CS-3334: Advanced Combinatorics **Fall 2022**

Lecture 4: October 8

Lecturer: Kuan Yang *Scribe: Weihao Zhu*

4.1 Applications of Crossing Lemma

Recall that we have introduced the crossing lemma and its proof which is based on the probabilistic method. Actually, the crossing lemma is not well-known in a long time, until 1997 when László Székely surprisingly applied it to some geometric problems. Today, we will introduce several applications of the crossing lemma.

4.1.1 Incidence Geometry

For a set of points P and a set of lines L, define $I(P,L) = |\{(p,\ell) \in P \times L : p \in \ell\}|$. For instance, let $P = [k] \times [2k^2]$ and $L = \{y = mx + b : m \in [k], b \in [k^2]\}$. In this way, each line contains exactly k points, which implies that $I(P,L) = O(k^4)$ while $|P| = 2|L| = 2k^3$. Can we make $I(P,L)$ larger?

Problem 4.1 *What is the maximal value of $|I(P,L)|$ when $|P| = O(n)$ and $|L| = O(n)$?*

The example above gives that $I(P,L) = O(n^{4/3})$. Is $O(n^{4/3})$ the upper bound?

A trivial bound is that $I(P,L) \le |P| \cdot |L| = O(n^2)$, but this does not rely on any properties of points and lines. Actually, notice that every pair of points determine at most one line, which helps us derive a better upper bound as follows:

$$\begin{aligned}
I(P,L)^2 &= \Big(\sum_{\ell \in L} \sum_{p \in P} [p \in \ell]\Big)^2 \\
&\le |L| \cdot \sum_{\ell \in L} \Big(\sum_{p \in P} [p \in \ell]\Big)^2 && \text{(Cauchy-Schwarz inequality)} \\
&= |L| \cdot \sum_{p_1, p_2 \in P} \sum_{\ell \in L} [p_1 \in \ell] \cdot [p_2 \in \ell] \\
&= |L| \cdot \Big(\sum_{p \in P} \sum_{\ell \in L} [p \in \ell] + \sum_{p_1 \ne p_2} \sum_{\ell \in L} [p_1 \in \ell \wedge p_2 \in \ell]\Big) \\
&\le |L| \cdot (I(P,L) + |P|^2),
\end{aligned}$$

It implies that $|I(P,L)| \lesssim |P| \cdot |L|^{1/2} + |L|$. When $|P| = O(n)$ and $|L| = O(n)$, we have $|I(P,L)| \le O(n^{3/2})$.

But this is not the best known bound. The following Szemerédi–Trotter Theorem shows that $O(n^{4/3})$ is indeed the correct upper bound.

Theorem 4.1 (Szemerédi–Trotter, 1983) $I(P,L) \lesssim |P|^{2/3}|L|^{2/3} + |P| + |L|$.

In 1997, László Székely found a simple proof of Szemerédi–Trotter theorem based on the crossing lemma as follows.

Proof: Construct a graph $G = (V, E)$ such that $V = P$, and $E = \{(p_1, p_2) : \exists \ell \in L, p_1, p_2 \in \ell \text{ are adjacent on } \ell\}$.

Obviously, we have $|E| = \sum_{\ell \in L}(|P \cap L| - 1) \geq \frac{1}{2}(I(P, L) - |L|)$. Note that two lines share at most a common point, so the crossing number of this graph is at most $|L|^2$. Based on the crossing lemma, we have

$$|L|^2 \geq cr(G) \gtrsim \frac{|E|^3}{|V|^2} \gtrsim \frac{I(P, L)^3}{|P|^2}$$

if $I(P, L) \geq 8|P|$, which helps us complete the proof of the theorem. ■

4.1.2 Randomness Extractor

Now, we will introduce the randomness extractor, which is an application of Szemerédi–Trotter theorem.

Suppose we have some inputs from a distribution over a set S of size 2^m, and we hope to extract n random bits. In other words, we would like to construct a function $ext : \{0, 1\}^m \to \{0, 1\}^n$.

However, it is sometimes impossible. Suppose $m = 8$ and $n = 7$. Also, the sampler is broken and it only produces half of the results. Given the function we construct, the adversary can always turn the results biased.

There are two possible solutions to solve the problem:

- Seeded extractor: $ext : \{0, 1\}^m \times \{0, 1\}^t \to \{0, 1\}^n$;
- Seedless extractor: $ext : (\{0, 1\}^m)^\ell \to \{0, 1\}^n$.

A seedless extractor gets inputs from several independent sources and produces unbiased random bits. Roughly speaking, we say ext is a (k, ε)-extractor, if every independent input source has 2^k possible results, and the distribution of the output of ext is ε-close to a uniform distribution over n $\{0, 1\}$ bits.The following theorem shows that the *inner product* function is a good seedless extractor with two sources.

