SZEMERÉDI'S REGULARITY LEMMA

YURI LIMA

Abstract. Szemerédi's regularity lemma is an important tool in discrete mathematics, specially in graph theory and additive combinatorics. It says that, in some sense, all graphs can be approximated by random-looking graphs. The lemma helps in proving theorems for arbitrary graphs whenever the corresponding result is easy for random graphs. One of its applications is the triangle removal lemma which, as observed by Ruzsa and Szemerédi in [12], gives a proof of Roth's theorem on the existence of arithmetic progressions of length 3 in subsets of the integers with positive density. It also implies the corner's theorem - which indeed is a strengthening of Roth's theorem. Our last application is the graph removal lemma.

1. Additive combinatorics

Additive combinatorics is the theory of
counting additive structures in sets.

T. Tao and V. Vu.

This theory has seen exciting developments and dramatic changes in direction in recent years, thanks to its connections with areas such as number theory, ergodic theory and graph theory. This section gives a brief historic introduction on the main results.

Van der Waerden's theorem, one of Kintchine's "Three Pearls of Number Theory", states that whenever the natural numbers are finitely partitioned (or, as it is customary to say, finitely colored), one of the cells of the partition contains arbitrarily long arithmetic progressions. In other words, the structure of the natural numbers can not be destroyed by partitions: arbitrarily large parts of $\mathbb{N}$ persist inside some component of the partition. This result was first proved in 1927 and represents the first great result on additive combinatorics. Afterwards, in the mid-thirties, Erdös and Turán [5] conjectured a density version of van der Waerden's theorem. To present it, let us define what is the notion of density in the natural numbers.

Definition 1.1. Given a set $A \subset \mathbb{N}$, the *upper density* of A is

$$\overline{\mathrm{d}}(A) \doteq \limsup_{n\to\infty} \frac{|A \cap \{1,2,\ldots,n\}|}{n}\,.$$

If the limit exists, we say that A has density, and denote it by $\mathrm{d}(A)$. As pointed out by Erdös and Turán, having positive upper density is a notion of largeness and it is natural to ask if sets with this property have arbitrarily long arithmetic progressions. This quite recalcitrant question was only settled in 1975 by Szemerédi [15]. Meanwhile, the first partial result was obtained by Roth [11] in 1953.

Date: January, 29, 2012.

Key words and phrases. corner's theorem, graph removal lemma, Roth's theorem, Szemerédi's regularity lemma, Szemerédi's theorem, triangle removal lemma.

Theorem 1.2 (Roth)**.** *If $A \subset \mathbb{N}$ has positive upper density, then it contains an arithmetic progression of length* 3.

His proof relied on a Fourier-analytic argument of energy increment for functions: one decomposes a function f as $g+b$, where g is good and b is bad in a specific sense[1]. If the effect of b is large, it is possible to break it into good and bad parts again and so on. In each step, the "energy" of b increases a fixed amount. Being bounded, it must stop after a finite number of steps. At the end, g controls the behavior of f and for it the result is straightforward. See [10] for further details.

After this, in 1969, Szemerédi [14] extended Roth's theorem to

Theorem 1.3 (Szemerédi)**.** *If $A \subset \mathbb{N}$ has positive upper density, then it contains an arithmetic progression of length* 4.

Finally, six years later, Szemerédi settled the conjecture in its full generality.

Theorem 1.4 (Szemerédi)**.** *If $A \subset \mathbb{N}$ has positive upper density, then it contains arbitrarily long arithmetic progression.*

His proof required a complicated combinatorial argument and relied on a graph-theoretical result, known as **Szemerédi's regularity lemma**, which turned out to be an important result in graph theory. It asserts, roughly speaking, that any graph can be decomposed into a relatively small number of disjoint subgraphs, most of which behave pseudo-randomly. This is the main topic of these notes.

It is worth to mention Erdös and Turán also conjectured that if $A \subset \mathbb{N}$ satisfies

$$\sum_{n\in A}\frac{1}{n}=\infty\,,$$

then it contains arbitrarily long arithmetic progressions. This question is wide open: nobody knows even if A contains arithmetic progressions of length 3. On the other hand, a remarkable result of Green and Tao states the conjecture for the particular case[2] of the prime numbers.

Theorem 1.5 (Green and Tao)**.** *The prime numbers contain arbitrarily long arithmetic progressions.*

2. Setting notation

$G=(V,E)$ is a graph, where V is a finite set of *vertices* and E is the set of *edges*, each of them joining two distinct elements of V. For disjoint $A,B\subset V$, $e(A,B)$ is the number of edges between A and B and

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

is the *density* of the pair (A,B).

Definition 2.1. For $\varepsilon>0$ and disjoint subsets $A,B\subset V$, the pair (A,B) is *ε-regular* if, for every $X\subset A$ and $Y\subset B$ satisfying

$$|X|\geq\varepsilon\cdot|A| \text{ and } |Y|\geq\varepsilon\cdot|B|$$

we have

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

[1]This follows the same philosophy of Calderón-Zygmund's theory on harmonic analysis.

[2]The sum of the inverse of the primes diverges.

A partition $\mathcal{U} = \{V_0, V_1, \ldots, V_k\}$ of V into pairwise disjoint sets in which V_0 is called the *exceptional set* is an *equipartition* if $|V_1| = \cdots = |V_k|$. We view the exceptional set as $|V_0|$ distinct parts, each consisting of a single vertex, and its role is purely technical: to make all other classes have exactly the same cardinality.

Definition 2.2. An equipartition $V = V_0 \cup V_1 \cup \cdots \cup V_k$ is ε*-regular* if

(a) $|V_0| \leq \varepsilon \cdot |V|$,
(b) all but at most εk^2 of the pairs (V_i, V_j) are ε-regular.

The classes V_i are called *clusters* or *groups*. Given two partitions $\mathcal{U}, \mathcal{W}$ of V, we say $\mathcal{U}$ *refines* $\mathcal{W}$ if every cluster of $\mathcal{W}$ is equal to the union of some clusters of $\mathcal{U}$.

