each $\langle\cdot,\cdot\rangle$, we are trying to show that

$$\begin{aligned}&\mathbb{E}_{x,y,z} f_\wedge(x,y) f(x,z) f(y,z)\\ &\quad\approx \mathbb{E}_{x,y,z} g_\wedge(x,y) g(x,z) g(y,z).\end{aligned}$$



However, this is just another instance of the sparse triangle counting lemma! And importantly, this instance is easier than the one we started with. Indeed, we have $\|f_\wedge - g_\wedge\|_\square \le \varepsilon^{\Omega(1)}$ (this can be proved by invoking the induction hypothesis). Furthermore, the first factor $f_\wedge(x,y)$ now behaves more like a bounded function (corresponding to a dense graph rather than a sparse graph). Let us pretend for a second that $f_\wedge \le 1$, ignoring the negligible part of $f_\wedge$ exceeding 1. Then we have reduced the original problem to a new instance of the triangle counting lemma, except that now $f \le \nu$ on $X \times Y$ has been replaced by $f_\wedge \le 1$ (this is the key point where **densification** occurs). Lemma 9.5.3 then follows from the induction hypothesis as we have reduced the sparsity of the pseudorandom host graph.

Coming back to the proof, as discussed earlier, while $f_\wedge$ is not necessarily ≤ 1, it is almost so. We need to handle the error term arising from replacing $f_\wedge$ by its capped version $\overline{f_\wedge}\colon X \times Y \to [0,1]$ defined by

$$\overline{f_\wedge} = \min\{f_\wedge, 1\} \quad \text{pointwise.}$$

We have

$$0 \le f_\wedge - \overline{f_\wedge} = \max\{f_\wedge - 1, 0\} \le \max\{\nu_\wedge - 1, 0\} \le |\nu_\wedge - 1|. \tag{9.7}$$

Also,

$$(\mathbb{E}|\nu_\wedge - 1|)^2 \le \mathbb{E}[(\nu_\wedge - 1)^2] = \mathbb{E}\nu_\wedge^2 - 2\mathbb{E}\nu_\wedge + 1 \le 3\varepsilon \tag{9.8}$$

by the 3-linear forms condition, since $\mathbb{E}\nu_\wedge^2$ and $\mathbb{E}\nu_\wedge$ are both within ε of 1. So

$$\begin{aligned}\left|\langle f_\wedge, f_\wedge\rangle - \langle \overline{f_\wedge}, f_\wedge\rangle\right| = \left|\langle f_\wedge - \overline{f_\wedge}, f_\wedge\rangle\right| &\le \mathbb{E}\,|\nu_\wedge - 1|\,\nu_\wedge\\ &= \mathbb{E}\,|\nu_\wedge - 1|\,(\nu_\wedge - 1) + \mathbb{E}\,|\nu_\wedge - 1|\\ &\le \mathbb{E}[(\nu_\wedge - 1)^2] + \mathbb{E}\,|\nu_\wedge - 1|\\ &\le \varepsilon^{\Omega(1)}. \qquad \text{[by (9.8)]}\end{aligned} \tag{9.9}$$

Lemma 9.5.4 (Cut norm between codegrees)
With the same assumptions as Lemma 9.5.3,

$$\|\overline{f_\wedge} - g_\wedge\|_\square \le \varepsilon^{\Omega(1)}.$$

Proof. Indeed, for any $A \subseteq X$ and $B \subseteq Y$, we have

$$\langle \overline{f_\wedge} - g_\wedge, 1_{A\times B}\rangle = \langle \overline{f_\wedge} - f_\wedge, 1_{A\times B}\rangle + \langle f_\wedge - g_\wedge, 1_{A\times B}\rangle.$$

By (9.7) followed by (9.8)

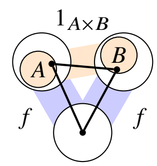
$$\langle \overline{f_\wedge} - f_\wedge, 1_{A\times B}\rangle \le \mathbb{E}|\overline{f_\wedge} - f_\wedge| \le \mathbb{E}|\nu_\wedge - 1| \le \varepsilon^{\Omega(1)}.$$

So it remains to show that

$$|\langle f_\wedge - g_\wedge, 1_{A\times B}\rangle| \le \varepsilon^{\Omega(1)}.$$

This is true since

$$\langle f_\wedge, 1_{A\times B}\rangle = \mathbb{E}_{x,y,z} 1_{A\times B}(x,y) f(x,z) f(y,z)$$
$$\text{and } \langle g_\wedge, 1_{A\times B}\rangle = \mathbb{E}_{x,y,z} 1_{A\times B}(x,y) g(x,z) g(y,z)$$



satisfy the hypothesis of the sparse counting lemma with f, g, ν on $X \times Y$ replaced by $1_{A\times B}, 1_{A\times B}, 1$, thereby decreasing the sparsity of ν by 1, and hence we can apply the induction hypothesis. □

