1 Arithmetic Progressions

Our story begins with the following very famous result in Ramsey Theory:

Theorem (van der Waerden, 1927). *For any $r, k \in \mathbb{N}$, and any coloring of $\mathbb{N}$ with r colors (namely, for any function $f : \mathbb{N} \to \{1, \ldots, r\}$), there is a monochromatic k-term arithmetic progression, namely a sequence $a, a+d, a+2d, \ldots, a+(k-1)d$ such that*

$$f(a) = f(a+d) = \cdots = f(a+(k-1)d)$$

We won't prove this theorem, but the basic idea is that you apply a very clever induction argument on both r and k.

This theorem guarantees that whenever we partition $\mathbb{N}$ into subsets $S_1, \ldots, S_r$ (these are just $S_i = f^{-1}(i)$), then for any k, some S_i will contain a k-term arithmetic progression. A natural question to ask is: which one? Additionally, a natural guess is that the "biggest" one will be the one that contains a k-term arithmetic progression. To formalize this, we make the following definition.

Definition. Given a set $S \subseteq \mathbb{N}$, its *density* is defined as

$$d(S) = \lim_{N \to \infty} \frac{|S \cap [N]|}{N}$$

where $[N] = \{1, 2, \ldots, N\}$, assuming that this limit exists.

Example. The set of even numbers has density $1/2$, as does the set of odd numbers. The set of squares has density 0, since if $\mathbb{S}$ is the set of squares, then

$$|\mathbb{S} \cap [N]| \approx \sqrt{N}$$

so

$$d(\mathbb{S}) \approx \lim_{N \to \infty} \frac{\sqrt{N}}{N} = \lim_{N \to \infty} \frac{1}{\sqrt{N}} = 0$$

Similarly, the set $\mathbb{P}$ of primes also has density 0; this is because the Prime Number Theorem says that

$$|\mathbb{P} \cap [N]| \approx \frac{N}{\log N}$$

and thus

$$d(\mathbb{P}) \approx \lim_{N \to \infty} \frac{N/\log N}{N} = \lim_{N \to \infty} \frac{1}{\log N} = 0$$

Finally, $\mathbb{N}$ itself has density 1.

Lemma. *If $S, T \subseteq \mathbb{N}$ are disjoint sets, then*

$$d(S \cup T) = d(S) + d(T)$$

Proof.

$$d(S \cup T) = \lim_{N \to \infty} \frac{|(S \cup T) \cap [N]|}{N} = \lim_{N \to \infty} \frac{|S \cap [N]| + |T \cap [N]|}{N} = d(S) + d(T)$$

□

One consequence of this lemma is that when we color $\mathbb{N}$ with r colors, then one of the color classes S_i must have strictly positive density. So one way of phrasing our "biggest" conjecture above is the following:

Conjecture (Erdős–Turán, 1936). *If $S \subseteq \mathbb{N}$ has positive density, then it contains a k-term arithmetic progression for any $k \in \mathbb{N}$.*

The first progress towards this theorem was made almost 20 years later:

Theorem (Roth, 1953)**.** *If $S \subseteq \mathbb{N}$ has positive density, then it contains a 3-term arithmetic progression.*

Finally, the full Erdős–Turán Conjecture was resolved by Szemerédi:

Theorem (Szemerédi 1969, Szemerédi 1975)**.** *If $S \subseteq \mathbb{N}$ has positive density, then it contains a k-term arithmetic progression for any $k \in \mathbb{N}$. The $k = 4$ case was proven first, with the full case coming six years later.*

A key component of Szemerédi's proof is the so-called Szemerédi Regularity Lemma, which is actually a statement about graphs. In addition to being an extremely deep and important statement in and of itself, it also demonstrates a remarkable and surprising connection between number theory and graph theory. We won't state or prove the Regularity Lemma yet, but will begin with one of its most important consequences.

Before that, it is worthwhile to mention two more major ideas related to Szemerédi's Theorem. The first is the following result, which is considered one of the most important advances in number theory of recent years:

Theorem (Green–Tao, 2004)**.** *For every k, there is a k-term arithmetic progression in the primes, namely some $a, d \in \mathbb{N}$ such that $a, a + d, a + 2d, \ldots, a + (k - 1)d$ are all prime.*

This was a real breakthrough, and took many years and several hundred pages to prove. Note that this is not at all implied from Szemerédi's Theorem, since the primes have density 0, as discussed above. However, both Szemerédi's Theorem and the Green–Tao Theorem are implied by the following conjecture, which is considered the biggest open problem in this entire field.