3. Szemerédi's regularity lemma

Szemerédi's regularity lemma says that every graph with many vertices can be partitioned into a small number of clusters with the same cardinality, most of the pairs being ε-regular, and a few leftover edges. In my point of view, this result allows the decomposition of every graph with a sufficiently large number of vertices into many components uniformly (every component has the same number of vertices) in such a way the relation of the clusters is at the same time

uniform: the densities do not vary too much, and

randomic: even controlling the density, nothing can be said about the distribution of the edges.

As a toy model, let $0 \leq p \leq 1$ and consider the complete random graph $G = (V, E)$ with n vertices in which every edge belongs to E with probability p. If A, B are disjoint subsets of V, the expected value of $d(A, B)$ is p, and the same happens for subsets $X \subset A$, $Y \subset B$. Szemerédi's regularity lemma says that, approximately, this is indeed the universal behavior.

Theorem 3.1 (Szemerédi's regularity lemma)**.** *For every $\varepsilon > 0$ and every integer t, there exist integers $T(\varepsilon, t)$ and $N(\varepsilon, t)$ for which every graph with at least $N(\varepsilon, t)$ vertices has an ε-regular equipartition $(V_0, V_1, \ldots, V_k)$, where $t \leq k \leq T(\varepsilon, t)$.*

Note the importance of having an upper bound for the number of clusters. Otherwise, we could just take each of them to be a singleton.

The idea in the proof is similar to Roth's approach. Start with an arbitrary partition of V into t disjoint classes $V_1, \ldots, V_t$ of equal sizes. Proceed by showing that, as long as the partition is not ε-regular, it can be refined in a way to distribute the density deviation. This is done by introducing a bounded *energy function* that increases a fixed amount every time the refinement is made. After a finite number of steps, the resulting partition is ε-regular.

We now discuss what should be the energy function. The natural way of looking for it is to identify the obstruction for a pair (U, W) to be ε-regular. This means there are subsets $U_1 \subset U$ and $W_1 \subset W$ such that $|U_1| \geq \varepsilon \cdot |U|$, $|W_1| \geq \varepsilon \cdot |W|$ and

$$|d(U_1, W_1) - d(U, W)| > \varepsilon.$$

Consider the partitions $\mathcal{U} = \{U_1, U\backslash U_1\}$ and $\mathcal{W} = \{W_1, U\backslash W_1\}$. The above inequality has the following probabilistic interpretation. Consider the random variable Z defined on the product $U \times W$ by: let u be a uniformly random element of U and w a uniformly random element of W, let $A \in \mathcal{U}$ and $B \in \mathcal{W}$ be those members of the respective partitions for which $u \in A$ and $w \in B$, and take

$$Z(u, w) \doteq d(A, B)\,.$$

The expectation of Z is equal to

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{\substack{A\in\mathcal{U}\\ B\in\mathcal{W}}} \frac{|A|}{|U|}\cdot\frac{|B|}{|W|}\cdot d(A,B) \\ &= \frac{1}{|U|\cdot|W|}\sum_{\substack{A\in\mathcal{U}\\ B\in\mathcal{W}}} e(A,B) \\ &= d(U,W). \end{aligned}$$

By assumption, Z deviates from $\mathbb{E}[Z]$ at least ε whenever $u \in U_1$, $w \in W_1$ and this event has probability

$$\frac{|U_1|}{|U|}\cdot\frac{|W_1|}{|W|} \geq \varepsilon^2.$$

Then $\mathrm{Var}[Z] \geq \varepsilon^4$. Noting that the expectation of Z^2 is

$$\begin{aligned} \mathbb{E}[Z^2] &= \sum_{\substack{A\in\mathcal{U}\\ B\in\mathcal{W}}} \frac{|A|}{|U|}\cdot\frac{|B|}{|W|}\cdot d^2(A,B) \\ &= \frac{1}{|U|\cdot|W|}\sum_{\substack{A\in\mathcal{U}\\ B\in\mathcal{W}}} \frac{e^2(A,B)}{|A|\cdot|B|}\,, \end{aligned}$$

we conclude that

$$\begin{aligned} \mathbb{E}[Z^2] &\geq \mathbb{E}[Z]^2+\varepsilon^4 \\ \frac{1}{|U|\cdot|W|}\sum_{\substack{\mathbf{A}\in\mathcal{U}\\ \mathbf{B}\in\mathcal{W}}} \frac{\mathbf{e^2(A,B)}}{|\mathbf{A}|\cdot|\mathbf{B}|} &\geq \frac{1}{|U|\cdot|W|}\cdot\frac{\mathbf{e^2(U,W)}}{|\mathbf{U}|\cdot|\mathbf{W}|}+\varepsilon^4. \end{aligned} \tag{3.1}$$

The boldface terms above represent the energy function we are looking for: given two disjoint subsets $A, B \subset V$, define

$$q(A,B) = \frac{1}{n^2}\cdot\frac{e^2(A,B)}{|A|\cdot|B|} = \frac{|A|\cdot|B|}{n^2}\cdot d^2(A,B)\,.$$

For partitions $\mathcal{U}, \mathcal{W}$, let

$$q(\mathcal{U},\mathcal{W}) = \sum_{\substack{A\in\mathcal{U}\\ B\in\mathcal{W}}} q(A,B)\,.$$

Definition 3.2. Given a partition $\mathcal{U}$ of V with exceptional set V_0, the *index* of $\mathcal{U}$ is

$$q(\mathcal{U}) = \sum_{A,B\in\mathcal{U}} q(A,B),$$

where the sum ranges over all unordered pairs of distinct parts A, B of $\mathcal{U}$, with each vertex of V_0 forming a singleton part in its own.

Note that $q(\mathcal{U})$ is a sum of $\binom{k+|V_0|}{2}$ terms of the form $q(A,B)$. The first good property it must have is boundedness.

Property 1. $q(\mathcal{U}) \leq 1/2$.

