Combinatorial Number Theory

Contents

Chapter 1
Roth's Theorem via Szemerédi Regularity

ch:sz

In this chapter we prove Roth's Theorem using graph theoretic methods, namely, the Szemerédi Regularity Lemma. Our main steps are as follows.

1. Prove Szemerédi Regularity via an energy increment argument (Section 2).

2. Prove the Triangle Removal Lemma using Szemerédi Regularity (Section 3).

3. Rephrase Roth's Theorem as a statement about a graph, and apply the Triangle Removal Lemma (Section 4).

Note that each section can be read as a black box, using just the statement of the theorem in the previous section. Thus, the reader anxious to see how Roth's Theorem is related to graph theory may wish to read Section 4 first.

We follow Diestel [1, §7.4] for the proof of Szemerédi's Theorem.

§1 Szemerédi Regularity

sec:sz-reg The Szemerédi Regularity Lemma gives a way to describe a graph using a "structured" component of bounded complexity, with small error (the error being the "pseudorandom" component of the graph). What do we mean by this?

- By *structured of bounded complexity*, we mean that we can group the vertices into a constant number of parts, so that the graph is well-described by the density of edges between these parts.

- Here, *pseudorandom* means that the density of edges between two different parts of the partition doesn't vary too much within subsets of the partition, as we'll describe below. In other words, the edges between two different parts are distributed fairly randomly.

We need to make precise the notion of what it means for the graph to be well-approximated by its structured component. Thus we define the notion of ε-*regularity*.

First, we need to make some preliminary definitions.

Definition 1.1: Let G be a graph and A, B be two subsets of vertices. Let $e(A, B)$ denote the number of edges between A and B (counting edges twice if both their endpoints are in $A \cap B$):

$$e(A, B) := |\{(v, w) : v \in A, w \in B, vw \text{ is an edge}\}|$$

The **edge density** between A and B is

$$d(A, B) := \frac{e(A, B)}{|A||B|}.$$

Definition 1.2: Let V be a set. A **partition** of V is a family of subsets $\mathcal{P} = \{V_1, \ldots, V_k\}$ whose disjoint union equals V.

A **equipartition** of V is a partition $\mathcal{P} = \mathcal{V}_0 \cup \{V_1, \ldots, V_k\}$ such that $|V_1| = \cdots = |V_k|$ and $\mathcal{V}_0 = \{\{v_{0,1}\}, \ldots, \{v_{0,j}\}\}$. We call $\mathcal{V}_0$ or $V_0 = \cup \mathcal{V}_0$ the exceptional set of the partition; it consists of singletons. We let $|\mathcal{P}| = k$ denote the size of $\mathcal{P}$.

For a graph G, a (equi)partition of G simply means a (equi)partition of the vertex set V of G.

Let $\mathcal{P}, \mathcal{Q}$ be two partitions. We say $\mathcal{Q}$ **refines** $\mathcal{P}$, written $\mathcal{P} \leq \mathcal{Q}$, if each set in $\mathcal{P}$ is the union of sets in $\mathcal{Q}$.

For two partitions $\mathcal{P}, \mathcal{Q}$, define the **join** $\mathcal{P} \vee \mathcal{Q}$ to be the smallest partition refining both of them. If P, Q are partitions of V and W, define $\mathcal{P} \otimes \mathcal{Q} = \{A \times B : A \in \mathcal{P}, B \in \mathcal{Q}\}$ as a partition of $V \times W$.

The role of the exceptional set in an equipartition is simply to ensure that all other sets are the same size.

Definition 1.3: Let G be a graph and let $0 < \varepsilon \leq 1$.

We say that subsets A, B are ε-**regular** if for all subsets $A' \subseteq A$ and $B' \subseteq B$ such that $|A'| \geq \varepsilon|A|$ and $|B'| \geq \varepsilon|B|$, we have

$$|d(A', B') - d(A, B)| < \varepsilon.$$

Let $\mathcal{P} = \mathcal{V}_0 \cup \{V_1, \ldots, V_k\}$ be an equipartition of G with exceptional set V_0. We say that G is ε-**regular** with respect to $\mathcal{P}$ if

1. $|V_0| < \varepsilon|G|$, and

2. for all but εm^2 pairs (A_i, A_j) with $1 \leq i < j \leq k$, (V_i, V_j) is ε-regular.[1]

In plain language, a graph G is ε-regular with respect to a partition $\mathcal{P}$ is for most pairs of sets A, B in the partition, if we take two not-too-small subsets A', B' of these sets, the the edge density for the sets and the subsets is approximately the same.

---

[1] Note that there are different definitions of ε-regularity; for instance some authors define an ε-regular partition to be a partition $\{V_1, \ldots, V_m\}$ such that the V_i have sizes varying by at most 1.

We also require that all except one set of $\mathcal{P}$ have the same size, and that the exceptional set is small (it may be empty).

We can now state Szemerédi's Regularity Lemma.

Theorem 1.4 (Szemerédi Regularity)**:** lem:sz Let m be a positive integer constant, and let $0 < \varepsilon < 1$ be constant. There exists an integer $C(m, \varepsilon)$ such that every graph G with at least m vertices has an ε-regular partition $\mathcal{P}$, such that

$$m \leq |\mathcal{P}| \leq C(m, \varepsilon).$$

Remark: The proof gives a value of $C(m, \varepsilon)$ that works; see (1.5).