Conjecture (Erdős)**.** *If $S = \{s_1, s_2, \ldots\} \subseteq \mathbb{N}$, and*

$$\sum_{i=1}^{\infty} \frac{1}{s_i} = \infty$$

then for every k, S contains a k-term arithmetic progression.

This implies the Green–Tao Theorem because Euler proved that the sum of the reciprocals of the primes diverges. Additionally, it implies Szemerédi's Theorem: intuitively, a set of density δ "should" be just a set of numbers that are each roughly $1/\delta$ apart, so we expect that

$$\sum_{i=1}^{\infty} \frac{1}{s_i} \approx \sum_{n=1}^{\infty} \frac{1}{n/\delta} = \delta \sum_{n=1}^{\infty} \frac{1}{n} = \infty$$

Thus, if Erdős's Conjecture were proved, then it would imply both Szemerédi's Theorem and the Green–Tao Theorem.

Exercise 1. Make this argument formal.

2 Triangle Removal

In order to both prove Roth's Theorem and demonstrate the powerful and surprising connection between graphs and arithmetic progressions, we will begin with a very important consequence of the Regularity Lemma, known as the Triangle Removal Lemma. Recall that a *triangle* in a graph G is a collection of three vertices that are all connected by edges.

Lemma (Ruzsa–Szemerédi, 1978)**.** *For every $\varepsilon > 0$, there exists a $\delta > 0$ such that every n and every graph G on n vertices with at most δn^3 triangles, we may remove εn^2 edges from G in order to make it triangle-free.*

Note that G may have up to $\binom{n}{2} \approx n^2/2$ edges, and up to $\binom{n}{3} \approx n^3/6$ triangles. Thus, what the Triangle Removal Lemma says is that if our graph has some small constant fraction of all possible triangles, then we can remove some small constant fraction of the edges to make it triangle-free. This is very surprising, because *a priori* it looks like we might need to remove m edges to remove m triangles—one for each triangle. However, this lemma says that once n is very large, we can in fact get away with removing *way* fewer edges.

Using this, we can prove Roth's Theorem, namely Szemerédi's Theorem for $k = 3$. First, we give a different (but equivalent) formulation of the theorem:

Theorem (Roth)**.** *For every $\varepsilon > 0$, there is some integer $N_0(\varepsilon) \in \mathbb{N}$ such that for all $N \geq N_0$ and any subset $S \subseteq [N]$ with $|S| \geq \varepsilon N$, S contains a 3-term arithmetic progression.*

This is equivalent to the previous formulation for the following reason: if $T \subseteq \mathbb{N}$ is a set of positive density, then its density is greater than some ε, so by setting $S = T \cap [N_0(\varepsilon)]$, we conclude that it has a 3-term arithmetic progression. Conversely, by gluing translated copies of a set $S \subseteq [N]$ with $|S| \geq \varepsilon N$, we get a set of positive density in $\mathbb{N}$, which must contain a 3-term arithmetic progression, and we can conclude that S contains one as well.

Exercise 2. Make this argument formal.

Proof. We will pick N_0 later. From such a set $A \subseteq [N]$, we construct a graph G as follows. Let X, Y, Z be three copies of the set $[3N]$, and then the vertices of G will be $X \cup Y \cup Z$. We put no edges inside X or Y or Z. Additionally, we connect $x \in X$ to $y \in Y$ if and only if $y - x \in A$, and we connect $y \in Y$ and $z \in Z$ if and only if $z - y \in A$. Finally, we connect $x \in X$ and $z \in Z$ if and only if $z - x \in 2A$, namely $z - x = 2a$ for some $a \in A$.

Now, for every $x \in [N]$ and $a \in A$, we automatically get a triangle in G, namely the triangle $x \in X, x + a \in Y, x + 2a \in Z$; indeed, by definition, all three of these vertices are pairwise adjacent. Therefore, each $a \in A$ yields at least N triangles in G, so we have at least $N|A| \geq \varepsilon N^2$ triangles in G. Moreover, all of these triangles are edge-disjoint, so in order to eliminate all of them, we'd need to delete at least εN^2 edges. Since G is a graph on $9N$ vertices, we can apply the contrapositive of the Triangle Removal Lemma (for $\varepsilon/9$) to conclude that there is some $\delta > 0$ such that G has at least δN^3 triangles. Now, we choose N_0 large enough that $\delta N_0^3 > \varepsilon N_0^2$. Then we conclude that if $N \geq N_0$, there must be some triangle in G that we haven't yet accounted for.

