Proof of regularity lemma

We now prove the regularity lemma. Idea is to define a function q measuring quality of a partitioning. Value of the function will be in $[0,1]$ and if there are enough pairs, that are not ϵ-regular, then the partition could be refined and q will grow by a constant depending only on ϵ. Hence after finitely many steps, we obtain an ϵ-regular partition.

Let G be an n-vertex graph with vertex partition $\mathcal{P} = \{V_1, \cdots, V_k\}$, the **mean square density** of a pair of partition classes V_i, V_j is

$$q(V_i, V_j) = \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2 = \frac{e(V_i, V_j)}{|V_i||V_j|n^2},$$

where $e(V_i, V_j)$ is the number of edges between V_i and V_j. The **mean square density** of the partition is

$$q(\mathcal{P}) = \sum_{i<j} q(V_i, V_j) = \sum_{i<j} \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2.$$

Observe that the mean square density of a partition is always between 0 and 1.

1: Show that $0 \leq q(\mathcal{P}) \leq 1$.

Solution:

$$q(\mathcal{P}) = \sum_{i<j} q(V_i, V_j) = \sum_{i<j} \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2 \leq \frac{1}{n^2} \sum_{i<j} |V_i||V_j| \leq \frac{1}{n^2} \cdot \frac{1}{2} \left(\sum_i |V_i| \right)^2 < 1$$

Given a partition $\mathcal{P}$, a **refinement** of $\mathcal{P}$ is a partition $\mathcal{P}'$ of the same underlying set such that each class of $\mathcal{P}'$ is contained in a class of $\mathcal{P}$. In this way, if X is a partition class of $\mathcal{P}$ we can also use the term **refinement** of X to refer to the classes of $\mathcal{P}'$ whose disjoint union is X.

Lemma 1. *If X, Y are disjoint vertex sets with refinements $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$, then*

$$q(X, Y) \leq \sum_{1 \leq i,j \leq 2} q(X_i, Y_j).$$

Furthermore, the mean square density of partition $\mathcal{P}$ is at most the mean square density of a refinement of $\mathcal{P}$.

Proof. Let X, Y be disjoint sets with refinements $X = X_1, X_2$ and $Y = Y_1, Y_2$.

2: Use Cauchy-Schwarz inequality, i.e. $\left(\sum_i a_i b_i\right)^2 \leq (\sum_i a_i^2)(\sum_i b_i^2)$ to show $\left(\sum_i c_i d_i\right)^2 \leq (\sum_i c_i)(\sum_i c_i d_i^2)$

Solution: Use $a_i = \sqrt{c_i}$ and $b_i = \sqrt{c_i} \cdot d_i$.

3: Apply this to the following with $d_i = d(X_i, Y_j)$

$$d(X,Y)^2 = \left(\sum_{1 \leq i,j \leq 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i, Y_j) \right)^2 \leq \left(\sum_{i,j} \frac{|X_i||Y_j|}{|X||Y|} \right) \left(\sum_{i,j} \frac{|X_i||Y_j|}{|X||Y|} d(X_i, Y_j)^2 \right).$$

4: Show $q(X, Y) \leq \sum_{i,j} q(X_i, Y_j)$.

Solution:

$$q(X,Y) = \frac{|X||Y|}{n^2} d(X,Y)^2 \leq \sum_{i,j} \frac{|X_i||Y_j|}{n^2} d(X_i, Y_j)^2 = \sum_{i,j} q(X_i, Y_j).$$

5: Finish the proof by showing $q(\mathcal{P}) \le q(\mathcal{P}')$ if $\mathcal{P}'$ is a refinement of $\mathcal{P}$.

Solution: We also get

$$q(X,Y) \le \sum_{i,j} q(X_i,Y_j) \le q(X_1,X_2) + q(Y_1,Y_2) + \sum_{i,j} q(X_i,Y_j).$$

Thus if $\mathcal{P}'$ is a refinement of $\mathcal{P}$ we have $q(P) \le q(P')$.

□

Lemma 2. *Suppose X and Y are partition classes of G such that X,Y is not ϵ-regular, then there is a refinement $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$ such that*

$$q(X,Y) + \frac{|X||Y|}{n^2}\epsilon^4 \le \sum_{1\le i,j\le 2} q(X_i,Y_j).$$

Proof. Because X,Y is not ϵ-regular, there are sets $X_1 \subset X$ and $Y_1 \subset Y$ such that $|X_1| \ge \epsilon|X|$ and $|Y_1| \ge \epsilon|Y|$ such that

$$|d(X,Y) - d(X_1,Y_1)| \ge \epsilon.$$

Thus,

$$\epsilon^2 \le (d(X,Y) - d(X_1,Y_1))^2$$

and

$$\epsilon^2 \le \frac{|X_1||Y_1|}{|X||Y|}.$$

Put $X_2 = X - X_1$ and $Y_2 = Y - Y_1$.