Proof of Lemma 9.5.3. We need to show that each of the four terms on the right-hand side of (9.6) is within $\varepsilon^{\Omega(1)}$ of $\langle g_\wedge, g_\wedge\rangle$. Let us show that

$$|\langle f_\wedge, f_\wedge\rangle - \langle g_\wedge, g_\wedge\rangle| \le \varepsilon^{\Omega(1)}.$$

By (9.9), $\langle f_\wedge, f_\wedge\rangle$ differs from $\langle \overline{f_\wedge}, f_\wedge\rangle$ by $\le \varepsilon^{\Omega(1)}$, and thus it suffices to show that

$$\langle \overline{f_\wedge}, f_\wedge\rangle = \mathbb{E}_{x,y,z} \overline{f_\wedge}(x,y) f(x,z) f(y,z)$$

and

$$\langle g_\wedge, g_\wedge\rangle = \mathbb{E}_{x,y,z} g_\wedge(x,y) g(x,z) g(y,z)$$

differ by $\le \varepsilon^{\Omega(1)}$. To show this, we apply the induction hypothesis to the setting where f, g, ν on $X \times Y$ are replaced by $\overline{f_\wedge}, g, 1$ (recall from Lemma 9.5.4 that $\|\overline{f_\wedge} - g\|_\square \le \varepsilon^{\Omega(1)}$), which reduces the sparsity of ν by 1. So the induction hypothesis implies

$$\left|\langle \overline{f_\wedge}, f_\wedge\rangle - \langle g_\wedge, g_\wedge\rangle\right| \le \varepsilon^{\Omega(1)}.$$

Thus $|\langle f_\wedge, f_\wedge\rangle - \langle g_\wedge, g_\wedge\rangle| \le \varepsilon^{\Omega(1)}$. Likewise, the other terms on the right-hand side of (9.9) are within $\varepsilon^{\Omega(1)}$ of $\langle g_\wedge, g_\wedge\rangle$ (Exercise!). The conclusion $\mathbb{E}[(f_\wedge - g_\wedge)^2] \le \varepsilon^{\Omega(1)}$ then follows. □

Exercise 9.5.5. State and prove a generalization of the sparse counting lemma to count an arbitrary but fixed subgraph (replacing the triangle above). How about hypergraphs?

9.6 Proof of the Relative Roth Theorem

Now we combine the dense model theorem and the sparse triangle counting lemma to prove the relative Roth theorem:

Theorem 9.2.5 (restated). For every $\delta > 0$, there exist $\varepsilon > 0$ and N_0 so that for all $N \geq N_0$, if $S \subseteq \mathbb{Z}/N\mathbb{Z}$ satisfies the 3-linear forms condition with tolerance ε, then every 3-AP-free subset of S has size less than $\delta\,|S|$.

Recall that with $x_0, x_1, y_0, y_1, z_0, z_1 \in \mathbb{Z}/N\mathbb{Z}$ chosen independently and uniformly at random, the set $S \subseteq \mathbb{Z}/N\mathbb{Z}$ with $|S| = pN$ satisfies the ***3-linear forms condition with tolerance ε*** if the probability that

$$\left\{\begin{array}{lll} -y_0 - 2z_0, & x_0 - z_0, & 2x_0 + y_0, \\ -y_1 - 2z_0, & x_1 - z_0, & 2x_1 + y_0, \\ -y_0 - 2z_1, & x_0 - z_1, & 2x_0 + y_1, \\ -y_1 - 2z_1, & x_1 - z_1, & 2x_1 + y_1 \end{array}\right\} \subseteq S$$

lies in the interval $(1 \pm \varepsilon)p^{12}$, and furthermore the same holds if we erase any subset of the above 12 linear forms and also change the "12" in p^{12} to the number of linear forms remaining.

The proof follows the strategy outlined in Section 9.3 on the transference principle. We need a counting version of Roth's theorem. As in Chapter 6, we define, for $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$, its 3-AP density by

$$\boldsymbol{\Lambda_3(f)} := \mathbb{E}_{x,d \in \mathbb{Z}/N\mathbb{Z}} f(x) f(x+d) f(x+2d).$$

Theorem 9.6.1 (Roth's theorem, functional/counting version)
For every $\delta > 0$, there exists $c = c(\delta) > 0$ such that every $f : \mathbb{Z}/N\mathbb{Z} \to [0,1]$ with $\mathbb{E}f \geq \delta$,

$$\Lambda_3(f) \geq c.$$

Exercise 9.6.2. Deduce the above version of Roth's theorem from the existence version (namely that every 3-AP-free subset of $[N]$ has size $o(N)$.)

