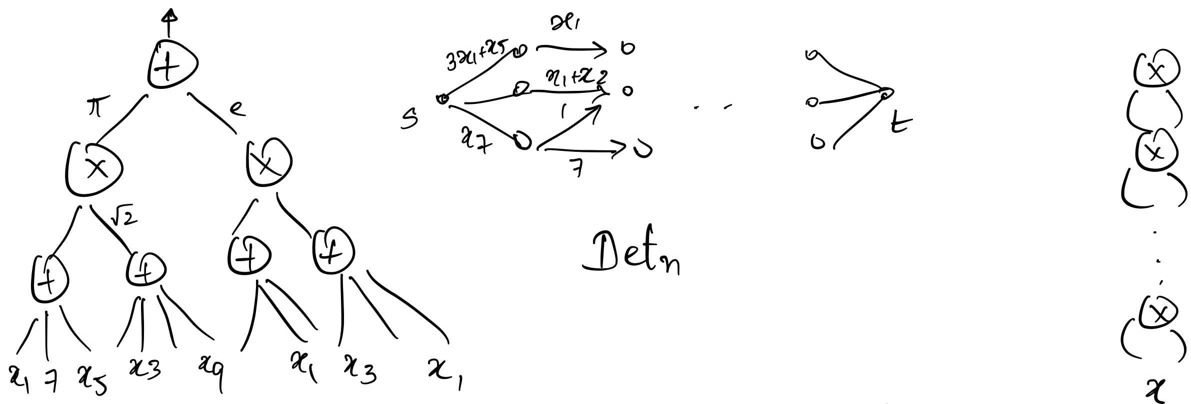
GCT Online Lecture series – Lower Bounds I.

Ramprasad Saptharishi
2021-11-16.

Agenda:
- Some recap of models,
- Typical lb proof templates.
- Some basic lower bounds, techniques

Models of computation.

Formulas $\subseteq$ ABPs $\subseteq$ Circuits.



(Don't panic... we'll work only with $d = \mathrm{poly}(n)$)

Constant depth circuits/formulas.

- $\Sigma\Pi$: $\sum_{i=1}^{s} \prod_{j=1}^{d} x_{ij}$ – sparse polynomials.
- $\Sigma\Pi\Sigma$: $\Sigma\Pi\Sigma\, x_{ijk}$ – depth 3 circuits.
  $\Sigma\Pi\, l_{ij}$ (surprisingly powerful)
- $\Sigma\wedge\Sigma$: $\sum l_i^{e_i}$ depth 3 powering circuits (or "diagonal" circuits)

... and so on.

Goal: Find an explicit family $\{f \in \mathbb{F}[x_1, .., x_n]\}_n$ of polynomials such that any circuit C from the class $\mathcal{C}$ computing f must have "large" size.

Ideally: $\mathcal{C}$ = alg. circuits. | Explicit: "you'll know it when you see it."
"large" = $\exp(n)$.

How do lower bounds "typically" proceed?

- Identify a weakness of the class $\mathcal{C}$.
- Quantify this weakness using a "complexity measure".
- Upper bound this measure for any small $C \in \mathcal{C}$
- Heuristically convince yourself that the measure is "large" for a "random polynomial".
- Find a hay in the haystack.

In a lot of cases, LBs follow the following template: (sub-additive measures)

- Show that any $C \in \mathcal{C}$ can be written as a sum of "building blocks" $\mathcal{B}$.

$$C = B_1 + \dots + B_t \quad \text{where each } B_i \in \mathcal{B} \qquad t = \text{fn of } s.$$

- Find a $\Gamma : \mathbb{F}[\bar{x}] \to \mathbb{R}$ that is subadditive

$$\Gamma(f+g) \leq \Gamma(f) + \Gamma(g)$$

and show that $\Gamma(B) \leq$ "small" for any $B \in \mathcal{B}$.

- Heuristically convince yourself that $\Gamma(f)$ is "large" if f is "random".

