* Class Next week!

Manditory to come to class for presentations.

Ask questions & be an active participant.

* Homework will be posted by next week.

Today: Applications of CC Lower Bounds

We'll use the following CC LBs:

(1) $BPP^{cc}(DISJ_n) = \Omega(n)$

also same LB for $UDISJ_n$ (unique DISJOINTNESS)

(2) There is a constant sized gadget g s.t. for every NP search problem $\{S_n\}$, $P^{cc}(S_n \circ g^n) = \Omega(n)$ ← Deterministic Lifting Thm

(3) There is a gadget $g: \{0,1\}^{O(\log n)} \times \{0,1\}^{O(\log n)} \to \{0,1\}$ s.t. $\forall \{S_n\}$, $BPP^{cc}(S_n \circ g^n) = \Omega(n)$ ← Randomized Lifting Thm

APPLICATIONS

- Property Testing
- game theory

} Presentations next week

- Streaming
- TIME/SPACE Turing Machine LBs

] LB by reduction (e.g. DISJ → Streaming)

- Proof complexity
- Circuit complexity
- Extension Complexity
- Clique/Coclique, Graph Theory, Learning Partial Functions

] LB via Lifting

APPLICATIONS

- Property Testing
- game theory

} Presentations next week

- Streaming
- TIME/SPACE Turing Machine LBs

] LB by reduction (e.g. DISJ → Streaming)

- Proof complexity
- Circuit complexity
- Extension Complexity
- Clique/Coclique, Graph Theory, Learning Partial Functions

] LB via Lifting

STREAMING LOWER BOUNDS

$S \in [n]^m$ is a length m stream

computing frequency moments of S:

Let $M_i = |\{j \in [m] \mid S_j = i\}|$

The k^{th} frequency moment of S, $F_k = \sum_{i=1}^{n} M_i^k$

F_0 = # distinct elements in stream

F_1 = length of stream

F_∞ = # occurrences of most frequent item

STREAMING LOWER BOUNDS

$S \in [n]^m$ is a length m stream

computing frequency moments of S:

Let $M_i = |\{j \in [m] \mid S_j = i\}|$

The k^{th} frequency moment of S, $F_k = \sum_{i=1}^{n} M_i^k$

F_0 = # distinct elements in stream

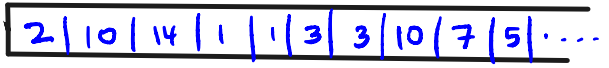
F_1 = length of stream

F_∞ = # occurences of most frequent item

We'll see: F_0, F_2 have low space approx. algs, whereas F_∞ requires large space

STREAMING LOWER BOUNDS

$S \in [n]^m$ is a length m stream



computing frequency moments of S:

Let $M_i = |\{j \in [m] \mid S_j = i\}|$

The k^{th} frequency moment of S, $F_k = \sum_{i=1}^{n} M_i^k$

F_0 = # distinct elements in stream

F_1 = length of stream

F_∞ = # occurrences of most frequent item

Theorem computing F_∞ requires $\Omega(\min\{m, n\})$ space

Stronger: any randomized alg for F_∞ to within $(1 \pm .2)$ factor wp $\geq \frac{2}{3}$ requires space $\Omega(\min\{m, n\})$.

Theorem computing F_∞ requires $\Omega(n)$ space/memory $(m=n)$

Pf Reduction from DISJ $\rightarrow$ low-space streaming alg for F_∞

Let A be space c streaming alg

Alice: $x \rightsquigarrow$ stream $a_x = \{i \mid x_i = 1\}$ $\quad 011011 \rightsquigarrow 2, 3, 5, 6$

Bob: $y \rightsquigarrow$ stream $b_x = \{j \mid y_j = 1\}$ $\quad 100100 \rightsquigarrow 1, 4$

Fact $\mathrm{DISJ}(x,y) = 1 \Rightarrow F_\infty(a_x b_x) = 2$

$\mathrm{DISJ}(x,y) = 0 \Rightarrow F_\infty(a_x b_x) = 1$

Theorem computing F_∞ requires $\Omega(n)$ space/memory

Pf Reduction from DISJ $\rightarrow$ low-space streaming alg for F_∞

Let A be space c streaming alg

Alice: $x \rightsquigarrow$ stream $a_x = \{i \mid x_i = 1\}$ $\quad 011011 \rightsquigarrow 2, 3, 5, 6$

Bob: $y \rightsquigarrow$ stream $b_x = \{j \mid y_j = 1\}$ $\quad 100100 \rightsquigarrow 1, 4$

Fact $\mathrm{DISJ}(x,y) = 1 \Rightarrow F_\infty(a_x b_x) = 2$
$\mathrm{DISJ}(x,y) = 0 \Rightarrow F_\infty(a_x b_x) = 1$

Simulation Alice simulates A on a_x & sends content of memory (c bits) to Bob; then Bob simulates rest of computation on b_y, and outputs "1" (not disjoint) iff $F_\infty(a_x b_y) = 2$

(c cost = $O(c)$.