Theorem 4.2 (Benny Chor & Oded Goldreich, 1988) *Let $F(x, y) = (-1)^{\langle x, y \rangle}$. Then for any $S, T \subseteq \{0, 1\}^n$, we have*

$$|\mathbf{E}_{x \sim S, y \sim T}[F(x, y)]| \leq \sqrt{\frac{2^n}{|S| \cdot |T|}}.$$

Note that $dist(F, U_2) = \frac{\mathbf{E}[F]}{2}$, which implies that inner product is a (k, ε)-extractor for $k > n/2 + \log(1/\varepsilon)$.

For 2-source extractor, this is almost the best so far. The best known result is $(1/2 - \varepsilon)n$, which was given by Jean Bourgain in 2005. However, the probabilistic argument shows that $k \sim \log n$.

For multi-source extractor, the following theorem gives a $(\delta n, \varepsilon)$-extractor, where the min-entropy $H^\infty(X)$ is given by

$$H^\infty(X) = \min_{\Pr[X = x] > 0} -\log \Pr[X = x].$$

Theorem 4.3 (Boaz Barak, Russell Impagliazzo & Avi Wigderson, 2006) *For every constant $\delta > 0$ there exists a constant $\ell = (1/\delta)^{O(1)}$ and a polynomial-time algorithm to compute a function $f : \{0,1\}^{n\cdot\ell} \to \{0,1\}^n$ such that for every independent random variables $X_1, X_2, \ldots, X_\ell \in \{0,1\}^n$ with min-entropy $H^\infty(X_i) \geqslant \delta n$,*

$$\mathsf{dist}(f(X_1, X_2, \ldots, X_\ell), U_n) < 2^{-\Omega(n)} .$$

We briefly introduce the sketch of a proof based on Szemerédi–Trotter theorem here. First, we need the following theorem which generalizes Szemerédi–Trotter theorem to finite fields.

Theorem 4.4 (Jean Bourgain, Nets Katz & Terence Tao, 2004) *Suppose $\mathbb{F} = \mathbb{F}_p$ is a finite field with prime cardinality. Let L be a set of n lines in $\mathbb{F}^2$ and P be a set of n points in $\mathbb{F}^2$. If there exists some $\delta > 0$ such that $p^\delta < n < p^{2-\delta}$, we have*

$$I(P, L) = O(n^{3/2-\varepsilon})$$

for some $\varepsilon > 0$ which depends only on δ.

We won't introduce the proof of Theorem 4.3 in detail. But the following lemma plays an important role in its proof, which is based on the finite-field Szemerédi–Trotter theorem.

Lemma 4.5 *Let $\delta > 0$ and $\mathbb{F}$ be a finite field in which Szemerédi–Trotter theorem holds.. Suppose $A, B, C \subseteq \mathbb{F}$ of size n where $|\mathbb{F}|^\delta < n < |\mathbb{F}|^{1-\delta}$. Then, there exists $\varepsilon > 0$ such that $|A + BC| > n^{1+\varepsilon}$. Here, $A + BC$ is defined as $\{a + bc : a \in A, b \in B, c \in C\}$.*

Proof: Let

$$S(x) = |\{(a, b, c) \in A \times B \times C : a + bc = x\}|$$

denote the number of x in all results generated by elements from A, B and C. So we have

$$\sum_{x \in A+BC} S(x) = n^3.$$

If $|A + BC| < n^{1+\varepsilon}$, then according to Cauchy-Schwarz inequality,

$$\sum S(x)^2 \geq \frac{(\sum S(x))^2}{|A + BC|} \geq n^{5-\varepsilon} .$$

Also, for every $b \in B$ and $c \in C$, there is at most one solution to the equation $a + bc = x$ for some fixed x. So $S(x) \leq n^2$ for all $x \in \mathbb{F}$.

Now define $T = \{x : S(x) > n^{2-2\varepsilon}\}$. Then we have $n^{1-2\varepsilon} \leq |T| \leq n^{1+2\varepsilon}$.

