PROOF COMPLEXITY AS A COMPUTATIONAL LENS

SW I

LECTURE 7

Today study $\boxed{\text{WIDTH}}$, i.e., the (max) size of clauses in resolution refutation.
Can help understand other fundamental proof complexity measures such as
- length (which we will see today)
- clause space (later)
- total space (later)

Need some convenient technical tools/notions

DEF (WEAKENING)
The weakening rule $\dfrac{B}{B \lor C}$
allows to derive weaker clause from stronger clause

PROPOSITION 1 Any resolution refutation $\pi : F \vdash \bot$ using weakening can be transformed to refutation $\pi' : F \vdash \bot$ without weakening in at most same length.

Proof sketch Given resolution refutation $\pi = (C_1, \ldots, C_L)$
Inductively build $\pi' = (C_1', \ldots, C_L')$
such that $C_i' \subseteq C_i$
π' possibly contains copies — remove later. And terminate π' first time $\bot$ is derived

<u>Base case</u> For $C_i \in F$, set $C_i' := C_i$

<u>Inductive step</u> If C_i derived from C_j by weakening, set $C_i' = C_j'$

If
$$\frac{C_j = C \vee x \qquad C_k = D \vee \overline{x}}{C_i = C \vee D}$$

have three cases

$x \notin C_j'$: Then set $C_i' = C_j'$

$\overline{x} \in C_k'$: Then set $C_i' = C_k'$,

Otherwise: Resolve C_j' and C_k' and let C_i' be resolvent

This proof shows that weakening rule also doesn't affect other cplx measures we will care about later

We still consider "resolution derivations" to <u>not</u> use weakening unless specified otherwise, but in view of Prop 1 can be relaxed about this

<u>DEF (RESTRICTION)</u>

Restriction ρ = partial truth value assignment.

Also represent $\rho = \{$ set of literals ℓ s.t. $\rho(\ell) = T \}$

$\boxed{\rho\text{-RESTRICTION}}$ of clause C is

$$C\restriction_\rho = \begin{cases} 1 & \text{if } \rho \cap Lit(C) \neq \emptyset \text{ i.e. } \rho \text{ satisfies } C \\ C \setminus \{\overline{\ell} \mid \ell \in \rho\} & \text{otherwise} \end{cases}$$

For CNF formula $F = \bigwedge_{i=1}^m C_i$ $\quad F\restriction_\rho = \bigwedge_{i=1}^m C_i\restriction_\rho$ (except 1-clauses removed)

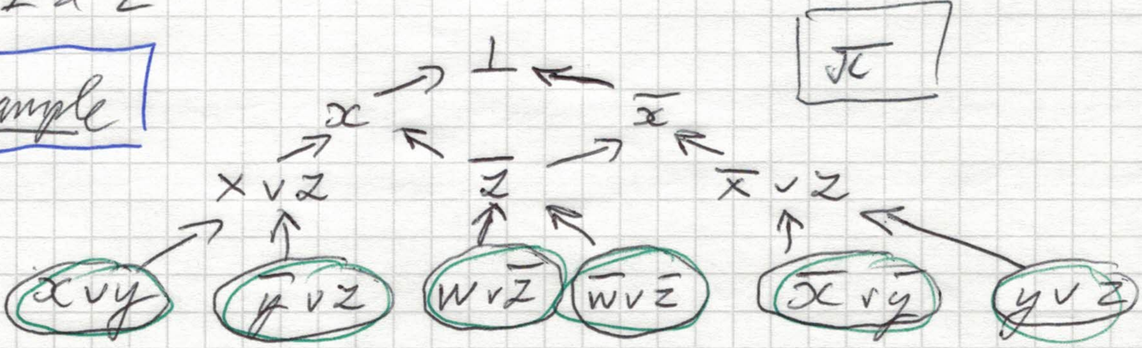
For derivation $\pi = (C_1, \dots, C_L)$ $\quad \pi\restriction_\rho = (C_1\restriction_\rho, \dots, C_L\restriction_\rho)$ (except 1-clauses removed)

PROPOSITION 2 If $\pi : F \vdash A$ resolution derivation and ρ restriction that does not satisfy A, then $\pi\restriction_\rho$ is a derivation of $A\restriction_\rho$ from $F\restriction_\rho$ (possibly using weakening) in at most same length

Same kind of inductive proof over resolution derivation as for Prop 1

Can then apply Prop 1 to get restricted refutation without weakening. In what follows, we might subconsciously use $\pi\restriction_\rho$ to refer to this cleaned up proof without weakening — OK because of Props 1 & 2

