## 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 $A, B \subseteq V(\mathcal{H})$ 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 $e(\mathcal{H}[A]) \geq \varepsilon \cdot e(\mathcal{H})$ for every $A \in \mathcal{F}$. Here $e(\mathcal{H}[A])$ is the number of edges contained in $A$. Let $\mathcal{F}_\varepsilon := \{A \in \mathcal{F} : e(\mathcal{H}[A]) \geq \varepsilon \cdot e(\mathcal{H})\}$. 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 $\deg_H(T) := |\{F \in E(\mathcal{H}) : T \subseteq F\}|$, the number of edges of $\mathcal{H}$, counted with multiplicities, which contain $T$. Let $\Delta_\ell(\mathcal{H}) := \max\{\deg_H(T) : T \subseteq V(\mathcal{H}), |T| = \ell\}$.

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 $|A| \geq \varepsilon \cdot v(\mathcal{H})$ for all $A \in \mathcal{F}$. Suppose $\mathcal{H}$ is $(\mathcal{F}, \varepsilon)$-dense and $q \in (0, 1)$ such that for every $\ell \in [k]$,

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

Then there exists a family $\mathcal{S} \subseteq \binom{V(\mathcal{H})}{\leq C q^{v(\mathcal{H})}}$ and functions $f : \mathcal{S} \rightarrow \overline{\mathcal{F}}$ and $g : \mathcal{I}(\mathcal{H}) \rightarrow \mathcal{S}$ 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
1.1 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 $E(\mathcal{H}_k) = \{k$-APs in $[n]\}$. We can make a few observations about this hypergraph:

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

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

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

We will conclude that $|f(S)| \leq \delta n$ (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 $\delta$ and $k \in [n]$, there exists $\varepsilon > 0$ such that every $A \subseteq [n]$ with at least $\delta n$ elements contains at least $\varepsilon n^2$ $k$-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 $I \subseteq f(S)$:

$$\#\{k$-AP-free sets\} \leq \left(\frac{n}{Cn^{1-1/(k-1)}}\right) \cdot 2^{\delta n} = 2^{2\delta n}$$

for $\delta$ 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 $(\delta, k)$-Szemerédi if every subset $S \subseteq D$ with $|S| \geq \delta |D|$ 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 $(\delta, k)$-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 $(\delta, k)$-Szemerédi.

It is a natural question to ask which sets are $(\delta, k)$-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 $[n]_p$ $(\delta, k)$-Szemerédi with high probability?

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

**Proof.** We use the Alteration Method. Note that $E[n]_p = pn$ and

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

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 $p \leq 0.1n^{-1/(k-1)}$ 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 $(\delta, k)$-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 $n \geq n_0$, if $m \geq Cn^{1-1/(k-1)}$, then there are at most $(\frac{\delta n}{m})^m$ $m$-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 $2^n \binom{n}{m}$ with $(\frac{\delta n}{m})^m$. Now $(\frac{n}{Cn^{1-1/(k-1)}})$ 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 $p \geq Cn^{-1/(k-1)}$ for all sufficiently large $n$, then

$$\lim_{n \to \infty} \mathbb{P}([n]_p \text{ is } (\delta, k)-\text{Szemerédi}) = 1.$$
Proof. Fix $k \in \mathbb{N}$ and $\delta \in (0, 1)$. Let $\beta = \delta/(2e) \cdot e^{-1/\delta}$, and let $C_{\beta}$ denote the constant which satisfies Theorem 1.8 for these values of $\beta$ and $k$. Set $C = 2C_{\beta}/\delta$. Assume that $p \geq Cn^{-1/(k-1)}$, 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

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

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

$$
P(\mathcal{A}) \leq P\left( \mathcal{A} \land |[n]_p| \geq \frac{pm}{2} \right) + P\left( |[n]_p| < \frac{pm}{2} \right) \leq P(X_m > 0) + e^{-pm/8} \leq 2e^{-pm/8}.
$$

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

References


