PROVING SZEMERÉDI'S REGULARITY LEMMA

GEON LEE

Abstract. This paper reviews the most common approach to the proof of Szemerédi's regularity lemma, as presented by R. Diestel [1].

1

We only consider simple graphs in this paper. A graph $G = (V, E)$ is a relational structure with the non-empty vertex set V and the edge set $E \subseteq V^2$. $|G|$ is the *order* or the number of vertices in G. $\|G\|$ is the total number of edges in G. For disjoint vertex sets C and D, $\|C, D\|$ is the number of edges between C and D. A *partition* $\mathcal{P} = \{V_1, V_2, \cdots, V_k\}$ of V is the set of nonempty subsets in V such that $\bigcup_i V_i = V$, and $V_i \cap V_j = \emptyset$ for all $i \neq j$. Any partition $\mathcal{P}'$ of V is a *refinement* of $\mathcal{P}$ if every element of $\mathcal{P}'$ is a subset of some element of $\mathcal{P}$.

2

Szemerédi's regularity lemma is an immensely powerful tool in extremal graph theory. Endre Szemerédi introduced the weaker version of the lemma to prove the Erdös-Turán conjecture (1936) that any sequence of natural numbers with positive density contains a (long) arithmetic progression. He proved the full lemma in his later paper (1978). Roughly speaking, regularity lemma states that the vertex set of any (large) graph can be divided into subsets of equal size such that edges between the subsets are distributed randomly; we expect to find an edge connecting two such subsets with some probabilty p, just as in a randomly generated graph. The lemma thereby lets us apply some well-known properties of the random graph to all classes of graph.

In order to precisely state the regularity lemma, we require the following definitions. For two disjoint sets of vertices, C and D, define

$$d(C, D) := \frac{\|C, D\|}{|C||D|}$$

as the *density* of (C, D). The notion of regularity extends from here.

For some $\epsilon > 0$, a pair of sets (C, D) is *ϵ-regular* if for all subsets $X \subseteq C$ and $Y \subseteq D$ with $|X| \geqslant \epsilon|C|$, $|Y| \geqslant \epsilon|D|$, we have

$$|d(C, D) - d(X, Y)| \leqslant \epsilon$$

A partition $\mathcal{P} = \{V_0, V_1, \cdots, V_k\}$ of V is ϵ-regular if the following holds:

(1) $|V_0| < \epsilon|V|$
(2) $|V_1| = |V_2| = \cdots = |V_k|$
(3) All but at most ϵk^2 pairs (V_i, V_j), $1 \leq i < j \leq k$, are ϵ-regular.

Note that V_0 is ignored when determining the regularity of the partition. V_0 is the *exceptional set*, which allows all the other sets in $\mathcal{P}$ to be of equal size. In the proof of the regularity lemma, we will repeat the process of splitting the graph into smaller and smaller subsets until we obtain the desired partition; $|V_0|$ can be thought of as a bin that contains all vertices that are left over from the splitting operations.

Regularity lemma simply states that any graph with at least some number of vertices contains an ϵ-regular partition:

Theorem 2.1 (Szemerédi's regularity lemma)**.** *For every $\epsilon > 0$ and every integer $m \geq 1$, there exists an integer M such that any graph $G = (V, E)$ of order at least m admits an ϵ-regular partition $\mathcal{P} = \{V_0, V_1, \cdots, V_k\}$, with $m \leq k \leq M$.*

The proof of the theorem heavily relies on the potential function, f, which will be defined shortly. The function takes a partition (or sets of vertices) as input and returns a positive real number that measures 'how regular' the input is. f is defined as the following: for two disjoint vertex sets C and D in a graph of order n,

$$\begin{aligned} f(C, D) &:= \frac{|C||D|}{n^2} d^2(C, D) \\ &= \frac{\|C, D\|^2}{n^2 |C||D|} \end{aligned}$$

For two different partitions $\mathcal{C} = \bigcup_i C_i$ and $\mathcal{D} = \bigcup_i D_i$,

$$f(\mathcal{C}, \mathcal{D}) := \sum_{i,j} f(C_i, D_j)$$

For a partition $\mathcal{P} = \bigcup_i V_i$,

$$f(\mathcal{P}) := \sum_{i<j} f(V_i, V_j)$$

If a partition $\mathcal{P} = \{V_0, V_1, \cdots, V_k\}$ contains an exceptional set V_0, then consider V_0 as the set of singletons:

$$f(\mathcal{P}) := \sum_{1 \leqslant i < j} f(V_i, V_j) + \sum_{i \geqslant 1} f(\{v\}, V_i)$$

where $v \in V_0$.