Proof by example

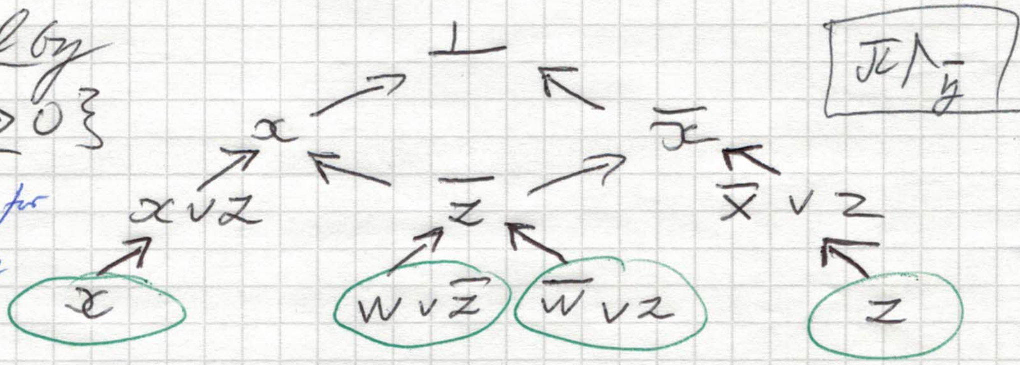


F

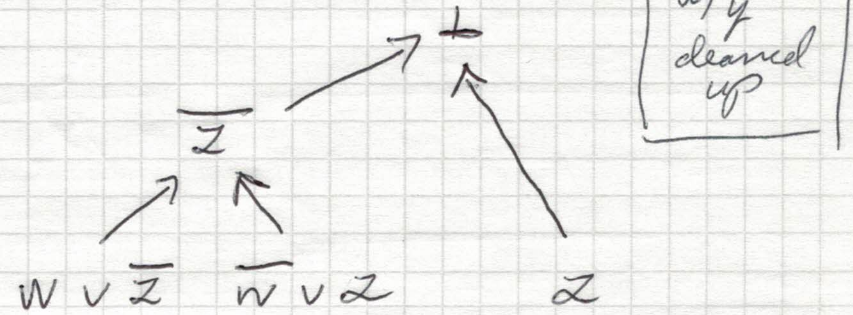
Restricted by $\rho = \{y \mapsto 0\}$

notation for simple restriction

$F\restriction_{\bar{y}}$



Cleaned up to remove weakening



DEF (WIDTH)

Width of a clause C $\quad W(C) = |C| = \#\text{literals in } C$

$$W(F) = \max_{C \in F} W(C)$$

$$W(\pi) = \max_{C \in \pi} W(C)$$

$$W(F \vdash A) = \min_{\pi : F \vdash A} \{W(\pi)\}$$

WIDTH OF DERIVING A FROM F

$W(F \vdash \perp)$ is the WIDTH OF REFUTING F

OBS 3 If F unsatisfiable CNF formula over n variables, then $W(F \vdash \perp) \leq n$

OBS 4 If $\pi : F \vdash \perp$ general resolution refutation for $|Vars(F)| = n$ and $W(\pi) = w$, then (w.l.o.g.)

$$L(\pi) = n^{O(w)}$$

Proof # ~~disnnect~~ clauses of width $\leq w$ is $\leq (2n+1)^w$

Is this simple counting bound right? Will see later: yes.

A narrow general resolution refutation is by necessity short

Today we want to prove:

$$\boxed{\text{If } \exists \text{ short resolution refutation of } F, \text{ then } \exists \text{ (somewhat) narrow refutation of } F}$$

Note that argument in proof for OBS 4 does not work for tree-like resolution (and is false)

LEMMA 5 If $W(F\restriction_x \vdash A) \leq w$, then $W(F \vdash A \vee \overline{x}) \leq \max\{w+1, W(F)\}$ possibly using weakening

Proof Suppose $\pi : F\restriction_x \vdash A$. Want to construct $\pi' : F \vdash A \vee \overline{x}$

First in π' list all clauses $C \in F$.

Then append $C_i \vee \overline{x}$ for all $C_i \in \pi$.

Width is clearly as claimed.

Straightforward to check by induction that π' valid ☒

LEMMA 6 If $W(F\restriction_x \vdash \bot) \leq w-1$
and $W(F\restriction_{\overline{x}} \vdash \bot) \leq w$
then $W(F \vdash \bot) \leq \max\{w, W(F)\}$

Proof Use Lemma 5 to derive $\overline{x}$ in width $\leq \max\{w, W(F)\}$.

Next resolve $\overline{x}$ with all $C \in F$ such that $x \in C$

This derives all clauses in $F\restriction_{\overline{x}} \setminus F$.

