Chapter 4

Regularity partitions

4.0.1 Regularity Lemmas

In order to formulate this application of geometric representations, we have to collect some facts about regularity lemmas. The original Regularity Lemma of Szemerédi has been the key to many proofs in extremal graph theory, graph algorithms, number theory, the theory of graph limits, and more. It has formulations not only in graph theory but in analysis and information theory as well. Its applications often depend on different versions of the Lemma, which are not equivalent. A version that is weaker than the original but gives much better bounds was found by Frieze and Kannan, and this is the version we need.

Let G be a graph on n nodes. For two sets $S, T \subseteq V$, let $e_G(S,T)$ denote the number of edges $ij \in E$ with $i \in S$ and $j \in T$. We need this definition in the case when S and T are not disjoint; in this case, edges induced by $S \cap T$ should be counted twice. We also use this notation in the case when G is edge-weighted, when the weights of edges connecting S and T should be added up. (An unweighted edge will be considered as an edge with weight 1.)

We come to an important definition of this section: We define, for two edge-weighted graphs G and H on the same set of n nodes, their *cut-distance* by

$$d_\square(G,H) = \max_{S,T\subseteq V} \frac{|e_G(S,T) - e_H(S,T)|}{n^2}.$$

This is, of course, not the only way to define a meaningful distance between two graphs; for example, the *edit distance* $|E(G)\triangle E(H)|$ is often used. For us, the cut-distance, which measures a certain global similarity, will be more important.

Let $\mathcal{P} = \{V_1, \dots, V_k\}$ be a partition of V into nonempty sets. We define the edge-weighted graph $G_\mathcal{P}$ on V by taking the complete graph and weighting its edge uv by $e_G(V_i,V_j)/(|V_i|\,|V_j|)$ if $u \in V_i$ and $v \in V_j$. The case $i = j$ takes some care: we then count edges twice, so $e_G(V_i,V_i) = 2|E(G[V_i])|$, and we include the case $u = v$, so $G_\mathcal{P}$ will have loops. Informally, $G_\mathcal{P}$ is obtained by averaging the adjacency matrix over sets $V_i \times V_j$.

The Regularity Lemma says, roughly speaking, that the node set of every graph has a partition $\mathcal{P}$ into a "small" number of classes such that $G_{\mathcal{P}}$ is "close" to G.

Lemma 4.0.1 (Weak Regularity Lemma) *For every simple graph G and every $k \geq 1$, the node set V has a partition $\mathcal{P}$ into k classes such that*

$$d_{\square}(G, G_{\mathcal{P}}) \leq \frac{4}{\sqrt{\log k}}. \qquad \square$$

We do not require here that $\mathcal{P}$ be an equitable partition; it is not hard to see that this version implies that there is also a k-partition in which in addition all partition classes are almost equal in the sense that they have $\lfloor n/k \rfloor$ or $\lceil n/k \rceil$ elements, and which satisfies the same inequality as in the lemma, just we have to double the error bound.

It will be useful to reformulate this lemma in terms of matrices. For two matrices $A, B \in \mathbb{R}^{n\times n}$, we define their *inner product*

$$\langle A, B\rangle = \frac{1}{n^2}\sum_{i,j=1}^{n} A_{ij}B_{ij}.$$

For $S, T \subseteq \{1, \dots, n\}$, we need the special matrix $\mathbb{1}_{S\times T}$ defined by

$$(\mathbb{1}_{S\times T})_{i,j} = \begin{cases} 1, & \text{if } i \in S \text{ and } j \in T, \\ 0, & \text{otherwise.} \end{cases}$$

This is a 0-1 matrix of rank 1. For a matrix $A \in \mathbb{R}^{n\times n}$ and $S, T \subseteq \{1, \dots, n\}$, we define

$$A(S, T) = \sum_{i\in S,\, j\in T} A_{i,j} = n^2\langle A, \mathbb{1}_{S\times T}\rangle$$

We define the norm

$$\|A\|_2 = \sqrt{\frac{1}{n^2}\sum_{i,j=1}^{n} A_{i,j}^2}.$$

The *cut-norm* is less standard, and is used mostly in combinatorics. It is defined as follows:

$$\|A\|_{\square} = \frac{1}{n^2}\max_{S,T\subseteq[n]} |A_{S,T}| = \max_{S,T\subseteq[n]} \langle A, \mathbb{1}_{S\times T}\rangle. \tag{4.1}$$