Proof of the relative Roth theorem (Theorem 9.2.5). Let $p = |S|/N$. Define

$$\nu : \mathbb{Z}/N\mathbb{Z} \to [0,\infty) \qquad \text{by} \qquad \nu = p^{-1} 1_S.$$

Let $X = Y = Z = \mathbb{Z}/N\mathbb{Z}$. Consider the associated edge-weighted tripartite graph

$$\nu' : (X \times Y) \cup (X \times Z) \cup (Y \times Z) \to [0,\infty)$$

defined by, for $x \in X$, $y \in Y$, and $z \in Z$,

$$\nu'(x,y) = \nu(2x+y), \qquad \nu'(x,z) = \nu(x-z), \qquad \nu'(y,z) = \nu(-y-2z).$$

Since ν satisfies the 3-linear forms condition (as a function on $\mathbb{Z}/N\mathbb{Z}$), ν' also satisfies the 3-linear forms condition in the sense of Section 9.5. Likewise,

$$\|\nu - 1\|_{\square} = \|\nu' - 1\|_{\square}$$

where $\|\nu - 1\|_\square$ on the left-hand side is in the sense of Section 9.4 and $\|\nu' - 1\|_\square$ is defined as in Section 9.5 where ν' is restricted to $X \times Y$. (The same would be true had we restricted to $X \times Z$ or $Y \times Z$.) Indeed,

$$\|\nu - 1\|_\square = \sup_{A \subseteq X, B \subseteq Y} \mathbb{E}(\nu(x+y) - 1)1_A(x)1_B(y)$$

whereas

$$\begin{aligned}\|\nu' - 1\|_\square &= \sup_{A \subseteq X, B \subseteq Y} \mathbb{E}(\nu'(x,y) - 1)1_A(x)1_B(y) \\ &= \sup_{A \subseteq X, B \subseteq Y} \mathbb{E}(\nu(2x+y) - 1)1_A(x)1_B(y),\end{aligned}$$

and these two expressions are equal to each other after a change of variables $x \leftrightarrow 2x$ (which is a bijection as N is odd).

By Lemma 9.5.2 (or simply two applications of the Cauchy–Schwarz inequality followed by the 3-linear forms condition), we obtain

$$\|\nu - 1\|_\square \le \varepsilon^{\Omega(1)}.$$

Now suppose $A \subseteq S$ and $|A| \ge \delta N$. Define $f \colon \mathbb{Z}/N\mathbb{Z} \to [0, \infty)$ by

$$f = p^{-1}1_A$$

so that $0 \le f \le \nu$ pointwise. Then by the dense model theorem (Theorem 9.4.6), there exists a function $g \colon \mathbb{Z}/N\mathbb{Z} \to [0,1]$ such that

$$\|f - g\|_\square \le \eta,$$

where $\eta = \eta(\varepsilon)$ is some quantity that tends to zero as $\varepsilon \to 0$.

Define the associated edge-weighted tripatite graphs

$$f', g' \colon (X \times Y) \cup (X \times Z) \cup (Y \times Z) \to [0, \infty)$$

where, for $x \in X$, $y \in Y$, and $z \in Z$,

$$\begin{aligned}f'(x,y) &= f(2x+y), & f'(x,z) &= f(x-z), & f'(y,z) &= f(-y-2z), \\ g'(x,y) &= g(2x+y), & g'(x,z) &= g(x-z), & g'(y,z) &= g(-y-2z).\end{aligned}$$

Note that g' takes values in $[0,1]$. Then

$$\|f' - g'\|_\square = \|f - g\|_\square \le \eta$$

when $f' - g'$ is interpreted as restricted to $X \times Y$ (and the same for $X \times Z$ or $Y \times Z$). Thus by the sparse triangle counting lemma (Theorem 9.5.1), we have

$$|t(K_3, f') - t(K_3, g')| \le \eta^{\Omega(1)}.$$

Note that

$$\begin{aligned}t(K_3, f') &= \mathbb{E}_{x,y,z} f'(x,y) f'(x,z) f'(y,z) \\ &= \mathbb{E}_{x,y,z \in \mathbb{Z}/N\mathbb{Z}} f(2x+y) f(x-z) f(-y-2z) \\ &= \mathbb{E}_{x,d \in \mathbb{Z}/N\mathbb{Z}} f(x) f(x+d) f(x+2d). \\ &= \Lambda_3(f)\end{aligned}$$