Finally refute $F\restriction_{\overline{x}}$ in width w. ☒

THEOREM 7 [BW01]

For tree-like resolution

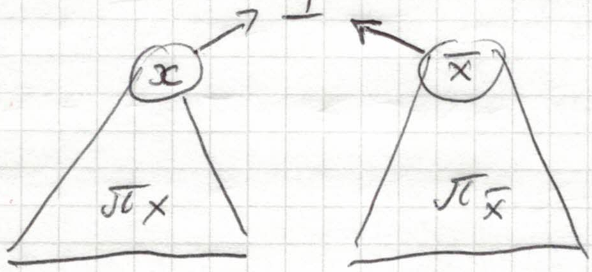
$$\boxed{W(F \vdash \bot) \leq W(F) + \log_2(L_{\tau}(F \vdash \bot))}$$

COROLLARY 8

For tree-like resolution

$$\boxed{L_{\tau}(F \vdash \bot) \geq 2^{(W(F \vdash \bot) - W(F))}}$$

SW VI



<u>Proof of Thm 7</u> Prove by induction that if
$L_T(F \vdash \bot) \leq 2^b$ then $W(F \vdash \bot) \leq W(F) + b$
Induction over b and #variables n

<u>Base cases:</u>

$b = 0$: Proof length $= 1 \Rightarrow \bot \in F$

$b = 1$: Formula is $x \wedge \bar{x} \Rightarrow$ Proof $x \rightarrow \bot \leftarrow \bar{x}$ has width 1

<u>Inductive step</u> Suppose $|Vars(F)| = n$
tree-like $\pi : F \vdash \bot$ in length $\leq 2^b$
Let x variable resolved over to get $\bot$
Fix $\pi_x : F \vdash x$ and $\pi_{\bar{x}} : F \vdash \bar{x}$
By tree-likeness $\boxed{L(\pi_x) + L(\pi_{\bar{x}}) + 1 = L(\pi)}$
Suppose w.l.o.g $L(\pi_{\bar{x}}) \leq 2^{b-1}$
Then by inductive hypothesis

$\pi_{\bar{x}}\restriction_x$ refutes $F\restriction_x$ in length $\leq 2^{b-1}$

so $W(F\restriction_x \vdash \bot) \leq W(F) + b - 1$

$\pi_x\restriction_{\bar{x}}$ refutes formula $F\restriction_{\bar{x}}$ with one fewer variable, so

$$W(F\restriction_{\bar{x}} \vdash \bot) \leq W(F) + b$$

Apply <u>Lemma 6</u> to get

$$W(F \vdash \bot) \leq W(F) + b$$

$\boxed{/\!/}$

Does this mean that "short proofs are narrow?" Not quite – every time we apply Lemma 6 to $\pi\restriction_x$ and $\pi\restriction_{\bar{x}}$ and then concatenate, proof length can double

So narrow proof can have grown exponentially in length

Such a blow-up cannot be avoided in the worst case [Ben-Sasson '09]

For general resolution, we get something more complicated

THEOREM 9 [BW01]

For unsatisfiable CNF formula F over n variables

$$\boxed{W(F \vdash \bot) \leq W(F) + \sqrt{8n \ln L_{\mathcal{R}}(F \vdash \bot)}}$$
$$= W(F) + O\left(\sqrt{n \log L_{\mathcal{R}}(F \vdash \bot)}\right)$$

This bound is geometric mean

$$\sqrt{\log(\text{worst case } 2^n) \cdot \log(\text{actual } L(F \vdash \bot))}$$

in case it helps to remember

COROLLARY 10

For unsatisfiable CNF formula F over n variables and general resolution, it holds that

$$\boxed{L_{\mathcal{R}}(F \vdash \bot) = \exp\left(\Omega\left(\frac{(W(F \vdash \bot) - W(F))^2}{n}\right)\right)}$$

SW VIII

Not so nice bound as for tree-like resolution

Focus on k-CNF formulas

Superpolynomial length lower bounds for

$W(F \vdash \perp) = \omega(\sqrt{n \log n})$

No implication below this

But Thm 9 & Cor 10 essentially tight

[Bonet & Galesi '01]

<u>Let us sketch proof</u>

Want to find good variable x, study $\pi\restriction_x$ & $\pi\restriction_{\bar{x}}$, and piece together with Lem 6.