- Find hay in the haystack.

$$\Gamma(f) = \dim\{ L(f) : L \in \mathcal{L} \}$$

This lecture : - Some examples of LBs that follow this template.
- Some examples that don't. (in some sense)

Starting with the basics :

- $\Sigma\Pi$ circuits (I know ...)

$\mathcal{B}$ = monomials. $\Gamma(f)$ = sparsity.

$\Gamma(B) \leq 1$ $\Gamma(f+g) \leq \Gamma(f) + \Gamma(g)$ (duh)

A random poly probably has lots of monomials in it.

$f = (x_1+1) \cdots (x_n+1)$ is one of them hay.

$\Gamma(f) = 2^n$ $\Rightarrow$ we have our lower bound.

- Depth 3 powering circuits : $\ell_1^{d_1} + \cdots + \ell_s^{d_s}$

$$\mathcal{B} = \{ \ell^e : \ell \text{- linear, } e \in \mathbb{N} \}.$$

What is an exploitable weakness for such building blocks?

Obs: $\frac{\partial^2}{\partial x_i x_j} \ell^d = \alpha \cdot \ell^{d-2}$

$$\partial^{=k}(f) = \{ \partial_m(f) : m - \text{mon of deg } k \}$$

Lemma: $\dim \partial^{=k}(\ell^d) \leq 1$.

Set $\Gamma_k(f) = \dim. \partial^{=k}(f)$ (where k will be chosen later)

Cor: $f \in \overset{s}{\Sigma}\wedge\Sigma$, then for any k, $\Gamma_k(f) \leq s$.

If f is a random n-var, deg d poly, what do we expect this to be?

There are $\approx \binom{n+k}{k}$ vectors in $\partial^{=k}(f)$

each with $\binom{n+d-k}{d-k}$ coordinates.

$M(f) =$ 

Wishful heuristic:

$\Gamma_k(f) \approx \min\left(\binom{n+k}{k}, \binom{n+d-k}{d-k}\right)$ for a random f.

$f = x_1 \cdots x_n$. $\qquad \Gamma_k(f) = \binom{n}{k}$.

$\therefore$ If $f \in \Sigma^{s} \wedge \Sigma$, then $\Gamma_k(f) \le s$ and $\Gamma_k(f) \ge \binom{n}{k}$

$\Rightarrow \quad s \ge \binom{n}{k}$. $\qquad k = n/2 \Rightarrow s = 2^{\Omega(n)}$

– LBs for depth 3 circuits.

$$f = \sum_{i=1}^{s} \prod_{j=1}^{d_i} l_{ij} \longrightarrow \text{surprisingly powerful.}$$

Hom. depth 3 circuits.

$$f = \sum_{i=1}^{s} \prod_{j=1}^{d} l_{ij}$$ each l_{ij} is a linear form (no constant term).

$\mathcal{B} = \left\{ \prod_{i=1}^{d} l_{ij} \ : \ l_{ij}\text{'s linear forms} \right\}$. $\qquad \partial_{x_1}(l_1 \cdots l_d)$

What is an exploitable weakness? $\qquad \alpha l_2 \cdots l_d + \alpha \cdot l_1 l_3 \cdots l_d \cdots$

Claim: $\Gamma_k(l_1 \cdots l_d) = \dim \underbrace{\partial^{=k}(l_1 \cdots l_d)}_{\subseteq \operatorname{span}\left\{ \prod_{i \in S} l_i \ : \ |S| = d-k \right\}} \le \binom{d}{k}$

$\therefore$ If $f = \sum_{i=1}^{s} l_{i1} \cdots l_{id}$ and f is "random", then

we can hope $\quad s \ge \dfrac{\min\left\{ \binom{n+k}{k}, \binom{n+d-k}{d-k} \right\}}{\binom{d}{k}}$.

Say $k = d/2$

Numerator. $\approx \binom{n}{d/2}$

Denom: $\approx 2^{d}$

$n = 100d$

Fact: If $f = ESYM_{n,d}$, $\dim \partial^{=k}(f) \approx$ wishful bound.

