Proof of Regularity Lemma

Zdeněk Dvořák

January 11, 2024

We are going to use the following well-known inequality.

Lemma 1 (Cauchy-Schwarz inequality). *For any real numbers $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$, we have $\left(\sum_{i=1}^n u_i^2\right)\left(\sum_{i=1}^n v_i^2\right) \ge \left(\sum_{i=1}^n u_i v_i\right)^2$.*

Our goal is to prove the Regularity Lemma.

Theorem 2. *For any positive integer m_0 and real number $\varepsilon > 0$, there exists an integer $M \ge m_0$ such that the following holds. Every graph G with at least m_0 vertices has an ε-regular partition of order at least m_0 and at most M.*

Without loss of generality, we can assume that $\varepsilon < 1/2$. Let $n \ge m_0$ be the number of vertices of G. Let us recall that an ε-regular partition consists of parts $V_1, \ldots, V_m$ of the same size, and an exceptional part V_0 of size at most εn. For the purposes of the proof, it will be more convenient to further break up V_0 to single-element parts. Hence, we say that a partition $\mathcal{P}$ of $V(G)$ is an *ε-regular partition of order m* if $\mathcal{P}$ contains distinct parts $V_1, \ldots, V_m$ such that

(i) $|V_1| = |V_2| = \ldots = |V_m|$,

(ii) for all but at most εm^2 values of $1 \le i < j \le m$, the pair (V_i, V_j) is ε-regular, and

(iii) $\mathcal{P} \setminus \{V_1, \ldots, V_m\}$ consists of at most εn parts of size one.

The proof proceeds by starting with an initial partition of order m_0 satisfying (i) and (iii) and proceeds to refine it, increasing a suitably defined "quality" of the partition, so that the following conditions hold:

- each refinement increases the quality by at least $\frac{\varepsilon^5}{4}n^2$,

- each refinement increases the order of the partition at most exponentially, and
- once further refinement is not possible, (ii) holds.

Moreover, the quality is bounded from above by n^2, guaranteeing that the process ends after at most $\frac{4}{\varepsilon^5}$ iterations, at which point the order is bounded by the tower function of exponentials of height $\frac{4}{\varepsilon^5}$ and ending in m_0, i.e., a constant M depending only on ε and m_0.

Let us now show the details of this argument, starting with the definition of the quality function. For disjoint sets $A, B \subseteq V(G)$, let $q(A, B) = |A||B|d^2(A, B)$. For a partition $\mathcal{P} = \{P_1, \ldots, P_k\}$ of $V(G)$, let

$$q(\mathcal{P}) = \sum_{i<j} q(P_i, P_j).$$

Note that $q(A, B) \leq |A||B|$ and $q(\mathcal{P}) \leq \sum_{i<j} |P_i||P_j| < n^2$. Next, let us show that the quality does not decrease when we refine the partition.

Lemma 3. *If $\{A_1, \ldots, A_s\}$ is a partition of A and $\{B_1, \ldots, B_s\}$ is a partition of B, then $\sum_{i,j} q(A_i, B_j) \geq q(A, B)$.*

Proof. We have $e(A, B) = \sum_{i,j} e(A_i, B_j)$ and $|A||B| = \sum_{i,j} |A_i||B_j|$.

Let $u_{i,j} = \sqrt{|A_i||B_j|}d(A_i, B_j)$ and $v_{i,j} = \sqrt{|A_i||B_j|}$. Therefore,

$$\begin{aligned}
\sum_{i,j} u_{i,j}^2 &= \sum_{i,j} q(A_i, B_j) \\
\sum_{i,j} v_{i,j}^2 &= \sum_{i,j} |A_i||B_j| = |A||B| \\
\sum_{i,j} u_{i,j}v_{i,j} &= \sum_{i,j} d(A_i, B_j)|A_i||B_j| = \sum_{i,j} e(A_i, B_j) = e(A, B).
\end{aligned}$$

Cauchy-Schwarz inequality gives

$$|A||B| \sum_{i,j} q(A_i, B_j) \geq e^2(A, B) = |A|^2|B|^2d^2(A, B),$$

and thus

$$\sum_{i,j} q(A_i, B_j) \geq |A||B|d^2(A, B) = q(A, B).$$

□

We say that a partition $\mathcal{P}'$ is a *refinement* of a partition $\mathcal{P}$ if every part of $\mathcal{P}'$ is a subset of a part of $\mathcal{P}$. The *common refinement* of two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ is the partition $\{A \cap B : A \in \mathcal{P}_1, B \in \mathcal{P}_2, A \cap B \neq \emptyset\}$.

