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

Lecture 9: November 1

Lecturer: Kuan Yang *Scribe: Weihao Zhu*

9.1 Algorithmic Lovász Local Lemma, Revisit

Recall that last time we have introduced the algorithmic Lovász Local Lemma. We discussed the algorithm given by Robin Moser and Gábor Tardos in 2010. In this section, we will introduce the original proof of the algorithmic local lemma, which was given by Moser in 2009.

Consider a k-SAT formula:

$$\varphi = C_1 \wedge C_2 \wedge \ldots \wedge C_m$$

of which each clause has exactly k literals. Robin Moser gave a fix-it algorithm to find a valid assignment as follows:

```
Algorithm 1: Moser's fix-it algorithm for $k$-SAT model
Input: A $k$-SAT formula: $\varphi = C_1 \wedge C_2 \wedge \ldots \wedge C_m$.
Function Solve($\varphi$):
    randomly initialize $x_1, \ldots, x_n$
    while there exists unsatisfied clause $C$ do
        Fix($C$)                    // Call Fix($C$) to adjust assignments of variables in $C$
Function Fix($C$):
    randomly re-sample $vbl(C)$
    while there exists unsatisfied clause $D$ overlapping with $C$ do
        Fix($D$)
```

Robin Moser proved that when each caluse does not intersect with too many other clauses, `Solve(`φ`)` can find a satisfying assignment in polynomial time with high probability. Precisely, the theorem is as follows:

Theorem 9.1 (Robin Moser, 2009 STOC best paper award) *Let d be the maximum degree of clauses, i.e., each cluase intersects with at most d clauses (including itself). Then,* `Solve(`φ`)` *finds a satisfying assignment in polynomial time with high probability as long as $d \leq 2^{k-3}$.*

Proof: Consider the recursion tree.

Suppose there are T times of `Fix` calls before terminating. Clearly, `Solve(`φ`)` used $n + kT$ random bits in total. We now argue that if T is sufficiently large, the number of random bits used by the recursion tree is smaller than kT.

A key observation is that for each sub-tree rooted at some `Fix(`C`)`, all satisfied clauses before `Fix(`C`)` cannot become unsatisfied after all `Fix` calls in the sub-tree have been executed. Thus we can see that all clauses in the first level are distinct. Denote by m 0/1 bits whether each clause is fixed at the first level.

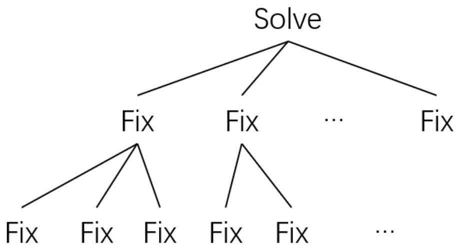

Figure 9.1: The graph shows an example of the recursion tree.

For any other node in the recursion tree, it is clear that each `Fix` call has at most d children. So we only need to record it by its index in the children of its parent node. Denote by $\lceil \log_2 d \rceil$ bits.

To record the structure of the recursion tree, consider its DFS sequence. For each node, use "1" to denote that the node is pushed in stack, and use "0" to record the outing. Overall, the number of bits we need is at most $m + (\lceil \log_2 d \rceil + 2)T \leq m + (k-1)T$. Finally, we use another n bits to record the final assignment. Note that every random bit used in Moser's algorithm is determined uniquely by the final assignment and the recursion tree, since for each `Fix(`C`)` call, the assignment of $\mathrm{vbl}(C)$ is determined before `Fix(`C`)`. Therefore, we use $m + (k-1)T + n$ bits to uniquely encode a sequence of $n + kT$ random bits.

Now, we need to apply the incompressibility theorem, which is as follows:

Theorem 9.2 (Incompressibility theorem) *N uniform random bits cannot be encoded to no more than $N - l$ bits with probability at least $1 - O(2^{-l})$.*

In Moser's proof, $n + kT$ random bits are encoded to $n + m + (k-1)T$ bits. However, T is not a fixed integer, which implies that we cannot use the incompressibility theorem directly. We need to find some other methods.

Let's fix $t = m + \log_2 n$, and we provide only $n + kt$ random bits in total. The algorithm will be forced to terminate if all random bits have been used up. If the algorithm succeeds after T steps of `Fix` calls, $n + kt$ random bits are decoded into $n + m + (k-1)T + k(t-T)$ bits. Otherwise when the algorithm fails to find a satisfying assignment, then $n + kt$ random bits are encoded into at most $n + m + (k-1)t$ bits. According the incompressibility theorem, the probability that the algorithm fails is at most

$$2^{(n+m+(k-1)t)-(n+kt)} = 2^{-\log_2 n} = 1/n,$$

which completes the proof. ■

Notice that we used the incompressibility theorem in the proof, which is an important theorem in coding theory. Another essential theorem in coding theory is Shannon's source coding theorem. We will give its simplified version here.