Note that the upper bound $C(m, \varepsilon)$ is essential because otherwise, the partition consisting of single vertices is trivially an ε-regular partition for any ε. This upper bound is the reason we say that G can be described by a structure of *bounded* complexity.

1.1 The characteristic function

Before we proceed with the proof of Theorem 1.4, we give another way to interpret Szemerédi Regularity. We said repeatedly that we're "decomposing" G into a structured and pseudorandom part, but haven't explicitly made this into a decomposition. We'll do this now.

It doesn't really make sense to "decompose" a graph into two graphs. Instead, we should look at the decomposition from another viewpoint: interpret a graph as a *function*. It is clear what we mean by a decomposition of a function: a way to write it as $f = f_1 + f_2$.

Definition 1.5: Let G be a graph and V its vertex set. Define the **characteristic function** of G to be the function $V \times V \to \{0, 1\}$, with

$$1_G(a, b) = \begin{cases} 0, & ab \text{ is not an edge,} \\ 1, & ab \text{ is an edge.} \end{cases}$$

Then you can think of Szemerédi Regularity as saying we have a decomposition

$$1_G = \underbrace{\mathbb{E}(1_G | \mathcal{P} \otimes \mathcal{P})}_{\text{structured}} + \underbrace{(1_G - \mathbb{E}(1_G | \mathcal{P} \otimes \mathcal{P}))}_{\text{pseudorandom}}.$$

We explain the notation. If A is finite, $f : A \to \mathbb{C}$ is a function, and $\mathcal{P}$ is a partition of A, then define the function $\mathbb{E}(f | \mathcal{P}) : A \to \mathbb{C}$ by

$$\mathbb{E}(f | \mathcal{P})(x) = \frac{1}{|A|} \sum_{a \in A} f(a), \text{ where } x \in A \in \mathcal{P}.$$

[2] Because by definition

$$eq : e(A, B) \mathbb{E}(1_G | \mathcal{P} \otimes \mathcal{P})(a, b) = d(A, B), \qquad a \in A, b \in B, \tag{1.1}$$

---

[2]For those familiar with measure theory, we are essentially saying the following:

pictorially you can think of $\mathbb{E}(1_G|\mathcal{P} \otimes \mathcal{P})$ as dividing the vertices of the graph G into the parts of $\mathcal{P}$, and writing a single number between every two sets A, B giving the edge density between A, B. This takes much less space to write down (or draw) than the graph G, and we choose $\mathcal{P}$ so this is a good approximation of the structure of G.

A picture here!

While we won't bother to give an actual definition of "pseudorandom" in the context of functions, note that this can be done so that it corresponds to the definition of ε-regularity, but perhaps with a different constant.

Also note that by thinking of a graph as a *function* $1_G : V \times V \to \{0, 1\}$, we can generalize the statement of Szemerédi Regularity to *generalized graphs*, which are functions $f : V \times V \to [0, 1]$.

§2 Proof of Szemerédi Regularity

sec:sz-reg-pf

2.1 Defining the energy

The basic idea is simple: start with any partition $\mathcal{P}$ into $k + 1$ approximately equal parts. If this partition is ε-regular, then we are done. Otherwise, we look at every pair $A, B \in \mathcal{P}$ that is not ε-regular, and refine the sets A, B, in order to separate the subsets that cause the problem. This is essentially a "greedy" algorithm.

The main problem, however, is that we need to make sure we *stop* in a bounded number of steps (depending on k, ε). We know we'll eventually terminate—because the partition of singletons is ε-regular—but this is not good enough for us.

We need to *quantify* the progress we've made towards ε-regularity. To do this, we define a quantity called the **energy** of $\mathcal{P}$: the higher the energy, the closer that G is to being described completely by $\mathcal{P}$ (and the edge densities between the sets of $\mathcal{P}$). Now we show the following.

- The energy is *bounded above* (Lemma 2.2).

- In our "greedy" algorithm, each step we must increase the energy of $\mathcal{P}$ by at least a certain amount.

This shows that we will only need to take a bounded number of steps, and hence that when we finish, we have a bounded number of sets in $\mathcal{P}$.

How do we define the energy? Here it is again helpful to think of a graph G as a function; we want to know how close 1_G is to $\mathbb{E}(1_G|\mathcal{P} \otimes \mathcal{P})$, our "approximation" of 1_G. Probability

---

1. Let $(A, \mathcal{A}, \mu)$ be a measure space, and $\mathcal{B}$ be a sub-σ-algebra of A. Define $\mathbb{E}(f|\mathcal{B}) : A \to \mathbb{C}$ as the projection of f onto the space of $\mathcal{B}$-measurable functions.

