1 Lecture 7: Hypergraph Container Lemma

Before stating the Hypergraph Container Lemma, we need to define a few new terms.

We say that a family of sets $\mathcal{F} \subseteq 2^{V(\mathcal{H})}$ is an *upset* (or an *increasing family*) if for every $A, B \subset V(\mathcal{H})$ with $A \in \mathcal{F}$ and $A \subseteq B$, we have $B \in \mathcal{F}$ as well.

Given a k-uniform hypergraph $\mathcal{H}$, an upset $\mathcal{F} \subseteq 2^{V(\mathcal{H})}$, and $\varepsilon \in (0,1]$, we say $\mathcal{H}$ is $(\mathcal{F},\varepsilon)$-*dense* if $e(\mathcal{H}[A]) \geqslant \varepsilon \cdot e(\mathcal{H})$ for every $A \in \mathcal{F}$. Here $e(\mathcal{H}[A])$ is the number of edges contained in A. Let $\mathcal{F}_\varepsilon := \{A \in \mathcal{F} : e(\mathcal{H}[A]) \geqslant \varepsilon \cdot e(\mathcal{H})\}$. Note that $\mathcal{F}_\varepsilon$ is an upset and $\mathcal{H}$ is $(\mathcal{F}_\varepsilon, \varepsilon)$-dense. The increasing subfamilies of $\mathcal{F}_\varepsilon$ are exactly the families of $\mathcal{F}$ for which $\mathcal{H}$ is $(\mathcal{F},\varepsilon)$-dense.

For each $T \subseteq V(\mathcal{H})$, let $\deg_{\mathcal{H}}(T) := |\{F \in E(\mathcal{H}) : T \subseteq F\}|$, the number of edges of $\mathcal{H}$, counted with multiplicities, which contain T. Let

$$\Delta_\ell(\mathcal{H}) := \max\{\deg_{\mathcal{H}}(T) : T \subseteq V(\mathcal{H}), |T| = \ell\}.$$

Note that $\Delta_1(\mathcal{H}) = \Delta(\mathcal{H})$.

Recall that we use $\mathcal{I}(\mathcal{H})$ to denote the family of all independent sets in $\mathcal{H}$ and $\overline{\mathcal{F}}$ is the family of sets not in $\mathcal{F}$.

Theorem 1.1 (**Hypergraph Container Lemma** [2])**.** *For every $k \in \mathbb{N}$ and $c, \varepsilon > 0$, there exists a positive constant C for which the following holds. Let $\mathcal{H}$ be a k-uniform hypergraph and $\mathcal{F} \subseteq 2^{V(\mathcal{H})}$ be an upset such that $|A| \geqslant \varepsilon \cdot v(\mathcal{H})$ for all $A \in \mathcal{F}$. Suppose $\mathcal{H}$ is $(\mathcal{F},\varepsilon)$-dense and $q \in (0,1)$ such that for every $\ell \in [k]$,*

$$\Delta_\ell(\mathcal{H}) \leqslant c \cdot q^{\ell-1} \frac{e(\mathcal{H})}{v(\mathcal{H})}.$$

Then there exists a family $\mathcal{S} \subseteq \binom{V(\mathcal{H})}{\leqslant Cq \cdot v(\mathcal{H})}$ and functions $f : \mathcal{S} \to \overline{\mathcal{F}}$ and $g : \mathcal{I}(\mathcal{H}) \to \mathcal{S}$ such that for every $I \in \mathcal{I}(\mathcal{H})$,

$$g(I) \subseteq I \text{ and } I \subseteq f(g(I)) \cup g(I).$$

The overall idea is that if $\mathcal{H}$ satisfies certain natural conditions, then we can label each independent set I with a small subset $g(I)$ in such a way that all sets labeled with the same $S \in \mathcal{S}$ are essentially contained in a single set $f(S)$ which contains very few edges of $\mathcal{H}$.

1.1 Example: k-AP Free Sets

Let $\mathcal{H}$ be the k-uniform hypergraph of k-APs in $[n]$ which has vertex set $V(\mathcal{H}) = [n]$ and edges $E(\mathcal{H}_k) = \{k\text{-APs in } [n]\}$.

