Szemerédi's regularity lemma

notes by

Marcel K. Goh

12 January 2020

1. Introduction

[Put some kind of introduction here.]

Fix a graph $G = (V, E)$ and let X and Y be subsets of V (not necessarily disjoint). Let $e(X, Y) = \{xy \in E : x \in X,\ y \in Y\}$ denote the number of edges that have an endpoint in each of X and Y. We define the *edge density* between X and Y to be the ratio

$$d(X, Y) = \frac{e(X, Y)}{|X||Y|}. \tag{1}$$

If X and Y are disjoint, then this is the fraction of all possible edges between X and Y that are actually present in the graph (and in the case that they are not disjoint, it isn't awfully far off anyway).

We say that a pair of vertex subsets (X, Y) is *ϵ-regular* if for all subsets $A \subseteq X$ and $B \subseteq Y$ with $|A| \ge \epsilon|X|$ and $|B| \ge \epsilon|Y|$, we have $|d(A, B) - d(X, Y)| \le \epsilon$. This means that if we zoom in to look at the edges between a subset of X and a subset of Y, we find that the picture is sort of a "scale-model" of the whole of X and the whole of Y in the sense that the number of edges that we see is proportional to the sizes of the subsets, unless the subsets are taken to be very small. If the pair (X, Y) is *not* ϵ-regular, then there must be some $A \subseteq X$ and $B \subseteq Y$, with $|A| \ge \epsilon|X|$ and $|B| \ge \epsilon|Y|$, such that $|d(A, B) - d(X, Y)| > \epsilon$. The pair (A, B) is said to *witness* the irregularity.

A *partition* $\mathcal{P}$ is a collection $\{V_1, \ldots, V_k\}$ of disjoint subsets of V whose union is all of V. We will say that a partition is *equitable* if the sizes of any two parts do not differ by more than 1. A partition is said to be *ϵ-regular* if the sum of $|V_i||V_j|$, taken over all pairs (V_i, V_j) that are not ϵ-regular, is less than $\epsilon|V|^2$. If the partition is equitable, then this is equivalent to saying that at most ϵk^2 of the pairs (V_i, V_j) are not ϵ-regular. Szemerédi's regularity lemma says that every graph admits an ϵ-regular equitable partition into a number of parts depending only on ϵ (and not the size of the graph).

References for the original instances of the proofs are listed at the bottom of this document; a large part of these notes are based on lectures given by Yufei Zhao in 2019.

2. The regularity lemma

In this section we will prove Szemerédi's regularity lemma via a sequence of auxiliary ones. The idea runs as follows. We begin with a partition of $G = (V, E)$ that is given to us (e.g., the trivial partition $\mathcal{P} = \{V\}$) and while the partition is not ϵ-regular, we iteratively refine it by subdividing each element of the partition into further parts. Using an "energy increment argument", we show that this process terminates after a bounded number of steps, and therefore the number of parts in the final partition is bounded. For vertex sets U and W, we define the *energy* of (U, W) to be the quantity

$$q(U, W) = \frac{|U||W|}{n^2} d(U, W)^2, \tag{2}$$

where $n = |V|$, and for partitions $\mathcal{P}_U = \{U_1, U_2, \ldots, U_k\}$ and $\mathcal{P}_W\{W_1, W_2, \ldots, W_l\}$ of U and W respectively, we will define the *energy* of the two partitions to be the sum

$$q(\mathcal{P}_U, \mathcal{P}_W) = \sum_{i=1}^{k} \sum_{j=1}^{l} q(U_i, W_j). \tag{3}$$

We will write $q(\mathcal{P}) = q(\mathcal{P}, \mathcal{P})$ when the two partitions are equal. Note that if $\mathcal{P}$ is a partition of V into k parts, we have

$$q(\mathcal{P}) = \sum_{i=k}^{k} \sum_{j=1}^{k} \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2 \le 1, \tag{4}$$

since edge density is at most 1. The first lemma states that the energy of a pair of refined partitions is at least the energy of the original pair.

Lemma A. *Let $G = (V, E)$ be a graph, let $U, W \subseteq V$ and suppose that $\mathcal{P}_U = \{U_1, U_2, \ldots, U_k\}$ and $\mathcal{P}_W\{W_1, W_2, \ldots, W_l\}$ are partitions of U and W respectively. Then $q(\mathcal{P}_U, \mathcal{P}_W) \geq q(U, W)$.*