Theorem 9.3 (Shannon's source coding theorem, simplified version) *For a discrete random variable X with finite support Ω. Let $f : \Omega \to \{0,1\}^*$ be a uniquely decodable code. Then*

$$\mathbf{E}[|f(X)|] \geq H(X) = -\sum_x \mathbf{Pr}[X = x] \log(\mathbf{Pr}[X = x]).$$

9.2 Entropy Function

In the proof above, we use two different strategies to encode the recursion tree, one by the algorithm and the other by the DFS sequence. Why the number of bits we need in these two strategies are different? Given the same support Ω, do some random variables give us more "information", while some give us less? If so, can we define a function that describes how much "information" the random variable provides us? In this section, we will introduce the entropy function and its applications.

9.2.1 Definitions and Properties

Definition 9.1 (Entropy) *Given a discrete random variable $X \in \Omega$. Define the entropy*

$$H(X) = \sum_{\omega \in \Omega} -\mathbf{Pr}[X = \omega] \cdot \log_2 \mathbf{Pr}[X = \omega].$$

(By convention, define $0 \cdot \log 0 = 0$.) Usually, denote $p(\omega) = \mathbf{Pr}[X = \omega]$.

Based on the defintion of the entropy function, we have several basic properties as follows.

Proposition 9.4

$$H(X) \geq 0.$$

Proposition 9.5 (Uniform bound) *Let $\Omega = \{\omega : \Pr[X = \omega] > 0\}$ be the support of X. Then,*

$$H(X) \leq \log_2 |\Omega|$$

with equality if and only if X is uniformly distributed.

Proof: Let $f(x) = -x \log_2 x$, $x \in [0, 1]$. Since f is concave, by Jensen's inequality, we have

$$H(X) = \sum_{\omega} f(\mathbf{Pr}[X = \omega]) \leq |\Omega| \cdot f(1/|\Omega|) = \log_2 |\Omega|,$$

which completes the proof. ■

Proposition 9.6 (Concentration) *If $H(X) \leq t$, then there exists $\omega \in \Omega$ such that*

$$P(\omega) = \mathbf{Pr}[X = \omega] \geq 2^{-t}.$$

Similar to conditional expectation, we can also define conditional entropy as follows.

Definition 9.2 (Conditional entropy) *Let E be an event, define*

$$H(X|E) = \sum_{x} -\mathbf{Pr}[X = x|E] \cdot \log_2 \mathbf{Pr}[X = x|E].$$

If Y is another random variable, define

$$H(X|Y) = \sum_{y} H(X|Y = y) \cdot \mathbf{Pr}[Y = y] = \mathbf{E}_y[H(X|Y = y)].$$

Lemma 9.7 (Chain rule) *Let $H(X,Y)$ be the entropy of joint r.u.s. (X,Y), i.e.,*

$$H(X,Y)=\sum_{x,y}-\mathbf{Pr}[X=x,Y=y]\cdot\log_2\mathbf{Pr}[X=x,Y=y].$$

Then,

$$H(X,Y)=H(X)+H(Y|X).$$

In particular,

$$H(X,Y)=H(X)+H(Y)$$

if these two random variables are independent.

Proof: Let $p(x,y)$ denote $\mathbf{Pr}[X=x,Y=y]$ and $p(x|y)$ denote $\mathbf{Pr}[X=x|Y=y]$. Then by Bayes' rule,

$$p(x,y)=p(y)p(x|y).$$

So, we have

$$\begin{aligned}H(X|Y)&=\mathbf{E}_y[H(X|Y=y)]\\&=\sum_y -p(y)\sum_x p(x|y)\log_2 p(x|y)\\&=\sum_{x,y}-p(x,y)\cdot\log_2\frac{p(x,y)}{p(y)}\\&=\sum_{x,y}-p(x,y)\cdot\log_2 p(x,y)+\sum_{x,y}p(x,y)\log_2 p(y)\\&=H(X,Y)+\sum_y p(y)\log_2 p(y)\\&=H(X,Y)-H(Y).\end{aligned}$$

■

Intuitively, $H(X|Y)$ measures the amount of additional information in X, given a particular value of Y, averaged over all values of Y. Here are some important properties of the conditional entropy.

