LECTURE 3. REGULARITY LEMMA

1. Statement and proof

Szemerédi's regularity lemma is one of the most powerful tools in extremal combinatorics. It was first used in proving the following celebrated result known as Szemerédi's theorem.

Theorem 1. *For all $k \geq 3$ and $\delta > 0$, there exists n_0 such that for all $N \geq n_0$, if $A \subseteq [N]$ and $|A| \geq \delta N$ then A contains an arithmetic progression of length k.*

The $k = 3$ case had been proven earlier by Roth and is known as Roth's theorem. Since then researchers realized that it is an extremely useful lemma. The statement requires a few definitions. For a graph G and disjoint subsets of vertices X, Y, define $e(X, Y)$ as the number of edges intersecting both X and Y. Define $d(X, Y) = \frac{e(X,Y)}{|X||Y|}$ as the *density* between the pair (X, Y).

Definition 2. *Let G be a graph. A disjoint pair (X, Y) of sets of vertices is ε-*regular *if for every $X' \subseteq X$ and $Y' \subseteq Y$ of sizes $|X'| \geq \varepsilon|X|$ and $|Y'| \geq \varepsilon|Y|$,*

$$\left|d(X, Y) - d(X', Y')\right| \leq \varepsilon.$$

ε-regularity captures the notion that the edges of a graph is well-distributed, or 'random-like', across the two parts. For example, a random bipartite graph will be ε-regular with high probability. Note that every pair satisfying $d(X, Y) \leq \varepsilon^3$ is trivially ε-regular since all subsets $X' \subseteq X$ and $Y' \subseteq Y$ with $|X'| \geq \varepsilon|X|$ and $|Y'| \geq \varepsilon|Y|$ satisfy $d(X', Y') \leq \varepsilon$. Hence ε-regularity is meaningful only when the density of the pair is sufficiently large.

We say that a vertex partition $V = V_0 \cup V_1 \cup \cdots \cup V_k$ is *equitable* if $|V_i| = |V_j|$ for all distinct indices $i, j \geq 1$ (this is slightly non-standard since we are allowing V_0 to have different size from others). We refer to the set V_0 as the *exceptional set*.

Definition 3. *Let G be a graph. A vertex partition $V_0 \cup V_1 \cup \cdots \cup V_k$ is ε-*regular *if (i) it is equitable, (ii) $|V_0| \leq \varepsilon n$, and (iii) all but at most εk^2 pairs (V_i, V_j) satisfying $1 \leq i < j \leq k$ are ε-regular.*

A partition being ε-regular is a very strong condition. In some sense it asserts that a partition can be 'approximated' by a 'random-like' blow-up of a graph (both of these can be formalized but we will not dive into this discussion here). The fascinating amazing beautiful Szemerédi's regularity lemma asserts that all (large enough) graphs have an ε-regular partition.

Theorem 4. *(Szemerédi's regularity lemma) For every ε and t, there exists N and T such that for every $n \geq N$, every n-vertex graph G admits an ε-regular partition $V_0 \cup V_1 \cup \cdots \cup V_k$ satisfying $t \leq k \leq T$.*

Note that the statement is trivially true if G has at most $\varepsilon^4 n^2$ edges. To see this, take an arbitrary partition $V_1 \cup \cdots \cup V_k$ satisfying $|V_1| = \cdots = |V_k|$ (suppose that n is divisible by k). As we have seen above, if $d(V_i, V_j) \leq \varepsilon^3$, then (V_i, V_j) is ε-regular. Hence the only pairs that are possibly not ε-regular are those satisfying $d(V_i, V_j) > \varepsilon^3$. Note that for such pairs,

$$e(V_i, V_j) > \varepsilon^3 |V_i||V_j| = \frac{\varepsilon^3}{k^2} n^2.$$

Since there are at most $\varepsilon^4 n^2$ edges in the graph, there are less than εk^2 such pairs. Therefore all equitable partitions are ε-regular! In this context, people often refer to the theorem above as 'dense regularity lemma'. There is a 'sparse regularity lemma' that applies to sparse graphs, but the implications is significantly weaker than in the dense case. I will discuss this in more detail later in the course if time allows.