In fact, as $d(A,B) \leq 1$,

$$\begin{aligned} q(\mathcal{U}) &\leq \frac{1}{n^2} \sum_{\substack{A,B\in\mathcal{U} \\ A\neq B}} |A|\cdot|B| \\ &\leq \frac{1}{2n^2}\cdot\left(\sum_{A\in\mathcal{U}}|A|\right)\cdot\left(\sum_{B\in\mathcal{U}}|B|\right) \\ &= \frac{1}{2}\,. \end{aligned}$$

It is also monotone increasing with respect to refinements. This is the content of the next two properties.

Property 2. If U, W are subsets of V and $\mathcal{U}, \mathcal{W}$ are partitions of U, V, respectively, then

$$q(\mathcal{U},\mathcal{W}) \geq q(U,W)\,.$$

This property follows easily from Cauchy-Schwarz inequality[3], but this analytical argument is not so clear. A soft way of proving it is to consider the probabilistic point of view, with the aid of the random variable Z. According to the above calculations,

$$\mathbb{E}[Z]^2 = \frac{n^2}{|U|\cdot|W|}\cdot q(U,W) \quad \text{and} \quad \mathbb{E}[Z^2] = \frac{n^2}{|U|\cdot|W|}\cdot q(\mathcal{U},\mathcal{W})$$

and so, by Jensen's inequality,

$$\begin{aligned} & \mathbb{E}[Z^2] &\geq& \ \mathbb{E}[Z]^2 \\ \Longrightarrow \quad & q(\mathcal{U},\mathcal{W}) &\geq& \ q(U,W)\,. \end{aligned}$$

Property 3. If $\mathcal{U}'$ refines $\mathcal{U}$, then

$$q(\mathcal{U}') \geq q(\mathcal{U})\,.$$

[3]The interested reader may check it in [9].

This is a direct consequence of Property 2 by breaking $q(\mathcal{U}')$ according to $\mathcal{U}$:

$$\begin{aligned} q(\mathcal{U}') &= \sum_{A',B'\in\mathcal{U}'} q(A',B') \\ &= \sum_{A,B\in\mathcal{U}} \sum_{\substack{A'\subset A\\ B'\subset B}} q(A',B') \\ &= \sum_{A,B\in\mathcal{U}} q(\mathcal{U}'\cap A, \mathcal{U}'\cap B) \\ &\geq \sum_{A,B\in\mathcal{U}} q(A,B) \\ &= q(\mathcal{U}). \end{aligned}$$

The next property grows the index of non ε-regular partitions and reflects the right choice of the energy function. In a few words, it says that

"The lack of uniformity implies energy increment"

and this idea permeates many results in recent developments in combinatorics, harmonic analysis, ergodic theory and others areas. Actually, all known proofs of Szemerédi's theorem use this principle at some stage. To mention some of them:

1. the original proof of Roth considers good and bad parts of functions.
2. Furstenberg's approach [7]: every non-compact system has a weak mixing factor.
3. the Fourier-analytic proof of Gowers [7] identifies arithmetic progressions via the nowadays called *Gowers norms*.
4. the construction of characteristic factors for multiple ergodic averages uses the *Gowers-Host-Kra seminorms*.

These two last results are still being developed to generate what is being called *higher-order Fourier analysis*. Going back to what matters, let's prove the

Proposition 3.3 (Lack of uniformity implies energy increment 1)**.** *Suppose $\varepsilon > 0$ and U, W are disjoint nonempty subsets of V and the pair (U, W) is not ε-regular. Then there are partitions $\mathcal{U} = \{U_1, U_2\}$ of U and $\mathcal{W} = \{W_1, W_2\}$ of W such that*

$$q(\mathcal{U},\mathcal{W}) > q(U,W) + \varepsilon^4 \cdot \frac{|U|\cdot|W|}{n^2}.$$

Proof. The reader must convince himself that this is exactly relation (3.1). For those still not convinced, let's do it again. Assume $U_1 \subset U$ and $W_1 \subset W$ are such that $|U_1| \geq \varepsilon \cdot |U|$, $|W_1| \geq \varepsilon \cdot |W|$ and

$$|d(U_1,W_1) - d(U,W)| > \varepsilon.$$

Consider $\mathcal{U} = \{U_1, U\backslash U_1\}$ and $\mathcal{W} = \{W_1, U\backslash W_1\}$. The evaluation of the variation $\mathrm{Var}[Z]$ will prove the proposition. On one hand, by the calculations in Property 2,

$$\mathrm{Var}[Z] = \frac{n^2}{|U|\cdot|W|}\cdot\left(q(\mathcal{U},\mathcal{W}) - q(U,W)\right). \tag{3.2}$$

On the other, Z deviates from $\mathbb{E}[Z]$ at least ε whenever $u \in U_1$, $w \in W_1$ and this event has probability

$$\frac{|U_1|}{|U|}\cdot\frac{|W_1|}{|W|} \geq \varepsilon^2.$$

Then $\mathrm{Var}[Z] \geq \varepsilon^4$ which, together with (3.2), gives that

$$\begin{aligned} q(\mathcal{U},\mathcal{W}) - q(U,W) &\geq \varepsilon^4 \cdot \frac{|U|\cdot|W|}{n^2} \\ \Longrightarrow \qquad q(\mathcal{U},\mathcal{W}) &\geq q(U,W) + \varepsilon^4 \cdot \frac{|U|\cdot|W|}{n^2} \,. \end{aligned}$$

□

Proposition 3.4 (Lack of uniformity implies energy increment 2)**.** *Suppose* $0 < \varepsilon < 1/4$ *and let* $\mathcal{U} = \{V_0, V_1, \ldots, V_k\}$ *be a non* ε*-regular equipartition of* V*, where* V_0 *is the exceptional set. Then there exists a refinement* $\mathcal{U}' = \{V_0', V_1', \ldots, V_l'\}$ *of* $\mathcal{U}$ *with the following properties:*