Since there is no edge within X, Y, or Z, this additional triangle must consist of some $x \in X, y \in Y, z \in Z$. Additionally, we necessarily have that $y - x \neq z - y$, for if these were equal to some a, then this triangle would just be one of the "simple" triangles we've already considered.

Therefore, we can define $a = y - x, b = \frac{z-x}{2}, c = z - y$. Then by the definition of the edges of G, we know that $a, b, c \in A$. On the other hand, we have that

$$b - a = \frac{z - x}{2} - (y - x) = \frac{z - x - 2y + 2x}{2} = \frac{z + x - 2y}{2}$$

and

$$c - b = (z - y) - \frac{z - x}{2} = \frac{2z - 2y - z + x}{2} = \frac{z + x - 2y}{2}$$

Thus, a, b, c forms a 3-term arithmetic progression contained entirely in A, as desired. □

3 Regularity

Definition. Let $G = (V, E)$ be a graph, and let $A, B \subseteq V$ be disjoint sets of vertices. By $e(A, B)$, we denote the number of edges with one endpoint in A and the other in B, namely

$$e(A, B) = |\{(u, v) \in V : u \in A, v \in B\}|$$

Additionally, the *edge density* between A and B is defined as

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

Thus, $d(A, B)$ measures what fraction of all possible edges between A and B are actually present.

Definition. Given a graph $G = (V, E)$, some $\varepsilon > 0$, and two disjoint subsets of vertices $A, B \subseteq V$, we say that the pair (A, B) is *ε-regular* if for every $A' \subseteq A, B' \subseteq B$ with $|A'| \geq \varepsilon|A|, |B'| \geq \varepsilon|B|$, we have

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

In other words, (A, B) is ε-regular if all the edges between A and B are "well-distributed" throughout A and B; no matter where we look in A and B, we see roughly the same density of edges.

Example (Basically the only example)**.** Suppose we fix some parameter $p \in (0, 1)$, and we put edges between A and B by picking them randomly: for every $a \in A, b \in B$, we connect a to b by flipping a p-biased coin and connecting them if and only if it comes up Heads. Then one can check that if ε isn't too small (namely $\varepsilon \gtrsim 1/\sqrt{|A| + |B|}$), then the pair (A, B) will be ε-regular in this graph we've defined (with very high probability). Intuitively, this is because the edges are indeed "well-distributed"—they were placed randomly, so how could they not be?

Indeed, this is more or less the only example: if a pair (A, B) is ε-regular, then we can pretty much pretend that it was gotten by putting edges in randomly with probability $p = d(A, B)$.

As it turns out, ε-regularity is a very strong and useful condition. For instance, it allows us to compute a huge number of quantities associated to the graph, by pretending that our graph is random and counting the associated quantities there. For instance, we have the following result:

Lemma (Counting Lemma)**.** *Let $A, B, C \subseteq V$ be disjoint sets of vertices, and suppose that each of the three pairs $(A, B), (B, C), (A, C)$ is ε-regular. Let*

$$r = d(A, B) \qquad s = d(B, C) \qquad t = d(A, C)$$

Then the number of triangles with one vertex in A, one in B, and one in C is approximately what we'd expect in a random graph, namely

$$rst|A||B||C|$$

More formally, if $r, s, t \geq 2\varepsilon$, then the number of such triangles is at least

$$(1 - 2\varepsilon)(r - \varepsilon)(s - \varepsilon)(t - \varepsilon)|A||B||C|$$

and one can also prove a similar upper bound.

Proof. For every $a \in A$, let $d_B(a), d_C(a)$ denote the number of neighbors of a in B, C, respectively. Then we first claim that the number of $a \in A$ such that $d_B(a) \leq (r - \varepsilon)|B|$ is at most $\varepsilon|A|$. For if not, the set of such a would form a subset $A' \subseteq A$ with $|A'| \geq \varepsilon|A|$ and such that $d(A', B) \leq r - \varepsilon$, which contradicts the ε-regularity. Similarly, there are at most $\varepsilon|A|$ vertices $a \in A$ such that $d_C(a) \leq (t - \varepsilon)|C|$.