$\therefore$ By $\Omega(n)$ LB for UDISJ, $\quad c = \Omega(n)$

MORE STREAMING LOWER BOUNDS

Previous LB actually showed something stronger:

Thm Any randomized streaming alg that for any stream S of length m computes F_0 to within $(1 \pm .2)$ factor (with prob $> \frac{2}{3}$) requires space $\Omega(\min\{m, n\})$.

Thm For $k \neq 1$ every randomized streaming alg for computing F_k exactly requires space $\Omega(\min\{m, n\})$

In our reduction F_∞ is 1 vs 2
so a factor of 2 difference.
For $k \neq 1$ the correct value will still be different in the 2 cases

MORE STREAMING LOWER BOUNDS

In contrast, we have very low space approx. algs for F_0 and F_2

Thm F_0, F_2 can be approx'd to within $(1 \pm \varepsilon)$ factor with prob $\geq (1-\delta)$ using space $O\left(\varepsilon^{-2}(\log n + \log m)\log\frac{1}{\delta}\right)$

APPLICATIONS

- Property Testing
- game theory

} Presentations next week

- Streaming
- TIME/SPACE Turing Machine LBs

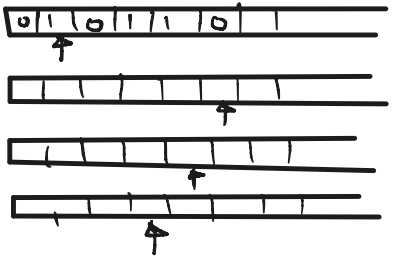
] LB by reduction (e.g. DISJ → Streaming)

- Proof complexity
- Circuit complexity
- Extension Complexity
- Clique/Coclique, Graph Theory, Learning Partial Functions

] LB via Lifting

TM TIME/SPACE LOWER BOUNDS

Multitape TMs: Read only input tape
plus $O(1)$ Read/Write tapes



Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$

We say that M recognizes/computes f if

$$\forall (x,y) \in \{0,1\}^{2n} \quad f(x,y)=1 \Rightarrow M(x0^n y) = 1$$
$$f(x,y)=0 \Rightarrow M(x0^n y) = 0$$

Theorem Let M compute f.

Then $$P^{cc}(f) \leq O\left(\frac{\text{Time}(M,n) \cdot \text{Space}(M,n)}{n}\right)$$

ie. if $P^{cc}(f) = \Omega(n)$ then any M computing f requires $\text{Time} \cdot \text{Space} = \Omega(n^2)$

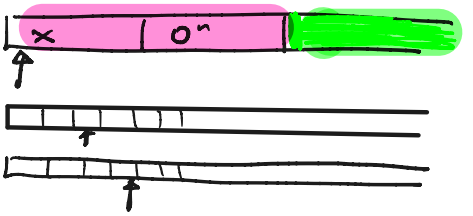
Proof

Let M be a TM that computes f. in Time $T(n)$, space $S(n)$

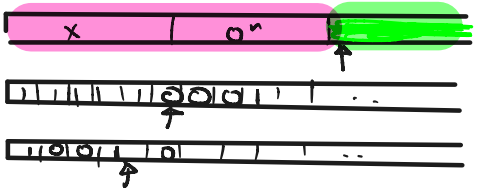
Then we will construct a CC protocol for f of cost $\leq T(n) \cdot S(n)$

Alice has x, Bob y.

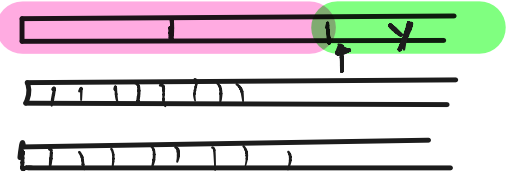
Alice simulates M on $x\,0^n$ until input head moves to green part



Then Alice sends entire content of RW tape and head locations to Bob



Bob continues simulation with y on green part until input head moves to pink



⋮

Comm. complexity :