(i) $\mathcal{U}'$ *is an equipartition of* V*,*
(ii) $k < l < k \cdot 8^k$*,*
(iii) $|V_0'| \leq |V_0| + n/2^k$ *and*
(iv) $q(\mathcal{U}') \geq q(\mathcal{U}) + \varepsilon^5/2$*.*

Proof. The idea is to apply the previous proposition to every non-regular pair. As there are at least εk^2 of them, the index will increase the fixed amount. Let c be the cardinality of every V_i, $i = 1, \ldots, k$. Saying that $\mathcal{U}$ is not ε-regular means that, for at least εk^2 pairs (i,j), $1 \leq i < j \leq k$, (V_i, V_j) is not ε-regular. For each of these, let $\mathcal{U}_{ij}$, $\mathcal{U}_{ji}$ be the partitions of V_i, V_j, respectively, given by Proposition 3.3 and consider $\mathcal{W}$ the smallest partition that refines $\mathcal{U}$ and all $\mathcal{U}_{ij}$, $\mathcal{U}_{ji}$. By Proposition 3.3,

$$\begin{aligned} q(\mathcal{W}) &\geq q(\mathcal{U}) + \varepsilon k^2 \cdot \left(\varepsilon^4 \cdot \frac{c^2}{n^2}\right) \\ &= q(\mathcal{U}) + \varepsilon^5 \cdot \left(\frac{kc}{n}\right)^2 \\ &\geq q(\mathcal{U}) + \frac{\varepsilon^5}{2} \end{aligned}$$

as $kc = n - |V_0| \geq n/2$. This proves that $\mathcal{W}$ (and any of its refinements) satisfies (iv). The problem is that $\mathcal{W}$ is not necessarily an equipartition. We adjust this by defining $b = \lfloor c/4^k \rfloor$, splitting every part of $\mathcal{W}$ arbitrarily into disjoint sets of size b and throwing the remaining vertices of each part, if any, to the exceptional set. This new partition $\mathcal{U}'$ satisfies (i), (ii) and (iii), as we'll verify below.

(i) $\mathcal{U}'$ is an equipartition by definition.

(ii) To get $\mathcal{W}$, every cluster of $\mathcal{U}$ is divided in at most 2^{k-1} parts. After, every element of $\mathcal{W}$ is divided in at most 4^k non-exceptional parts. This implies that

$$l \leq k \cdot 2^{k-1} \cdot 4^k < k \cdot 8^k.$$

(iii) Each cluster of $\mathcal{W}$ contributes with at most b vertices to V_0' and so

$$|V_0'| \leq |V_0| + b \cdot \left(k \cdot 2^{k-1}\right) \leq |V_0| + kc \cdot \frac{2^{k-1}}{4^k} < |V_0| + \frac{n}{2^k} \,.$$

□

Finally, we are able to prove the regularity lemma.

Proof of Theorem 3.1. First, note that if the result is true for (ε, t) and $\varepsilon' > \varepsilon$, $t' < t$, then the result is also true for the pair (ε', t'). This allows us to assume that $\varepsilon < 1/4$ and t/ε is arbitrarily large.

Begin with an arbitrary partition $\mathcal{U} = \{V_0, V_1, \dots, V_t\}$ of V such that $|V_0| \leq \lfloor n/t \rfloor$ and $|V_1| = \cdots = |V_t| = \lfloor n/t \rfloor$. Apply Proposition 3.4 at most ε^{-5} times to obtain an equipartition $\mathcal{U}'$. Let $T(\varepsilon, t)$ be the largest number obtained by iterating the map $x \mapsto x \cdot 6^x$ at most ε^{-5} times, starting from t. Then $\mathcal{U}'$ has at most $T(\varepsilon, t)$ clusters. In addition, the cardinality of its exceptional set V_0' is bounded by

$$|V_0'| \leq |V_0| + \frac{1}{\varepsilon^5} \cdot \frac{n}{2^t} \leq \left\lfloor \frac{n}{t} \right\rfloor + \frac{n}{2^t \varepsilon^5}\,,$$

which is smaller than εn if t is large. This concludes the proof. □

4. Triangle removal lemma

Most applications of Szemerédi's regularity lemma deal with monotone problems, when throwing in more edges can only help. In these applications, one starts applying the original form of the regularity lemma to create a regular partition, then gets rid of all edges within the clusters of the partition, also the edges of non-regular pairs as well as those of regular pairs with small density. The leftover "pure" graph is much easier to handle and still contains most of the original edges. This is what happens in proving the triangle removal lemma.

The triangle removal lemma is the (intuitive, yet nontrivial) fact that if one has to delete at least εn^2 edges of a graph with n vertices to destroy all triangles in it, then the graph must contain at least δn^3 triangles, where $\delta = \delta(\varepsilon) > 0$. If one only thinks naively, the conclusion is that the graph contains at least εn^2 triangles, and the strength of the triangle removal lemma is that, instead of quadratic, the number of triangles is cubic. It was first proved by Ruzsa and Szemerédi [12], who also observed it implies Roth's theorem, as we shall see in the next section.

Definition 4.1. Given $\varepsilon > 0$, a graph $G = (V, E)$ is *ε-far from being triangle free* if one has to delete at least $\varepsilon \cdot |V|^2$ edges of G to destroy all triangles in it.

In particular, every graph that is ε-far from being triangle-free has at least one triangle (indeed, at least $\varepsilon \cdot |V|^2$ of them).