The proof of the regularity lemma is based on a technique known as 'density increment argument'. It proceeds by showing that as long as the existing partition is not ε-regular, it can be refined so that some quantity increases by a fixed amount. We can then show that the process ends in a finite number of steps since this quantity will be bounded from above.

Definition 5. *Let G be a n-vertex graph. The* mean square density *of a partition $\Pi = \{V_i\}_{i=1}^k$ is*

$$d_2(\Pi) = \sum_{1 \leq i < j \leq k} \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2.$$

We say that a partition Π' refines another partition Π if each part of Π is a union of some parts of Π'. The following lemma collects several useful facts about mean square density of partitions.

Lemma 6. *Let G be a graph on n vertices.*

(i) *For all vertex partitions Π, $0 \leq d_2(\Pi) \leq \frac{1}{2}$.*
(ii) *If Π' refines Π, then $d_2(\Pi') \geq d_2(\Pi)$.*

Proof. (i) Note that since $\sum_{1 \leq i < j \leq k} \frac{|V_i||V_j|}{n^2} \leq \frac{1}{2}$ and $0 \leq d(V_i, V_j) \leq 1$, the mean square density is always between 0 and 1.

(ii) If $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$, then

$$e(X, Y) = \sum_{i,j=1}^{2} e(X_i, Y_j).$$

Therefore,

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

Therefore by Cauchy-Schwartz inequality,

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

In particular, when $Y_2 = \emptyset$, This implies that if Π' is obtained from Π by splitting one part into two, then $d_2(\Pi') \ge d_2(\Pi)$. Therefore the statement follows by repeatedly applying this inequality.

The next lemma describes the density increment part of the proof.

Lemma 7. *Suppose that $X = X_1 \cup X_2$ and $Y = Y_1 \cup Y_2$ where $|X_1| \ge \varepsilon|X|$, $|Y_1| \ge \varepsilon|Y|$ and $|d(X_1,Y_1) - d(X,Y)| \ge \varepsilon$. Then*

$$\sum_{i,j=1}^{2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i,Y_j)^2 \ge d(X,Y)^2 + \frac{1}{2}\varepsilon^5.$$

Proof. Note that by (1), we have

$$\begin{aligned}
&\sum_{i,j=1}^{2} \frac{|X_i||Y_j|}{|X||Y|} \left(d(X_i,Y_j) - d(X,Y)\right)^2 \\
&= \sum_{i,j=1}^{2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i,Y_j)^2 - 2d(X,Y) \sum_{i,j=1}^{2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i,Y_j) + d(X,Y)^2 \\
&= \left(\sum_{i,j=1}^{2} \frac{|X_i||Y_j|}{|X||Y|} d(X_i,Y_j)^2\right) - d(X,Y)^2.
\end{aligned}$$

Note that the initial term of the above is at least ε^4 for $i = j = 1$ and at least 0 for all other terms. Therefore the lemma follows. $\square$

For a given partition $\Pi = V_0 \cup V_1 \cup \cdots \cup V_k$, we view it as a partition consisting of $|V_0|+k$ parts where each vertex in V_0 forms its own part. Define $|\Pi| = k$.

Lemma 8. *If $\Pi = V_0 \cup V_1 \cup \cdots \cup V_k$ is not an ε-regular partition, then there exists a refinement Π' of Π satisfying $d_2(\Pi') \ge d_2(\Pi) + \varepsilon^5$ and $|\Pi'| \le |\Pi| 2^{|\Pi|}$.*

Proof. Since $V_0 \cup V_1 \cup \cdots \cup V_k$ is not an ε-regular partition, there exists at least εk^2 pairs (V_i, V_j) for which there exist subsets $V_i^{(1)} \subseteq V_i$ and $V_j^{(1)} \subseteq V_j$ satisfying $|V_i^{(1)}| \ge \varepsilon|V_i|$, $|V_j^{(1)}| \ge \varepsilon|V_j|$, and

$$\left| d(V_i^{(1)}, V_j^{(1)}) - d(V_i, V_j) \right| > \varepsilon.$$

Define $V_i^{(2)} = V_i \setminus V_i^{(1)}$ and $V_j^{(2)} = V_j \setminus V_j^{(1)}$. Then by Lemma 7,

$$\sum_{k,\ell=1}^{2} \frac{|V_i^{(k)}||V_j^{(\ell)}|}{n^2} d(V_i^{(k)}, V_j^{(\ell)})^2 \geq \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2 + \frac{|V_i||V_j|}{n^2}\varepsilon^4. \tag{2}$$