Also note that the function f is bounded above by 1:

$$\begin{aligned} f(\mathcal{P}) &= \sum_{i<j} f(V_i, V_j) \\ &= \sum_{i<j} \frac{|V_i||V_j|}{n^2} \cdot d^2(V_i, V_j) \\ &\leq \sum_{i<j} \frac{|V_i||V_j|}{n^2} \\ &\leq 1 \end{aligned}$$

Given a partition with enough non-ϵ-regular pairs in it, our main goal is to construct its refinement that is substantially 'more regular' than before (as measured by f). Then we may choose any arbitrary partition in G with its sets of equal size except the exceptional set; and refine it again and again to obtain a sequence of

partitions with bigger f values. However only bounded number of such refinements is possible until f reaches its upper bound. Hence the sequence cannot continue infinitely, and we expect to obtain a partition that satisfies the conditions of the theorem some time along this process. The detailed proof follows in the next section.

3

Using the Cauchy-Schwartz inequality, $\sum(a_ib_i)^2 \leqslant \sum a_i^2 \sum b_i^2$, we note some useful properties of the potential function f. When two disjoint sets C and D are partitioned into $\mathcal{C} = \bigcup C_i$ and $\mathcal{D} = \bigcup D_i$ respectively, f always increases:

$$\begin{aligned} f(\mathcal{C},\mathcal{D}) &= \sum_{i,j} \frac{\|C_i, D_j\|^2}{n^2|C_i||D_j|} \\ &\geqslant \frac{(\sum_{i,j} \|C_i, D_j\|)^2}{n^2 \sum |C_i| \sum |D_j|} \\ &= f(C, D) \end{aligned}$$

Also, if a partition $\mathcal{P} = \bigcup C_i$ is refined to $\mathcal{P}' = \bigcup \mathcal{C}_i$, where $\mathcal{C}_i$ is the set of subsets of C_i, then f increases as well:

$$\begin{aligned} f(\mathcal{P}) &= \sum_{i<j} f(C_i, C_j) \\ &\leqslant \sum_{i<j} f(\mathcal{C}_i, \mathcal{C}_j) \\ &\leqslant f(\mathcal{P}') \end{aligned}$$

since $f(\mathcal{P}') = \sum_{i<j} f(\mathcal{C}_i, \mathcal{C}_j) + \sum_i f(\mathcal{C}_i)$

Now we shall see how far f may increase for a pair of two vertex sets if each set is partitioned into two subsets. The increment is yet less than any constant.

Lemma 3.1. *Let C and D be disjoint sets of vertices in a graph of order n. If the pair (C, D) is not ϵ-regular, then for all $\epsilon > 0$, there exist partitions $\mathcal{C} = \{C_1, C_2\}$ and $\mathcal{D} = \{D_1, D_2\}$ such that*

$$f(\mathcal{C},\mathcal{D}) \geqslant f(C, D) + \epsilon^4 \frac{|C||D|}{n^2}$$

Proof. Fix ϵ. Since (C, D) is ϵ-regular, there exist subsets $C_1 \subseteq C$ and $D_1 \subseteq D$, with $|C_1| \geqslant \epsilon|C|$ and $|D_1| \geqslant \epsilon|D|$, and $|d(C_1, D_1) - d(C, D)| \geqslant \epsilon$. Let $C_2 := C \setminus C_1$, $D_2 := D \setminus D_1$; and $\mathcal{C} := \{C_1, C_2\}$, $\mathcal{D} := \{D_1, D_2\}$. Lastly, define $\delta := d(C_1, D_1) - d(C, D)$. It follows that