Corollary 4. *If a partition $\mathcal{P}'$ is a refinement of a partition $\mathcal{P}$, then $q(\mathcal{P}') \geq q(\mathcal{P})$.*

Proof. Let $\mathcal{P} = \{P_1, \ldots, P_s\}$. For $1 \leq i \leq s$, let $\mathcal{P}_i = \{Q \in \mathcal{P}' : Q \subseteq P_i\}$, so that $\mathcal{P}' = \mathcal{P}_1 \cup \ldots \cup \mathcal{P}_s$. Then

$$q(\mathcal{P}') \geq \sum_{1 \leq i < j \leq s} \sum_{Q \in \mathcal{P}_i, R \in \mathcal{P}_j} q(Q, R) \geq \sum_{1 \leq i < j \leq s} q(P_i, P_j) = q(\mathcal{P})$$

by Lemma 3. □

We also need to know that if (A, B) is not an ε-pair, then we can partition it with a substantial increase in quality.

Lemma 5. *Suppose (A, B) is a non-ε-regular pair of disjoint subsets of vertices of a graph G; i.e., there exist sets $A_1 \subseteq A$ and $B_1 \subseteq B$ such that $|A_1| \geq \varepsilon|A|$, $|B_1| \geq \varepsilon|B|$, and $|d(A_1, B_1) - d(A, B)| > \varepsilon$. For $A_2 = A \setminus A_1$ and $B_2 = B \setminus B_1$, we have $\sum_{1 \leq i,j \leq 2} q(A_i, B_j) > q(A, B) + \varepsilon^4|A||B|$.*

Proof. Let $d = d(A, B)$. For $1 \leq i, j \leq 2$, $u \in A_i$, and $v \in B_j$, let $d(u, v) = d(A_i, B_j)$ and $\delta(u, v) = d(u, v) - d$. Note that $|\delta(u, v)| > \varepsilon$ for $u \in A_1$ and $v \in B_1$. We have

$$\begin{aligned} d|A||B| = e(A, B) &= \sum_{1 \leq i,j \leq 2} e(A_i, B_j) \\ &= \sum_{1 \leq i,j \leq 2} d(A_i, B_j)|A_i||B_j| = \sum_{u \in A, v \in B} d(u, v) \\ &= d|A||B| + \sum_{u \in A, v \in B} \delta(u, v), \end{aligned}$$

and thus

$$\sum_{u \in A, v \in B} \delta(u, v) = 0.$$

Moreover, $q(A,B) = |A||B|d^2$ and

$$\begin{aligned}
\sum_{1\le i,j\le 2} q(A_i,B_j) &= \sum_{1\le i,j\le 2} |A_i||B_j|d^2(A_i,B_j) = \sum_{u\in A, v\in B} d^2(u,v)\\
&= \sum_{u\in A,v\in B} (d+\delta(u,v))^2 = \sum_{u\in A,v\in B} (d^2 + 2d\delta(u,v) + \delta^2(u,v))\\
&= |A||B|d^2 + 2d \sum_{u\in A,v\in B} \delta(u,v) + \sum_{u\in A,v\in B} \delta^2(u,v)\\
&= q(A,B) + \sum_{u\in A,v\in B} \delta^2(u,v) \ge q(A,B) + \sum_{u\in A_1,v\in B_1} \delta^2(u,v)\\
&> q(A,B) + |A_1||B_1|\varepsilon^2 \ge q(A,B) + \varepsilon^4|A||B|.
\end{aligned}$$

□

Let $\mathcal{P} = \{V_1,\dots,V_s,V_{s+1},\dots,V_{s+t}\}$ be a partition of $V(G)$ such that $t \le \varepsilon n$, $|V_i| = 1$ for $i \ge s+1$ and $|V_1| = \ldots = |V_s| = N$. Note that $sN = n - t \ge (1-\varepsilon)n \ge n/2$. Suppose that $\mathcal{P}$ is not an ε-regular partition of order s, i.e., that it does not satisfy (ii). Let X be the set of pairs (i,j) such that $1 \le i < j \le s$ and (V_i,V_j) is not ε-regular; we have $|X| \ge \varepsilon s^2$. For each $(i,j) \in X$, let $\mathcal{P}_{ij} = \{V_{i,j,1}, V_{i,j,2}, V_{j,i,1}, V_{j,i,2}, V(G)\setminus(V_i\cup V_j)\}$ be the partition of $V(G)$ where

$$\begin{aligned}
V_i = V_{i,j,1}\cup V_{i,j,2}, V_j = V_{j,i,1}\cup V_{j,i,2},\\
|V_{i,j,1}| \ge \varepsilon|V_i|, |V_{j,i,1}| \ge \varepsilon|V_j|, \text{ and}\\
|d(V_{i,j,1},V_{j,i,1}) - d(V_i,V_j)| > \varepsilon.
\end{aligned}$$

Let $\mathcal{P}'$ be the common refinememt of $\mathcal{P}$ and all partitions $\mathcal{P}_{ij}$ for $(i,j) \in X$.