For each non-ε-regular pair (V_i, V_j), partition the sets V_i and V_j as above, and let Π' be the common refinement of Π and all these partitions. Suppose that for each i, we have $V_i = V_{i,1} \cup V_{i,2} \cup \cdots \cup V_{i,k_i}$ where the sets on the right-hand-side consist of parts of Π'. Note that $k_i \leq 2^k$ for all i and thus $|\Pi'| \leq |\Pi| 2^{|\Pi|}$. The equation (2) applies to all pairs (V_i, V_j) that were not ε-regular. Lemma 6 (ii) therefore implies that for all such pairs (V_i, V_j),

$$\sum_{a=1}^{k_i}\sum_{b=1}^{k_j} \frac{|V_{i,a}||V_{j,b}|}{n^2} d(V_{i,a}, V_{j,b})^2 \geq \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2 + \frac{|V_i||V_j|}{n^2}\varepsilon^4.$$

Since there are at least εk^2 non-ε-regular pairs, we see that

$$d_2(\Pi') \geq d_2(\Pi) + \varepsilon k^2 \cdot \frac{((1-\varepsilon)n/k)^2}{n^2}\varepsilon^4 \geq d_2(\Pi) + \frac{1}{2}\varepsilon^5. \qquad \square$$

We now collect the facts to prove the regularity lemma.

Proof of the regularity lemma. Start with an arbitrary equitable partition $\Pi_0 = V_0 \cup V_1 \cup \cdots \cup V_t$ where $|V_0| \leq t-1$. If Π_0 is not an ε-regular partition, then there exists a refinement Π_1' of Π_0 for which $d_2(\Pi_1') \geq d_1(\Pi_0) + \frac{1}{2}\varepsilon^5$ and $A = |\Pi_1'| \leq |\Pi_0| 2^{|\Pi_0|}$. We can make Π_1' into an equitable partition by further partitioning each part into parts of sizes exactly $\frac{1}{2}\varepsilon^6 \frac{n}{A}$ and at most one part of size less than that. We move all the small parts into the exceptional set. This way we obtain an equitable partition Π_1 where the exceptional set increases by at most $\frac{1}{2}\varepsilon^6 n$. By Lemma 6 (ii) we see that $d_2(\Pi_1) \geq d_2(\Pi_1') \geq d_2(\Pi_0) + \frac{1}{2}\varepsilon^5$.

Repeat the process. Note that the k-th partition Π_k satisfies $d_2(\Pi_k) \geq \frac{1}{2}\varepsilon^5(k-1)$. On the other hand by Lemma 6 (i) we have $d_2(\Pi_k) \leq \frac{1}{2}$. Hence the process ends in at most ε^{-5} steps. Therefore we obtain an ε-regular partition. Moreover the exceptional set has size at most $\frac{1}{2}\varepsilon^6 n \cdot \varepsilon^{-5} < \varepsilon n$. $\square$

Note that at each step the number of parts in the partition increases exponentially. Therefore the bound on T, the number of parts, that we obtain through this proof is quite poor (a tower type bound where the height of the tower is about ε^{-5}). Surprisingly, Gowers proved that such behavior is unavoidable, i.e., that there are graphs whose you need a 'tower-type' number of parts in a smallest ε-regular partition.

The regularity lemma produces exceptional sets and exceptional pairs. One can easily remove the exceptional sets at the cost of slightly increasing the value of ε. On the other hand, the existence of exceptional pairs is unavoidable, i.e., there are graphs for which every ε-regular partition must have exceptional pairs.

2. Counting lemma

Regular partition is useful in embedding graphs.