Proposition 9.8 $H(X|Y)=0$ *if and only if* $X=f(Y)$ *for some function* f.

Proposition 9.9 $H(X|Y)=H(X)$ *if and only if* X *and* Y *are independent.*

Theorem 9.10 (Subadditivity)

$$H(X,Y)\leq H(X)+H(Y).$$

Proof:

$$\begin{aligned} H(X)+H(Y)-H(X,Y) &= -\sum_{x} p(x)\log_2 p(x) - \sum_{y} p(y)\log_2 p(y) + \sum_{x,y} p(x,y)\log_2 p(x,y) \\ &= -\sum_{x,y} p(x,y)\log_2 \frac{p(x)p(y)}{p(x,y)} \\ &\geq -\log_2 \Big(\sum_{x,y} p(x,y)\cdot \frac{p(x)p(y)}{p(x,y)}\Big) \qquad \text{(by convexity of } -\log_2 x\text{)} \\ &= -\log_2 \sum_{x,y} p(x)p(y) \\ &= 0. \end{aligned}$$

■

Remark. More generally, we have

$$H(X_1,\ldots,X_n) \leq H(X_1,\ldots,X_{n-1}) + H(X_n) \leq \ldots \leq H(X_1)+\ldots+H(X_n).$$

Proposition 9.11 (dropping condition)

$$H(X|Y,Z) \leq H(X|Y).$$

Proof: As $H(X|Y) = H(X,Y) - H(Y) \leq H(X)$, for any $z \in \Omega_Z$, we have

$$H(X|Y,Z=z) \leq H(X|Z=z).$$

Taking expectation of z, it yields

$$H(X|Y,Z) \leq H(X|Z).$$

■

9.2.2 Chung-Frankl-Graham-Shearer Theorem

Let's start from a special case of Shearer's lemma.

Theorem 9.12 (Shearer's lemma, special case)

$$2H(X,Y,Z) \leq H(X,Y) + H(X,Z) + H(Y,Z).$$

Proof: As

$$\begin{aligned} H(X,Y) &= H(X) + H(Y|X), \\ H(X,Z) &= H(X) \qquad\qquad\quad + H(Z|X), \\ H(Y,Z) &\geq \qquad\qquad H(Y|X) + H(Z|X,Y), \end{aligned}$$

we have

$$H(X,Y)+H(X,Z)+H(Y,Z) \geq 2H(X) + 2H(Y|X) + 2H(Z|X,Y) = 2H(X,Y,Z).$$

■

Now, let's come to the Chung-Frankl-Graham-Shearer theorem, which can be proved in a similar way.

Theorem 9.13 (Chung, Frankl, Graham & Shearer, 1986) *Let $S_1, \ldots, S_m$ be subsets of $[n]$ such that every $i \in [n]$ belongs to at least k of $S_1, \ldots, S_m$. Then*

$$kH(X_1, \ldots, X_n) \leq \sum_{i=1}^{m} H(X_{S_i}).$$

This theorem leads to many useful and interesting results.

Corollary 9.14 *Let $\mathcal{F}$ be a family of subsets of $[n]$ and p_i denote the fraction of sets if $\mathcal{F}$ that contain i. Then,*

$$\log_2 |\mathcal{F}| \leq \sum_{i=1}^{n} H_b(p_i),$$

where $H_b(p) = H(Bernoulli(p)) = -p \log_2 p - (1-p) \log_2 (1-p)$.

Proof: For each $S \in \mathcal{F}$, let v_S be its indicator vector of length n. Let $X = (X_1, \ldots, X_n)$ be a random variable taking values in $\{0,1\}^n$ and set $\mathbf{Pr}[X = v_S] = \frac{1}{|\mathcal{F}|}$ for all $S \in \mathcal{F}$.

Clearly, $H(X) = |\mathcal{F}| \cdot (-1/|\mathcal{F}| \cdot \log_2 1/|\mathcal{F}|) = \log_2 |\mathcal{F}|$ and $H(X_i) = H_b(p_i)$. So the result follows from subadditivity. ■

Corollary 9.15

$$\sum_{i \leq np} \binom{n}{i} \leq 2^{nH_b(p)}$$

when $0 < p \leq 1/2$.

Proof: (by Frankl) Let $\mathcal{F} = \binom{[n]}{\leq pn}$ be the family of all subsets of $[n]$ of size no larger than pn, and p_i be the fraction of containing i.