# of Rounds $= \frac{T(n)}{n}$ (since they have to spend n steps going thru middle zone)

Cost per round $\leq O(S(n))$

$$\therefore CC(f) = O\left(\frac{T(n) \cdot S(n)}{n}\right)$$

Note 0^n in middle is kind of cheating

If we instead gave input $\boxed{x \mid y}$,

the cost of protocol would be $O(\#\text{-of-Reversals} * S(n))$

APPLICATIONS

- Property Testing
- game theory

} Presentations next week

- Streaming
- TIME/SPACE Turing Machine LBs

] LB by reduction (e.g. DISJ → Streaming)

- Proof complexity
- Circuit complexity
- Extension Complexity
- Clique/Coclique, Graph Theory, Learning Partial Functions

] LB via Lifting

Proof complexity LBs

Cutting Planes Proof System:

- Lines are Linear inequalities
- Start with unsat 3CNF $C = C_1 \wedge \dots \wedge C_m$

  Convert each clause into equiv linear inequality:

  $C_i = x_1 \vee \bar{x}_2 \vee x_3 \implies \mathcal{L}(C_i) = x_1 + (1 - x_2) + x_3 \geq 1$

  Axioms: $\mathcal{L}(C_i)$ $\forall$ initial clauses C_i

  plus: $\{x_i \geq 0,\ 1 \geq x_i\}$ $\forall i = 1, \dots, n$

- Rules: ① $\ell_1 \geq 0 \quad \ell_2 \geq 0 \implies a\ell_1 + b\ell_2 \geq 0$, $a, b \geq 0$

  ② $a_1 x_1 + a_2 x_2 + \dots + a_n x_n \geq k \implies x_1 + \dots + x_n \geq \lceil \frac{k}{2} \rceil$

  a_i divisible by 2 $\forall i$

CPs Proof

Let $C = C_1 \wedge \cdots \wedge C_m$ be UNSAT CNF over $X_1, \ldots, X_n$

A CP refutation π of C: a DAG where each

- sources are labelled with an axiom ($\mathcal{L}[C_j]$ for some j, or $X_i \geq 0$ or $X_i \leq 1$)
- all other vertices labelled with an inequality that is derivable from parents by one of the 2 rules
- Last line (sink) is $0 \geq 1$

Size(π) = # of lines in π

depth(π) = depth of dag (length of longest path in π)

Theorem (Depth LB for CPs)

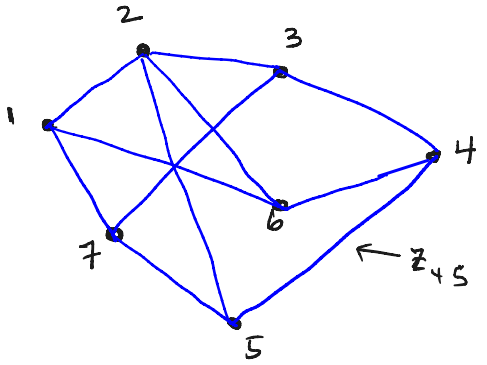
There exists a family of UNSAT kCNFs ($k \sim 10$) $\{C_n\}$, where C_n has n variables, $O(n)$ clauses such that any CPs refutation of C_n requires depth $\Omega\left(\frac{n}{\log n}\right)$.

* Above theorem can be strengthened using stronger lifting theorem to prove exponential size LBs for CPs refutations

Tseitin Formulas

Let G_n be an undirected graph on n vertices, n odd, degree K.

Variables correspond to edges in G_n



Tseitin (G_n) constraints

For every vertex v:
"sum of edges incident to v is odd"

(1) $z_{12} \oplus z_{16} \oplus z_{17} \equiv 1$

(2) $z_{12} \oplus z_{23} \oplus z_{26} \equiv 1$

$\vdots$

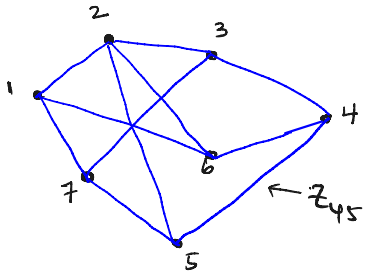
(7) $z_{17} \oplus z_{37} \oplus z_{57} \equiv 1$

Number of clauses $\approx n \cdot 2^{K-1} = O(n)$ for constant K

Tseitin Formulas

Let G_n be an undirected graph on n vertices, n odd, degree K.

