V. Szemerédi's regularity lemma

1. Preliminaries

The 'regularity lemma' of Endre Szemerédi [5] roughly asserts that, for each $\varepsilon > 0$, there exists a number k such that the vertex set V of any graph $G = (V, E)$ can be partitioned into at most k *almost* equal-sized classes so that between *almost* any two classes, the edges are distributed *almost* homogeneously. Here the meaning of *almost* depends on ε. We will make this precise and prove it in Section 2. First, some 'ε-free' preliminaries.

Let $G = (V, E)$ be a graph. For nonempty $A, B \subseteq V$, define

$$(1) \qquad e(A,B) := \text{number of adjacent pairs in } A \times B,$$

$$d(A,B) := \frac{e(A,B)}{|A||B|} \qquad \text{and} \qquad c(A,B) := \frac{e(A,B)^2}{|A||B|}.$$

Moreover, if P and Q are collections of nonempty sets, define

$$(2) \qquad c(P,Q) := \sum_{X \in P, Y \in Q} c(X,Y) \qquad \text{and} \qquad c(P) := c(P,P).$$

A *partition* of a set X is a collection of pairwise disjoint nonempty sets with union X. Observe that, if $\lambda_1, \dots, \lambda_n, \alpha_1, \dots, \alpha_n \in \mathbb{R}$ with $\lambda_1 + \cdots + \lambda_n = 1$ and $\alpha := \sum_{i=1}^n \lambda_i \alpha_i$, then

$$(3) \qquad \sum_{i=1}^n \lambda_i \alpha_i^2 = \alpha^2 + \sum_{i=1}^n \lambda_i (\alpha_i - \alpha)^2.$$

This implies, if P and Q are partitions of the nonempty sets A and B respectively,

$$(4) \qquad c(P,Q) = c(A,B) + \sum_{X \in P, Y \in Q} |X||Y|(d(X,Y) - d(A,B))^2.$$

Indeed, for $X \in P$ and $Y \in Q$, define $\lambda_{X,Y} := |X||Y|/|A||B|$ and $\alpha_{X,Y} := d(X,Y)$. Then $\sum_{X,Y} \lambda_{X,Y} = 1$ and $d(A,B) = \sum_{X,Y} \lambda_{X,Y} d(X,Y)$ (as $e(A,B) = \sum_{X,Y} e(X,Y)$). Appropriate substitution in (3) and multiplying both sides by $|A||B|$ gives (4).

Equality (4) implies $c(P,Q) \geq c(A,B)$, which in turn implies the following. Call a partition P' of a set A a *refinement* of partition P of A if each set in P' is contained in some set in P. Then, if P' and Q' are refinements of P and Q respectively,

$$(5) \qquad c(P', Q') \geq c(P,Q).$$

2. Szemerédi's regularity lemma

Let $\varepsilon > 0$ and let V be any set. Call a partition P of V *ε-balanced* if P contains a subcollection $\mathcal{C}$ such that all sets in $\mathcal{C}$ have the same size and such that $|V \setminus \bigcup \mathcal{C}| \leq \varepsilon |V|$.

Lemma 1. *Each partition P of V has an ε-balanced refinement Q with $|Q| \leq (1+\varepsilon^{-1})|P|$.*

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 $|Q| \leq |P| + |V|/t = (1+\varepsilon^{-1})|P|$. 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. ∎

Now, let $\varepsilon > 0$ and let $G = (V, E)$ be a graph. Call a pair (A, B) of subsets $A, B \subseteq V$ *ε-regular* if for all $X \subseteq A$ and $Y \subseteq B$:

$$\text{(6)} \qquad \text{if } |X| > \varepsilon|A| \text{ and } |Y| > \varepsilon|B| \text{ then } |d(X,Y) - d(A,B)| \leq \varepsilon.$$

Call a partition P of V *ε-regular* if

$$\text{(7)} \qquad \sum_{\substack{A,B \in P \\ (A,B)\ \varepsilon\text{-irregular}}} |A||B| \leq \varepsilon|V|^2.$$

Define $f_\varepsilon(x) = (1+\varepsilon^{-1})x4^x$ for $x \in \mathbb{R}$. For $n \in \mathbb{N}$, f_ε^n is the n-th iterate of f_ε.

Theorem 1 (Szemerédi's regularity lemma). *For each $\varepsilon > 0$ and graph $G = (V, E)$, each partition P of V has an ε-balanced ε-regular refinement Q with $|Q| \leq f_\varepsilon^{\lceil \varepsilon^{-5} \rceil}(|P|)$.*

Proof. Set $P_0 := P$. For $i \geq 0$, if P_i has been set, let P_{i+1} be an ε-balanced refinement of P_i with $|P_{i+1}| \leq f_\varepsilon(|P_i|)$ and with $c(P_{i+1})$ maximal. Using (5) we know $0 \leq c(P_i) \leq c(T) = 2|E| \leq |V|^2$ for each i, where T is the trivial partition of V into singletons. Hence $c(P_{i+1}) \leq c(P_i) + \varepsilon^5|V|^2$ for some i with $1 \leq i \leq \lceil \varepsilon^{-5} \rceil$. Set $Q := P_i$. So $|Q| \leq f_\varepsilon^i(|P|) \leq f_\varepsilon^{\lceil \varepsilon^{-5} \rceil}(|P|)$. As $Q = P_i$ is ε-balanced, it suffices to prove that Q is ε-regular.