Applying this notation to the adjacency matrices of two graphs G and H on the same set of n nodes, we get

$$d_{\square}(G, H) = \|A_G - A_H\|_{\square}.$$

For a matrix $A \in \mathbb{R}^{n\times n}$ and partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of $\{1, \dots, n\}$, we define the matrix $A_{\mathcal{P}} \in \mathbb{R}^{n\times n}$ by

$$(A_{\mathcal{P}})_{u,v} = \frac{A(V_i, V_j)}{|V_i|\,|V_j|} \qquad \text{for } u \in V_i, v \in V_j.$$

In words, the matrix $A_{\mathcal{P}}$ is obtained by averaging the entries of A over every block $V_i \times V_j$. If A is the adjacency matrix of a simple graph G, then $A_{\mathcal{P}}$ is the weighted adjacency matrix of $G_{\mathcal{P}}$.

Using the cut norm, we can re-state the Weak Regularity Lemma as follows. In fact, we state it in three versions, to make the proof easier to follow. Let as say that $B \in \mathbb{R}^{n\times n}$ is a $\mathcal{P}$-matrix (where $\mathcal{P} = \{U_1, \dots, U_k\}$ is a partition of $\{1, \dots, n\}$), if B is constant on every block $U_i \times U_j$, i.e., $B_{\mathcal{P}} = B$.

Lemma 4.0.2 *Let $A \in \mathbb{R}^{n\times n}$ and $k \ge 1$.*

(a) There are $2k$ sets $S_1, \dots, S_k, T_1, \dots, T_k \subseteq [n]$ and k real numbers a_i such that

$$\Big\| A - \sum_{i=1}^{k} a_i \mathbb{1}^{\mathsf{T}}_{S_i\times T_i} \Big\|_{\square} \le \frac{1}{\sqrt{k}} \|A\|_2.$$

(b) There is a partition $\mathcal{P}$ of $\{1, \dots, n\}$ into k classes, and a $\mathcal{P}$-matrix B, for which

$$\|A - B\|_{\square} \le \frac{2}{\sqrt{\log k}} \|A\|_2.$$

(c) There is a partition $\mathcal{P}$ of $\{1, \dots, n\}$ into k classes for which

$$\|A - A_{\mathcal{P}}\|_{\square} \le \frac{4}{\sqrt{\log k}} \|A\|_2.$$

Note that $\mathbb{1}_{S_i\times T_i}$ is just a 0-1 matrix of rank 1, showing that Lemma 4.0.2(a) gives a low-rank approximation of A. The important fact about version (a) of the lemma is that the number of terms (k) and the error bound ($1/\sqrt{k}$) are polynomially related.

Proof. (a) Let A be an arbitrary $n \times n$ matrix, and let $S, T \subseteq \{1, \dots, n\}$ be two nonempty sets such that

$$\|A\|_{\square} = \frac{1}{n^2} |A(S, T)| = |\langle A, \mathbb{1}_{S\times T}\rangle|.$$

We claim that

$$\|A - a\mathbb{1}_{S\times T}\|_2^2 \le \|A\|_2^2 - \|A\|_{\square}^2, \tag{4.2}$$

where

$$a = \frac{A(S, T)}{|S|\,|T|} = \frac{n^2}{|S|\,|T|} \langle A, \mathbb{1}_{S\times T}\rangle$$

(the value of a will not be relevant in the sequel). Indeed,

$$\begin{aligned} \|A - a\mathbb{1}_{S\times T}\|_2^2 &= \|A\|^2 + a^2\|\mathbb{1}_{S\times T}\|^2 - 2a\langle A, \mathbb{1}_{S\times T}\rangle \\ &= \|A\|^2 - \frac{n^2}{|S|\,|T|} \langle A, \mathbb{1}_{S\times T}\rangle^2 \le \|A\|_2^2 - \|A\|_{\square}^2. \end{aligned}$$