2. To recover the above, give A the constant measure, and identify $\mathcal{P}$ with the σ-algebra that it generates.

theory (or analysis) tells us that a good way to measure the size of a function is using its L^2 norm, so we consider

$$\mathbb{E}(\mathbb{E}(1_G|\mathcal{P}\otimes\mathcal{P})^2).$$

Note that this tells us how close 1_G is to $\mathbb{E}(1_G|\mathcal{P}\otimes\mathcal{P})^2$, as we have the following inequalities, with the differences between successive terms being the *variances*:

$$\mathbb{E}(1_G)^2 \underbrace{\leq}_{\mathrm{Var}(\mathbb{E}(1_G|\mathcal{P}\otimes\mathcal{P}))} \mathbb{E}(\mathbb{E}(1_G|\mathcal{P}\otimes\mathcal{P})^2) \underbrace{\leq}_{\mathrm{Var}(1_G-\mathbb{E}(1_G|\mathcal{P}\otimes\mathcal{P}))} \mathbb{E}(1_G^2) = \mathbb{E}(1_G). \qquad (1.2)$$

As we refine $\mathcal{P}$, we expect $\mathbb{E}(1_G|\mathcal{P}\otimes\mathcal{P})$ to be a better approximation, so it will get further from $\mathbb{E}(1_G)$ and closer to 1_G in the L^2 sense, i.e., the variance at the left will decrease and the variance at the right will increase.

> 🔑 The **energy increment argument** goes as follows: define a quantity called the energy, usually as an expectation of a square. Show that the energy is bounded above, and that in each step the energy must increase by at least a certain amount, so that an algorithm ends after a finite number of steps.

Definition 2.1: df:energy-P

1. Let A be a set and $\mathcal{P}$ be a partition of A. Define the **energy** of $\mathcal{P}$ to be
$$E_f(\mathcal{P}) = \mathbb{E}(\mathbb{E}(f|\mathcal{P})^2).$$

2. Let G be a graph, V_1, V_2 be two subsets of its vertices, and $\mathcal{P}, \mathcal{Q}$ be partitions of V_1, V_2. Define
$$E_G(\mathcal{P},\mathcal{Q}) := E_{1_G|_{V_1\times V_2}}(\mathcal{P}\otimes\mathcal{Q}) = \mathbb{E}(\mathbb{E}(1_G|_{V_1\times V_2}|\mathcal{P}\otimes\mathcal{Q})^2).$$

3. Let G be a graph and $\mathcal{P}$ be a partition. Define the **energy** of the $\mathcal{P}$ to be
$$E_G(\mathcal{P}) := E_{1_G}(\mathcal{P}\otimes\mathcal{P}) = \mathbb{E}(\mathbb{E}(1_G|\mathcal{P}\otimes\mathcal{P})^2).$$

Let's unravel this formula, noting $\sum_{(x,y)\in V\times V} = \sum_{A,B\in\mathcal{P}}\sum_{a\in A,b\in B}$ and using (1.1):

$$E_G(\mathcal{P}) = \frac{1}{|V|^2}\sum_{A,B\in\mathcal{P}}\sum_{a\in A,b\in B}d(A,B)^2$$

$$= \frac{1}{|V|^2}\sum_{A,B\in\mathcal{P}}|A||B|d(A,B)^2 eq:energy-formula \qquad (1.3)$$

For convenience, we let $E_G(A,B) := E_G(\{A\},\{B\}) = d(A,B)$.

We have the following.

Lemma 2.2: lem:energy-bound Let $\mathcal{P}$ be a partition of graph G. Then
$$0 \leq E_G(\mathcal{P}) \leq 1.$$

Proof. This is clear from Definition 2.1 and the fact that $\mathbb{E}(1_G|\mathcal{P}\otimes\mathcal{P})(x)\in[0,1]$. $\qquad\square$

2.2 Refining a partition

What happens to the expectation when we refine a partition?

Lemma 2.3: lem:refine-1 Let A be a finite set and let $f : A \to \mathbb{R}_{\geq 0}$ be a function. Suppose $\{B, C\}$ is a partition of A, and suppose

$$\mathbb{E}(f|_B) - \mathbb{E}(f) = \varepsilon.$$

Then

$$E_f(\{B, C\}) = E_f(\{A\}) + \varepsilon^2 \frac{|B|}{|A|}\left(1 + \frac{|B|}{|C|}\right) \geq E_f(\{A\}).$$

Note the fact that $E_f(\{B, C\}) \geq E_f(\{A\})$ actually follows from the Quadratic Mean–Arithmetic Mean inequality. Here we want more than inequality, though: we want to know that if $\mathbb{E}(f|_B)$ is far away from $\mathbb{E}(f)$ then the energy of $\{B, C\}$ is significantly more than the energy of $\{A\}$.

Later we will apply this to a subset that violates ε-regularity.

Proof. We calculate

$$E_f(\{B, C\}) = \mathbb{E}(\mathbb{E}(f|\{B, C\})^2) = E(f)^2 + \mathrm{Var}(\mathbb{E}(f|\{B, C\}))$$

$$= E(f)^2 + \frac{|B|}{|A|}[\mathbb{E}(f|_B) - \mathbb{E}(f)]^2 + \frac{|C|}{|A|}[\mathbb{E}(f|_C) - \mathbb{E}(f)]^2$$

$$= E(f)^2 + \frac{|B|}{|A|}\varepsilon^2 + \frac{|C|}{|A|}\left(\frac{|B|}{|C|}\varepsilon\right)^2$$

$$= E_f(\{A\}) + \varepsilon^2 \frac{|B|}{|A|}\left(1 + \frac{|B|}{|C|}\right).$$

$\square$

Now we apply this to our setting.