Likewise, $t(K_3, g') = \Lambda_3(g)$. And so

$$|\Lambda_3(f) - \Lambda_3(g)| \le \eta^{\Omega(1)}. \tag{9.10}$$

We have

$$\mathbb{E}g \ge \mathbb{E}f - \eta \ge \delta - \eta.$$

Provided that ε is chosen to be small enough so that η is small enough (say, so that $\mathbb{E}g \ge \delta/2$), we deduce from Roth's theorem (the functional version, Theorem 9.6.1) $\Lambda_3(g) \gtrsim_\delta 1$. Therefore

$$p^{-3}N^{-2}\,|\{(x,d) : x, x+d, x+2d \in A\}| = \Lambda_3(f) \overset{(9.10)}{\ge} \Lambda_3(g) - \eta^{\Omega(1)} \gtrsim_\delta 1$$

provided that η is sufficiently small. We can now conclude that A must have a nontrivial 3-AP if N is large enough. Indeed, if A were 3-AP-free, then

$$|\{(x,d) : x, x+d, x+2d \in A\}| = |A| \le |S| = pN,$$

and so the above inequality would imply $p \lesssim_\delta N^{-1/2}$. However, this would be incompatible with the 3-linear forms condition on S, since the probability that random $x_0, x_1, y_0, y_1, z_0, z_1 \in \mathbb{Z}/N\mathbb{Z}$ satisfy

$$\left\{\begin{matrix} -y_0 - 2z_0, & x_0 - z_0, & 2x_0 + y_0, \\ -y_1 - 2z_0, & x_1 - z_0, & 2x_1 + y_0, \\ -y_0 - 2z_1, & x_0 - z_1, & 2x_0 + y_1, \\ -y_1 - 2z_1, & x_1 - z_1, & 2x_1 + y_1 \end{matrix}\right\} \subseteq S$$

lies in the interval $(1 \pm \varepsilon)p^{12}$, but this probability is at least $|S|/N^5 = p/N^4$ (the probability that all 12 terms above are equal to the same element of S). So $(1+\varepsilon)p^{12} \ge pN^{-4}$, and hence $p \gtrsim N^{-4/11}$, which would contradict the earlier $p \lesssim_\delta N^{-1/2}$ if N is large enough. □

***Remark* 9.6.3.** The above proof generalizes to a proof of the relative Szemerédi theorem, assuming Szemerédi's theorem as a black box.

All the arguments in this chapter can be generalized to deduce the relative Szemerédi theorem (Theorem 9.2.7) from Szemerédi's theorem. The ideas are essentially the same, although the notation gets heavier.

Further Reading

The original paper by Green and Tao (2008) titled *The primes contain arbitrarily long arithmetic progressions* is worth reading. Their follow-up paper *Linear equations in primes* (2010a) substantially strengthens the result to asymptotically count the number of k-APs in the primes, though the proof was conditional on several claims that were subsequently proved, most notably the inverse theorem for Gowers uniformity norms (Green, Tao, and Ziegler 2012).

A number of expository articles were written on this topic shortly after the breakthroughs: Green (2007b, 2014), Tao (2007b), Kra (2006), Wolf (2013).

The graph-theoretic approach taken in this chapter is adapted from the article *The Green–Tao theorem: an exposition* by Conlon, Fox, and Zhao (2014). The article presents a full proof of the Green–Tao theorem that incorporates various simplifications found since the original work. The analytic number theoretic arguments, which were omitted from this chapter, can also be found in that article.

Chapter Summary

- **Green–Tao theorem**. The primes contain arbitrarily long arithmetic progressions. Proof strategy:
  - Embed the primes in a slightly larger set, the "almost primes," which enjoys certain pseudorandomness properties.
  - Show that every k-AP-free subset of such a pseudorandom set must have negligible size.
- **Relative Szemerédi theorem.** If $S \subseteq \mathbb{Z}/N\mathbb{Z}$ satisfies a ***k*-linear forms condition**, then every k-AP-free subset of S has size $o(|S|)$.
  - The 3-linear forms condition is a pseudorandomness hypothesis. It says that the associated tripartite graph has F-density close to random whenever $F \subseteq K_{2,2,2}$.
- Proof of the relative Szemerédi theorem uses the **transference principle** to transfer Szemerédi's theorem from the dense setting to the sparse pseudorandom setting.
  - First approximate $A \subseteq S$ by a dense set $B \subseteq \mathbb{Z}/N\mathbb{Z}$ (**dense model theorem**).
  - Then show that the normalized count of k-APs in A and B are similar (**sparse counting lemma**).
  - Finally conclude using Szemerédi's theorem that B has many k-APs, and therefore so must A.
- **Dense model theorem.** If a sparse set S is close to random in normalized cut norm, then every subset $A \subseteq S$ can be approximated by some dense $B \subseteq \mathbb{Z}/N\mathbb{Z}$ in normalized cut norm.
- **Sparse counting lemma.** If two graphs (one sparse and one dense) are close to normalized cut norm, then they have similar triangle counts, provided that the sparse graph lies inside a sparse pseudorandom graph satisfying the 3-linear forms condition (which says that the densities of $K_{2,2,2}$ and its subgraphs are close to random).