Definition 9. *Given a partition $\Pi = V_1 \cup V_2 \cup \cdots \cup V_t$ we define its (ε, δ)-*reduced graph $R(\Pi)$ *as a graph on vertex set $[t]$ where a pair $\{i, j\}$ is adjacent if (V_i, V_j) is an ε-regular pair of density at least δ.*

The following theorem is known as the counting lemma.

Theorem 10. *For every graph H and positive real number δ, there exists c and ε_0 such that the following holds for every $\varepsilon \le \varepsilon_0$. Let G be a graph and Π be a vertex partition of G. If there exists a copy of H in $R(\Pi)$, then there exist a copy of H in G. Moreover if the copy of H is over the vertices $1, 2, \cdots, r$, then there are at least $c \prod_{i=1}^{r} |V_i|$ copies of H in G.*

To prove the theorem we need the following basic lemma.

Lemma 11. *Let (X, Y) be an ε-regular pair of density d. Then for every $Y' \subseteq Y$ of size $|Y'| \ge \varepsilon |Y|$, there exists less than $\varepsilon |X|$ vertices in X that have less than $(d - \varepsilon)|Y|$ neighbors in Y.*

Proof. Let X' be the set of vertices in X having less than $(d-\varepsilon)|Y|$ neighbors in Y. Then

$$e(X', Y) < |X'| \cdot (d - \varepsilon)|Y|$$

and thus $d(X', Y) < d - \varepsilon$. Since the pair (X, Y) is ε-regular, this can happen only if $|X'| < \varepsilon |X|$. □

We now prove Theorem 10.

Proof of Theorem 10. For simplicity, we assume that $H = K_r$ for some $r \ge 2$. Without loss of generality, we may assume that $V_1, V_2, \cdots, V_r$ are disjoint vertex sets such that for all distinct pairs i, j, (V_i, V_j) is ε-regular of density at least δ.

Denote the vertices of K_r as $v_1, v_2, \cdots, v_r$. For each $i \in [r]$, we will embed v_i to V_i one vertex at a time. After embedding the t-th vertex, the following condition will be satisfied for each $j > t$:

$$\left| V_j \cap \bigcap_{i=1}^{t} N(v_i) \right| \ge (\delta - \varepsilon)^t |V_j|.$$

Note that the condition is trivially satisfied for $t = 0$.

Suppose that we embedded the vertices $v_1, v_2, \cdots, v_t$ for some $t \ge 0$. For each $j > t$, define $W_j = V_j \cap \bigcap_{i=1}^{t} N(v_i)$. As mentioned above, we have $|W_j| \ge (d - \varepsilon)^t |V_j|$ for all $j > t$. Since $\varepsilon < (d - \varepsilon)^t$, for each $j > t + 1$, there are at most $\varepsilon |V_{t+1}|$ vertices in V_{t+1} having less than $(d - \varepsilon)|W_j|$ neighbors in W_j.

Thus we can find a set $W'_{t+1} \subseteq W_{t+1}$ of size at least

$$|W'_{t+1}| \ge |W_{t+1}| - r \cdot (d - \varepsilon)^t |V_{t+1}| \ge \frac{1}{2}(d - \varepsilon)^t |V_{t+1}|$$

such that each vertex in W'_{t+1} has at least $(d-\varepsilon)|W_j|$ neighbors in W_j for all $j > t+1$. This means that we may embed v_{t+1} to an arbitrary vertex in W'_{t+1}. Therefore the number of copies of K_r is at least

$$\prod_{i=1}^{r} \frac{1}{2}(d-\varepsilon)^i |V_i| = \frac{1}{2^r}(d-\varepsilon)^{\binom{r}{2}} \prod_{i=1}^{r} |V_i|. \qquad \square$$

The proof above can easily be modified to show that it suffices to have a homomorphisms from H to $R(\Pi)$ as long as G is sufficiently large.

Erdős-Stone theorem follows as an immediate corollary of the regularity lemma and the counting lemma.