Lemma 2.4: lem:refine-2 Let G be a graph and $A, B \in \mathcal{P}$ be a non-ε-regular pair. Let A', B' be such that

$$|d(A', B') - d(A, B)| > \varepsilon.$$

Then

$$E_G(A, B) \geq E_G(\{A', A\backslash A'\}, \{B', B\backslash B'\}) + \varepsilon^4.$$

Proof. Apply Lemma 2.3 to $f = 1_G|_{A \times B}$ with $A \times B$ in place of A and $A' \times B'$ in place of B to get

$$E_{1_G|_{A,B}}(\{A' \times B', A \times B \backslash A' \times B'\}) \geq E_{1_G|_{A,B}}(\{A \times B\}) + \varepsilon^2 \frac{|A' \times B'|}{|A \times B|}$$

$$\geq E_{1_G|_{A,B}}(\{A \times B\}) + \varepsilon^4.$$

Now note $\{A', A\backslash A'\} \otimes \{B', B\backslash B'\}$ is a further refinement of $\{A' \times B', A \times B \backslash A' \times B'\}$. $\square$

2.3 Energy Increment Argument

We now prove Szemerédi's Theorem.

Proof of Theorem 1.4. We will iterate the following construction.

Lemma 2.5 (Energy increment argument): lem:energy-increment Let G be a graph with n vertices and let $0 < \varepsilon \leq \frac{1}{4}$.

Suppose $\mathcal{P} = \mathcal{V}_0 \cup \{V_1, \ldots, V_k\}$ is a non-ε-regular equipartition of the vertices of G with $|V_0| \leq \varepsilon n$. Then there exists $k \leq \ell \leq k4^k$ and an equipartition

$$\mathcal{Q} = \mathcal{W}_0 \cup \{W_1, \ldots, W_\ell\}$$

such that $|\mathcal{W}_0| \leq |\mathcal{V}_0| + \frac{n}{2^k}$ and

$$E_G(\mathcal{Q}) \geq E_G(\mathcal{P}) + \varepsilon^5.$$

Proof. Because $\mathcal{P}$ is not ε-regular, for at least $2\varepsilon k^2$ pairs $(i,j) \in [1,n]^2$, (V_i, V_j) is not ε-regular. Call these *bad pairs*. Let $V_{ij} \subset V_i$ and $V_{ji} \subset V_j$ be such that

$$|d(V_{ij}, V_{ji}) - d(V_i, V_j)| > \varepsilon.$$

<u>Step 1:</u> We refine $\mathcal{P}$ using the sets that obstruct ε-regularity. Consider the partition

$$\mathcal{P}' = \mathcal{P} \vee \bigvee \{V_{ij}, V_{ij}^c\}.$$

We first show that

$$eq:sz-1 \qquad E_G(\mathcal{P}') \geq E_G(\mathcal{P}) + \varepsilon^5. \tag{1.4}$$

Write

$$\mathcal{P}' = \mathcal{V}_0' \cup \mathcal{V}_1' \cup \cdots \cup \mathcal{V}_k'$$

where $\mathcal{V}_i'$ are partitions of V_i. Let $\mathcal{V}_i = \{V_i\}$ for $i \geq 0$; we have

$$\mathcal{V}_i \leq \{V_{ij}, V_i \backslash V_{ij}\} \leq \mathcal{V}_i',$$

we have

$$
\begin{aligned}
E_G(\mathcal{P}') &= \sum_{i,j} \frac{|V_i||V_j|}{n^2} E_G(\mathcal{V}_i', \mathcal{V}_j') \\
&= \sum_{(i,j) \text{ bad}} \frac{|V_i||V_j|}{n^2} E_G(\mathcal{V}_i', \mathcal{V}_j') + \sum_{(i,j) \text{ not bad}} \frac{|V_i||V_j|}{n^2} E_G(\mathcal{V}_i', \mathcal{V}_j') \\
&\geq \sum_{(i,j) \text{ bad}} \frac{|V_i||V_j|}{n^2} E_G(\{V_{ij}, V_i \backslash V_{ij}\}, \{V_{ji}, V_j \backslash V_{ji}\}) + \sum_{(i,j) \text{ not bad}} \frac{|V_i||V_j|}{n^2} E_G(\mathcal{V}_i, \mathcal{V}_j) \\
&= \sum_{(i,j) \text{ bad}} \frac{|V_i||V_j|}{n^2} \left[E_G(V_i, V_j) + \varepsilon^4 \right] + \sum_{(i,j) \text{ not bad}} E_G(\mathcal{V}_i, \mathcal{V}_j) \qquad \text{by Lemma 2.4} \\
&\geq E_G(\mathcal{P}) + 2\varepsilon^5 \frac{k^2 |V_1|^2}{n^2}. \qquad (\geq \varepsilon k^2 \text{ bad pairs})
\end{aligned}
$$

Finally note that $\frac{k^2|V_1|^2}{n^2} = \left(\frac{|V_1 \cup \cdots \cup V_k|}{n^2}\right)^2 \geq (1-\varepsilon)^2 > \frac{1}{2}$ since $\varepsilon \leq \frac{1}{4}$.

<u>Step 2</u>: We modify $\mathcal{P}'$ so that all except one set has the same number of elements. To do this, we chop up all the sets in $\mathcal{P}$ into even smaller pieces of equal size, and then throw the leftovers into W_0.