Suppose Q is not ε-regular. For each ε-irregular pair $(A, B) \in Q^2$, we can choose $X \subseteq A$ and $Y \subseteq B$ with $|X| > \varepsilon|A|$, $|Y| > \varepsilon|B|$, and $|d(X,Y) - d(A,B)| > \varepsilon$. Define partitions $U_{A,B} := \{X, A \setminus X\}$ of A and $W_{A,B} := \{Y, B \setminus Y\}$ of B. Then (4) implies:

$$\text{(8)} \qquad c(U_{A,B}, W_{A,B}) \geq c(A,B) + |X||Y|(d(X,Y) - d(A,B))^2 > c(A,B) + \varepsilon^4|A||B|.$$

Now for each fixed $A \in Q$, all partitions $U_{A,B}$ and $W_{B,A}$ of A (over all B with (A, B) ε-irregular) have a common refinement R_A with $|R_A| \leq 2^{2|Q|}$ (as $|U_{A,B}| = |W_{B,A}| = 2$). Let $R := \bigcup_{A \in Q} R_A$. So $|R| \leq |Q|4^{|Q|}$. Let S be an ε-balanced refinement of R with $|S| \leq (1+\varepsilon^{-1})|R|$ (which exists by Lemma 1). So $|S| \leq (1+\varepsilon^{-1})|Q|4^{|Q|} = f_\varepsilon(|Q|)$.

Moreover, for ε-irregular $(A, B) \in Q^2$, $c(R_A, R_B) \geq c(U_{A,B}, W_{A,B}) > c(A,B) + \varepsilon^4|A||B|$ (by (5) and (8)). So, as $c(R_A, R_B) \geq c(A, B)$ for any $A, B \in Q$ by (5), negating (7) gives

$$\text{(9)} \qquad c(S) - c(Q) \geq c(R) - c(Q) = \sum_{A,B \in Q} c(R_A, R_B) - c(A,B) \geq \varepsilon^4 \sum_{\substack{A,B \in Q \\ (A,B)\ \varepsilon\text{-irregular}}} |A||B| > \varepsilon^5|V|^2.$$

Therefore, $c(S) > c(Q) + \varepsilon^5|V|^2 \geq c(P_{i+1})$, contradicting the maximality of P_{i+1}. ∎

It is important to observe that the bound on $|Q|$, though generally huge, only depends on ε and $|P|$, and not on the size of the graph. Gowers [1] showed that the bound necessarily is huge (at least a tower of powers of 2's of height proportional to $\varepsilon^{-1/16}$).

Exercise

2.1. Let P be an ε-balanced ε-regular partition of V, and let $\mathcal{C}$ be as above. Prove that at most $(\varepsilon/(1-\varepsilon)^2)|\mathcal{C}|^2$ pairs in $\mathcal{C}^2$ are ε-irregular.

3. Δ-graphs

Call a graph $G=(V,E)$ a Δ-*graph* if each edge belongs to a unique triangle. For any n, let $\tau(n)$ be the maximum number of edges of any Δ-graph on n vertices.[1]

Theorem 2. $\tau(n)=o(n^2)$.

Proof. Let $\varepsilon>0$ with $\varepsilon<\frac{1}{2}$. Set $k_\varepsilon := f_\varepsilon^{\lfloor \varepsilon^{-5}\rfloor}(1)$. It suffices to prove:

(10) Let $G=(V,E)$ be a Δ-graph with $n:=|V|\geq k_\varepsilon/\varepsilon^3$. Then $|E|\leq 12\varepsilon n^2$.

By Szemerédi's regularity lemma, V has an ε-regular partition P with $|P|\leq k_\varepsilon$. Let F be the set of edges uv for which there exists $(X,Y)\in P$ with $u\in X$, $v\in Y$ such that (X,Y) is ε-irregular, or $d(X,Y)\leq 2\varepsilon$, or $|X|<\varepsilon^{-2}$. We claim

(11) $E\setminus F$ contains no triangles.