Corollary 12. *For all H and δ, there exists n_0 such that if G is a graph on $n \geq n_0$ vertices with at least $\left(1 - \frac{1}{\chi(H)-1} + \delta\right)\frac{n^2}{2}$ edges, then G contains a copy of H.*

Proof. Let ε be small enough depending on δ. Let $t = \frac{1}{\varepsilon}$ and apply the regularity lemma to find an ε-regular partition $\Pi = V_0 \cup V_1 \cup \cdots \cup V_k$ where $t \leq k \leq T$. Remove the following edges:

(i) edges incident to V_0 (at most εn^2 edges),
(ii) edges inside V_i for all i (at most $k \cdot (\frac{n}{k})^2 \leq \varepsilon n^2$ edges),
(iii) edges bewteen irregular pairs (at most $\varepsilon k^2 \cdot (\frac{n}{k})^2 \leq \varepsilon n^2$ edges).
(iv) edges between regular pairs of density at most $\frac{\delta}{2}$ (at most $\frac{\delta}{2}(\frac{n}{k})^2\binom{k}{2} < \frac{\delta}{4}\frac{n^2}{2}$ edges).

The number of edges in the remaining graph is at least $\left(1 - \frac{1}{\chi(H)-1} + \frac{\delta}{2}\right)\frac{n^2}{2}$. Therefore the (ε, δ)-reduced graph $R(\Pi)$ contains at least $\left(1 - \frac{1}{\chi(H)-1} + \frac{\delta}{2}\right)\frac{k^2}{2}$ edges. By Turán's theroem, we can find a copy of $K_{\chi(H)}$ in the reduced graph. Then by the counting lemma, we can find a copy of H in G. $\square$

This provides a 'conceptual proof' of the Erdős-Stone theorem, although the bound we obtain is rather poor.

3. Removal Lemma

A *graph property* is a family of graphs. A graph property $\mathcal{P}$ is *monotone decreasing* if $G \in \mathcal{P}$ and $G' \subseteq G$ implies $G' \in \mathcal{P}$. For a monotone decreasing graph property $\mathcal{P}$, define the *distance* from G to $\mathcal{P}$ as $\min_{H \in \mathcal{P}, H \subseteq G} |E(G) - E(H)|$. In other words, it is the minimum number k such that one can obtain a graph in $\mathcal{P}$ by removing at most k edges from G.

Example 13. *Here are some examples of graph properties.*

(i) *Let $\mathcal{P}$ be the family of disconnected graphs. Then $\mathcal{P}$ is monotone decreasing. The distance from G to $\mathcal{P}$ is the MINCUT of G.*
(ii) *For a graph H, let $\mathcal{P}_H$ be the family of H-free graphs. Then $\mathcal{P}_H$ is monotone decreasing.*

(iii) *For a graph H, let $\mathcal{Q}_H$ be the family of induced-H-free graphs. Then $\mathcal{Q}_H$ is not monotone decreasing (nor increasing).*

The removal lemma asserts that for every graph H, every sufficiently large graph that is far from being H-free contains many copies of H. For a graph H, we say that G is *k-far from being H-free* if the distance of G from $\mathcal{P}_H$ (defined above) is at least k.

Theorem 14. *(Removal lemma) For all ε there exists δ such that the following holds for sufficiently large n. If G is a n-vertex graph that is εn^2 far from being H-free, then G contains at least $\delta n^{|V(H)|}$ copies of H.*

Proof. Let ε' be small enough depending on ε and let $t = \frac{4}{\varepsilon}$. Apply the regularity lemma to find an ε-regular partition $\Pi = V_0 \cup V_1 \cup \cdots \cup V_k$ where $t \leq k \leq T$. Remove the following edges:

(i) edges incident to V_0 (at most $\varepsilon' n^2$ edges),
(ii) edges inside V_i for all i (at most $k \cdot (\frac{n}{k})^2 \leq \frac{\varepsilon}{4} n^2$ edges),
(iii) edges between irregular pairs (at most $\frac{\varepsilon}{4} k^2 \cdot (\frac{n}{k})^2 \leq \frac{\varepsilon}{4} n^2$ edges).
(iv) edges between regular pairs of density at most $\frac{\varepsilon}{2}$ (at most $\frac{\varepsilon}{2} (\frac{n}{k})^2 \binom{k}{2} < \frac{\varepsilon}{4} n^2$ edges).