Note that $|\mathcal{P}'| \leq k2^k$ because we defined at most k sets $V_{ij} \subset V_i$, and these sets split up V_i into at most 2^k different subsets.

Now cut up each nonexceptional set in $\mathcal{P}'$ into parts of size $\left\lfloor \frac{n}{k4^k} \right\rfloor$, and let these be the sets of $\mathcal{Q}$: there are at most $k4^k$ of them. There are at least k such sets, because the sets of $\mathcal{P}'$ contained in $\mathcal{V}_i$ number at most 2^k; hence one of these sets contains at least $\frac{(1-\varepsilon)n}{k2^k} > \frac{n}{k4^k}$ elements.

Place the leftover elements into W_0: there are at most $|V_0| + k2^k \left\lfloor \frac{n}{k4^k} \right\rfloor \leq |V_0| + \frac{n}{2^k}$ of these.

Because $\mathcal{Q}$ refines $\mathcal{P}'$, we have

$$E_G(\mathcal{Q}) \geq E_G(\mathcal{P}').$$

Together with (1.4), this proves the lemma. $\qquad\square$

Suppose G has n vertices, and let $\mathcal{P}$ be any equipartition of G with sets of size $\left\lfloor \frac{n}{k} \right\rfloor$, where $k \geq m$ is to be chosen.

Now we iterate the energy increment argument, Lemma 2.5. Because $0 \leq E_G(\mathcal{Q}) \leq 1$ for any partition $\mathcal{Q}$ (Lemma 2.2), we know that after at most $\frac{1}{\varepsilon^5}$ iterations we will have an ε-regular partition.[3]

Let $s := \left\lfloor \frac{1}{\varepsilon^5} \right\rfloor$ (or as we noted, $\left\lfloor \frac{1}{4\varepsilon^5} \right\rfloor$). It remains to choose k (the number of sets to start with) and $C(m, \varepsilon) \geq f^s(m)$ appropriately. Note that whatever we choose for $C(m, \varepsilon)$, the theorem is trivial for $n \leq C(m, \varepsilon)$, because we can just divide G into singleton sets. Hence, we will just have to make sure the energy increment argument works for $n > C(m, \varepsilon)$.

Note that initially the exceptional set has size less than k. After all the iterations, the exceptional set has size less than $k + \frac{sn}{2^k}$. We have $k + \frac{sn}{2^k} \leq \varepsilon n$ if

$$n \geq \frac{2k}{\varepsilon} \text{ and } 2^k \geq \frac{s}{2\varepsilon}.$$

- First, choose k such that

$$k \geq \max\left(\log_2\left(\frac{s}{2\varepsilon}\right), m\right).$$

- Let $f(x) := x4^x$. We would like

$$C(m, \varepsilon) \geq f^{(s)}(m)$$

because this makes sure that the final partition has size at most $C(m, \varepsilon)$. Since we want $n \geq \frac{2k}{\varepsilon}$, we let

$$eq: sz - constant \quad C(m, \varepsilon) := \max\left(f^{(s)}(m), \frac{2k}{\varepsilon}\right). \tag{1.5}$$

---

[3]We can do better: (1.2) tells us we can need at most $\frac{\mathbb{E}(1_G) - \mathbb{E}(1_G)^2}{\varepsilon^5} = \frac{d(V,V) - d(V,V)^2}{\varepsilon^5} \leq \frac{1}{4\varepsilon^5}$ iterations.

Iterating the lemma repeatedly then gives the theorem. $\square$

§3 Triangle Removal Lemma

sec:tri-rem

Using the Szemerédi Regularity Lemma, we can prove the Triangle Removal Lemma.

Lemma 3.1: lem:tri-removal For all $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ such that for every graph G with n vertices, at least one of the following is true:

1. G can be made triangle-free by removing less than εn^2 edges.

2. G has at least $\geq \delta n^3$ triangles.

In other words, if G has a small number of triangles compared to n^3, then it is possible to delete all of them by removing few edges compared to n^2; if G cannot be made triangle-free by removing a small number of edges compared to n^2, then G must have a lot of triangles compared to n^3.

Remark: Our proof of the Triangle Removal Lemma using Szemerédi Regularity gives a very weak bound for δ. A direct (but more difficult) proof of Triangle Removal shows we can take

$$\delta^{-1} \leq 2 \uparrow O\left(\ln \frac{1}{\varepsilon}\right)$$

where $a \uparrow b = \overbrace{2^{\cdot^{\cdot^{\cdot^2}}}}^{b}$. Instead of measuring the energy using squares, $d(A,B)^2$, one measures energy using the *entropy* function $d(A,B) \ln d(A,B)$. On the other hand, Behrend's construction shows that $\delta^{-1} \succeq \left(\frac{1}{\varepsilon}\right)^{\ln\left(\frac{1}{\varepsilon}\right)}$.

The main idea of the proof is as follows. First, we choose an ε'-regular partition $\mathcal{P}$ for appropriate ε, then remove all the edges that obstruct ε-regularity (as well as edges between low-density pairs), so that the resulting graph is very "random." Because we can choose ε small, we can make sure we remove very few edges. It there still exists a triangle with edges between 3 sets $A, B, C \in \mathcal{P}$—i.e., we failed to make G triangle-free—then actually there must exist a lot of triangles, since edges between A, B, C are distributed fairly randomly, and there is a positive proportion of edges between each pair among A, B, C.