Fix k-CNF formula F over n variables

Fix refutation $\pi : F \vdash \perp$ $\quad L(\pi) < 2^{n+1} \ll e^{2n}$

Set

$$\boxed{d = \sqrt{2n \ln L}} \qquad (*)$$

$$\boxed{a = \left(1 - \frac{d}{2n}\right)^{-1}} > 1 \qquad (**)$$

since $\ln L < 2n$

Say clause D $\boxed{\text{FAT}}$ if $W(D) \geq d$

Let

$$\boxed{fat(\pi)} = \left| \{ D \in \pi \mid W(D) \geq d \} \right|$$

Theorem 9 follows from next lemma

LEMMA 11 Let G be k-CNF formula over $n' \leq n$ variables and suppose $\pi' : G \vdash \perp$ is such that $fat(\pi') < a^b$ (with parameters as in $(*)$ and $(**)$ defined for F). Then

$$\boxed{W(G \vdash \perp) \leq k + d + b - 1}$$

Proof of Thm 9 assuming Lem 11

Observe $\quad k = W(F)$

$$d = \sqrt{2n \ln L}$$

so we only need to work on b.

Fix $\pi : F \vdash \bot$ of minimal length $L_R(F \vdash \bot) = L$

$$fat(\pi) < L \le a^{\lceil \log_a L \rceil} \le a^{\log_a L + 1}$$

$$\boxed{\log_a x = \frac{\ln x}{\ln a}}$$

Set $b = \boxed{\log_a L + 1} = \boxed{1 + \frac{\ln L}{\ln a}}$

$$\ln a = \ln\left(\left(1 - \frac{d}{2n}\right)^{-1}\right) = -\ln\left(1 - \frac{d}{2n}\right) \ge \frac{d}{2n}$$

using $\boxed{\ln(1+x) \le x \text{ for } x > -1}$

By (*) $\frac{d}{2n} = \sqrt{\frac{\ln L}{2n}}$ so

$$\boxed{\frac{\ln L}{\ln a} \le \sqrt{2n \ln L}}$$

So $k + d + b - 1 \le W(F) + 2\sqrt{2n \ln L}$ $\boxed{\text{Vi}}$

Proof sketch for Lemma 11

Nested induction over b and #variables n'

Base cases a bit convoluted if we want

$$W(G \vdash \bot) - W(G) \le d + b - 1$$

exactly. Let's ignore them.

Induction step

Consider $\pi : G \vdash \bot$ with $fat(\pi) < a^b$

G has $2n'$ distinct literals all in all

At least $d \cdot fat(\pi)$ literals in fat clauses

So some literal appears in at least

$$\frac{d}{2n'} fat(\pi) \geq \frac{d}{2n} fat(\pi)$$

fat clauses. Suppose w.l.o.g that x is such a literal. Then setting $x = 1$ makes these fat clauses go away and

$$\pi\restriction_x : G\restriction_x \vdash \bot$$

has less than

$$\left(1 - \frac{d}{2n}\right) fat(\pi) \leq \left(1 - \frac{d}{2n}\right) a^b = a^{b-1}$$

fat clauses. By induction hypothesis

$$W(G\restriction_x \vdash \bot) \leq k + d + b - 2$$

$G\restriction_{\bar{x}}$ has one fewer variable, so by IH

$$W(G\restriction_{\bar{x}} \vdash \bot) \leq k + d + b - 1$$

Apply Lemma 6 to get

$$W(G \vdash \bot) \leq k + d + b - 1$$

Again the narrow proof obtained can be exponentially longer. That minimizing width can cause exponential increase in length in worst case in unavoidable [Thapen '16].

Most resolution length lower bounds can be proven via Ben-Sasson & Wigderson size-width relation

- Prove strong lower bound on $W(F \vdash \bot)$
- Use $\boxed{L(F \vdash) = \exp\left(\Omega\left(\frac{(W(F \vdash \bot) - W(F))^2}{|Vars(F)|}\right)\right)}$

Width lower bounds (and also lower bounds on other cplx measures) can be proven via graph expansion

High-level idea: Given CNF formula F over variables X, create (bipartite) CLAUSE-VARIABLE INCIDENCE GRAPH (CVIG) with $G(F)$

- Left vertex set = clauses C
- Right vertex set = variables x
- Edge between C and x if x or $\bar{x}$ occurs in C

Want to show that if $G(F)$ well-connected, then F hard

DEF (BIPARTITE BOUNDARY EXPANDER)
Bipartite graph $G = (U \dot\cup V, E)$ is a bipartite (s, δ)-BOUNDARY EXPANDER if $\forall U' \subseteq U, |U'| \leq s$, it holds that $|\partial(U')| \geq \delta |U'|$, where the BOUNDARY $\partial(U') = \{ v \in V : |N(v) \cap U'| = 1 \}$ consists of all vertices on right-hand side V with unique neighbour in U' on left-hand side

Problem: Often there is higher level structure that is not exactly captured by CNF encoding