The number of edges removed is less than εn^2 and therefore there still exists a copy of H in G. Since all remaining edges are between ε'-regular pairs of density at least $\frac{\varepsilon}{2}$, it implies that there exists a homomorphisms from H to the $(\varepsilon', \frac{\varepsilon}{2})$-reduced graph $R(\Pi)$. When ε' is sufficiently small, the counting lemma shows that there are at least $\delta n^{|V(H)|}$ copies of H (where δ is a constant depending on ε'). □

Note that there is a tower-type dependency between δ and ε in the removal lemma where the tower of the height is roughly $\frac{1}{\varepsilon^5}$.

Roth's theorem follows from the removal lemma as a simple corollary.

Theorem 15. *(Roth) For all δ, there exists n_0 such that the following holds for all $n \geq n_0$. If $A \subseteq [n]$ and $|A| \geq \delta n$, then A contains a 3-term arithmetic progression.*

Proof. Let $A \subseteq [n]$ be a set given as above.

Consider a 3-partite graph G with $3n$ vertices in each part $V_1 \cup V_2 \cup V_3$ each labelled by $[3n]$. A pair $(v, w) \in V_1 \times V_2$ or $\in V_2 \times V_3$ is adjacent if and only if $w - v \in A$. A pair $(v, w) \in V_1 \times V_3$ is adjacent if and only if $w - v \in 2A$ (i.e. $w - v = 2a$ for some $a \in A$).

We call a triangle in G *trivial* if it consists of $v_i \in V_i$ for $i = 1, 2, 3$, where $v_2 - v_1 = v_3 - v_2 = \frac{1}{2}(v_3 - v_1) = a \in A$. Note that for each $a \in A$, there are at least n trivial triangles 'associated' with a. Since the trivial triangles are edge-disjoint, we see that G is δn^2-far from being triangle-free.

Hence by the triangle removal lemma, there exists at least εn^3 triangles in G. Since there are at most $3n^2$ trivial triangles, if n is sufficiently large, then we can find a non-trivial triangle whose vertices are $v_i \in V_i$ for $i = 1, 2, 3$.

Denote $a = v_2 - v_1$, $b = v_3 - v_2$ and $c = \frac{1}{2}(v_3 - v_1)$. Then $a, b, c \in A$ and $c = \frac{1}{2}(a + b)$. Moreover not all values a, b, c are equal. This implies that a, b, c forms a 3-term arithmetic progression. $\square$

One can prove the whole Szemerédi's theorem using a similar approach, but instead of graph regularity lemma, it requires the hypergraph regularity lemma. The statement of the hypergraph regularity lemma is technical and will not be covered in this course.

A careful analysis of the proof shows that for large enough n, if $A \subseteq [n]$ and $|A| \geq \frac{n}{\log_* n}$, then A contains a 3-term arithmetic progression. Where $\log_* n$ is the number of times one must apply the logarithmic function to n in order to make the outcome less than some fixed number, say e.

The current best known result in this direction says that $|A| \geq n\frac{(\log\log n)^4}{\log n}$ suffices (see Bloom [1], Sanders [4]). The proof is based on Fourier analytical techniques. On the other hand, the following construction discovered by Behrend gives the best known lower bound.

Theorem 16. *There exists a constant c such that for all large enough n, there exists a set $A \subseteq [n]$ of size at least $|A| \geq \frac{n}{e^{c\sqrt{\log n}}}$ that contains no 3-term arithmetic progression.*

Proof. For integers m and d, let $n > 2^d m^d$. Consider $X = [m]^d$ and note that for each $(x_1, \cdots, x_d) \in X$, we have $\sum_{i=1}^d x_i^2 \leq dm^2$. Therefore X can be covered by at most dm^2 spheres, i.e. surfaces of the form $\sum_{i=1}^d x_i^2 = C$. Take the sphere S that contains the maximum number of points and define $Y = X \cap S$. By the pigeonhole principle, we see that $|Y| \geq \frac{m^d}{dm^2}$. Consider the map $f : Y \to [n]$ defined as