This proof shows the general pattern of proofs involving Szemerédi regularity: by removing (or ignoring) a few edges, we can make the distribution of edges in G very random (uniform, regular) with respect to densities between components. The theorem is then easy to prove using under this regularity hypothesis.

Proof of Theorem 3.1. By the Szemerédi Regularity Lemma 1.4, given $\varepsilon' > 0$ and m, there exists $M := C(\varepsilon', m)$ such that for any graph G, there exists an ε'-regular partition $\mathcal{P}$ of G into k components, for some $m \leq k \leq M$.

Given $\varepsilon > 0$, take $\varepsilon' = \frac{\varepsilon}{4}$ and $m = \frac{2}{\varepsilon} = \frac{1}{2\varepsilon'}$.
We now remove the following edges from $\mathcal{P}$:

1. Edges within one component: There are k components and each component has at most $\binom{n/k}{2}$ edges, so there are at most $k\binom{n/k}{2} < \frac{n^2}{2k} \leq \boxed{\frac{1}{2m}n^2}$ such edges.

2. Edges between non-ε-regular pairs: There are at most $\varepsilon' k^2$ non-ε-regular pairs and each has at most $\left(\frac{n}{k}\right)^2$ edges, so there are at most $\varepsilon' k^2 \left(\frac{n}{k}\right)^2 \leq \boxed{\varepsilon' n^2}$ such edges.

3. Edges involving the exceptional set V_0: V_0 has at most $\varepsilon' n$ vertices, so there are at most $\boxed{\varepsilon' n^2}$ such edges.

4. Edges between pairs with edge density at most $d := \frac{\varepsilon}{2}$: There are at most $\binom{k}{2}$ pairs and at most $d\left(\frac{n}{k}\right)^2$ edges between each pair, so less than $\boxed{\dfrac{d}{2}n^2}$ such edges.

In total we've removed less than
$$\left(\frac{1}{2m} + 2\varepsilon' + \frac{d}{2}\right)n^2 \leq \left(\frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{4}\right)n^2 = \varepsilon n^2$$
edges. (The choice of ε', m, and d should be clear from this inequality.)

If G is triangle-free, we are done. Otherwise, at least one triangle still remains. Since we've removed all edges within pairs and edges to V_0, it must go between three different components A, B, and C. Because A, B, C still have edges between them, all pairs among A, B, C are regular with density at least d. We show that there are many triangles between A, B, and C. Note we will not use the existence of a triangle again.

Let p be the size of each component in $\mathcal{P}$. For a vertex v, let $N_B(v)$, $N_C(v)$ denote the neighbors in B and C, respectively. Consider
$$A_B = \{v \in A : N_B(v) \geq (d - \varepsilon')p\}.$$
Note that $d(A \backslash A_B, B) < d - \varepsilon'$ because every vertex in $A \backslash A_B$ has less than $d - \varepsilon'$ neighbors in B. Since A, B is a regular pair, we must have $|A \backslash A_B| < \varepsilon'|A|$ and
$$|A_B| > (1 - \varepsilon')|A| = (1 - \varepsilon')p.$$
Similarly defining A_C, the same inequality holds for A_C, and by inclusion-exclusion,
$$|A_B \cap A_C| \geq (1 - 2\varepsilon')p.$$
For a vertex v, let $N_B(v)$, $N_C(v)$ denote the neighbors in B and C, respectively. For every $v \in A_B \cap A_C$, we have that
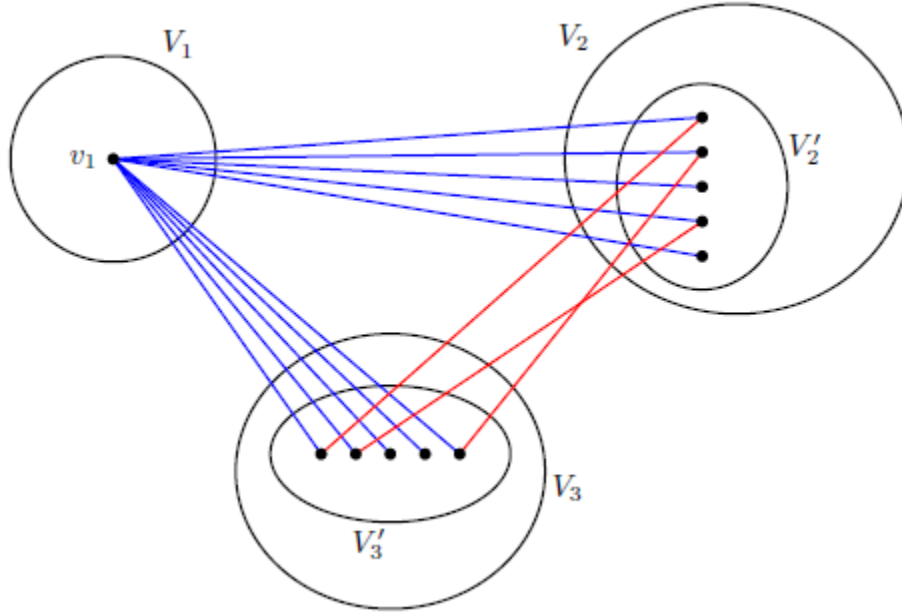$$|N_B(v)|, |N_C(v)| \geq (d - \varepsilon')p \geq \varepsilon' p,$$
so by the regularity hypothesis,
$$e(N_B(v), N_C(v)) = d(N_B(v), N_C(v)) \cdot |N_B(v)||N_C(v)| \geq (d(B,C) - \varepsilon') \cdot [(d - \varepsilon')p]^2 > (d - \varepsilon')^3 p^2,$$
and v is part of at least $(d - \varepsilon')p^2$ triangles.[4]

---

[4]Figure taken from [2].