Ex For Tseitin formula, think of clauses

$$\left\{\begin{array}{l} x \vee y \vee z \\ x \vee \overline{y} \vee \overline{z} \\ \overline{x} \vee y \vee \overline{z} \\ \overline{x} \vee \overline{y} \vee z \end{array}\right\}$$ as single constraint $x \oplus y \oplus z = 1$

Ex For PHP, we only studied assignments corresponding to matchings, so hole axioms $\overline{x}_{i,j} \vee \overline{x}_{i',j}$ never really played a part in the argument

Want to allow "generalized CVIG" that captures combinatorial structure behind CNF formula

Given formula F over variables V, we choose

(1) Partition $F = E \cup \bigcup_{i=1}^{m} F_i$

E and F_i are disjoint sets of clauses that should be satisfiable

(2) Division $V = \bigcup_{i=1}^{n} V_i$

V_i and V_j need not be disjoint

$\boxed{\text{OVERLAP}}$ $ol(V) = \max_x |\{V_i \in V : x \in V_i\}|$

Overlap will be small constant for us

Today it will be $= 1$

Generalized CVIG for $(\mathcal{F}, \mathcal{V})_{\mathcal{E}}$

- Left vertex set $\mathcal{U} = \{F_1, F_2, \ldots, F_m\}$
- Right vertex set $\mathcal{V} = \{V_1, V_2, \ldots, V_n\}$
- Edge (F_i, V_j) if $Vars(F_i) \cap V_j \neq \emptyset$

Special clause set $\mathcal{E}$ not part of graph, but "filters" with truth value assignments used in lower bound arguments

Describe two-person game between Adversary and Satisfier played on $(\mathcal{F}, \mathcal{V})_{\mathcal{E}}$

Good Satisfier strategy = resolution width lower bound

DEF (RESOLUTION EDGE GAME)

Fix $(\mathcal{F}, \mathcal{V})_{\mathcal{E}}$-graph for CNF formula F over variables $\mathcal{V}$. $\boxed{\text{RESOLUTION EDGE GAME}}$ on edge (F_i, V_j) with respect to filtering set $\mathcal{E}$:

(1) Adversary chooses total assignment α s.t. $\alpha(\mathcal{E}) = \top$

(2) Satisfier modifies α on V_j to get α'

(3) Satisfier wins if $\alpha'(F_i \wedge \mathcal{E}) = \top$

Satisfier wins resolution edge game on (F_i, V_j) if has winning strategy

Satisfier wins resolution edge game on $(\mathcal{F}, \mathcal{V})_{\mathcal{E}}$ if wins on all edges in graph

DEF (RESOLUTION EXPANDER)

$(\mathcal{F}, \mathcal{V})_{\mathcal{E}}$-graph is $\boxed{(s, \delta, \mathcal{E})\text{- RESOLUTION EXPANDER}}$ if

- $(\mathcal{F}, \mathcal{V})_{\mathcal{E}}$ graph is (s, δ) - boundary expander
- Satisfier wins resolution edge game on $(\mathcal{F}, \mathcal{V})_{\mathcal{E}}$.

Note expansion measured in terms of vertices in $(\mathcal{F}, \mathcal{V})_{\mathcal{E}}$-graph. Every vertex can correspond to many clauses or variables

$\forall \mathcal{F}' \subseteq \mathcal{F} \setminus \{\mathcal{E}\}$ s.t. $|\mathcal{F}'| = \#\text{subsets } F_i \in \mathcal{F}' \leq s$

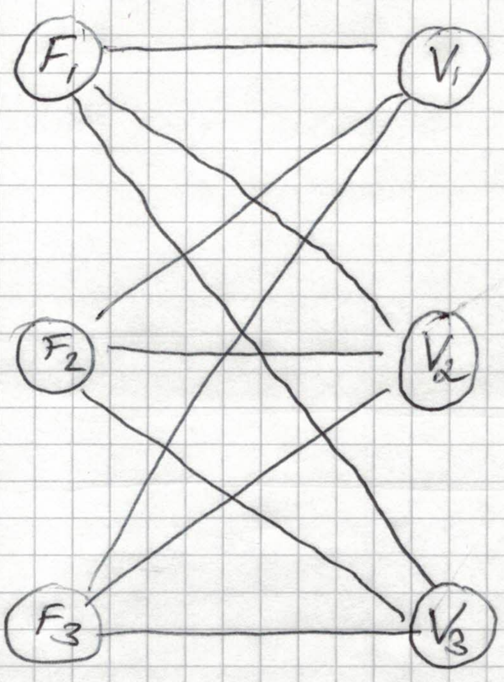
it holds that $|\partial(\mathcal{F}')| = \#\text{unique neighbors } V_j \geq \delta |\mathcal{F}'|$