Variables correspond to edges in G_n



Tseitin (G_n) constraints

For every vertex v:
"sum of edges incident to v is odd"

(1) $z_{12} \oplus z_{16} \oplus z_{17} \equiv 1$

(2) $z_{12} \oplus z_{23} \oplus z_{26} \equiv 1$

$\vdots$

(7) $z_{17} \oplus z_{37} \oplus z_{57} \equiv 1$

- Tseitin (G_n) UNsat. (for n odd)
- search (Tseitin (G_n)): given assignment α to variables output a clause falsified by α

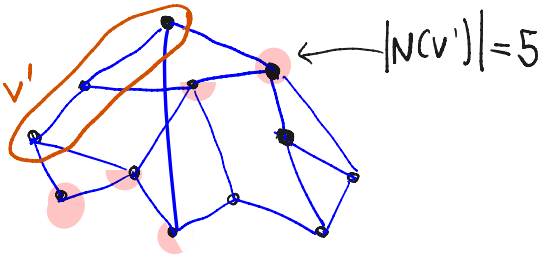
Search (G_n)

Tseitin Formulas

Let G_n be an undirected graph on n vertices, n odd, degree K.

Defn: G_n is (a,b)-expanding if $\forall V' \subseteq V$, $|V'| \leq a \cdot n$

$$|N(V')| \geq b \cdot |V'|$$



Lemma: Let $K=20$ there exists G_n (degree k) such that G_n is $(\frac{1}{3}, 1)$-expanding

Lemma Assume G_n is $(\frac{1}{3}, 1)$-expanding, degree k.
Then any decision tree for Search(G_n) has depth $\Omega(n)$.

Tseitin Formulas

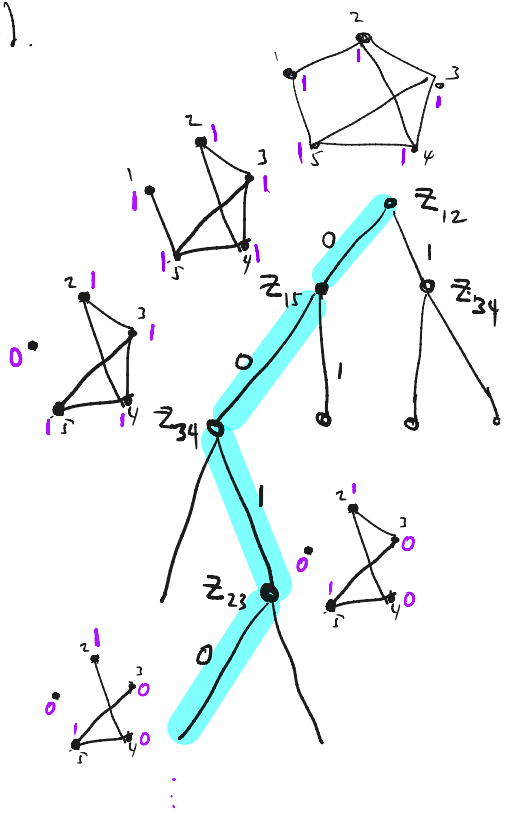
Let G_n be an undirected graph on n vertices, n odd, degree K.

Lemma Assume G_n is $(\frac{1}{3}, 1)$-expanding, degree K.
Then any decision tree for Search(G_n) has depth $\Omega(n)$.

Proof sketch
Let T be a dec. tree for Search(G_n). $\rho = \phi$
Starting at root, for every variable x_{ij} queried:

If z_{ij} breaks largest connected component of $T|_\rho$
set z_{ij} to 0/1 to force contradiction in larger piece & update ρ.

Otherwise set z_{ij} to a consistent value,
& update ρ



Tseitin Formulas

Lemma Assume G_n is $(\frac{1}{3}, 1)$-expanding, degree k.
Then any decision tree for Search(G_n) has depth $\Omega(n)$.

Proof sketch

Let T be a dec. tree for Search(G_n). $\rho = \phi$

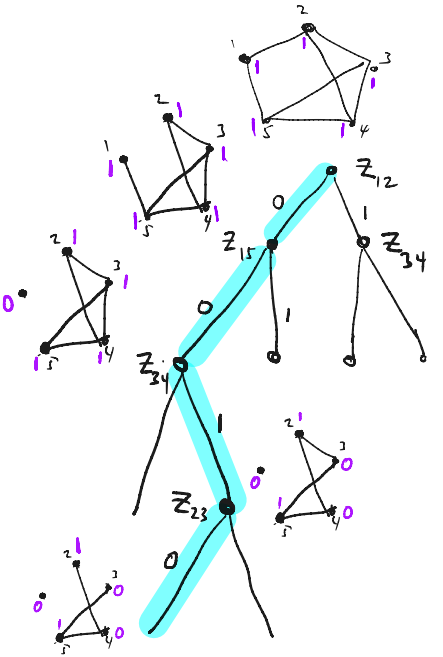
Starting at root, for every variable x_{ij} queried:

If z_{ij} breaks largest connected component of $T|_\rho$
set z_{ij} to 0/1 to force contradiction in
larger piece & update ρ.