$$f(y_1, \cdots, y_d) = \sum_{i=1}^{d} (y_i - 1)(2m)^{i-1}.$$

Since $y_i \in [m]$ for each i, we see that f is injective and $f(\vec{y}) + f(\vec{y'}) = 2f(\vec{y''})$ if and only if $\vec{y} + \vec{y'} = 2\vec{y''}$. However since S is a convex set, there are no three points $\vec{y}, \vec{y'}, \vec{y''} \in S$ satisfying $\vec{y} + \vec{y'} = 2\vec{y''}$. Thus $f(Y)$ contains no 3-term arithmetic progression. Moreover

$$|f(Y)| = |Y| \geq \frac{m^d}{dm^2} \geq \frac{n}{d2^d m^2}.$$

We may take $m = e^{\sqrt{\log n} - 1}$ and $d = \sqrt{\log n}$ and this gives $|f(Y)| \geq \frac{n}{e^{c\sqrt{\log n}}}$ for some constant c. $\square$

Establishing the 'correct' density is an interesting open problem. For general k-term progressions, Erdős offered a \$3,000 prize for a proof of the following conjecture.

Conjecture 17. *Let A be a set of positive integers. If $\sum_{n \in A} \frac{1}{n} = \infty$, then A contains arithmetic progressions of any given length.*

If true, then the conjecture implies the celebrated Fields medal winning result of Green and Tao as a corollary.

Theorem 18. *(Green-Tao [3]) Primes contain arithmetic progressions of any given length.*

As mentioned above, even though the original proof of Szemerédi's theorem was based on a graph theoretic approach, the bounds obtained through the method is comparably worse than the bounds obtained through other methods. The following question drawn a considerable amount of interest in this context.

Question 19. *Improve the dependency between δ and ε in the removal lemma to a tower of bounded height.*

The best known result in this direction was proved by Fox [2], who improved the height of the tower from a polynomial-type dependency on $\frac{1}{\varepsilon}$ to a logarithmic-type dependency on $\frac{1}{\varepsilon}$.

4. Induced matching lemma and $(6,3)$-theorem

Ruzsa and Szemerédi proved the following theorem known as the induced matching lemma.

Theorem 20. *Let G be a graph on n vertices that consists of the union of n induced matchings. Then $e(G) = o(n^2)$.*

Proof. Let δ be a fixed real and n be large enough depending on δ. Let G be an n-vertex graph with at least δn^2 edges consisting of the union of n matchings. We will show that not all matchings can be induced.

Take an ε-regular partition $V_0 \cup V_1 \cup \cdots \cup V_k$ of G where $\frac{1}{\varepsilon} \le k \le T$. Remove the following edges:

(i) edges incident to V_0 (at most εn^2 edges),
(ii) edges inside V_i for all i (at most $k \cdot (\frac{n}{k})^2 \le \varepsilon n^2$ edges),
(iii) edges between irregular pairs (at most $\varepsilon k^2 \cdot (\frac{n}{k})^2 \le \varepsilon n^2$ edges).
(iv) edges between regular pairs of density at most ε (at most $\varepsilon(\frac{n}{k})^2\binom{k}{2} < \varepsilon n^2$ edges).
(v) for each matching M, the edges that are incident to V_i satisfying $|V_i \cap M| < \varepsilon|V_i|$ (at most $n \cdot \varepsilon n = \varepsilon n^2$ edges).

The number of edges removed is less than $5\varepsilon n^2$. Hence one of the initial matchings M have at least $(\delta - 5\varepsilon)n$ remaining edges. Take an edge $e \in M$ and suppose that it intersects V_i and V_j. By (v), we know that $|V_i \cap V(M)| \ge \varepsilon|V_i|$ and $|V_j \cap V(M)| \ge \varepsilon|V_j|$. Therefore by ε-regularity, of the pair (V_i, V_j), there exists at least εn^2 edges between $V_i \cap V(M)$ and $V_j \cap V(M)$. If n is sufficiently large, then we can find an edge between $V_i \cap V(M)$ and $V_j \cap V(M)$ that is not in M. Therefore M is not an induced matching. □

Induced matching lemma provides an alternative proof of Roth's theorem. Indeed given a set $A \subseteq [n]$ of size $|A| \ge \delta n$, consider a bipartite graph G

whose two parts X and Y are labelled by $[2n]$ and $[3n]$ respectively, and edge set consists of $M_x = \{(x+a, x+2a) : a \in A\}$ for $x \in [n]$. Then G is a graph on $5n$ vertices with at least δn^2 edges. Therefore not all matchings M_x are induced. This can easily be seen to imply that A contains a 3-term arithmetic progression.