$\therefore$ Any hom. $\Sigma\Pi\Sigma$ ckt computing $ESYM_{n,d}$ requires size $\left(\frac{2n}{d}\right)^{d/2} / 2^d$.

On the other hand, there are non-hom. $\Sigma\Pi\Sigma$ cks for $ESYM_{n,d}$ of size just $O(n^2)$ [Ben-Or]

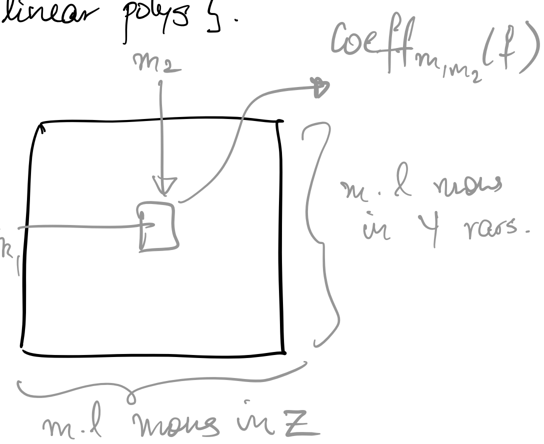
- Multilinear $\Sigma\Pi\Sigma$ circuits.

$$f = \sum_{i=1}^{s} l_{i1} \cdots l_{id}$$ where $l_{i1}, \ldots, l_{id}$ are variable disjoint.

$\mathcal{B} = \{ l_1 \cdots l_d$: variable disjoint linear polys $\}$.

$$\underbrace{l_1 \cdots}_{\text{Alice}} \Big| \underbrace{\cdots l_d}_{\text{Bob}}$$

$$X = Y \sqcup Z$$

$M_{Y,Z}(f) =$ 

Qn: Is $\mathrm{rank}(M(f))$ a good measure?

$(Y_1 + Z_1) \cdots (Y_d + Z_d)$

[Raz] If you choose the partition at random, then it is a good measure.

$$\text{rel-rank}_{Y,Z}(f) = \frac{\mathrm{rank}(M_{Y,Z}(f))}{\sqrt{\#\mathrm{Rows} \cdot \#\mathrm{Cols}}} = \frac{\mathrm{rank}(M_{Y,Z}(f))}{2^{(|Y|+|Z|)/2}}.$$

Fact: f, g var disjoint $\Rightarrow \underset{Y,Z}{\mathrm{relRank}}(fg) \le \underset{Y_1, Z_1}{\mathrm{relRank}}(f) \cdot \underset{Y_2, Z_2}{\mathrm{relRank}}(g)$

$\therefore M(fg) = M(f) \otimes M(g)$

$n = 2d$

What is $\underset{X = Y \sqcup Z}{\mathbb{E}}\left[\mathrm{relRank}(l_1 \cdots l_k)\right]$

$$\le \prod_{i=1}^{d/2} \mathbb{E}[\mathrm{relRank}(l_i)]$$

Suppose $\mathrm{Var}(l_i) \ge 3. \Rightarrow \mathrm{relRank}(l) \le \frac{2}{2^{3/2}} = \frac{1}{\sqrt{2}}$

$\underset{x_1 + x_2}{\mathrm{Var}(l_i) = 2}$. $\quad \mathbb{E}[\mathrm{relRank}] \le \frac{1}{2} \cdot \frac{2}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$

$$\therefore \mathbb{E}[\mathrm{relRank}(l_1 \cdots l_k)] \le \left(\frac{3}{4}\right)^{d/2}$$

For a random poly, we expect $M(f)$ to be full rank

$\therefore$ relRank (random poly) ≈ 1.

$\Rightarrow$ We have a $2^{\Omega(d)}$ lower bound.

[Raz] $n^{\Omega(\log n)}$ lb for ML formulas computing Det/Perm.