Let $R = \{(a, b, c, x) \in A \times B \times C \times T : a + bc = x\}$, then we have

$$|R| \geq n^{3-4\varepsilon}.$$

But R can be seen as the incidence set of lines and points over $\mathbb{F}^2$. Let $P = C \times T$ be the point set and $L = \{\ell_{a,b} : Y = bX + a\}$ be lines. We have $|P| \leq n^{2+2\varepsilon}$ and $|L| \leq n^2$, but $I(P, L) \geq n^{3-4\varepsilon}$, which contradicts the finite-field Szemerédi–Trotter theorem if ε is sufficiently small. ■

Given this lemma, we can construct the randomness extractor: Let $X_1, X_2, \ldots, X_\ell \in \{0,1\}^n$ be inputs from ℓ independent sources with min-entropy δn. Then we take the functions:

$$f_1(X_1, X_2, X_3) = X_1 + X_2 X_3$$
$$f_2(X_1, \ldots, X_9) = (X_1 + X_2 X_3) + (X_4 + X_5 X_6) \cdot (X_7 + X_8 X_9)$$
$$\cdots\cdots$$
$$f_i(X_1, \ldots, X_{3^i}) = f_{i-1}(X_1, \ldots, X_{3^{i-1}}) + f_{i-1}(X_{3^{i-1}+1}, \ldots, X_{2\cdot 3^{i-1}}) \cdot f_{i-1}(X_{2\cdot 3^{i-1}+1}, \ldots, X_{3^i})$$
$$\cdots\cdots$$

After a constant $\ell = (1/\delta)^{O(1)}$ steps, the size of support of f_i will be almost $\mathbb{F}$.

However, although the support of f can be almost the whole set of $\mathbb{F}$, we still don't know if we have a distribution close to uniform.

In 2005, Barak, Impagliazzo and Wigderson proved the statistical analog of Lemma 4.5:

Lemma 4.6 *Let $\mathbb{F}$ be a prime field, then there exists some constant $\varepsilon > 0$ such that for all distribution A, B, C over $\mathbb{F}$ with min-entropy at least m, the distribution $A + BC$ is $2^{-\varepsilon m}$-close to having min-entropy at least $\min\{(1+\varepsilon)m, 0.9 \log |\mathbb{F}|\}$.*

The distribution $A + BC$ is defined as the distribution of $a + bc$ where a, b, c are chosen independently at random with distribution A, B and C respectively.

Finally, Barak, Impagliazzo and Wigderson proved that if $X_1, X_2, \ldots, X_9$ are chosen from 9 independent distribution over $\mathbb{F}$ with min-entropy at least $0.9 \log |\mathbb{F}|$, then the distribution of $f_2(X_1, X_2, \ldots, X_9)$ is $|\mathbb{F}|^{-0.01}$-close to uniform over $\mathbb{F}$. By the union bound, the statistical distance of distribution of the eventual outputs to uniform distribution over n bits will be at most $2^{-\Omega(n)}$, which completes the proof of Theorem 4.3.

4.1.3 Unit Distance Problem

We now consider another geometric problem. Given n points on a plane, how many pairs of points are at distance 1?

For example, in a grid of size $\sqrt{n} \times \sqrt{n}$, there are $O(n)$ such pairs.

In 1946, Paul Erdős made the following conjecture:

Conjecture 4.7 (Paul Erdős, 1946) *For any n points, the number of unit distances is at most $n^{1+o(1)}$.*

However, Paul Erdős only derived an upper bound of $O(n^{1.5})$. In 1973, Józsa and Szemerédi improved this bound to $o(n^{1.5})$. In 1984, Beck and Spencer gave a bound of $O(n^{1.44})$. The best known bound is $O(n^{4/3})$, which was given by Spencer, Szemerédi and Trotter in 1984.

Here, we will introduce the proof by Székely in 1997 as follows.

Proof: For each point p, if there are at least 2 points having unit distance to p, draw a unit circle centered at p. Then, for any pair (p, q), if there exists more than one arc between them, we only keep one and erase others.

Now, we have a drawing of a graph with n vertices and m edges, where $m \geq (\#\text{unit distance} - n)/2$. Since any two circles intersect at most two points, we have $cr \leq 2 \cdot \binom{n}{2}$, which implies $\#\text{unit distance} \lesssim n^{4/3}$. ■

4.2 Probabilistic Method in Set Systems

In this section, we will introduce several applications of the probabilistic method in set systems.

4.2.1 Sperner Theorem & LYM Inequality

Let $\mathcal{F}$ be a family of subsets of $[n]$. $\mathcal{F}$ is called an anti-chain if $\forall S, T \in \mathcal{F}$, we have $S \not\subseteq T$. For instance, $\mathcal{F} = \binom{[n]}{k}$ is an anti-chain. The problem is how large an anti-chain can be?

Recall that we have introduced Sperner's theorem last semester.

Theorem 4.8 (Sperner, 1928) *If $\mathcal{F}$ is an anti-chain of $[n]$, then $|\mathcal{F}| \leq \binom{n}{[n/2]}$.*

Here, we will introduce LYM inequality as follows, which implies Sperner's theorem immediately.