Proof. We will define a random variable Z as follows. We select vertices $u \in U$ and $w \in W$ uniformly at random; suppose that $U_i \in \mathcal{P}_U$ contains u and $W_j \in \mathcal{P}_W$ contains w. We let $Z = d(U_i, W_j)$. We compute the first moment

$$\mathbf{E}\{Z\} = \sum_{i=1}^{k} \frac{|U_i|}{|U|} \sum_{j=1}^{l} \frac{|W_j|}{|W|} d(U_i, W_j) = \frac{e(U, W)}{|U||W|} = d(U, W) \tag{5}$$

and the second moment

$$\mathbf{E}\{Z^2\} = \sum_{i=1}^{k} \frac{|U_i|}{|U|} \sum_{j=1}^{l} \frac{|W_j|}{|W|} d(U_i, W_j)^2 = \frac{n^2}{|U||W|} q(\mathcal{P}_U, \mathcal{P}_W) \tag{6}$$

of Z, where n denotes the size of V. By Jensen's inequality, we have $\mathbf{E}\{Z^2\} \geq \mathbf{E}\{Z\}^2$ and therefore

$$q(\mathcal{P}_U, \mathcal{P}_W) \geq \frac{|U||W|}{n^2} d(U, W) = q(U, W). \quad \blacksquare \tag{7}$$

In particular, if $\mathcal{P}$ is a partition of V and $\mathcal{P}'$ refines $\mathcal{P}$, then we can apply Lemma A to every pair (V_i, V_j) of sets in $\mathcal{P}$ to conclude that $q(\mathcal{P}') \geq q(\mathcal{P})$. The next lemma shows that the inequality in Lemma A is sometimes strict, a fact we will need for the energy increment argument.

Lemma B. *With the same definitions as in Lemma A, suppose furthermore that for some $\epsilon > 0$, the pair (U, W) is not ϵ-regular and the irregularity is witnessed by $U_1 \subseteq U$ and $W_1 \subseteq W$. Then*

$$q(\{U_1, U \setminus U_1\}, \{W_1, W \setminus W_1\}) \geq q(U, W) + \epsilon^4 \frac{|U||W|}{n^2}, \tag{8}$$

where $n = |V|$.

Proof. Define the random variable Z as in the proof of Lemma *A*. Note that the variance of Z is

$$\begin{aligned}
\mathbf{V}\{Z\} &= \mathbf{E}\{Z^2\} - \mathbf{E}\{Z\}^2 \\
&= \frac{n^2}{|U||W|} q(\{U_1, U \setminus U_1\}, \{W_1, W \setminus W_1\}) - d(U, W)^2 \\
&= \frac{n^2}{|U||W|} \big(q(\{U_1, U \setminus U_1\}, \{W_1, W \setminus W_1\}) - q(U, W) \big).
\end{aligned} \tag{9}$$

But we also have the formula

$$\begin{aligned}
\mathbf{V}\{Z\} &= \mathbf{E}\left\{ (Z - \mathbf{E}\{Z\})^2 \right\} \\
&= \frac{|U_1||W_1}{|U||W|} \big(d(U_1, W_1) - d(U, W) \big)^2 + \frac{|U_1||W \setminus W_1|}{|U||W|} \big(d(U_1, W \setminus W_1) - d(U, W) \big)^2 \\
&\quad + \frac{|U \setminus U_1||W_1|}{|U||W|} \big(d(U \setminus U_1, W_1) - d(U, W) \big)^2 + \frac{|U \setminus U_1||W \setminus W_1|}{|U||W|} \big(d(U \setminus U_1, W \setminus W_1) - d(U, W) \big)^2 \\
&\geq \frac{|U_1|}{|U|} \cdot \frac{|W_1|}{|W|} \cdot \big(d(U_1, W_1) - d(U, W) \big)^2 \\
&\geq \epsilon \cdot \epsilon \cdot \epsilon^2,
\end{aligned} \tag{10}$$

where the final inequality follows from the fact that (U_1, W_1) was the witness for the non-ϵ-regularity of (U, W). Combining both calculations for the variance proves the inequality we need. $\blacksquare$

We are now able to formulate the step in the inner loop of our regularisation procedure.

Lemma C. *Let $G = (V, E)$ be a graph, let $\mathcal{P} = \{V_1, V_2 \ldots, V_k\}$ be a partition of V, and let $\epsilon > 0$. If the partition $\mathcal{P}$ is not ϵ-regular, then there exists a refinement $\mathcal{Q}$ of $\mathcal{P}$ in which every V_i is partitioned into at most 2^k parts and such that*

$$q(\mathcal{Q}) \geq q(\mathcal{P}) + \epsilon^5. \tag{11}$$

Proof. If $\mathcal{P}_1$ and $\mathcal{P}_2$ are refinements of $\mathcal{P}$ that subdivide V_i into $V_{i1} \cup V'_{i1}$ and $V_{i2} \cup V'_{i2}$ respectively, then the *common refinement* of $\mathcal{P}_1$ and $\mathcal{P}_2$ divides V_i into the union