Ex

$w \vee x \vee y$
$\overline{u} \vee x$
$u \vee \overline{x}$

$x \vee \overline{y}$
$x \vee \overline{z}$
$\overline{x} \vee y \vee z$

$x \vee y$
$x \vee z$
$\overline{x} \vee \overline{y} \vee \overline{z}$



$\{w, y\}$

$\{x, y\}$

$\{u, w, z\}$

$$\mathcal{E} = \{\overline{y} \vee z\}$$

Edge (F_1, V_1) Adversary wins. Play any total assignment α_1 extending $\rho_1 = \{u \mapsto \bot, x \mapsto \top, y \mapsto \bot\}$
Flipping w or y cannot satisfy $u \vee \overline{x}$

Edge (F_1, V_2) Adversary again wins by playing any extension α_2 of $\rho_2 = \{u \mapsto \bot, w \mapsto \bot, y \mapsto \bot, z \mapsto \bot\}$
Satisfier cannot flip y because of $\mathcal{E}$
But $F_2 \restriction_{\{\overline{y}, \overline{z}\}} = x \wedge \overline{x}$

Edge (F_2, V_2) Satisfier wins. Given any α_3 s.t. $\alpha_3(\mathcal{E}) = T$, set $\alpha'(x) = \alpha_3(y \vee z)$ and leave other variables unchanged

Say that CNF formula F over variables V ADMITS AN $(s, \delta, \mathcal{E})$-RESOLUTION EXPANDER if there is a partition of the clauses $F = \mathcal{E} \cup \bigcup_{j=1}^{m} F_j$ and division of the variables $V = \bigcup_{j=1}^{n} V_j$ such that the resulting $(F, V)_{\mathcal{E}}$-graph is an $(s, \delta, \mathcal{E})$-resolution expander.

Then this is sufficient to prove resolution width lower bounds

THEOREM (ESSENTIALLY [BW01], BUT FORMULATED AS IN [MN24])
If a CNF formula F admits an $(s, \delta, \mathcal{E})$-resolution expander with overlap ℓ, then $\boxed{W(F \vdash \bot) > \frac{\delta s}{2\ell}}$

What we want to do now:

(a) Prove this theorem

(b) Show how PHP and Tseitin lower bounds immediately follow

For the proof, define "progress measure" $\mu : \{\text{clauses}\} \to \mathbb{N}_0$ that assesses how hard it is to derive a clause

PROPERTY A: For axiom clause $A \in F$ $\mu(A) = O(1)$

PROPERTY B: For $\frac{C \vee x \quad D \vee \bar{x}}{C \vee D}$ $\boxed{\mu(C \vee D) \leq \mu(C \vee x) + \mu(D \vee \bar{x})}$

PROPERTY C: For empty clause $\bot$ $\mu(\bot) > s$.

From this, immediately get:

CLAIM D

In any resolution refutation $\pi : F \vdash \bot$ there is a clause $C \in \pi$ with $\mu(C) \in (s/2, s]$

The theorem then follows from combining this with:

CLAIM E

Any clause C with $\mu(C) = \sigma \leq s$ has width $W(C) \geq \delta\sigma / \ell$

Now, let's fill in details

Sets of clauses $\mathcal{F}$ imply clause C, $\mathcal{F} \vDash C$, if $\alpha(\mathcal{F}) = T$ implies $\alpha(C) = T$

Given $(\mathcal{F}, V)_E$-graph, define

$$\boxed{\mu(C) = \min\{ |\mathcal{F}'| : \textstyle\bigwedge_{F \in \mathcal{F}'} F \wedge E \vDash C \}}$$

i.e., size of smallest collection of clause sets F_i on left-hand side that together with E imply C

Let us verify properties

(A) $A \in F$ means $A \in E \Rightarrow \mu(A) = 0$

$A \in F_i \Rightarrow \mu(A) \leq 1$

(B) Suppose $\mathcal{F}_1 \wedge E \vDash C \vee x$

$\mathcal{F}_2 \wedge E \vDash D \vee \overline{x}$

Then $(\mathcal{F}_1 \wedge \mathcal{F}_2) \wedge E \vDash C \vee D$

so $\mu(C \vee D) \leq \mu(C \vee x) + \mu(D \vee \overline{x})$

(c) Suppose $\mu(\bot) \leq s$. Then $\exists \mathcal{F}' \subseteq \mathcal{F}, |\mathcal{F}'| \leq s$,

such that $\bigwedge_{F \in \mathcal{F}'} F \wedge E \vDash \bot$, i.e.,
$\bigwedge_{F \in \mathcal{F}'} F \wedge E$ is unsatisfiable. We show for any $\mathcal{F}' \subseteq \mathcal{F}$, $|\mathcal{F}'| = s$, that $\bigwedge_{F \in \mathcal{F}'} F \wedge E$ is satisfiable, so $\mu(\bot) > s$. Fix any such $\mathcal{F}'$, $|\mathcal{F}'| = s$