Theorem 4.9 (LYM inequality; Bollobás 1965, Lubell 1966, Meshalkin 1963 & Yamamoto 1954) *If $\mathcal{F}$ is an anti-chain of $[n]$, then*

$$\sum_{S \in \mathcal{F}} \frac{1}{\binom{n}{|S|}} \leq 1.$$

Proof: Consider a random permutation π of $[n]$. Construct a chain

$$C_\pi = \{\emptyset, \{\pi(1)\}, \{\pi(1), \pi(2)\}, \ldots, \{\pi(1), \ldots, \pi(n)\}\}.$$

For any $S \in \mathcal{F}$, let X_S be 1 if $S \in C_\pi$ and 0 otherwise. Therefore,

$$\mathbf{E}[X_S] = \mathbf{Pr}[S \in C_\pi] = \frac{|S|!\,(n - |S|)!}{n!} = \frac{1}{\binom{n}{|S|}}.$$

Since $\mathcal{F}$ is an anti-chain, we have $\sum X_S = |\mathcal{F} \cap C_\pi| \leq 1$, which implies that $\sum \mathbf{E}[X_S] \leq 1$. This completes the proof. ■

4.2.2 Intersecting Set Families

We say $\mathcal{F}$ is an intersecting family of subsets of $[n]$, if $\forall S, T \in \mathcal{F}$, $S \cap T \neq \emptyset$. For instance, $\mathcal{F}$ = all subsets of size k containing 1 is intersecting.

We would also like to know how large an intersecting family can be. However this problem is trivial. Consider $\mathcal{F}$ be the family of all subsets containing 1. Clearly, $\mathcal{F}$ is intersecting, and $|\mathcal{F}| = 2^{n-1}$. On the other hand, for any subset $S \subseteq [n]$, pair it and its complement. Then an intersecting family $\mathcal{F}$ can only take at most one subset from each pair. The pigeonhole principle shows that the size of $\mathcal{F}$ is at most 2^{n-1}.

So we turn to consider the following problem. Suppose that $\mathcal{F}$ only takes k-element subsets for some fixed k. How large can $\mathcal{F}$ be? Note that if $2k > n$, then every two subsets of size k are intersecting. So we assume $2k \leq n$. The following theorem shows an upper bound of the size of such intersecting families.

Theorem 4.10 (Erdős–Ko–Rado, 1961; proved in 1938) *If $n \geq 2k$ and $\mathcal{F}$ is an intersecting family of k-element subsets, then $|\mathcal{F}| \leq \binom{n-1}{k-1}$.*

Proof: (by Gyula Katona) Let π be a random permutation of $[n]$. Consider a circle and all contiguous blocks of size k. That is, $C_\pi = \{\{\pi((i+j) \bmod n) : j \in [k]\} : i \in [n]\}$. (We assume that $\pi(0) = \pi(n)$ here.)

For any $S \in \mathcal{F}$, let X_S be 1 if $S \in C_\pi$ and 0 otherwise. Therefore,

$$\mathbf{E}[X_S] = \mathbf{Pr}[S \in C_\pi] = \frac{n}{\binom{n}{k}}.$$

Since $\mathcal{F}$ is an intersecting family, we have $\sum X_S = |\mathcal{F} \cap C_\pi| \leq k$. To prove this, note that for every $S \in C_\pi$, there exists $2(k-1)$ other subsets in C_π intersecting S, but they can be paired off into $k-1$ distinct pairs, and two subsets in each pair are disjoint. So $\sum X_S = |\mathcal{F} \cap C_\pi| \leq k$, and thus $\sum \mathbf{E}[X_S] \leq k$, which completes the proof. ■

4.3 Turán's Theorem Revisit, and Derandomization

Theorem 4.11 (Caro 1979, Wei 1981) *For any graph G,*

$$\alpha(G) \geq \sum_{v \in V(G)} \frac{1}{deg(v)+1}.$$

.

Proof: Consider a random permutation π of V. Let I be the set of vertices that appear before all its neighbours. Obviously, I is an independent set.

For any vertex v, $\mathbf{Pr}[v \in I] = \frac{1}{deg(v)+1}$, which implies that $\mathbf{E}[|I|] = \sum_v (deg(v)+1)^{-1}$. This completes the proof. ■

Notice that if we take the component of graph G, we have the following corollary.

Corollary 4.12 *For any graph G,*

$$\omega(G) \geq \sum_{v \in V(G)} \frac{1}{n - deg(v)} \geq \frac{1}{1 - \frac{2m}{n^2}}.$$

The last inequality above is due to Jensen's inequality. After rearranging, we have $m \leq (1 - \frac{1}{r}) \cdot \frac{n^2}{2}$ if graph G is K_r-free, which is the same as Turán's Theorem.

Derandomization *to be continued...*