Deriving Szemerédi's regularity lemma for graphs from the compactness of the graphon space

Notes for our seminar — Lex Schrijver

Abstract. We give a derivation of Szemerédi's regularity lemma for graphs from the compactness of the graphon space (Lovász and Szegedy [2] — see Lovász [1]).

Let W be the space of symmetric measurable functions $[0,1]^2 \to [0,1]$. Let Π be the collection of partitions of $[0,1]$ into finitely many measurable sets, each of positive measure. For any $P \in \Pi$, let L_P be subspace of W spanned by the functions $\chi^{C\times D}$ with $C, D \in P$. For any $w \in W$, let w_P be the orthogonal projection of w onto L_P. A partition P is *balanced* if all classes have the same measure. All norms are L_2-norms.

Lemma 1. *Let $P, Q \in \Pi$, with P balanced. Then for each $\varepsilon > 0$ there exists an $R \in \Pi$ with $Q \le R$ and $|R| \le (1 + 2|P|^2/\varepsilon)^{|P|}$, such that $\|x_Q - x_R\| \le \varepsilon\|x\|$ for each $x \in L_P$.*

Proof. Let $p := |P|$ an $\delta := \varepsilon/2p^2$. For $C \in P$ and $A \in Q$ set $\alpha_{C,A} := \mu(C \cap A)/\mu(A)$, and let $\alpha'_{C,A}$ be obtained by rounding $\alpha_{C,A}$ down to an integer multiple of δ.

Let R be the partition with $Q \le R$ given as follows. Sets $A, B \in Q$ are in the same class of R if and only if $\alpha'_{C,A} = \alpha'_{C,B}$ for each $C \in P$. As $[0,1]$ contains at most $1 + 1/\delta$ integer multiples of δ, $|R| \le (1 + 1/\delta)^p$.

Consider any $x \in L_P$. So x can be written as $x = \sum_{C,D\in P} \lambda_{C,D}\chi^{C\times D}$. Hence

$$(1) \qquad x_Q = \sum_{C,D\in P} \lambda_{C,D}(\chi^{C\times D})_Q = \sum_{C,D\in P} \lambda_{C,D} \sum_{A,B\in Q} \alpha_{C,A}\alpha_{D,B}\chi^{A\times B}.$$

Define

$$(2) \qquad w := \sum_{C,D\in P} \lambda_{C,D} \sum_{A,B\in Q} \alpha'_{C,A}\alpha'_{D,B}\chi^{A\times B}.$$

As w has the same value on $A \times B$ as on $A' \times B'$ whenever $\alpha'_{C,A} = \alpha'_{C,A'}$ and $\alpha'_{C,B} = \alpha'_{C,B'}$ for all $C \in P$, we know that w belongs to L_R. Moreover, using Cauchy-Schwarz and $|\alpha_{C,A}\alpha_{D,B} - \alpha'_{C,A}\alpha'_{D,B}| \le |(\alpha_{C,A} - \alpha'_{C,A})\alpha_{D,B}| + |\alpha'_{C,A}(\alpha_{D,B} - \alpha'_{D,B})| \le 2\delta$,

$$(3) \qquad \|x_Q - w\|^2 = \sum_{A,B\in Q} \Big(\sum_{C,D\in P} \lambda_{C,D}(\alpha_{C,A}\alpha_{D,B} - \alpha'_{C,A}\alpha'_{D,B})\Big)^2 \mu(A)\mu(B) \le \sum_{A,B\in Q} p^2 \sum_{C,D\in P} \lambda^2_{C,D}(2\delta)^2\mu(A)\mu(B) = 4p^2\delta^2 \sum_{C,D\in P} \lambda^2_{C,D} = 4p^4\delta^2 \sum_{C,D\in P} \lambda^2_{C,D}\mu(C)\mu(D) = 4p^4\delta^2\|x\|^2 = \varepsilon^2\|x\|^2.$$

∎

Call a partition P of a finite set V *ε-balanced* if $P \setminus P'$ is balanced for some $P' \subseteq P$ with

$|\bigcup P'| \le \varepsilon |V|$.