Now, fix some $a \in A$ with $d_B(a) \geq (r - \varepsilon)|B|, d_C(a) \geq (t - \varepsilon)|C|$. Then let $B' \subseteq B$ be the set of neighbors of a in B, and define $C' \subseteq C$ similarly. Then since we assumed that $r, t \geq 2\varepsilon$, we get that $|B'| \geq \varepsilon|B|, |C'| \geq \varepsilon|C|$, so by regularity of the pair (B, C), we know that

$$|d(B', C') - s| < \varepsilon$$

and thus

$$d(B', C') \geq s - \varepsilon$$

and thus

$$e(B', C') \geq (s - \varepsilon)|B'||C'| \geq (s - \varepsilon)(r - \varepsilon)(t - \varepsilon)|B||C|$$

However, every edge between B' and C' yields a triangle containing a, since a is adjacent to all vertices in B', C'. Now, we sum over the $\geq (1 - 2\varepsilon)|A|$ vertices $a \in A$ that have $d_B(a) \geq (r - \varepsilon)|B|, d_C(a) \geq (t - \varepsilon)|C|$. Doing this, we get that the number of triangles going between A, B, C is at least

$$(1 - 2\varepsilon)(s - \varepsilon)(r - \varepsilon)(t - \varepsilon)|A||B||C|$$

□

Hopefully you have been convinced that ε-regularity is a very strong and very useful notion: if we know that some pair of vertex-sets in our graph is ε-regular, then we can more or less pretend that part of the graph is random. This makes the following result, the long-awaited Szemerédi Regularity Lemma, so surprising.

Theorem (Szemerédi)**.** *Let $\varepsilon > 0$. Then there exist integers M such that we may partition the vertices of G as*

$$V = C_1 \cup C_2 \cup \cdots \cup C_k$$

where $k \leq M$ so that

1. *Each C_i has the same size, plus or minus 1.*
2. *At most $\varepsilon\binom{k}{2}$ of the pairs (C_i, C_j) for $1 \leq i < j \leq k$ are **not** ε-regular.*

In other words, we can always partition any graph into a collection of "clusters" of essentially equal size, in such a way that essentially all of the pairs of clusters are ε-regular. Crucially, the number k of clusters is bounded by M, which depends *only* on ε; thus, the "complexity" of the partition depends only on how regular we require our partition to be. In particular, once N is much larger than M, then all graphs on N vertices are basically the same: they are composed of M clusters, and look like they were randomly generated from these clusters. In other words, all big graphs are basically the same.

Before talking about the proof of the lemma, let's see how it implies the Triangle Removal Lemma we proved yesterday. Recall that that lemma said that for any $\varepsilon > 0$, there is some $\delta > 0$ such that if a graph has $\leq \delta n^3$ triangles, then we can remove $\leq \varepsilon n^2$ edges to make it triangle-free.

Proof. First, using the Regularity Lemma, we can find some $(\varepsilon/2)$-regular partition of G, namely we can write

$$V(G) = C_1 \cup \cdots \cup C_k$$

such that all but $\varepsilon\binom{k}{2}/2$ of the pairs (C_i, C_j) are not $(\varepsilon/2)$-regular. Moreover, since refining a partition only makes more pairs more regular, we can assume that $k \geq 2/\varepsilon$. Now, we're going to remove a bunch of edges from G:

1. For every non-ε-regular pair (C_i, C_j), we will remove all edges between C_i and C_j. Since there are at most $\varepsilon\binom{k}{2}/2$ such pairs, and each pair contains at most $(n/k)^2$ edges (since $|C_i| = n/k \pm 1$), this step removes at most
$$\frac{\varepsilon}{2}\binom{k}{2}\left(\frac{n}{k}\right)^2 \leq \frac{\varepsilon}{2}\frac{k^2}{2}\frac{n^2}{k^2} = \frac{\varepsilon}{4}n^2$$
edges.
2. Within each cluster C_i, we remove all edges (namely the edges that have both endpoints in C_i). Since $|C_i| \approx n/k$, this is at most $\binom{n/k}{2}$ edges per cluster, and thus at most
$$k\binom{n/k}{2} \leq k\frac{(n/k)^2}{2} = \frac{n^2}{2k} \leq \frac{\varepsilon}{4}n^2$$
edges.

3. Between every pair of clusters (C_i, C_j) with $d(C_i, C_j) \leq \varepsilon$, we remove all edges. Since there are at most $\binom{k}{2}$ such pairs, and each pair has at most $\varepsilon(n/k)^2$ edges, this step removes at most
$$\binom{k}{2}\frac{\varepsilon n^2}{k^2} \leq \frac{k^2}{2}\frac{\varepsilon n^2}{k^2} = \frac{\varepsilon}{2}n^2$$
edges.