We can make a few observations about this hypergraph:

- $v(\mathcal{H}) = n$,
- $e(\mathcal{H}) = \sum_{i=1}^{n} \left\lfloor \frac{n-i+1}{k-1} \right\rfloor \approx \frac{n^2}{2(k-1)}$,
- $\Delta_1(\mathcal{H}) \leqslant k \cdot \frac{n}{k-1} \leqslant 2n \leqslant c \cdot q^{1-1} \frac{e(\mathcal{H})}{v(\mathcal{H})}$,
- $\Delta_\ell(\mathcal{H}) \leqslant \Delta_2(\mathcal{H}) \leqslant \binom{k}{2} \leqslant 2n^{1/(k-1)} \leqslant c \cdot q^{\ell-1} \frac{e(\mathcal{H})}{v(\mathcal{H})}$ for every $\ell \in \{2, \dots, k-1\}$,
- $\Delta_k(\mathcal{H}) = 1 \leqslant c \cdot q^{k-1} \frac{e(\mathcal{H})}{v(\mathcal{H})}$.

Choose $c = 2k^2$ and $q = n^{-1/(k-1)}$.

Since independent sets in $\mathcal{H}$ are k-AP-free sets, the Hypegraph Container Lemma 1.1 says that for each independent set I, there exist S and $f(S)$ such that $|S| \leqslant Cqn = Cn^{1-1/(k-1)}$, $f(S)$ contains at most εn^2 k-APs, and $I \subseteq S \cup f(S)$.

We will conclude that $|f(S)| \leqslant \delta n$ (small). This requires a supersaturation-type result which is a robust version of Szemerédi's Theorem [7].

Theorem 1.2 (Varnavides 1959 [8]). *For every positive integer δ and $k \in [n]$, there exists $\varepsilon > 0$ such that every $A \subseteq [n]$ with at least δn elements contains at least εn^2 k-APs.*

Using the Container Method to count the number of independent sets in $\mathcal{H}$ gives us that the number of k-AP-free sets in $[n]$ is bounded by the number of certificates times the number of ways to choose $I \subseteq f(S)$:

$$\#\{k\text{-AP-free sets}\} \leqslant \binom{n}{Cn^{1-1/(k-1)}} \cdot 2^{\delta n} = 2^{2\delta n}$$

for δ arbitrarily small when n is sufficiently large. This is not too useful, so instead we will use the Container Method to prove a random version of Szemerédi's Theorem.

We say that a set $D \subseteq [n]$ is (δ, k)-*Szemerédi* if every subset $S \subseteq D$ with $|S| \geqslant \delta |D|$ contains a k-AP. We can use this notation to restate several earlier results:

Theorem 1.3 (Szemerédi 1975 [7]). *For every $\delta > 0, k$, for n sufficiently large, the set $[n]$ is (δ, k)-Szemerédi.*

Theorem 1.4 (Green–Tao 2004 [4])**.** *For every $\delta > 0, k$, for n sufficiently large, the following holds. Let $p_1, \ldots, p_n$ be the first n primes. Then $\{p_1, \ldots, p_n\}$ is (δ, k)-Szemerédi.*

It is a natural question to ask which sets are (δ, k)-Szemerédi, but to answer such a general question is likely hopeless. Instead, we can look at what happens with random subsets of the integers.

For $p \in [0,1]$, let $[n]_p$ denote the *p-random subset of* $[n]$ formed by adding each $i \in [n]$ to $[n]_p$ independently with probability p.

Question 1.5. *For which p is $[n]_p$ (δ, k)-Szemerédi with high probability?*

Observation 1.6. *For $p \ll n^{-1/(k-1)}$, $[n]_p$ is not (δ, k)-Szemerédi with high probability. Even $p \leqslant 0.1n^{-1/(k-1)}$ is sufficient.*

Proof. We use the Alteration Method. Note that $\mathbb{E}[n]_p = pn$ and

$$\mathbb{E}(\#k\text{-APs in } [n]_p) = (\#k\text{-APs}) \cdot p^k \approx \frac{n^2}{2k} \cdot p^k.$$

