SZEMERÉDI REGULARITY LEMMA

YUCHENG HUA

Abstract. The *Szemerédi Regularity Lemma* (SzRL) was first introduced to prove the *Erdős-Turán Conjecture* regarding upper densities. This work introduces the concept of regularity in graphs and SzRL, then provides a proof for the *Triangle Removal Lemma* with the SzRL, and eventually concludes with a proof for *Roth's theorem*, which states that all subsets of N with positive natural density contains an arithmetic progression of length three.

The introduction of the *Szemerédi Regularity Lemma* (SzRL) is motivated by the study of number theory. The density of certain subsets of natural numbers N has always been an interesting subject for studying. The conjecture of Erdős and Turán asserts the relation between upper density $\overline{d}(A) = \limsup_{n\to\infty} \frac{A\cap[n]}{n}$ of a set $A \subset \mathbb{N}$ and arbitrarily long arithmetic progression.

Conjecture 0.1 (Erdős & Turán). *If $\overline{d}(A) > 0$, then A contains arbitrarily long arithmetic progression.*

This conjecture was proved true by Endre Szemerédi in 1975; now it is commonly known as *Szemerédi's Theorem*. A key part of his proof is the *Szemerédi Regularity Lemma* (SzRL).

For the setup of the SzRL, we start by defining the density of a pair of subsets of vertices for some graph G as well as the concept of ϵ-regularity.

Definition 0.2. *For a graph G and two disjoint subsets $A, B \subset V(G)$ of the vertices, denote the number of edges between A and B as*

$$e(A, B) := |\{(a, b) \in A \times B : ab \in E(G)\}|.$$

*Define the **density** of the pair $\{A, B\}$ as*

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

Further, the pair (A, B) is said to be ϵ-regular for some $\epsilon > 0$ if

$$|d(A, B) - d(A', B')| = \left| \frac{e(A, B)}{|A||B|} - \frac{e(A', B')}{|A'||B'|} \right| \leq \epsilon$$

for every $A' \subset A$ and $B' \subset B$ with $|A'| \geq \epsilon|A|$ and $|B'| \geq \epsilon|B|$.

Date: July 3, 2023.

In other words, For all pairs of sufficiently large (A', B'), the induced bipartite graph $G[A, B]$ can be well approximated if (A, B) is ϵ-regular. The density of pairs can be seen as the probability of hitting an edge between a pair of subsets, and regularity can be seen as a description on the distribution. With this concept, we are ready to state the SzRL.

Theorem 0.3 (The Szemerédi Regularity Lemma)**.** *Let $\epsilon > 0$ and $m \in \mathbb{N}$. There exists a constant M dependent on m and ϵ such that for any graph G, there exists a partition $V(G) = A_0 \cup \cdots \cup A_k$ of the vertex set with $m \leq k \leq M$ such that*

(1) $|A_1| = \cdots = |A_k|$
(2) $|A_0| \leq \epsilon |V(G)|$,
(3) *all but ϵk^2 of the pairs (A_i, A_j) are ϵ-regular.*

The SzRL can be proved by starting with a random equipartition and defining an index that keeps track of the total regularity. Upon refining the partition the index is improved by a constant at each step and eventually hits the upper bound within finite steps of refinement.

Notice that the SzRL only gives information regarding 'most' of the edges in G. It does not provide anything regarding The edges with parts of the partition and the edges between irregular parts. However, one can ensure that there are only small amounts of edges within each A_i by choosing a large m such as $\lceil 1/\epsilon \rceil$. Furthermore, the SzRL is only useful for large and relatively dense graphs since for small/sparse graphs, the regularity requirement is rather trivial with too few edges between parts.

Another lemma that concerns regularity is the *Embedding Lemma*, which attempts to find copies of some graph H in another graph G. Given a graph H and an integer m, replace each vertex u_i of H by a set U_i of m vertices and define edge set $\{rs : r \in U_i, s \in U_j, u_i u_j \in E(H)\}$. Essentially, we 'blow up' each vertex of H to size m along with the edge set.

For $\delta > \epsilon > 0$, let $\mathcal{G}(H, m, \epsilon, \delta)$ denote the family of graphs G such that $V(G) = V(H(m))$, $G \subset H(m)$, and the induced bipartite graph $G[U_i, U_j]$ is ϵ-regular and has density at least δ whenever $u_i u_j \in E(H)$. With the above settings, we can state the lemma.