By expansion $|\partial(\mathcal{F}')| \geq \delta |\mathcal{F}'| > 0$.
i.e., $\exists$ right vertex set V_s and left vertex set F_s (possibly after relabelling) such that
V_s unique neighbour of F_s, i.e.

$$V_s \in N(F_s) \setminus N(\textstyle\bigcup_{j=1}^{s-1} F_j)$$

By the same argument, find

$$V_{s-1} \in N(F_{s-1}) \setminus N(\textstyle\bigcup_{j=1}^{s-2} F_j)$$

$$V_{s-2} \in N(F_{s-2}) \setminus N(\textstyle\bigcup_{j=1}^{s-3} F_j)$$

etcetera — referred to as PEELING ARGUMENT
For $i = 1, 2, \ldots, s$

$$\boxed{V_i \in N(F_i) \setminus N(\textstyle\bigcup_{j=1}^{i-1} F_j)}$$

V_i is not a neighbour of any F_j for $j < i$.

Now take any α_0 s.t. $\alpha_0(E) = T$ (which exists, since E should be satisfiable. Play Satisfier strategy on $F_1 \leftrightarrow V_1$ if needed to get α_1 s.t. $\alpha_1(F_1 \wedge E) = T$.
Now play Satisfier strategy on $F_2 \leftrightarrow V_2$ if needed to get α_2 s.t. $\alpha_2(F_2 \wedge E_2) = T$. But note that $V_2 \cap Vars(F_1) = \emptyset$, so $\alpha_2(F_1) = \alpha_1(F_1) = T$ and $\alpha_2(F_1 \wedge F_2 \wedge E) = T$

By induction, for $F_i \longleftrightarrow V_i$ have α_{i-1} s.t. $\alpha_{i-1}\left(\bigwedge_{j=1}^{i-1} F_j \wedge E \right) = \top$ and playing Satisfier strategy on $F_i \longleftrightarrow V_i$ yields α_i satisfying $\bigwedge_{j=1}^{i} F_j \wedge E$. This shows that $\bigwedge_{j=1}^{s} F_j \wedge E$ satisfiable as claimed.

It remains to prove Claim E, i.e., that $\mu(C) = \sigma \leq s$ implies $W(C) \geq \delta\sigma / \ell$.

Fix $\mathcal{F}_C \subseteq \mathcal{F}$ witnessing that $\mu(C) = |\mathcal{F}_C| = \sigma$

Then $\bigwedge_{F \in \mathcal{F}_C} \wedge E \vDash C$ $\quad (\dagger)$

but for all $\mathcal{F}' \subsetneq \mathcal{F}_C$ it holds that

$$\bigwedge_{F \in \mathcal{F}'} \wedge E \nvDash C \qquad (\ddagger)$$

Claim E now follows from another claim

CLAIM F For all $V \in \partial(\mathcal{F}_C)$ it holds that $V \cap Vars(C) \neq \emptyset$

Given claim F, since every variable x occurs in at most ℓ sets V, we get $W(C) \geq |\partial(\mathcal{F}_C)| / \ell \geq$ $\geq \delta\sigma / \ell$

Proof of Claim F Fix $V \in \partial(\mathcal{F}_C)$ and its unique neighbour $F_V \in \mathcal{F}_C$. By $(\ddagger)$ there exists α s.t. $\alpha\left(\bigwedge_{F \in \mathcal{F}_C \setminus \{F_V\}} F \wedge E \right) = \top$ but $\alpha(C) = \bot$

Satisfier strategy on $F_V \longleftrightarrow V$ yields α' such that $\alpha'\left(\bigwedge_{F \in \mathcal{F}_C} \wedge E \right) = \top$. If $V \cap Vars(C) = \emptyset$, then $\alpha'(C) = \bot$, but that contradicts $(\dagger)$.