6: Show that

$$\sum_{1\le i,j\le 2} \frac{|X_i||Y_j|}{|X||Y|} = 1 \qquad \text{and} \qquad d(X,Y) = \sum_{1\le i,j\le 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i,Y_j)$$

Solution: Count pairs (x,y), where $x \in X$ and $y \in Y$ in two different ways - the second one is counting them partitioned.

$$d(X,Y) = \frac{e(X,Y)}{|X||Y|} = \sum_{1\le i,j\le 2} \frac{e(X_i,Y_j)}{|X||Y|} = \sum_{1\le i,j\le 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i,Y_j)$$

7: Show

$$\epsilon^4 \le -d(X,Y)^2 + \sum_{1\le i,j\le 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i,Y_j)^2$$

Hint start with $\epsilon^4 \le \epsilon^2 \cdot \epsilon^2 \le \cdots$ and when you have term for indices 1, just add all other for $1 \le i,j \le 2$.

Solution: Now

$$\epsilon^4 \le \frac{|X_1||Y_1|}{|X||Y|}(d(X,Y) - d(X_1,Y_1))^2 \le \sum_{1\le i,j\le 2} \frac{|X_i||Y_j|}{|X||Y|}(d(X,Y) - d(X_i,Y_j))^2$$

$$= d(X,Y)^2 \sum_{1\le i,j\le 2} \frac{|X_i||Y_j|}{|X||Y|} - 2d(X,Y) \sum_{1\le i,j\le 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i,Y_j) + \sum_{1\le i,j\le 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i,Y_j)^2.$$

The first sum in the previous line is at most 1 and the second sum is $d(X,Y)$, so we get

$$\epsilon^4 \le -d(X,Y)^2 + \sum_{1\le i,j\le 2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i,Y_j)^2$$

8: Finish the proof.

Solution: Multiplying both sides by $|X||Y|/n^2$ gives the result.

□

Lemma 3. *Suppose G is an n-vertex graph with partition $\mathcal{A}$ that includes a class V of at least two elements. If $\mathcal{B}$ is a refinement that comes from refining V into a single element X and $Y = V - X$, then*

$$q(\mathcal{A}) \ge q(\mathcal{B}) - \frac{2}{n}.$$

Proof. Let V be the class of $\mathcal{P}$ that is refined into a single element X and $Y = V - X$. Let $V_i \neq V$ be a partition class of $\mathcal{P}$, Observe that

$$q(\mathcal{P}') - q(\mathcal{P}) = \sum_{V_i \neq V} q(X,V_i) + q(Y,V_i) - q(V,V_i).$$

9: Use the definition of $q(A,B)$ to estimate that each $q(X,V_i)$ and $q(Y,V_i) - q(V,V_i)$ is upper bounded by $\frac{|V_i|}{n^2}$. Then finish the proof.

Solution: For each V_i we have

$$q(X,V_i) \le \frac{|V_i|}{n^2}$$

and

$$\begin{aligned} q(Y,V_i) - q(V,V_i) &= \frac{1}{n^2}\frac{e(Y,V_i)^2}{|Y||V_i|} - \frac{1}{n^2}\frac{e(V,V_i)^2}{|V||V_i|} \le \frac{1}{n^2}\frac{e(Y,V_i)^2(|V|-|Y|)}{|Y||V||V_i|} \\ &\le \frac{1}{n^2}\frac{e(Y,V_i)^2}{|Y||V||V_i|} \le \frac{1}{n^2}\frac{(|Y||V_i|)^2}{|Y||V||V_i|} \le \frac{|V_i|}{n^2} \end{aligned}$$

Summing both of the above equations for all $V_i \neq V$ gives the lemma,

□

Lemma 4. *If $\mathcal{P}$ is an equipartition into $k \ge 4\epsilon^{-6}$ parts such that more than ϵk^2 pairs of classes are not ϵ-regular, then there is an equipartition $\mathcal{R}$ of at most $k^2 2^{k-1}$ parts such that,*

$$q(\mathcal{R}) \ge q(\mathcal{P}) + \epsilon^5/2.$$

Proof. Proof outline. First we split pairs on $\mathcal{P}$ that are not ϵ-regular. This gives $\mathcal{P}'$ and Lemma 2 gives a boost in q. Then we create a further refinement $\mathcal{P}''$, which has parts of equal sizes and some leftover. Finally, we move the leftover back in to classes of $\mathcal{P}''$, resulting in $\mathcal{R}$. In this step, q may decrease but Lemma 3 will help us control the decrease.

For a pair of classes X, Y of the partition that are not ϵ-regular, there is a refinement given by Lemma 2 that increases the mean square density. Let $\mathcal{P}'$ be the resulting refinement of $\mathcal{P}$ when we split every pair that is not ϵ-regular.

10: What is the upper bound on the number of classes in $\mathcal{P}'$?
Hint: One class on $\mathcal{P}$ will be split into at most how many classes of $\mathcal{P}'$?

Solution: Let us apply these refinements to every class of $\mathcal{P}$. Each class is in at most $k-1$ pairs that are not ϵ-regular, so each class is refined at most $k-1$ times. Therefore, each class of $\mathcal{P}$ is split into at most 2^{k-1} new parts. Let $\mathcal{P}'$ be the resulting refinement of $\mathcal{P}$ and note that $\mathcal{P}'$ has $k' = k2^{k-1}$ total classes.