$$V_i = (V_{i1} \cap V_{i2}) \cup (V'_{i1} \cap V_{i2}) \cup (V_{i1} \cap V'_{i2}) \cup (V'_{i1} \cap V'_{i2}); \tag{12}$$

by induction, we can similarly define the common refinement of any finite number of partitions that refine $\mathcal{P}$. For every pair (i, j) for which (V_i, V_j) is not ϵ-regular, we can find $A_{ij} \subseteq V_i$ and $A_{ji} \subseteq V_j$ that witnesses the irregularity. Lemma B will produce a refinement of $\mathcal{P}$ for each (i, j) that divides V_i and V_j each into two new parts, and we can let $\mathcal{Q}$ be the common refinement of these partitions, as defined above. Note that we have constructed $\mathcal{Q}$ such that it does not have more than 2^k parts.

Let $\mathcal{R}$ be the set of all $(i, j) \in [1, k]^2$ such that (V_i, V_j) is ϵ-regular. For each i, let $\mathcal{Q}_{V_i}$ denote the subdivision of V_i given by $\mathcal{Q}$. By Lemma A, we have

$$\begin{aligned} q(\mathcal{Q}) &= \sum_{i=1}^{k} \sum_{j=1}^{k} q(\mathcal{Q}_{V_i}, \mathcal{Q}_{V_j}) \\ &\geq \sum_{(i,j) \in \mathcal{R}} q(V_i, V_j) + \sum_{(i,j) \notin \mathcal{R}} q(\{A_{ij}, V_i \setminus A_{ij}\}, \{A_{ji}, V_j \setminus A_{ji}\}). \end{aligned} \tag{13}$$

Applying Lemma B, we find that

$$\begin{aligned} q(\mathcal{Q}) &\geq \sum_{(i,j) \in \mathcal{R}} q(V_i, V_j) + \sum_{(i,j) \notin \mathcal{R}} q(V_i, V_j) + \epsilon^4 \frac{|V_i||V_j|}{n^2} \\ &\geq q(\mathcal{P}) + \epsilon^5, \end{aligned} \tag{14}$$

as desired. ▮

With Lemma C in hand, we can state and prove Szemerédi's regularity lemma without too much further effort.

Theorem R (*Szemerédi,* 1978). *For all $\epsilon > 0$ there exists an M with such that every graph admits an ϵ-regular partition of its vertices into no more than M parts. The constant M depends only on ϵ (not on the size of the graph) and we have the upper bound*

$$M \leq 4^{4^{\cdot^{\cdot^{\cdot^{4}}}}} \tag{15}$$

where the tower of 4s consists of ϵ^{-5} repeated exponents.

Proof. We start with the trivial partition $\mathcal{P} = \{V\}$ and apply Lemma C while the current partition is not ϵ-regular. Since $0 \leq q(\mathcal{P}) \leq 1$, and the energy of the partition increases by at least ϵ^5 with each iteration, the algorithm terminates after at most ϵ^{-5} steps. At any particular step, if $\mathcal{P}$ has k parts, Lemma C outputs a partition with $k2^k \leq 4^k$ parts, and this observation yields the upper bound in the theorem statement. ▮

The bound we stated above is not tight, but it turns out that the tower of exponents is inescapable. It was shown by Gowers (1997) that there exists a $c > 0$ such that for all $\epsilon > 0$ small enough, one can construct a graph whose ϵ-regular partition requires more than M parts, where M is a exponential tower of 2s that is ϵ^{-c} high.

Equitable partitions. In Lemma C, one can require that the resulting partition $\mathcal{Q}$ be equitable (no two parts differ by more than one element), and the inequality would still hold, although with an increment that is possibly less than ϵ^5. The difference is negligible, in the sense that the final bound obtained for M will still be of the same order. It is important to note that it is *not* possible to obtain an ϵ-regular partition by proving the regularity lemma with Lemma C as stated above, and then further subdividing and merging the partitions afterwards, because

References

William Timothy Gowers, "Lower bounds of tower type for Szemerédi's uniformity lemma," *Geometric and Functional Analysis* **7** (1997), 322–397.

Imre Ruzsa and Endre Szemerédi, "Triple systems with no six points carrying three triangles," *Colloquia Mathematica Societatis János Bolyai* **18** (1978), 939–945.

Endre Szemerédi, "On sets of integers containing no k elements in arithmetic progression," *Acta Arithmetica* **27** (1975), 199–245.

Endre Szemerédi, "Regular partitions of graphs," *Colloques Internationaux C.N.R.S. n° 260—Problèmes Combinatoires et Théorie des Graphes, Orsay* (1978), 399–401.

Terence Tao and Van Ha Vu, *Additive Combinatorics* (Cambridge: Cambridge University Press, 2006).