Thus, all in all, we have removed $\leq \varepsilon n^2$ edges to get a subgraph; let's call this subgraph G'. If G' is triangle-free, then we're done. If not, then we want to prove that G must have started with many triangles, namely at least δn^3 of them.

Since G' has a triangle, and since we deleted all edges internal to every cluster, such a triangle must go between three clusters C_i, C_j, C_k, and we must have never deleted the edges between any of these pairs. Additionally, since we deleted all edges between irregular pairs, we get that all three of these pairs are $(\varepsilon/2)$-regular. So if we apply the counting lemma we proved earlier to these three $(\varepsilon/2)$-regular pairs, we find that they contain at least
$$(1-\varepsilon)\left(d(C_i, C_j) - \frac{\varepsilon}{2}\right)\left(d(C_j, C_k) - \frac{\varepsilon}{2}\right)\left(d(C_i, C_k) - \frac{\varepsilon}{2}\right)|C_i||C_j||C_k|$$
triangles. Since we removed all edges between pairs that had density $\leq \varepsilon$, we get that all these densities are at least ε. So we have at least
$$(1-\varepsilon)\left(\varepsilon - \frac{\varepsilon}{2}\right)^3\left(\frac{n}{k}\right)^3 = \frac{(1-\varepsilon)\varepsilon^3}{8k^3}n^3$$
triangles in G', and thus at least that many triangles in G. So setting $\delta = (1-\varepsilon)\varepsilon^3/8k^3$ gives us the desired result. □

Finally, we can sketch the proof of the Regularity Lemma. The proof, as it turns out, is surprisingly straightforward: more or less, we just use the greedy algorithm to build better and better partitions, until we finally get one that satisfies the conditions we want it to.

Proof. The most important tool that we need is the so-called *mean-square density* of a partition, which is basically a weighted average of the squares of the densities between all pairs of parts. Specifically, if $\mathcal{C} = (C_1, \ldots, C_k)$ is some partition of V, then we define
$$E(\mathcal{C}) = \sum_{i=1}^{k}\sum_{j=1}^{k}\frac{|C_i||C_j|}{n^2}d(C_i, C_j)^2$$
The three most important properties of the mean-square density are the following:

1. For any partition $\mathcal{C}$, $0 \leq E(\mathcal{C}) \leq 1$. This is because it is a weighted average of the numbers $d(C_i, C_j)^2$, which are all between 0 and 1.

2. If $\mathcal{C}'$ is a refinement of the partition $\mathcal{C}$ (namely every part of $\mathcal{C}'$ is a subset of some part of $\mathcal{C}$), then $E(\mathcal{C}') \geq E(\mathcal{C})$. This is basically an application of the Cauchy-Schwarz inequality.

3. Finally, if $\mathcal{C}$ is far away from having the desired properties (namely it has too many non-ε-regular pairs), then we can refine it to some partition $\mathcal{C}'$ so that
$$E(\mathcal{C}') \geq E(\mathcal{C}) + \varepsilon^4$$

With these three properties, the proof is quite straightforward. We start with any partition of the vertices. If it has the desired properties, we're done. If not, then we take a refinement as in property (3), and we've increased the mean-square density by at least ε^4. Since we started at some non-negative mean-square density, and our final mean-square density is always upper-bounded by 1, we can do this at most ε^{-4} times. So once we do this that many times, we must have reached a partition with the properties we wanted. In the end, in order to make all the clusters have roughly the same size, we take one final refinement with that property.

The one mysterious property is (3). The idea is as follows: if (C_i, C_j) is a pair of clusters that's not ε-regular, then we can find subsets $X \subseteq C_i, Y \subseteq C_j$ with $|X| \geq \varepsilon |C_i|, |Y| \geq \varepsilon |C_j|$ such that

$$|d(C_i, C_j) - d(X, Y)| \geq \varepsilon$$

Now, we replace C_i in our partition by the two clusters $X, C_i \setminus X$, and similarly for Y. Now, we actually do this for *all* pairs (C_i, C_j), thus splitting each cluster into 2^{k-1} subclusters by taking the intersections of all these splittings. One can show, again using the Cauchy-Schwarz inequality, that this actually produces the desired ε^4 increase in the mean-square density, which completes the proof. □