This is true for every $v \in A_B \cap A_C$, so we get that there are more than

$$\underbrace{(1 - 2\varepsilon') p}_{\frac{\varepsilon}{2}} \cdot \underbrace{(d - \varepsilon')^3 p^2}_{\frac{\varepsilon}{4}} = \frac{\varepsilon^3}{128} p^3$$

triangles.

Finally, note that $p \geq \frac{(1-\varepsilon')n}{M}$: p is bounded below because the number of components k is bounded above by M. Hence there are at least δn^3 triangles where

$$\delta := \frac{\varepsilon^3}{128} \cdot \left(\frac{1 - \varepsilon'}{M} \right)^3.$$

$\square$

§4 Proof of Roth's Theorem

sec:sz-roth

Theorem 4.1 (Varnavides's Theorem): thm:varn For every $\varepsilon > 0$ there exists $\delta > 0$ with the following property: for any $n \in \mathbb{N}$, if A is a subset of $\{1, 2, \ldots, n\}$ with density $\frac{|A|}{n} \geq \varepsilon$, then A has at least δn^2 3-term arithmetic progressions (including the trivial ones).

Corollary 4.2 (Roth's Theorem): thm:roth For every $\varepsilon > 0$ there exists L with the following property: for any $n > L$, if A is a subset of $\{1, 2, \ldots, n\}$ with density ε, then A contains a nontrivial 3-term arithmetic progression.

As a stepping stone to this theorem, we first prove the following.

Theorem 4.3 (Varnavides's Theorem for groups)**:** thm:varn-group For every $\varepsilon > 0$ there exists $\delta > 0$ with the following property: if H is a finite group of odd order n and $A \subseteq H$ with density $\frac{|A|}{n} \geq \varepsilon$, then A has at least δn^2 3-term arithmetic progressions.

We will translate this to a graph-theoretic problem. We would like each arithmetic progression $a, b, c \in H$ to correspond to a triangle. Because a, b, c do not play symmetric roles, and because we want to count trivial arithmetic progressions, we actually consider a graph G whose vertices are *three* copies of the vertices of H.

The presence of trivial arithmetic progressions forces there to be a lot of disjoint triangles, which forces us to remove a lot of edges to make G triangle-free. This in turn forces a lot of triangles (on the order of n^3), and hence arithmetic progressions (on the order of n^2). In particular, there is a *nontrivial* arithmetic progression, for n large. (Note how the presence of trivial AP's forces the presence of nontrivial AP's!)

4.1 Varnavides's Theorem for Groups

Proof. Step 1: Construct the graph G depending on A and H: Let X, Y, Z be 3 copies of the elements of H, and let the vertices of G be labeled by $X \sqcup Y \sqcup Z$. Now draw the following edges:

- Draw an edge between $x \in X$ and $y \in Y$ if $y - x \in A$.

- Draw an edge between $x \in X$ and $z \in Z$ if $\frac{z-x}{2} \in A$.

- Draw an edge between $y \in Y$ and $z \in Z$ if $z - y \in A$.

We claim that each arithmetic progression in A corresponds to n triangles. Indeed, consider the map

$$\varphi : H \times H \times H \mapsto H \times H \times H$$

$$(x, y, z) \mapsto (a, b, c) = \left(y - x, \frac{z - x}{2}, z - y \right).$$

We check:

1. $\varphi(x, y, z)$ is an arithmetic progression: $\left(y - x, \frac{z-x}{2}, z - y \right)$ has common difference $-\frac{x}{2} + \frac{z}{2}$.

2. Each arithmetic progression has n preimages: We have

$$\varphi(x, x + a, x + a + c) = \left(a, \frac{a+c}{2}, c \right)$$

   for every $x \in H$. There are n^2 arithmetic progressions as there are n choices for the first term and n choices for the common difference; hence this accounts for all possibilities for (x, y, z).

By the way we defined the edges, each AP (a, b, c) corresponds to exactly n triangles $\varphi^{-1}(a, b, c)$.

There are at least εn trivial arithmetic progressions in A,

$$\{(a, a, a) : a \in A\}.$$

These correspond to εn^2 edge disjoint triangles in G:

$$(x \in X, x + a \in Y, x + 2a \in Z).$$

Because these triangles do not share an edge in common—knowing any two elements of $(x, x + a, x + 2a)$ automatically determines the third one—*in order to make G triangle free we have to remove εn^2 edges.*

This is the reason we defined G the way we did: in order to draw the above conclusion we need each arithmetic progression in G to correspond to n triangles, so we needed a map φ that took n triangles into a single arithmetic progression in A. Moreover, a triangle is made up of edges, which are structures between *two* vertices, so we must be able to describe each component of $\varphi(x, y, z)$ with just *two* of the variables x, y, z. (Note there are other choices for φ that work.)