Theorem 4.2 (Triangle removal lemma)**.** *For any $0 < \varepsilon < 1$, there is $\delta = \delta(\varepsilon) > 0$ such that, whenever $G = (V, E)$ is ε-far from being triangle-free, then it contains at least $\delta \cdot |V|^3$ triangles.*

Proof. Let $G = (V, E)$ be an ε-far from being triangle-free graph and $|V| = n$. We can assume $n > N(\varepsilon/4, \lfloor 4/\varepsilon \rfloor)$ by just taking δ sufficiently small so that

$$\delta \cdot N(\varepsilon/4, \lfloor 4/\varepsilon \rfloor)^3 < 1\,.$$

Consider the $\varepsilon/4$-regular partition $\mathcal{U} = \{V_0, V_1, \dots, V_k\}$ given by Theorem 3.1. Let $c = |V_1| = \cdots = |V_k|$ and G' the graph obtained from G by deleting the following edges:

- All edges incident in V_0: there are at most $\varepsilon n^2/4$ edges.
- All edges inside the clusters $V_1, \dots, V_k$: the number of edges is at most

$$c^2 \cdot k < \frac{n^2}{k} < \frac{\varepsilon \cdot n^2}{4} \cdot$$

- All edges that lie in irregular pairs: there are less than

$$\left(\frac{\varepsilon}{4}\cdot k^2\right)\cdot c^2 < \frac{\varepsilon\cdot n^2}{4}$$

edges.
- All edges lying in a pair of clusters whose density is less than $\varepsilon/2$: their cardinality is at most

$$\binom{k}{2}\cdot\frac{\varepsilon\cdot c^2}{2} < \frac{\varepsilon\cdot n^2}{4}\cdot$$

The number of deleted edges is less than $\varepsilon\cdot n^2$ and so G' **contains a triangle**. The three vertices of such triangle belong to three remaining distinct clusters[4], let us say V_1, V_2, V_3. We'll show that in fact these clusters support many triangles.

Call a vertex $v_1 \in V_1$ *typical* if it has at least $\varepsilon c/4$ adjacent vertices in V_2 and at least $\varepsilon c/4$ adjacent vertices in V_3. As, by hypothesis,

$$d\left(V_i', V_j'\right) \geq \frac{\varepsilon}{4} \tag{4.1}$$

whenever $V_i' \subset V_i$, $V_j' \subset V_j$ have cardinality at least $\varepsilon c/4$, there are more than $c/2$ typical vertices in V_1. In fact, the number of vertices in V_1 with at least $\varepsilon c/4$ adjacent vertices in V_2 is greater than $(1-\varepsilon/4)\cdot c$. If this were not the case, the subset $V_1' \subset V_1$ of non-typical vertices would have more than $\varepsilon c/4$ elements and would satisfy

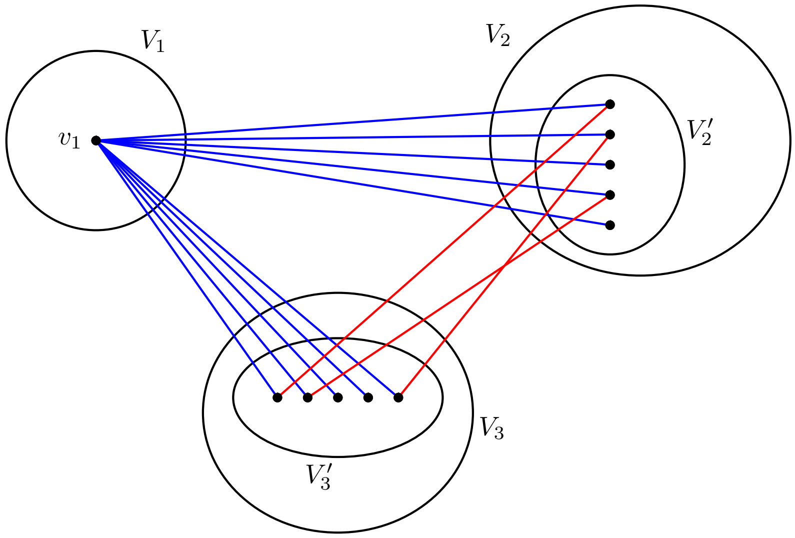
$$d\left(V_1', V_2\right) < \frac{|V_1'|\cdot\frac{\varepsilon c}{4}}{|V_1'|\cdot|V_2|} = \frac{\varepsilon}{4}\,,$$

contradicting (4.1). As the same argument holds to V_3, the number of typical vertices in V_1 is at least

$$\left(1-2\cdot\frac{\varepsilon}{4}\right)\cdot c > \frac{c}{2}\cdot$$

Let $v_1 \in V_1$ be one of them and consider $V_2' \subset V_2, V_3' \subset V_3$ the vertices adjacent to v_1.

[4]The existence of the triangle is merely used to guarantee the existence of V_1, V_2 and V_3.



Every edge between V_2' and V_3' generates a triangle. Observe that this number is at least

$$e(V_2', V_3') \geq \frac{\varepsilon}{4} \cdot |V_2'| \cdot |V_3'| \geq \frac{\varepsilon^3 \cdot c^2}{4^3} \cdot$$

Summing this up in $v_1 \in V_1$ typical, G' has at least $(\varepsilon c/4)^3$ triangles. Because $c > n/2T(\varepsilon/4, \lfloor 4/\varepsilon \rfloor)$, this quantity is greater or equal to

$$\left(\frac{\varepsilon}{4} \cdot \frac{n}{2 \cdot T(\varepsilon/4, \lfloor 4/\varepsilon \rfloor)} \right)^3 = \left(\frac{\varepsilon}{8 \cdot T(\varepsilon/4, \lfloor 4/\varepsilon \rfloor)} \right)^3 \cdot n^3 = \delta(\varepsilon) \cdot n^3.$$

□

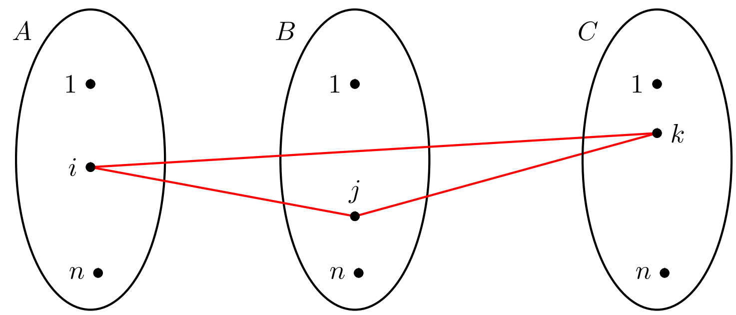
4.1. **Roth's theorem.** As an application of the triangle removal lemma, we prove Theorem 1.2.