The induced matching lemma was originally developed in order to prove the following $(6,3)$-theorem.

Theorem 21. *If H is a 3-uniform hypergraph with no 6 vertices spanning at least 3 edges, then $e(H) = o(n^2)$.*

Proof. It suffices to prove it for graphs in which all vertices have degree at least 2 since otherwise we can consider the graph obtained by removing all vertices of degree at most 1.

For each vertex $v \in V(H)$, define M_v as the set of pairs $\{x, y\}$ which together with v form a hyperedge of H. If M_v is not a matching, then there exist three vertices x, y, z for which $\{v, x, y\}$ and $\{v, x, z\}$ both form an edge in H. Since y has degree at least 2, by taking an arbitrary edge incident to y that is not $\{v, x, y\}$, we can find 6 vertices spanning at least 3 edges. Thus we may assume that M_v is a matching for each v.

Consider the graph G defined on $V(H)$ whose edge set consists of the union of M_v for all $v \in V(H)$. If there exists $v \in V(H)$ such that M_v does not form an induced matching in G, then we can find 6 vertices spanning at least 3 edges in H. Therefore by the induced matching lemma, we see that $e(G) = o(n^2)$. Since $e(G) = 3e(H)$, we see that $e(H) = o(n^2)$. □

By using Behrend's construction, we can find a 3-uniform hypergraph with at least $\frac{n^2}{e^{c\sqrt{\log n}}}$ edges with no 6 vertices spanning at least 3 edges. Thus together with Theorem 21, it shows that the degenerate Turán problems for hypergraphs can exhibit the following interesting behavior. Let $\mathcal{L}$ be the family of hypergraph on 6 vertices with 3 edges. We have shown that

$$\frac{n^2}{e^{c\sqrt{\log n}}} \le ex(n, \mathcal{L}) \le \frac{n^2}{\log_* n}.$$

Therefore for hypergraphs, we do not necessarily have $c_1 n^\gamma \le ex(n, \mathcal{L}) \le c_2 n^\gamma$.

The following generalization of the $(6,3)$-theorem is still wide open.

Conjecture 22. *(Brown-Erdős-Sós) Let $k \ge 3$ be an integer. If H is a 3-uniform hypergraph with no $k+3$ vertices spanning at least k edges, then $e(H) = o(n^2)$.*

5. Lucrative Erdős prize problems

Conjecture 23. *($10,000) For every real number C, the difference between the n-th prime and the $(n+1)$-th prime exceeds*

$$\frac{C \log n (\log\log n)(\log\log\log\log n)}{(\log\log\log n)^2}$$

infinitely often.

The conjecture above has been recently solved by Ford, Green, Konyagin, Tao, and independently by Maynard. Then \$3,000 problem mentioned above is still open. Following is a \$1,000 problem solved by Bob Hough.

Conjecture 24. *(\$1,000) A set of pairs of integers* $\{(a_1, b_1), (a_2, b_2), \cdots\}$ *is unavoidable if for every* $n \geq 1$*, there exists a pair in the set such that* $n = a_i (mod\, b_i)$ *Does there exist an integer* N *such that for every unavoidable set of congruences, either (i)* $b_i = b_j$ *for two distinct indices or (ii)* $b_i \leq N$ *for some index.*

As mentioned before, the hypergraph Turán problem is a \$1,000 problem. The sunflower problem is another famous open problem in combinatorics. I will introduce this conjecture later in the course.

References

[1] T. Bloom A quantitative improvement for Roth's theorem on arithmetic progressions.
[2] J. Fox, A new proof of the graph removal lemma, *Ann. of Math.* **174** (2011), 561–579.
[3] B. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions *Ann. of Math.* **167** (2008), 481–547.
[4] T. Sanders, On Roth's theorem on progressions, *Ann. of Math.* **174** (2011), 619–636.