Lemma 0.4 (The Embedding Lemma)**.** *Let $n \in \mathbb{N}$, $\delta > 0$, $\epsilon_0 = \delta^n/(n+2)$. Let R be a graph, let $m, t \in \mathbb{N}$ with $t \leq \epsilon_0 m$, and let $H \subset R(t)$ with maximum degree at most n. If $\epsilon_0 > \epsilon > 0$ and $G \in \mathcal{G}(H, m, \epsilon, \delta + \epsilon)$, then G contains at least $(\epsilon_0 m)^{|V(H)|}$ copies of H.*

One can prove the lemma by choosing the vertices one by one. With the SzRL and the Embedding Lemma, we can proved the *Triangle Removal Lemma*.

Theorem 0.5 (The Triangle Removal Lemma). *For every $\gamma > 0$, there exists a $\delta > 0$ such that if G is a graph with at most δn^3 triangles, then all the triangles in G can be destroyed by removing γn^2 edges.*

To prove the theorem, we start with choosing a sufficiently small ϵ to apply SzRL and obtain a regular partition $\mathcal{P} = \{A_0, \dots, A_k\}$. Then remove all edges inside parts, between irregular pairs, and between sparse pairs. The choice of ϵ ensures that there are no more than γn^2 of such edges. Then consider the reduced graph $R_{G'}$ of G' with the vertex set $[k]$ and edge set

$$\{ij : \text{the pair } (A_i, A_j) \text{ is dense and } \epsilon\text{-regular}\}$$

. Notice that if G' is not triangle-free, all triangles in G' corresponds to triangles in $R_{G'}$. Applying the *Embedding Lemma* by setting $H = K_3$, we obtain that G has more than γn^3 triangles.

With the *Triangle Removal Lemma*, we can prove a simple version of the Erdős-Turán Conjecture.

Theorem 0.6. *(Roth) If $A \in \mathbb{N}$ has positive upper density, then A contains an arithmetic progression of length three, i.e. there exists $a \in A$ and $b \in \mathbb{N}$ such that $\{a, a+b, a+2b\} \in A$.*

To prove it, choose $0 < \epsilon < \overline{d}(A)$ and n sufficiently large such that $|A \cap [n]| > \epsilon n$. Construct a graph G by defining $V(G)$ as disjoint copies $X = Y = Z = [n]$ and $E(G) = E_{XY} \cup E_{YZ} \cup E_{XZ}$ where

$$E_{XY} = \{xy : x \in X, y \in Y, y = x + a \text{ for some } a \in A\}$$
$$E_{YZ} = \{yz : y \in Y, z \in Z, z = y + a \text{ for some } a \in A\}$$
$$E_{XZ} = \{xz : x \in X, y \in Y, z = x + 2a \text{ for some } a \in A\}$$

Assume A contains no arithmetic progression of length 3 (3-AP). Then the only triangles are $\{x, x+a, x+2a\}$. Notice there are exactly $n|A \cap [n]|$ such triangles and they are edge-disjoint. Then we can remove at most ϵn^2 edges to destroy all of the triangles. This is a contradiction to our choice of ϵ.

Until now, this work has shown that SzRL offers an graph theoretic approach to problems in number theory with clever construction of graphs. The SzRL has other applications in different areas such as the *Erdős-Stone Theorem* or compactification of the space of graphs. More details of *Szemeŕedi's Theorem* and other applications of SzRL can be found in the references.

References

[1] R. Morris, R.I. Oliveira, *Extremal and Probablistic Combinatorics*, IMPA, 2011.

[2] T. Szabó, *Regularity Lemma and its Application*, FU Berlin, 2017.

[3] E. Szemerédi, On Sets of Integers Containing no k Elements in Arithmetic Progression, *ACTA Arithmetica*, **27**: 199–245, 1975.

[4] Szemerédi, Endre (1978), "Regular partitions of graphs", *Colloq. Internat. CNRS*, **260**: 399–401, 1978.

[5] T. Tao, *Szemerédi's Proof of Szemerédi's Theorem*, 2005.

[6] L. Lovász, B. Szegedy, "Szemerédi's lemma for the analyst", *Geometric and Functional Analysis*, **17**: 252–270, 2007.