Assume $pn \gg \frac{n^2}{2k} \cdot p^k$ (i.e. $p \ll n^{-1/(k-1)}$). Then we can find (in expectation) $pn - \frac{n^2}{2k} \cdot p^k = (1 - o(1))pn$ numbers in $[n]_p$ with no k-AP. When $p \leqslant 0.1n^{-1/(k-1)}$ then $pn - \frac{n^2}{2k} \cdot p^k = (1 - 0.1^{k-1}/k)pn$. □

Question 1.7. *What happens when $p >> n^{-1/(k-1)}$? Can we say that $[n]_p$ is (δ, k)-Szemerédi?*

Theorem 1.8 (Balogh–Morris–Samotij [2], Saxton–Thomason [5])**.** *For every $\beta > 0$ and $k \in \mathbb{N}$, there exist constants C and n_0 so that for every $n \in \mathbb{N}$ with $n \geqslant n_0$, if $m \geqslant Cn^{1-1/(k-1)}$, then there are at most $\binom{\beta n}{m}$ m-subsets of $[n]$ that contain no k-AP.*

Proof sketch. We adapt the method that we attempted earlier for counting the number of k-AP-free sets by replacing $2^{\delta n}$ with $\binom{\delta n}{m}$. Now $\binom{n}{\leqslant Cn^{1-1/(k-1)}}$ is a lower-order term. Choosing $\delta = \beta/2$ gives the result. □

The sparse random analogue of Szemerédi's theorem, originally proved by Schacht [6] and independently by Conlon and Gowers [3], follows as a corollary to Theorem 1.8.

Corollary 1.9 (Schacht 2016 [6]; Conlon–Gowers 2016 [3])**.** *For every $\delta \in (0,1)$ and every $k \in \mathbb{N}$, there exists a constant C such that if $p \geqslant Cn^{-1/(k-1)}$ for all sufficiently large n, then*

$$\lim_{n\to\infty} \mathbb{P}([n]_p \textit{ is } (\delta, k)\textit{-Szemerédi}) = 1.$$

Proof. Fix $k \in \mathbb{N}$ and $\delta \in (0,1)$. Let $\beta = \delta/(2e) \cdot e^{-1/\delta}$, and let $C_{1.8}$ denote the constant which satisfies Theorem 1.8 for these values of β and k. Set $C = 2C_{1.8}(\beta, k)/\delta$. Assume that $p \geqslant Cn^{-1/(k-1)}$, let $m = \delta p/2$, and let X_m denote the number of m-subsets of $[n]_p$ that contain no k-AP. By Theorem 1.8, we have

$$\mathbb{P}(X_m > 0) \leqslant \mathbb{E}[X_m] \leqslant \binom{\beta n}{m} p^m \leqslant \left(\frac{\beta e p n}{m}\right)^m = \left(\frac{2\beta e}{\delta}\right)^m = e^{-m/\delta}.$$

Let $\mathcal{A}$ denote the event that $[n]_p$ is not (δ, k)-Szemerédi, i.e., that $[n]_p$ contains a subset with $\delta|[n]_p|$ elements and no k-AP. By Chernoff's inequality, we have

$$\mathbb{P}(\mathcal{A}) \leqslant \mathbb{P}\left(\mathcal{A} \wedge |[n]_p| \geqslant \frac{pn}{2}\right) + \mathbb{P}\left(|[n]_p| < \frac{pn}{2}\right) \leqslant \mathbb{P}(X_m > 0) + e^{-pn/8} \leqslant 2e^{-pn/8}.$$

□

Note that both Theorem 1.8 and Corollary 1.2 1.9 are sharp up to the the constant C.

References

[1] J. Balogh, R. Morris and W. Samotij, *Independent sets in hypergraphs*, J. Amer. Math. Soc. 28 (2015), 669–709.

[2] D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, Ann. Math., Second Series, 184 no. 2 (2016), 367–454.

[3] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. Math., 167 no. 2 (2008), 481–547.

[4] D. Saxton and A.G. Thomason, *Hypergraph Containers*, Inventiones Mathematicae, 201 (2015), 925–992.

[5] M. Schacht, *Extremal results for random discrete structures*, Ann. Math., Second Series, 184 no. 2 (2016), 333-65.

[6] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. 27 (1975), 199-234.

[7] P. Varnavides, *On certain sets of positive density*, J. London Math. Soc. 34 (1959), 358-360.