11: Show that
$$q(\mathcal{P}') > q(\mathcal{P}) + \epsilon^5.$$
Hint: What is increase for one pair that is not ϵ-regular? How many increases we get? Use Lemma 2.

Solution: For any pair of classes X, Y of $\mathcal{P}$ that are not ϵ-regular there is a refinement $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$ such that
$$\sum_{1 \le i,j \le 2} q(X_i, Y_j) \ge q(X,Y) + \frac{|X||Y|}{n^2}\epsilon^4 \ge q(X,Y) + \frac{1}{k^2}\epsilon^4. \tag{1}$$
The sets X_1, X_2, Y_1, Y_2 are further refined in the construction of $\mathcal{P}'$. If $V_1', \dots, V_{k'}'$ are the classes of $\mathcal{P}'$, then
$$\sum_{V_i' \subset X, V_j' \subset Y} q(V_i', V_j') \ge q(X,Y) + \frac{1}{k^2}\epsilon^4.$$
Furthermore, for all other refinements to $\mathcal{P}$ the mean square density cannot decrease. Therefore, as there are more than ϵk^2 such pairs X, Y, we have
$$q(\mathcal{P}') > q(\mathcal{P}) + \epsilon^5.$$

Now it remains to convert $\mathcal{P}'$ into an equipartition. Split each of classes of $\mathcal{P}'$ into subclasses of size exactly $n/(k^2 2^{k-1})$ and a "leftover" class of size $< n/(k^2 2^{k-1})$. Furthermore, for simplicity, let us split all of the "leftover" vertices into classes of a single vertex. Let the resulting partition be $\mathcal{P}''$. Observe that at this point there are at most $k^2 2^{k-1}$ classes that are not singletons.

Now we distribute these "leftover" vertices as evenly as possible into the classes of size exactly $n/(k^2 2^{k-1})$ to get an equipartition $\mathcal{R}$. However, because $\mathcal{P}''$ is a refinement of $\mathcal{R}$ we have that $q(\mathcal{R}) \le q(\mathcal{P}'')$. Fortunately, the decrease is small.

12: Calculate an upper bound on the number of "leftover" vertices.

Solution: The total number of "leftover" vertices is less than
$$k2^{k-1}\frac{n}{k^2 2^{k-1}} = \frac{n}{k}.$$

We can arrive at $\mathcal{P}''$ from $\mathcal{R}$ by creating a new singleton class for each "leftover" vertex.

13: Use Lemma 3 to find an lower bound on $q(\mathcal{R}) - q(\mathcal{P}'')$. Make the bound ONLY in ϵ to some power.

Solution: Repeatedly applying Lemma 3 gives

$$q(\mathcal{R}) - q(\mathcal{P}'') \geq \frac{n}{k}\frac{2}{n} = \frac{2}{k} = \epsilon^6/2.$$

14: Show that $q(\mathcal{R}) \geq q(\mathcal{P}) + \epsilon^5/2$. Hint: Use $q(\mathcal{P}'')$.

Solution:

$$q(\mathcal{R}) \geq q(\mathcal{P}'') - \epsilon^6/2 \geq q(\mathcal{P}') - \epsilon^6/2 \geq q(\mathcal{P}) + \epsilon^5 - \epsilon^6/2 \geq q(\mathcal{P}) + \epsilon^5/2.$$

The partition $\mathcal{R}$ simply moved singleton partition classes to those that were not singletons. So the total number of classes is at most

$$k^2 2^{k-1},$$

which completes the result. □

Theorem 5 (Szemerédi regularity lemma, 1974)**.** *Given $\epsilon > 0$ and $m \geq 1$, there exists a constant $M = M(\epsilon, m)$ such that every graph on at least m vertices has an equipartition into r parts such that all but at most ϵr^2 pairs of classes are ϵ-regular and $m \leq r < M$.*

Proof. **15:** Finish the proof and give a upper bound on M not depending on n. Hint: Make sure you can apply Lemma 4 and keep applying it. How many application you need and how will this influence the number of parts?

Solution: Begin with an equipartition of G into $k \geq \max\{m, 4\epsilon^{-6}\}$ classes. If it is ϵ-regular we are done. Otherwise, the lemma above allows us to refine the partition into $k^2 2^{k-1}$ classes and increase the mean square density by at least $\epsilon^5/2$. We continue this process until we reach an ϵ-regular partition. Because the mean square density cannot exceed 1, this process must stop after at most $2\epsilon^{-5}$ steps. That is, we have an equipartition into r classes where all but at most ϵr^2 pairs of classes are ϵ-regular. Furthermore, for each application of the lemma the number of classes increases from k to at most $k^2 2^{k-1} \leq 2^{2k} = 4^k$ parts, so when the process stops we have $m < r < M$ where M is at most a tower of 4s of height at most $2\epsilon^{-5}$.

□