Otherwise set z_{ij} to a consistent value,
& update ρ

↑ stop when the largest connected component, C^* in $T|_\rho$
size $\leq \frac{n}{3}$. Then $\frac{n}{6} \leq |C^*| \leq \frac{n}{3}$
(since all other components are satisfiable, this must happen)

All edges in $N(C^*)$ must have been queried
and by expansion, $|N(C^*)| \geq |C^*| \geq \frac{n}{6}$.
∴ at least $\frac{n}{6} = \Omega(n)$ edges queried on blue path. ∎

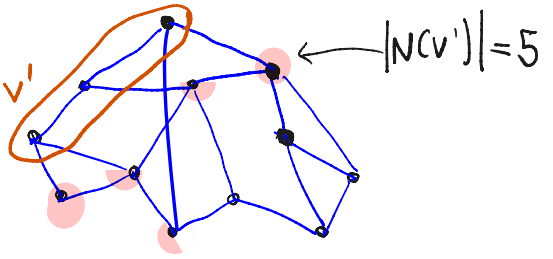


Tseitin Formulas

Let G_n be an undirected graph on n vertices, n odd, degree k.

Defn: G_n is (a,b)-expanding if $\forall V' \subseteq V$, $|V'| \leq a \cdot n$

$$|N(V')| \geq b \cdot |V'|$$



stronger

Lemma Assume G_n is $(\frac{1}{3}, 1)$-expanding, degree k.

Then any decision tree for Search(G_n) has depth $\Omega(n)$.

randomized

BPP Lifting Theorem [göös-P-Watson] [. . . .]

For any search problem S_N (N underlying vars)
There is a gadget $g: \{0,1\}^{O(\log n)} \times \{0,1\}^{O(\log n)} \to \{0,1\}$

Such that:

$$\text{Randomized CC}(S_N \circ g^N) = \Omega(\text{Randomized-DT}(S_N))$$

Combining BPP Lifting thm with stronger Lemma we have:

Lemma 1 (CC - Lowerbound)

$$\text{Randomized CC}(\text{Search}(G_n) \circ g^{kn}) = \Omega(n)$$

Lemma 2 (CP refutation $\Rightarrow$ CC protocol)

Let Π be a CP refutation of C, of depth d

Then for any partition of variables of C into 2 groups, $\vec{x}$ and $\vec{y}$

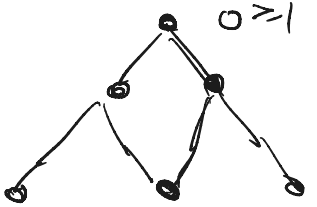
$\exists$ a randomized comm. protocol for Search(C) where
Alice gets $\vec{x}$, Bob gets $\vec{y}$
of cost $O(d \cdot \log^2 n)$

Lemma 2 (CP refutation $\Rightarrow$ CC protocol)

Let Π be a CP refutation of C, of depth d, over variables $\vec{x}, \vec{y}$

Then $\exists$ a randomized comm. protocol for Search(C) where Alice gets $\vec{x}$ Bob gets $\vec{y}$ of cost $O(d \cdot \log^2 n)$

Pf sketch: given assignment $\alpha = \alpha_0 \alpha_1$, Alice/Bob start at root of proof and evaluate each line to find a path from root to leaf s.t. all lines on path evaluate to false on α.



to evaluate each line requires cost $O(\log^2 n)$ by randomized protocol for evaluating $a \geq b$

$\underbrace{c_0 + c_1 x_1 + \dots + c_n x_n}_{\text{Alice: evaluate on } \alpha_0 \text{ to get } a} + \underbrace{d_1 y_1 + \dots + d_n y_n}_{\text{Bob: evaluate on } \alpha_1 \text{ to get } -b} \geq 0$

Theorem There exists a family of UNSAT KCNFs ($K \sim 10$) $\{C_n\}$, where C_n has n variables, $O(n)$ clauses such that any CPs refutation of C_n requires depth $\Omega\left(\frac{n}{\log^2 n}\right)$.

Pf Let $C_n = \underbrace{Tseitin(G_n) \circ g^{O(n)}}$ [n odd, G_n expanding]

$K \cdot O(\log n)$ CNF where each variable z_{ij} replaced by $g(x_{ij}, y_{ij})$

Assume Π is a depth $\ll \frac{n}{\log^2 n}$ CP refutation of C_n

By Lemma 2 (CP refutation $\Rightarrow$ CC protocol), this implies a randomised protocol for search$(G_n \circ g^{O(n)})$ of cost $\ll n/\log^2 n$.

This contradicts Lemma 1 (CC-lowerbound) ∎