WLB IX

Now we can get width lower bounds (and hence length lower bounds) for Tseitin formulas and pigeonhole principle formulas

TSEITIN FORMULAS

Bounded-degree undirected graph $G = (V, E)$

Charge function $\chi : V \to \{0,1\}$ such that $\sum_{v \in V} \chi(v)$ odd number

$PARITY_{v,\chi}$ = Clauses encoding $\sum_{e \ni v} x_e \equiv \chi(v) \pmod 2$

$Ts(G, \chi) = \bigwedge_{v \in V} PARITY_{v,\chi}$

Modifying our notion of edge expander slightly (only makes them easier to find)

<u>DEF</u> G is an (s, δ)-EDGE EXPANDER if $\forall V' \subseteq V$, $|V'| \le s$ it holds that $|E(V', V \setminus V')| \ge \delta |V'|$, where $E(U, W) = \{(u,w) \in E \mid u \in U, w \in W\}$

<u>THEOREM</u> If G is an (s, δ)-edge expander and χ is odd, then $W(Ts(G, \chi) \vdash \bot) > \delta s / 2$.

<u>Proof sketch</u> Left vertex sets $F_v := PARITY_{v,\chi}$. $E = \emptyset$. Every variable x_e, $e \in E$ forms singleton set $V_e = \{x_e\}$, so overlap $ol(\mathcal{V}) = 1$

If G is (s, δ)-edge expander, then $(\mathcal{F}, \mathcal{V})_E$ is resolution expander with same parameters — $\partial(\mathcal{F}')$ are precisely edges leaving vertex set. Satisfier can win on any $F_v \leftrightarrow x_e$ by flipping x_e to satisfy parity $\square$

$\boxed{\text{PIGEONHOLE PRINCIPLE FORMULAS}}$

Define formulas over general bipartite graphs
$G = (U \,\dot\cup\, V, E)$

PIGEON AXIOMS $\quad \underbrace{\bigvee_{v \in N(u)} x_{u,v}} \quad u \in U$

HOLE AXIOMS $\quad \underline{\overline{x}_{u,v} \vee \overline{x}_{u',v}} \quad v \in V,\ u \neq u' \in N(v)$

FUNCTIONALITY AXIOMS $\quad \underline{\overline{x}_{u,v} \vee \overline{x}_{u,v'}} \quad u \in U,\ v \neq v' \in N(u)$

ONTO AXIOMS $\quad \underbrace{\bigvee_{u \in N(v)} x_{u,v}} \quad v \in V$

Conjunction of all clauses yields formula $\boxed{\text{Onto-FPHP}(G)}$

PHP formulas from before are Onto-FPHP$(K_{n+1,n})$

Let $\rho_G = \{ x_{u,v} \mapsto \bot \mid u \in U, v \in V, (u,v) \notin E \}$

Then $\underbrace{\text{Onto-FPHP}(K_{n+1,n})}\restriction_{\rho_G} = \text{Onto-FPHP}(G)$

So sufficient to prove lower bounds for constant-left-degree graphs — #variables $\Theta(n)$

- For every pigeon u, pick uniformly and independently at random d holes for d suitable constant (≈ 5 or 50)
- For any isolated hole v, pick one pigeon u_v

Yields bipartite (s, δ)-boundary expander for $\delta > 0$ and $s = \Omega(n)$ with matching from V to U

THEOREM If $G = (U \dot\cup V, E)$ is bipartite (s, δ)-boundary expander with full matching of V into U*, then $W(\text{Onto-FPHP}(G) \vdash \bot) > \delta s/2$

(*) Condition can be significantly relaxed, but we keep it for purposes of exposition

Proof sketch Construct $(\mathcal{F}, \mathcal{V})_{\mathcal{E}}$-graph as follows

- Left sets $F_u = \{ \bigvee_{v \in N(u)} x_{u,v} \}$
- right sets $V_v = \{ x_{u,v} \mid u \in N(v) \}$
  so $\mathrm{d}(\mathcal{V}) = 1$
- filtering set $\mathcal{E}$ = all hole axioms, functionality axioms, and onto axioms

If G is (s, δ)-boundary expander, then $(\mathcal{F}, \mathcal{V})_{\mathcal{E}}$ resolution expander with same parameters.
Graphs are easily verified to be isomorphic

Suppose Adversary and Satisfier play on edge (F_u, V_v). Then α is full matching of V into U, since filtering set satisfied.
If $\alpha(F_u) = T$, Satisfier does nothing
Otherwise, change α on V_v so that
$\alpha(x_{u,v}) = T$ $\alpha(x_{u',v}) = \bot$ for $u' \neq u$.
This is new full matching of V into U, so filtering set satisfied ☒

Since Onto-FPHP(G) has $\Theta(n)$ variables if G has constant left degree, linear width lower bound $\Longrightarrow$ exponential length lower bound

By restriction argument, exponential lower bound holds for $K_{n+1,n}$ for all PHP flavours for resolution.

Most resolution length lower bounds can be established with size-width relation — but not all

Note that for k-clique formulas the resolution width is k, so get nothing for range of parameters we are interested in ($\ll \sqrt{n}$)