Proof of Theorem 1.2. Assume that

$$|A \cap \{1, \ldots, n\}| > \varepsilon n\,,\ \forall\, n \geq n_0.$$

Consider a graph $G = (V, E)$ in the following way:

- $V = V_1 \cup V_2 \cup V_3$, where V_1, V_2, V_3 have $3n$ vertices labeled from 1 to $3n$ each.
- There is an edge from $i \in V_1$ to $j \in V_2$ iff $j - i \in A$.
- There is an edge from $j \in V_2$ to $k \in V_3$ iff $k - j \in A$.
- There is an edge from $i \in V_1$ to $k \in V_3$ iff $(k - i)/2 \in A$.



Then i, j, k form a triangle iff

$$\begin{cases} j - i = a_1 \in A \\ k - j = a_3 \in A \\ \dfrac{k-i}{2} = a_2 \in A \end{cases} \implies (a_1, a_2, a_3) \text{ is an arithmetic progression in } A,$$

that is, the triangles identify arithmetic progressions of length 3 in A, including the trivial ones (a, a, a), $a \in A$. There are more than $\varepsilon n \cdot n = \varepsilon n^2$ of these trivial triangles $i, i + a, i + 2a$ and they are all disjoint. This mere disjointness implies G is ε-far from being triangle-free and so, by Theorem 4.2, G has at least δn^3 triangles of which at least $\delta n^3 - 81n^2$ are non-trivial. The proof is complete by taking $n > 81\delta^{-1}$. □

4.2. **Corner's theorem.** This result was first proved by Ajtai and Szemerédi [1]. The simpler proof we present here, using the triangle removal lemma, was obtained by Solymosi [13]. We point out, and leave the proof to the reader, that the corner's theorem is a strengthening of Roth's theorem.

Definition 4.3. A *corner* is an axis-aligned isosceles triangle of $\mathbb{Z}^2$, that is, it is a set of three different elements of $\mathbb{Z}^2$ of the form

$$(x, y), (x + h, y) \text{ and } (x, y + h).$$

The corner's theorem states that every set of positive density has a corner.

Theorem 4.4 (Corner's theorem). *For every $\varepsilon > 0$, there exists $n > 0$ such that any subset of $\{1, \dots, n\} \times \{1, \dots, n\}$ with at least εn^2 points has a corner.*

Proof. We proceed similarly to the proof of Roth's theorem. Let A be a subset of $\{1, \dots, n\} \times \{1, \dots, n\}$ with at least εn^2 points and consider the tripartite graph $G = (V, E)$ defined by:

- $V = V_1 \cup V_2 \cup V_3$, where V_1, V_2 and V_3 represent the horizontal, vertical and diagonal lines of $\{1, \dots, n\} \times \{1, \dots, n\}$ respectively.
- There is an edge from a line of V_i to a line of V_j iff the intersection of the two lines belongs to A.

G has $|V_1| + |V_2| + |V_3| = n + n + 2n = 4n$ vertices. The triangles of G correspond to the corners of A, including the trivial ones $(x, y), (x, y), (x, y)$. G has more than $|A| \geq \varepsilon n^2$ of these trivial triangles and they are all disjoint, so that G is $\varepsilon/16$-far from being triangle-free. By the triangle removal lemma, G has at least δn^3 triangles of which at least $\delta n^3 - n^2$ are non-trivial and give the required corner. □

5. Graph removal lemma

The triangle removal lemma asserts that every graph for which it is necessary to throw a positive fraction of edges in order to destroy all triangles indeed has a positive fraction of triangles. The fact is that, as proved by Erdös, Frankl and Rödl [4], instead of fixing the triangle configuration, the result is also true for any fixed configuration. More formally, let H be a finite graph with h vertices and, analogously, consider the following

Definition 5.1. Given $\varepsilon > 0$, a graph $G = (V, E)$ is *ε-far from being H-free* if one has to delete at least $\varepsilon \cdot |V|^2$ edges of G to destroy all copies of H in it.

We then have the

Theorem 5.2 (Graph removal lemma)**.** *For any $0 < \varepsilon < 1$, there is $\delta = \delta(\varepsilon) > 0$ such that, whenever $G = (V, E)$ is ε-far from being H-free, then it contains at least $\delta \cdot |V|^h$ copies of H.*

The proof of this theorem is more intricate than that of the triangle removal lemma. Actually, it depends on the structure of the graph H. If, for example, H is a four-cycle, then the argument applied in the proof of the triangle removal lemma does not work, mainly because, once the "impure" edges are discarded, the copy of H that remains may have two vertices in a same cluster. In other words, the connectivity properties of H influence the distribution of the vertices along the clusters in a potential candidate for copy of H in G. As in the triangle removal lemma, this problem does not occur if H is the complete graph K_r in r vertices. For this reason, the proof of the graph removal lemma will be accomplished in three parts:

Part 1. The establishment of the graph removal lemma for K_r.

Part 2. We observe that, for a general H, the application of the same idea in Part 1 only guarantees the existence of r clusters, where r is the chromatic number of H.

Part 3. If we apply the same idea as in Part 1, allowing the choice of more than one vertex in a same cluster, we obtain the result for any H.