$$\|C_1, D_1\| = \frac{\|C, D\||C_1||D_1|}{|C||D|} + \delta|C_1||D_1|$$

Via the Cauchy-Schwartz inequality and the above equation,

$$\begin{aligned}
n^2 f(\mathcal{C},\mathcal{D}) &= \sum_{i,j} \frac{\|C_i, D_j\|^2}{|C_i||D_j|} \\
&= \frac{\|C_1, D_1\|^2}{|C_1||D_1|} + \sum_{i+j>2} \frac{\|C_i, D_j\|^2}{|C_i||D_j|} \\
&\geqslant \frac{\|C_1, D_1\|^2}{|C_1||D_1|} + \frac{(\sum_{i+j>2} \|C_i, D_j\|)^2}{\sum_{i+j>2} |C_i||D_j|} \\
&= \frac{\|C_1, D_1\|^2}{|C_1||D_1|} + \frac{(\|C, D\| - \|C_1, D_1\|)^2}{|C||D| - |C_1||D_1|} \\
&= \frac{\|C, D\|^2}{|C||D|} + \delta^2 \left(|C_1||D_1| + \frac{|C_1|^2|D_1|^2}{\|C, D\| - \|C_1, D_1\|} \right) \\
&\geqslant \frac{\|C, D\|^2}{|C||D|} + \delta^2 |C_1||D_1| \\
&\geqslant n^2 f(C, D) + \epsilon^4 |C||D|
\end{aligned}$$

□

Using the above results, we will show that given a non-ϵ-regular partition, there exists its refinement that increases f by some constant.

Lemma 3.2. *$0 < \epsilon \leqslant 1/3$. Let $\mathcal{P} = \{V_0, V_1, \cdots, V_k\}$ be a partition of a graph G of order n, with $|V_0| \leqslant \epsilon n$ and $|V_1| = \cdots = |V_k| := d$. If $\mathcal{P}$ is not ϵ-regular, there exists another partition $\mathcal{P}' = \{V_0', V_1', \cdots, V_l'\}$ of G such that the following holds:*

(i) $k \leqslant l \leqslant k4^k$
(ii) $|V_0'| \leqslant |V_0| + n/2^k$
(iii) $|V_1'| = \cdots = |V_l'|$
(iv) $f(\mathcal{P}') \geqslant f(\mathcal{P}) + \epsilon^5/2$

Proof. For $1 \leqslant i \leqslant j \leqslant k$, define partitions $\mathcal{V}_{ij}$ of V_i and $\mathcal{V}_{ji}$ of V_j as follows: if the pair (V_i, V_j) is not ϵ-regular, let $\mathcal{V}_{ij} := \{V_i\}$ and $\mathcal{V}_{ji} := \{V_j\}$; if not, then we can find partitions $\mathcal{V}_{ij}$ of V_i and $\mathcal{V}_{ji}$ of V_j, with $|\mathcal{V}_{ij}| = |\mathcal{V}_{ji}| = 2$ and

$$f(\mathcal{V}_{ij}, \mathcal{V}_{ji}) \geqslant f(V_i, V_j) + \epsilon^4 d^2/n^2.$$

Consider two elements in V_i as equivalent whenever they are in the same set in $\mathcal{V}_{ij}$ for all $j \neq i$, and let $\mathcal{V}_i$ be the set of such equivalence classes. Then $\mathcal{V} := \{V_0\} \cup \bigcup_{i=1}^k \mathcal{V}_i$ is a partition of G that refines $\mathcal{P}$. Note that each V_i may be partitioned into two sets at most $k-1$ times. Due to our construction of $\mathcal{V}_i$, we have at most 2^{k-1} equivalance classes in it; hence $|\mathcal{V}_i| \leqslant 2^{k-1}$, and $k \leqslant |\mathcal{V}| \leqslant k2^k$.

Since $\mathcal{V}_i$, $\mathcal{V}_j$ each refines $\mathcal{V}_{ij}$ and $\mathcal{V}_{ji}$,

$$\begin{aligned} f(\mathcal{V}) &= \sum_i f(\mathcal{V}_i) + \sum_{i\geqslant 1} f(\mathcal{V}_0, \mathcal{V}_i) + \sum_{1\leqslant i<j} f(\mathcal{V}_i, \mathcal{V}_j) \\ &\geqslant f(\mathcal{V}_0) + \sum_{i\geqslant 1} f(\mathcal{V}_0, \{V_i\}) + \sum_{1\leqslant i<j} f(\mathcal{V}_{ij}, \mathcal{V}_{ji}) \\ &\geqslant f(\mathcal{V}_0) + \sum_{i\geqslant 1} f(\mathcal{V}_0, \{V_i\}) + \sum_{1\leqslant i<j} f(V_i, V_j) + \epsilon k^2 \frac{\epsilon^4 d^2}{n^2} \\ &= f(\mathcal{P}) + \epsilon^5 \frac{k^2 d^2}{n^2} \\ &\geqslant f(\mathcal{P}) + \frac{\epsilon^5}{2} \end{aligned}$$