Step 2: Choose $\delta = \delta\left(\frac{\varepsilon}{9}\right)$ as in the Triangle Removal Lemma 3.1. Because we cannot make G triangle-free by removing less than $\frac{\varepsilon}{9}(3n)^2$ edges, G has at least $\delta(3n)^3$ triangles, and hence A has at least $\delta 27 n^2$ AP's. $\qquad\square$

4.2 Varnavides's Theorem for Integers

Proof of Theorem 4.1. Consider A as a subset of $\mathbb{Z}/(2n-1)$. Note that the arithmetic progressions in A are the same whether A is considered as a subset of $\{1, \ldots, n\} \subseteq \mathbb{Z}$ or $\mathbb{Z}/(2n-1)$, because a 3-term AP cannot "wrap-around": if (a, b, c) is an arithmetic progression with $1 \leq a, b \leq n$, then $c \leq 2n - 1$.

Now A has density at least $\frac{\varepsilon n}{2n-1} > \frac{\varepsilon}{2}$ in $\mathbb{Z}/(2n-1)$, so the theorem follows from Theorem 4.3 with $\varepsilon \hookleftarrow \frac{\varepsilon}{2}$. $\qquad\square$

Note we could have adapted the proof of Theorem 4.3 to give the proof of Theorem 4.1 directly, without the intermediate theorem on groups. However, it is worth knowing that the theorem holds true for general groups.

Proof of Roth's Theorem. Let δ be as in Varnavides's Theorem 4.1. Let $L = \frac{\delta}{\varepsilon}$. Then A has at least δn^2 arithmetic progressions, and hence at least

$$\delta n^2 - \varepsilon n > 0$$

nontrivial arithmetic progressions. $\qquad\square$

§5 Problems

3.1 (Hypergraph removal) Just as we rephrased Roth's Theorem in the language of graphs, rephrase Szemerédi's Theorem in the language of hypergraphs.

3.2 Define a *corner* to be an axis-aligned isosceles triangle of $\mathbb{Z}^2$, that is, a set of three elements

$$\{(x, y), (x + h, y), (x, y + h)\}$$

with $h \neq 0$.

(a) Prove the following.

Theorem 5.1 (Corner Theorem)**:** For every $\varepsilon > 0$, there exists L such that for $n > L$, any subset A of $\{1, \ldots, n\} \times \{1, \ldots, n\}$ with at least εn^2 points has a corner.

(b) Deduce Roth's Theorem from Corner Theorem.

5.1 Solutions

3.1 Suppose we're looking for m-term AP's. We'll do the case $m = 4$. We can reduce to the problem for groups with order relatively prime to $(m - 1)! = 6$, just as we did for $m = 3$. Given such a group H, consider a graph whose vertices are 4 copies of H, labeled $H_1, \ldots, H_4$. Now draw a face between...

$$
\begin{aligned}
x, y, z, \text{ if} && 3x + 2y + z && \in H \\
x, y, w \text{ if} && 2x + y && - w \in H \\
x, w, z \text{ if} && x && - z - 2w \in H \\
y, w, z \text{ if} && -y - 2z - 3w \in H.
\end{aligned}
$$

Here, $x \in H_1$, $y \in H_2$, $z \in H_3$, and $w \in H_4$. Each m-term AP corresponds to $n^{m-2} = n^2$ simplices. The claim is that there is a simplex corresponding to a nontrivial AP.

3.2 (a) Let G be the graph with vertices coming from 3 sets L_h, L_v, and L_d, defined as follows. Let

- L_h consist of horizontal lines $y = i$ for $1 \le i \le n$,
- L_v consist of vertical lines $x = i$ for $1 \le i \le n$, and
- L_d consist of diagonal lines $y = x + i$ for $-(n - 1) \le i \le n - 1$.

Now connect two vertices if the lines that they represent intersect at a point in A. Each triangle then corresponds to a corner (possibly trivial) in A.

Note that every "trivial" corner corresponds to exactly one triangle, whose vertices are 3 lines that intersect at a point. Knowing any 2 of these lines determines the third. Hence the trivial corners correspond to εn^2 edge-disjoint triangles. To make G triangle-free, we must remove at least εn^2 edges.

Now there are $2n + (2n - 1) = 4n - 1$ vertices, and n^2 is quadratic in $4n - 1$. Hence by the triangle removal lemma, there is δ such that G actually has δn^3 triangles. This means there are $\delta n^3 - \varepsilon n^2$ triangles corresponding to nontrivial corners.

Let $L = \frac{\varepsilon}{\delta}$. Then for $n > L$, there is a nontrivial corner.

(b) Let A be a subset of $\{1, \ldots, n\}$ of density ε. Consider the subset A' of $\{1, \ldots, 2n\} \times \{1, \ldots, 2n\}$ consisting of (x, y) such that $y - x \in A + n$. Then A' has a corner iff A has a 3-AP. Moreover, each point of A corresponds to more than n points in A', so A' has at least $\frac{\varepsilon}{4} n^2$ edges, and hence has a corner for n large enough.

Bibliography

[1] R. Diestel. *Graph Theory*. Springer, 2005.

[2] Y. Lima. Szemerédi's regularity lemma. 2012.