Lemma 2. *Let $\varepsilon > 0$. Then each partition P of a finite set V has an ε-balanced refinement Q with $|Q| \le (1 + 1/\varepsilon)|P|$.*

Proof. Define $t := \varepsilon|V|/|P|$. Split each class of P into classes, each of size $\lceil t \rceil$, except for at most one of size less than t. This gives Q. Then $|Q| \le |P| + |V|/t = (1 + 1/\varepsilon)|P|$. Moreover, the union of the classes of Q of size less than t has size at most $|P|t = \varepsilon|V|$. So Q is ε-balanced. ∎

Given a graph $G = (V, E)$, a *rectangle* is a set $R = X \times Y$ with $X, Y \subseteq V$. If $R \ne \emptyset$, let $d(R) := e(R)/|R|$, where $e(R)$ is the number of adjacent pairs of vertices in R.

Theorem 1 (Szemerédi's regularity lemma)**.** *For each $\varepsilon > 0$ and $p \in \mathbb{N}$ there exists $k_{p,\varepsilon} \in \mathbb{N}$ such that for each graph $G = (V, E)$ and each partition P of V with $|P| = p$ there is an ε-balanced refinement Q of P with $|Q| \le k_{p,\varepsilon}$ and*

$$(4) \qquad \sum_{A,B \in Q} \max_{\substack{\emptyset \ne R \subseteq A \times B \\ R \text{ rectangle}}} |R| \cdot |d(R) - d(A \times B)| < \varepsilon |V|^2.$$

Proof. Fix $\varepsilon > 0$ and $p \in \mathbb{N}$. Define $g(t) := p(1 + 1/\varepsilon)(1 + 8t^2/\varepsilon)^t$ for each $t \in \mathbb{N}$. For each $w \in W$, let t_w be the minimum size of a balanced partition $T \in \Pi$ such that $\|w - w_T\| < \varepsilon/4$. By the compactness of the graphon space there exists a finite $F \subseteq W$ such that for each $w \in W$ there is an $f \in F$ and a measure-preserving measurable permutation ϕ of $[0, 1]$ such that for all measurable $X, Y \subseteq [0, 1]$:

$$(5) \qquad |(w - f^\phi)(X \times Y)| < \varepsilon/4g(t_w)^2.$$

Let $k_{p,\varepsilon} := \max\{g(t_f) \mid f \in F\}$. We show that $k_{p,\varepsilon}$ is as required.

Let $G = ([n], E)$ be a graph and let w be the element of W corresponding to G. Let N be the partition of $[0, 1]$ into n equal intervals.

By the above there exists an $f \in F$ and a measure-preserving measurable permutation ϕ of $[0, 1]$ such that (5) holds. Set $u := f^\phi$. So $t := t_u = t_f$. Hence there is a balanced partition $T \in \Pi$ with $|T| = t$ and $\|u - u_T\| < \varepsilon/4$. Define $x := u_T$.

By Lemma 1, there is a partition $U \in \Pi$ with $U \ge N$ such that $|U| \le (1 + 8t^2/\varepsilon)^t$ and $\|x_N - x_U\| \le \varepsilon/4$. Let $S := P \wedge U$. So $|S| \le |P||U| \le p(1 + 8t^2/\varepsilon)^t$. By Lemma 2, there is an ε-balanced refinement Q of S with $N \le Q \le S$ and $|Q| \le (1 + 1/\varepsilon)|S| \le g(t) \le k_{p,\varepsilon}$. We show that this Q gives the partition of the theorem.