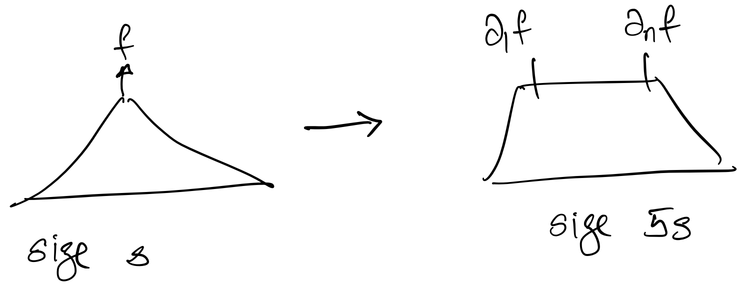
[CELS] Depth hierarchy thm for ML formulas.

Some lbs that <u>don't</u> follow this template:

- Baur-Strassen: $\Omega(n \log d)$ lb for general circuits computing $x_1^d + \cdots + x_n^d$.

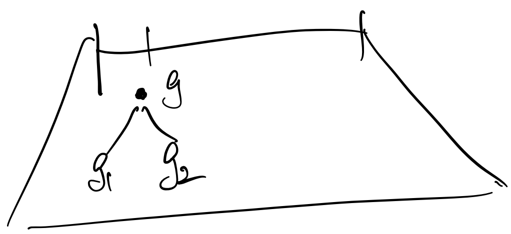
Sketch:

Lemma 1:



(Back propagation in the ML comm).

Can we fix an output $\bar{a} \in \mathbb{F}^n$ and ask how many input result in that output?



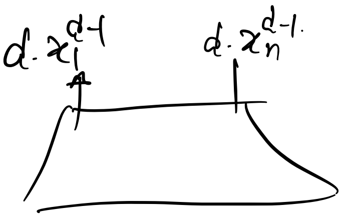
$g - g_1 g_2$ or $g - (g_1 + g_2)$

There is one quadratic eqn for each gate.

solns to s- quadratic eqns.

[Bezout] If this is finite, then $\leq 2^s$.

If $f = x_1^d + \cdots + x_n^d$,



$\bar{a} = (d, \ldots, d)$ How many inputs lead to this?

$\Rightarrow$ each x_i is a $(d-1)^{th}$- root of unity.

$\Rightarrow$ We have $(d-1)^n$ inputs.

$(d-1)^n \leq 2^s \Rightarrow s = \Omega(n \log d)$

- Determinental complexity LB. [Mignon-Ressayre]

$$f = \operatorname{Det}\begin{bmatrix} l_{11} & \cdots & l_{1m} \\ l_{m1} & \cdots & l_{mm} \end{bmatrix} \qquad l_i \in \mathbb{F}[x_1,\dots,x_n] \text{ linear.}$$

Want an LB for m.

$f(X) = \det(A(X))$ let X_0: $f(X_0) = 0$.

$$f(X+X_0) = \det(A(X+X_0)) = \det(\underbrace{L(X)}_{} + B)$$

$\hookrightarrow \det(B) = 0$ $\exists v: Bv = 0$.

Suppose X' s.t $L(X')v = 0$

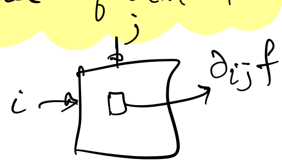
then $L(X') + B$ is also singular!

$\Rightarrow f(X' + X_0) = 0$

$V = \{X' : L(X') \cdot v = 0\} \to$ vector space of dim $\geq N - m$.

$$\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} v = 0$$

Claim: If $f(X_0) = 0$ and V is a vector space of dim $n - r$
$f(X_0 + X) = 0 \quad \forall X \in V$, then
$\operatorname{rank}(\operatorname{Hess}(f)(X_0)) \leq 2r$.



Claim: There is an $n \times n$ matrix X_0 s.t $\operatorname{Perm}(X_0) = 0$
and $\operatorname{rank}(\operatorname{Hess}(\operatorname{Perm})(X_0)) = n^2$.

Cor: $\mathrm{Det}_m(A(x)) = \mathrm{Perm}_n \Rightarrow m \geq \frac{n^2}{2}$.