As remarked above, Part 1 follows the same lines of the proof of the triangle removal lemma: we clean out the graph and the remaining copy of K_r is supported in r different clusters, which indeed contain many copies of K_r. The construction of many copies is again accomplished by the typicality of most of the vertices, and is given by the following

Lemma 5.3. *If (A, B) is ε'-regular and $d(A, B) > \varepsilon$, then at least $(1 - \varepsilon')|A|$ vertices of A are adjacent to at least $(\varepsilon - \varepsilon')|B|$ vertices of B.*

Proof. Let $A' = \{v \in A\,;\, v \text{ is adjacent to less than } (\varepsilon - \varepsilon')|B| \text{ vertices of } B\}$. Then

$$d(A', B) < \frac{|A'| \cdot (\varepsilon - \varepsilon')|B|}{|A'| \cdot |B|} = \varepsilon - \varepsilon'. \tag{5.1}$$

If $|A'| \geq \varepsilon'|A|$, the ε'-regularity guarantees that

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

thus contradicting (5.1). □

Proof of the graph removal lemma for K_r. Let $G = (V, E)$ be ε-far from being K_r-free graph with $|V| = n > N((\varepsilon/6)^r, (\varepsilon/6)^{-r})$, and consider the $(\varepsilon/6)^r$-regular partition $\mathcal{U} = \{V_0, V_1, \dots, V_k\}$ given by Szemerédi's regularity lemma. Let $c = |V_1| = \cdots = |V_k|$ and G' the graph obtained from G by deleting the following edges:

- All edges incident in V_0: there are at most $(\varepsilon/6)^r \cdot n^2$ edges.
- All edges inside the clusters $V_1, \dots, V_k$: the number of edges is at most
$$c^2 \cdot k < \frac{n^2}{k} < (\varepsilon/6)^r \cdot n^2.$$
- All edges that lie in irregular pairs: there are less than
$$\left((\varepsilon/6)^r \cdot k^2\right) \cdot c^2 < (\varepsilon/6)^r \cdot n^2$$
edges.
- All edges lying in a pair of clusters whose density is less than $\varepsilon/3$: their cardinality is at most
$$\binom{k}{2} \cdot \frac{\varepsilon \cdot c^2}{3} < \frac{\varepsilon \cdot n^2}{3} \cdot$$

The number of deleted edges is less than $\varepsilon \cdot n^2$ and so G' **contains** K_r. The vertices of such K_r belong to r distinct remaining clusters, say $V_1, V_2, \dots, V_r$. We'll show that these clusters support many copies of K_r. This is done in r steps, the step i consisting of choosing a vertex v_i from V_i in such a way that v_i is adjacent to each of the previously chosen vertices $v_1, \dots, v_{i-1}$. If there are $\delta_i c$ ways of choosing v_i, where δ_i is independent of n, then G contains at least
$$(\delta_1 c) \cdots (\delta_r c) > \left(\frac{\delta_1 \cdots \delta_r}{2^r \cdot T((\varepsilon/6)^r, (\varepsilon/6)^{-r})^r}\right) \cdot n^r =: \delta \cdot n^r$$
copies of K_r and we're done.

By Lemma 5.3, at least $(1 - r \cdot (\varepsilon/6)^r)|V_1|$ points in V_1 are joined to at least $(\varepsilon/3 - (\varepsilon/6)^r)|V_j|$ points of V_j for each $j = 2, \dots, r$. Take one such point v_1 and denote by V_j^1 the subset of V_j of all the adjacent vertices to v_1, for each $j = 2, \dots, r$. We have
$$|V_j^1| \geq \left(\frac{\varepsilon}{3} - \left(\frac{\varepsilon}{6}\right)^r\right)|V_j| > \left(\frac{\varepsilon}{6}\right)|V_j|$$
and hence any two of the clusters $V_2^1, \dots, V_r^1$ are $(\varepsilon/6)^{r-1}$-regular and have density at least $\varepsilon/3 - (\varepsilon/6)^r$. This concludes the first step.

We now proceed to step 2: again by Lemma 5.3, at least $(1 - r \cdot (\varepsilon/6)^{r-1})|V_2^1|$ points in V_2^1 are joined to at least $(\varepsilon/3 - (\varepsilon/6)^r - (\varepsilon/6)^{r-1})|V_j^1|$ points of V_j^1 for each $j = 3, \dots, r$. Take one such point $v_2 \in V_2^1$ and denote by V_j^2 the subset of V_j^1 of all the adjacent vertices to v_2, for each $j = 3, \dots, r$. We have
$$|V_j^2| \geq \left(\frac{\varepsilon}{3} - \left(\frac{\varepsilon}{6}\right)^r - \left(\frac{\varepsilon}{6}\right)^{r-1}\right)|V_j^1| > \left(\frac{\varepsilon}{6}\right)|V_j^1|$$
and hence any two of the clusters $V_3^2, \dots, V_r^2$ are $(\varepsilon/6)^{r-2}$-regular and have density at least $\varepsilon/3 - (\varepsilon/6)^r - (\varepsilon/6)^{r-1}$.

Assuming, without loss of generality, that $r\varepsilon < 1$ and $\varepsilon/6 + \cdots + (\varepsilon/6)^r < \varepsilon/3$, the above procedure can be repeated r times. This concludes the proof. □

Now let H be an arbitrary graph. The *chromatic number* of H is the smallest number of colors needed to paint the vertices of H in such a way that no two adjacent vertices have the same color. Equivalently, it is the smallest r for which H is r-partite, that is, for which one can divide the vertices of H into r disjoint subsets such that no two vertices on the same subset are adjacent. Let these subsets have cardinality $h_1, \ldots, h_r$. Let $K_{h_1,\ldots,h_r}$ be the complete r-partite graph whose subsets have cardinality $h_1, \ldots, h_r$. Obviously, $K_{h_1,\ldots,h_r}$ contains H and so the number of copies of H in a given graph is at least the number of copies of $K_{h_1,\ldots,h_r}$.

Observe that if we apply the same idea as in Part 1 to a ε-far from being H-free graph, the remaining copy of H has vertices in at least r clusters $V_1, \ldots, V_r$, and not necessarily in h different clusters. This is not a problem: instead of choosing one vertex in each V_i, we choose h_i of them. If the same procedure works, each of these choices generates a copy of $K_{h_1,\ldots,h_r}$ and thus of H. This is how we proceed below.