For each $A, B \in Q$, choose $R = X \times Y \subseteq A \times B$, where X and Y are unions of classes of N such that $|(w - w_Q)(R)|$ is maximized. Let $\mathcal{R}$ be the collection of these chosen R. By (5), $|(w - u)(R)| < \varepsilon/4g(t)^2$ for all $R \in \mathcal{R}$, and hence

$$(6) \qquad \sum_{R \in \mathcal{R}} |(w - u)(R)| < |\mathcal{R}|\varepsilon/4g(t)^2 \le \varepsilon/4.$$

Since $w_Q - u_Q$ is constant on $A \times B$, we also have for any $R \in \mathcal{R}$ with $R \subseteq A \times B$:

$$|(w_Q - u_Q)(R)| \leq |(w_Q - u_Q)(A \times B)| = |(w - u)(A \times B)| \leq \varepsilon/4g(t)^2. \tag{7}$$

Hence we obtain, similarly to (6), $\sum_{R\in\mathcal{R}} |(w_Q - u_Q)(R)| < \varepsilon/4$.

Finally, as $u(R) = u_N(R)$ for all $R \in \mathcal{R}$ and as $u_Q = (u_N)_Q$ is the nearest point on L_Q nearest to u_N, while $x_U \in L_U \subseteq L_Q$, with Cauchy-Schwarz we get (as $\| \sum_{R\in\mathcal{R}} \pm\chi^R\| \leq 1$)

$$\begin{aligned} \sum_{R\in\mathcal{R}} |(u - u_Q)(R)| &= \sum_{R\in\mathcal{R}} |(u_N - u_Q)(R)| \leq \|u_N - u_Q\| \leq \|u_N - x_U\| \leq \\ &\|(u-x)_N\| + \|x_N - x_U\| \leq \|u - x\| + \|x_N - x_U\| < \varepsilon/2. \end{aligned} \tag{8}$$

Hence $\sum_{R\in\mathcal{R}} |(w - w_Q)(R)| < \varepsilon$. For the graph G it means (4). ∎

To interpret (4), for $A, B \in Q$, let $m_{A,B}$ denote the maximum described in (4). Let Q' be such that $Q \setminus Q'$ is balanced and $|\bigcup Q'| \leq \varepsilon|V|$. Set $Q'' := Q \setminus Q'$, and let Z be the collection of pairs $(A, B) \in Q'' \times Q''$ with $m_{A,B} \geq \sqrt{\varepsilon}|A||B|$. Then (4) implies

$$\sum_{(A,B)\in Z} |A||B| \leq \sum_{(A,B)\in Z} \varepsilon^{-1/2} m_{A,B} \leq \sqrt{\varepsilon}|V|^2. \tag{9}$$

Moreover, as $|\bigcup Q'| < \varepsilon|V|$,

$$\sum_{A,B\in Q''} |A||B| \geq \sum_{A,B\in Q} |A||B| - 2\varepsilon|V|^2 = (1 - 2\varepsilon)|V|^2. \tag{10}$$

Hence, assuming $\varepsilon < 1/4$, $|Z| \leq \sqrt{\varepsilon}(1 - 2\varepsilon)^{-1}|Q''|^2 < 2\sqrt{\varepsilon}|Q''|^2$. For each $(A, B) \in (Q'' \times Q'') \setminus Z$ one has $m_{A,B} < \sqrt{\varepsilon}|A||B|$, implying that for each rectangle $R \subseteq A \times B$ with $|R|/|A \times B| \geq \sqrt[4]{\varepsilon}$ one has $|d(R) - d(A \times B)| < \sqrt[4]{\varepsilon}$. In other words, $A \times B$ is $\sqrt[4]{\varepsilon}$-regular.

References

[1] L. Lovász, *Large Graphs, Graph Homomorphisms and Graph Limits*, American Mathematical Society, Providence, R.I., to appear.

[2] L. Lovász, B. Szegedy, Limits of dense graph sequences, *Journal of Combinatorial Theory, Series B* 96 (2006) 933–957.