If not, there exist (not necessarily distinct) $X_1,X_2,X_3\in P$ such that (X_i,X_j) is ε-regular and $d(X_i,X_j)>2\varepsilon$ for all distinct $i,j=1,2,3$ and such that $|X_3|\geq\varepsilon^{-2}$. Let U be the set of vertices in X_1 with at most $\varepsilon|X_2|$ neighbours in X_2. Then $d(U,X_2)\leq\varepsilon<d(X_1,X_2)-\varepsilon$, so, since (X_1,X_2) is ε-regular, $|U|\leq\varepsilon X_1$. So less than half of the vertices in X_1 have at most $\varepsilon|X_2|$ neighbours in X_2. Similarly, less than half of the vertices in X_1 have at most $\varepsilon|X_3|$ neighbours in X_3. Hence there exists a vertex $u\in X_1$ with more than $\varepsilon|X_2|$ neighbours in X_2 and more than $\varepsilon|X_3|$ neighbours in X_3. Let U_i be the set of neighbours of u in X_i ($i=2,3$). As G is a Δ-graph, the edges spanned by $U_2\cup U_3$ form a matching, and so $e(U_2,U_3)\leq|U_2|$. So $|U_3|^{-1}\geq d(U_2,U_3)\geq d(X_2,X_3)-\varepsilon>\varepsilon$. Hence $|U_3|<\varepsilon^{-1}$. Therefore, $|X_3|<\varepsilon^{-1}|U_3|<\varepsilon^{-2}$, a contradiction. This proves (11).

Next we show:

(12) $|F|\leq 4\varepsilon n^2$.

The number of edges connecting any ε-irregular pair (X,Y) is at most εn^2, by (7). The number of edges spanned by those $(X,Y)\in P^2$ with $d(X,Y)\leq 2\varepsilon$ is at most $\sum_{X,Y\in P}2\varepsilon|X||Y|\leq 2\varepsilon n^2$. The number of edges intersecting those X_i with $|X_i|<\varepsilon^{-2}$ is at most $k_\varepsilon\varepsilon^{-2}n\leq\varepsilon n^2$. So we have (12).

By (11), each triangle of G contains an edge in F. Hence, by (12), G has at most $4\varepsilon n^2$ triangles, and hence, as G is a Δ-graph, at most $12\varepsilon n^2$ edges. This proves (10). ∎

Note that ε-balancedness of partition P of V is not used in this proof.

[1] $f(n)=o(g(n))$ means $\lim_{n\to\infty}f(n)/g(n)=0$.

4. Arithmetic progressions

An *arithmetic progression* of *length* k is a sequence of numbers $a_1, \ldots, a_k$ with $a_i - a_{i-1} = a_2 - a_1 \neq 0$ for $i = 2, \ldots, k$. For any k and n, let $\alpha_k(n)$ be the maximum size of a subset of $[n]$ containing no arithmetic progression of length k. (Here $[n] := \{1, \ldots, n\}$.)

We can now derive the theorem of Roth [3], which implies that any set X of natural numbers with $\limsup_{n\to\infty} |X \cap [n]|/n > 0$ contains an arithmetic progression of length 3.

Corollary 2a. $\alpha_3(n) = o(n)$.

Proof. We show that $\alpha_3(n) \leq \tau(9n)/3n$. Then Theorem 2 gives the corollary.

Choose $S \subseteq [n]$ with $|S| = \alpha_3(n)$ such that S contains no arithmetic progression of length 3. Let $V := [3n] \times [3]$ and for $i \in [n]$ and $s \in S$, let $T_{i,s}$ be the triangle spanned by $(i, 1)$, $(i + s, 2)$, $(i + 2s, 3)$. Let E be the set of edges spanned by these $T_{i,s}$. We show that $G = (V, E)$ is a Δ-graph.

Let T be any triangle in E. Let T be spanned by $(i, 1)$, $(j, 2)$ and $(k, 3)$. Then $j = i + s$, $k = j + t$, and $k = i + 2u$ for some $s, t, u \in S$. So $u = \frac{1}{2}(s + t)$. If $T \neq T_{i,s}$, then $t \neq s$, and hence s, u, t is an arithmetic progression of length 3, contradicting our assumption. So G is a Δ-graph.

Therefore, $3n\alpha_3(n) = 3n|S| = |E| \leq \tau(|V|) = \tau(9n)$. ∎

This was extended to $\alpha_k(n) = o(n)$ for any k by Szemerédi [4]. Recently, Green and Tao [2] proved that there exist arbitrarily long arithmetic progressions of primes.

References

[1] W.T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, *Geometric and Functional Analysis* 7 (1997) 322–337.

[2] B. Green, T. Tao, The primes contain arbitrarily long arithmetic progressions, *Annals of Mathematics* (2) 167 (2008) 481–547.

[3] K. Roth, Sur quelques ensembles d'entiers, *Comptes Rendus des Séances de l'Académie des Sciences Paris* 234 (1952) 388–390.

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

[5] E. Szemerédi, Regular partitions of graphs, in: *Problèmes combinatoires et théorie des graphes* (Proceedings Colloque International C.N.R.S., Paris-Orsay, 1976) [Colloques Internationaux du C.N.R.S. N^o 260], Éditions du C.N.R.S., Paris, 1978, pp. 399–401.