Repeated application of this inequality gives a sequence of matrices

$$A_r = A - \sum_{k=1}^{r} a_k \mathbb{1}_{S_i \times T_i} \qquad (r = 0, 1, \dots, k),$$

where $A_0 = A$ and

$$\|A_{r+1}\|_2^2 \le \|A_r\|_2^2 - \|A_r\|_\square^2.$$

Hence

$$\|A_k\|_2^2 \le \|A\|_2^2 - \sum_{r=0}^{k-1} \|A_r\|_\square^2.$$

Since the left side is nonnegative, there must be an index r for which

$$\|A_r\|_\square^2 = \Big| A - \sum_{k=1}^{r} a_k \mathbb{1}_{S_i \times T_i} \Big|_\square^2 \le \frac{1}{k} \|A\|_2^2.$$

Replacing the remaining coefficients $a_{r+1}, \dots, a_{k-1}$ by 0, we get the decomposition in (a).

(b) Consider the decomposition as in (a), and the partition into the atoms of the Boolean algebra generated by all the sets S_i and T_i.

(c) This follows by (b) and the following observation: for every matrix $A \in \mathcal{R}^n$, every partition $\mathcal{P}$ of $\{1, \dots, n\}$ and every $\mathcal{P}$-matrix B,

$$\|A - A_\mathcal{P}\|_\square \le 2|A - B|. \tag{4.3}$$

Indeed,

$$\|A - A_\mathcal{P}\|_\square \le |A - B| + |B - A_\mathcal{P}| = |A - B| + |B_\mathcal{P} - A_\mathcal{P}| = |A - B| + |(B - A)_\mathcal{P}| \le 2|A - B|.$$

□

Remarks. 1. There are many alternative ways to define the cut-norm, or norms that are closely related. We could define it by

$$\|A\|_\square = \frac{1}{n^2} \max\{|\mathbf{x}^\mathsf{T} A \mathbf{y}| : \ \mathbf{x}, \mathbf{y} \in [0,1]^n\}. \tag{4.4}$$

If we maximize instead over all vectors in $[-1,1]^n$, we obtain a norm that may be different, but only by a factor of 4:

$$\|A\|_\square \le \frac{1}{n^2} \max\{|\mathbf{x}^\mathsf{T} A \mathbf{y}| : \ \mathbf{x}, \mathbf{y} \in [-1,1]^n\} \le 4\|A\|_\square. \tag{4.5}$$

We can also play with the sets S and T in the definition. For example,

$$\frac{1}{2}\|A\|_\square \le \frac{1}{n^2} \max_{S \subseteq [n]} |A(S,S)| \le \|A\|_\square. \tag{4.6}$$

and

$$\frac{1}{4}\|A\|_\square \le \frac{1}{n^2} \max_{|S|,|T|\ge n/2} |A(S,T)| \le \|A\|_\square. \tag{4.7}$$

2. In (a), one could require in addition that $\sum_r a_r^2 \le 4$, at the cost of increasing the error bound to $(4/\sqrt{k})\|A\|_2$. The trick is to use maximizing sets in (4.7).

3. The original Regularity Lemma measures the error of the approximation $A \approx A_{\mathcal{P}}$ differently. Again there are different versions, closest to ours is the following: For a matrix $A \in \mathbb{R}^{n\times n}$ and partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of $\{1, \dots, n\}$, and for every $1 \le i, j \le k$, we choose the sets $S_{ij} \subseteq V_i$ and $T_{ij} \subseteq V_j$ which maximize the "local error" $\langle A, \mathbb{1}_{S_{ij}\times T_{ij}}\rangle$. The total error is the sum of these:

$$e(\mathcal{P}) = \sum_{i,j=1}^{n} |\langle A - A_{\mathcal{P}}, \mathbb{1}_{S_{ij}\times T_{ij}}\rangle| = \Big|\Big\langle A - A_{\mathcal{P}}, \sum_{i,j=1}^{n} \mathbb{1}_{S_{ij}\times T_{ij}}\Big\rangle\Big|.$$

A version of the proof above gives the following: There is a partition $\mathcal{P}$ of $\{1, \dots, n\}$ into k classes for which

$$e(\mathcal{P}) = O\Big(\frac{1}{\log^* k}\|A\|_2\Big).$$

(Here $\log^* k$ is a very slowly decreasing function, defined as the number of times we have to apply the log function to k to get a negative number.) This bound is also known to be best possible.