Clearly, $p_1 = p_2 = \ldots = p_n$ and

$$\sum_i p_i |\mathcal{F}| = \sum_{S \in \mathcal{F}} |S| \leq pn|\mathcal{F}|.$$

So $p_i \leq p$ and thus $H_b(p_i) \leq H_b(p)$. Applying above results finishes the proof. ■

Now, let's introduce another corollary of the Chung-Frankl-Graham-Shearer theorem, which is also known as the product theorem.

Corollary 9.16 (Product theorem) *Let $S_1, \ldots, S_m \subseteq \Omega$, where each $i \in \Omega$ appears in at least k subsets S_j. Then for every $\mathcal{F} \subseteq 2^{\Omega}$,*

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}|_{S_i}|,$$

where $\mathcal{F}|_S = \{F \cap S : F \in \mathcal{F}\}$.

Proof: Let $n = |\Omega|$. Each subset of Ω corresponds to a $\{0,1\}$ vector of length n. Let $X = (X_1, X_2, \ldots, X_n)$ be a random variable sampled uniformly from all corresponding vectors of elements in $\mathcal{F}$. For each set $S = \{a_1, a_2, \ldots, a_{|S|}\}$, let X_S be the joint random variable $(X_{a_1}, X_{a_2}, \ldots, X_{a_{|S|}})$. Clearly the support of X_S

is all corresponding vectors of elements in $\mathcal{F}|_S$. Then, applying Shearer's lemma and the uniform bound, we have

$$k\log_2|\mathcal{F}| = kH(X_1,\ldots,X_n) \leq \sum_{i=1}^{m} H(X_{S_i}) \leq \sum_{i=1}^{m} \log_2|\mathcal{F}|_{S_i}|.$$

■

We will show two applications of the product theorem as follows.

Theorem 9.17 *Let $\mathcal{F} \subseteq 2^{[n]}$ be a family of subsets of $[n]$. Suppose that for any $S,T \in \mathcal{F}$, $S \cap T$ contains a pair of consecutive numbers. Then, $|\mathcal{F}| \leq 2^{n-2}$.*

Remark. Note that the bound is optimal. Consider $\mathcal{F} = 2^{[n-2]} + \{n-1,n\}$.

Proof: Let N_0 and N_1 be the set of all even numbers and ood numbers in $[n]$, respectively. Consider the projections: $\mathcal{F}_i = \{F \cap N_i : F \in \mathcal{F}\}$.

Note that for any $S,T \in \mathcal{F}$, there exists a k such that $\{k,k+1\} \subseteq S \cap T$. So, $S \cap T \cap N_0 \neq \emptyset$ and $S \cap T \cap N_1 \neq \emptyset$. Namely, for any $S_i, T_i \in \mathcal{F}_i$, $S_i \cap T_i \neq \emptyset$.

It implies that $\mathcal{F}|_{N_i}$ is an intersecting set family, and thus $|\mathcal{F}|_{N_i}| \leq 2^{|N_i|-1}$. Finally by the product theorem we have

$$|\mathcal{F}_i| \leq 2^{|N_0|-1} \cdot 2^{|N_1|-1} = 2^{n-2}.$$

■

Theorem 9.18 (Chung, Frankl, Graham & Shearer, 1986) *Every triangle intersecting family of subgraphs of K_n has size smaller than $2^{\binom{n}{2}-2}$.*

Proof: Let $\mathcal{G}$ be a triangle intersecting family. For any $S \subseteq [n]$ with $|S| = \lceil \frac{n}{2} \rceil$, let $E_S = \binom{S}{2} \cup \binom{[n]\setminus S}{2}$ and $r = |E_S| = \binom{\lfloor n/2 \rfloor}{2} + \binom{\lceil n/2 \rceil}{2} \leq \frac{1}{2}\binom{n}{2}$.

For any S, each triangle has at least one edge in E_S. Thus, $\mathcal{G}$ restricted to E_S is an intersecting family. So $|\mathcal{G}|_{E_S}| \leq 2^{|E_S|-1} = 2^{r-1}$.

Each edge of K_n appears in at least $k = \frac{r}{\binom{n}{2}} \cdot \binom{n}{\lfloor n/2 \rfloor}$ different E_S's. Applying the product theorem, it gives that $|\mathcal{G}|^k \leq (2^{r-1})^{\binom{n}{\lfloor n/2 \rfloor}}$. Therefore,

$$|\mathcal{G}| \leq 2^{\binom{n}{2} \cdot \frac{r-1}{r}} < 2^{\binom{n}{2}-2}.$$

■