Let $\mathcal{V}_i = \{A \in \mathcal{P}' : A \subseteq V_i\}$. We have

$$q(\mathcal{P}') \ge \sum_{i<j} \sum_{A\in\mathcal{V}_i, B\in\mathcal{V}_j} q(A,B).$$

Lemma 3 implies

$$\sum_{A\in\mathcal{V}_i,B\in\mathcal{V}_j} q(A,B) \ge q(V_i,V_j)$$

for every $i < j$. Moreover, if $(i,j) \in X$, then

$$\sum_{A\in\mathcal{V}_i,B\in\mathcal{V}_j} q(A,B) \ge \sum_{1\le o,p\le 2} q(V_{i,j,o},V_{j,i,p}) > q(V_i,V_j) + \varepsilon^4|V_i||V_j|$$

by Lemma 5. We conclude that

$$q(\mathcal{P}') > \left(\sum_{i<j} q(V_i, V_j)\right) + |X|\varepsilon^4 N^2 \geq q(\mathcal{P}) + \varepsilon^5 s^2 N^2 \geq q(\mathcal{P}) + \frac{\varepsilon^5}{4} n^2.$$

The partition $\mathcal{P}'$ has at most $s2^s$ non-singleton parts (each of the parts $V_1, \ldots, V_s$ is contained in less than s pairs of X, and thus it is divided less than s times). However, the non-singleton parts can have different sizes. Let $f(s) = s4^s$. We divide every non-singleton part of $\mathcal{P}'$ into as many parts of size $\lceil \frac{N}{4^s} \rceil$ as possible, and their remainders to singleton parts. This gives a refinement $\mathcal{P}''$ of $\mathcal{P}'$ such that

- every non-singleton part of $\mathcal{P}''$ has size exactly $\lceil \frac{N}{4^s} \rceil$,
- the number of non-singleton parts of $\mathcal{P}''$ is at most $f(s)$, since $\mathcal{P}$ has s non-singleton parts of size N and at most 4^s parts of size $\lceil \frac{N}{4^s} \rceil$ fit into each of them, and
- the number of singleton parts of $\mathcal{P}''$ is by at most $s2^s \frac{N}{4^s} \leq s\frac{n/s}{2^s} = \frac{n}{2^s}$ larger than the number of singleton parts of $\mathcal{P}$.

Moreover, $q(\mathcal{P}'') \geq q(\mathcal{P}') > q(\mathcal{P}) + \frac{\varepsilon^5}{4} n^2$.

Proof of Regularity Lemma. Let

$$s_0 = \max\left(2m_0, \left\lceil \log \frac{8}{\varepsilon^6} \right\rceil\right), \text{ and}$$
$$M = \max\left(\left\lceil \frac{2}{\varepsilon} s_0 \right\rceil, f(f(\ldots f(s_0)\ldots))\right),$$

where f is iterated $\lfloor \frac{4}{\varepsilon^5} \rfloor$-times.

If G has at most M vertices, then G has an ε-regular partition to $|V(G)| \leq M$ singleton parts. Otherwise, let us start with an arbitrary partition of $V(G)$ to s_0 parts of the same size and less than $s_0 \leq \frac{\varepsilon}{2} n$ singleton parts. We repeat the refinement procedure described until we reach an ε-regular partition $\mathcal{P}$. Since $0 \leq q(\mathcal{Q}) \leq n^2$ for every partition $\mathcal{Q}$ of $V(G)$ and each iteration of the procedure increases the value of q by more than $\frac{\varepsilon^5}{4} n^2$, the number of iterations is at most $\frac{4}{\varepsilon^5}$, and thus $\mathcal{P}$ has at most M non-singleton parts.

We need to verify that the number of singleton parts does not increase over εn. Each iteration increases the number of singleton parts by at most $\frac{n}{2^s} \leq \frac{n}{2^{s_0}}$. Therefore, after $\lfloor \frac{4}{\varepsilon^5} \rfloor$ iterations, there are at most $\frac{4}{\varepsilon^5 2^{s_0}} n + \frac{\varepsilon}{2} n \leq \varepsilon n$ singleton parts.

Since the initial partition consists of parts of size at most n/s_0, this is also an upper bound on the size of the parts in $\mathcal{P}$. Consequently, the order of $\mathcal{P}$ is at least $\frac{(1-\varepsilon)n}{n/s_0} \geq s_0/2 \geq m_0$.

$\square$