We shall once again refine $\mathcal{V}$ by cutting each $\mathcal{V}_i$ into sets of equal size (that is much smaller than before) and adding all the remaining elements to the exceptional set V_0. The subsets must be small enough so that we do not obtain too many 'remainder' elements during the refinement; thereby the resulting exceptional set $|V_0'|$ is kept small, or less than $\epsilon|V|$. To obtain the desired partition, let $\{V_1', \cdots, V_l'\}$ be the maximal collection of disjoint sets of size $d' = \lfloor d/4^k \rfloor$, with each V_i' contained entirely within some set $V \in \mathcal{V}_j$, where $j \geqslant 1$. Lastly let $V_0' := \mathcal{V} \setminus \bigcup_{i=1}^{l} V_i'$; and $\mathcal{P}' := \{V_0', V_1', \cdots, V_l'\}$. Then, due to the choice of d',

$$k \leqslant l \leqslant k4^k$$

and

$$\begin{aligned} |V_0'| &\leqslant |V_0| + d'|\mathcal{V}| \\ &\leqslant |V_0| + \frac{d}{4^k}(k2^k) \\ &= |V_0| + \frac{dk}{2^k} \\ &\leqslant |V_0| + \frac{n}{2^k} \end{aligned}$$

Finally, since $\mathcal{P}'$ refines $\mathcal{V}$, we have

$$f(\mathcal{P}') \geqslant f(\mathcal{V}) \geqslant f(\mathcal{P}) + \frac{\epsilon^5}{2}$$

□

The proof of the regularity lemma easily follows from the repeated application of **Lemma 3.2**.

Proof of the regularity lemma. Fix $m \geqslant 1$ and $\epsilon > 0$. Let $G(V, E)$ of order $n \geqslant m$ be given. Also assume that $\epsilon \leqslant 1/3$ with no loss of generality. Before applying the above lemma to G, we must (i) require that the size of the exceptional set is bounded above by ϵn at all times; and (ii) determine M.

(i) f increases by at least $\epsilon^5/2$ per each iteration of **Lemma 3.2**, which therefore may be repeated at most $2/\epsilon^5$ times before f hits the upper bound of 1. To obtain an ϵ-regular partition with the size of its exceptional set less than or equal to ϵn,

choose $k \geqslant m$ large enough so that

$$\frac{2}{\epsilon^5} \cdot \frac{n}{2^k} \leqslant \frac{\epsilon n}{2}$$

or

$$2^{k-2} \geqslant 1/\epsilon^6$$

holds. Next, we require that the exceptional set in the initial partition does not exceed $\epsilon n/2$, by only considering n large enough. For $|V_0| < k$, we would like to have $k \leqslant \epsilon n/2$, or $n \geqslant 2k/\epsilon$. Note that for n small, the regularity lemma will produce trivial results. (See the final paragraph of the proof.)

(ii) Given the initial partition with k sets in it, the number of sets may grow up to $k4^k$ after applying **Lemma 3.2**. Let g is the map that takes k to $k \cdot 4^k$. Then we may choose M as:

$$M := max\left(g^{2/\epsilon^5}(k), \frac{2k}{\epsilon}\right)$$

Finally we are ready to construct an ϵ-regular partition $\{V_0, V_1, \cdots, V_k\}$ of G. If $n \leqslant M$, let $k := n$ and $V_0 = \emptyset$, so that the given graph is partitioned into k singletons. Otherwise, let V_0 be the minimal set such that k divides $|V \setminus V_0|$; hence $|V_0| < k \leqslant \epsilon n/2$. Now place the remaining vertices into sets of any equal size, and apply **Lemma 3.2** repeatedly until the desired partition is obtained. □

Acknowledgements. I wish to express my indebtedness to my mentors, John Wilmes and John Wiltshire-Gordon.

References

[1] R. Diestel. *Graph Theory*. Springer. 2010.
[2] A. Marcus. *The Szémeredi Regularity Lemma*. 2008.
http://cs-www.cs.yale.edu/homes/marcus/papers/srl.pdf