Proof of the graph removal lemma. Apply the same argument as in Part 1 to obtain clusters $V_1, \ldots, V_r$ such that any pair is $(\varepsilon/6)^h$-regular and has density at least $\varepsilon/3$. We can thus find $(1 - r \cdot (\varepsilon/6)^h)|V_1|$ points of V_1 which are joined to at least $(\varepsilon/3 - (\varepsilon/6)^h)|V_i|$ points of V_i for each $i = 2, \ldots, r$. Take one such point $v_1 \in V_1$ and denote by V_i^1 the set of all vertices of V_i which are joined to v_1, for each $i = 2, \ldots, r$. Also, set $V_1^1 = V_1 \backslash \{v_1\}$ (here is the difference: we don't discard V_1). We have

$$|V_i^1| \geq (\varepsilon/3 - (\varepsilon/6)^h)|V_i| > (\varepsilon/6)|V_i|$$

for every $i = 1, \ldots, r$ and hence each pair among $V_1^1, \ldots, V_r^1$ is $(\varepsilon/6)^{h-1}$-regular and has density at least $\varepsilon/3 - (\varepsilon/6)^h$.

Repeating this argument successively h_1 times in the first cluster, h_2 times in the second cluster, ... , r times in the r-th cluster, we construct vertices $v_1, \ldots, v_h$ forming a copy of $K_{h_1,\ldots,h_r}$. This completes the proof. □

We point out a recent proof of the graph removal lemma avoiding the use of Szemerédi's regularity lemma has been obtained by Fox [6]. Although it does not use the regularity lemma, its idea is similar. Instead of using the *mean square density* given by the index

$$q(\mathcal{U}) = \sum_{A,B \in \mathcal{U}} q(A, B) = \sum_{A,B \in \mathcal{U}} \frac{|A|}{n} \cdot \frac{|B|}{n} \cdot d^2(A, B),$$

it uses a mean entropy density

$$\sum_{A,B \in \mathcal{U}} \frac{|A|}{n} \cdot \frac{|B|}{n} \cdot f(d(A, B)),$$

where $f(x) = x \log x$ for $0 < x \leq 1$ and $f(0) = 0$. Like in the regularity lemma, whenever a partition does not supports many copies of H, Fox shows, using a Jensen defect inequality, that there is a refinement of the partition that increases the mean square entropy a fixed amount.

We finish this post mentioning another version of the graph removal lemma that counts the number of induced graphs. A subgraph H of a graph G is said to be *induced* if any pair of vertices of H are adjacent if and only if they are adjacent in G. For example, K_5 has an induced K_4 but does not have an induced four-cycle. This shows that induced graphs are harder to find, and actually that the excess of edges might prevent them to exist. In this setting, we have to consider a new definition of ε-far from being H-free, in which one can remove or include edges.

Definition 5.4. Given $\varepsilon > 0$, a graph $G = (V, E)$ is *ε-unavoidable for H* if any graph that differs from G in no more that $\varepsilon \cdot |V|^2$ edges has an induced copy of H.

We thus have the graph removal lemma, proved by Alon, Fischer, Krivelevich and Szegedy [2].

Theorem 5.5 (Graph removal lemma for induced graphs)**.** *For any $0 < \varepsilon < 1$, there is $\delta = \delta(\varepsilon) > 0$ such that, whenever $G = (V, E)$ is ε-unavoidable for H, then it contains at least $\delta \cdot |V|^h$ induced copies of H.*

We won't prove the theorem. Instead, we refer the reader to the original paper.

References

1. **M. Ajtai and E. Szemerédi**, *Sets of lattice points that form no squares*, Studia Sci. Math. Hungar. **9** (1974), 9-11.
2. **N. Alon, E. Fischer, K. Krivelevich and M. Szegedy**, *Efficient testing of large graphs*, Combinatorica **20** (2000), no. 4, 451-476.
3. **N. Alon and J. Spencer**, *The probabilistic method*, 3rd edition, Wiley (2008).
4. **P. Erdös, P. Frankl and V. Rödl** , *The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent*, Graphs Combin. **2** (1986), 113-121.
5. **P. Erdös and P. Turán**, *On some sequences of integers*, J. London Math. Soc. **11** (1936), 261-264.
6. **J. Fox**, *A new proof of the graph removal lemma*, Annals of Mathematics **174** (2011), 561-579.
7. **H. Furstenberg**, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204-256.
8. **T. Gowers**, *A new proof of Szemerédi's theorem*, GAFA **11** (2001), 465-588.
9. **J. Komlós and M. Simonovits**, *Szemerédi's regularity lemma and its applications in graph theory*, Combinatorics, Paul Erdös is eighty, Vol. 2 (1993), 295–352, Bolyai Soc. Math. Stud., 2, János Bolyai Math. Soc., Budapest, 1996.
10. **A. Arbieto, C. Matheus and C.G. Moreira**, *The remarkable effectiveness of Ergodic Theory in Number Theory - Part I*, Ensaios Matemáticos **17** (2009), 1–74.
11. **K.F. Roth**, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 245-352.
12. **Z. Ruzsa and E. Szemerédi**, *Triple systems with no six points carrying three triangles*, Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai 18, Volume II, 939-945.
13. **J. Solymosi**, *Note on a generalization of Roth's theorem*, Algorithms Combin. **25** (2003), 825-827.
14. **E. Szemerédi**, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. **20** (1969), 89-104.
15. **E. Szemerédi**, *Integer sets containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299-345.
16. **L. Trevisan**, `http://lucatrevisan.wordpress.com/2009/05/13/the-triangle-removal-lemma/`.
17. **B.L. van der Waerden**, *Beweis einer Baudetschen Vermutung*, Nieuw. Arch. Wisk. **15** (1927), 212-216.
18. **T. Tao and V. Vu**, *Additive Combinatorics*, Cambridge studies in advanced mathematics, Cambridge University Press (2006).

Weizmann Institute of Science, Faculty of Mathematics and Computer Science, POB 26, 76100, Rehovot, Israel.

E-mail address: `yuri.lima@weizmann.ac.il`