PIGEONHOLE PRINCIPLE

$m > n$ pigeons don't fit into n holes if each pigeon wants its own hole

Encode (opposite of) this statement in CNF

$x_{i,j} \Leftrightarrow$ "pigeon i flies to hole j"

$i \in [m] = \{1, 2, \ldots, m\}$

$j \in [n] = \{1, 2, \ldots, n\}$

Pigeon axioms $\boxed{P^i \doteq \bigvee_{j=1}^{n} x_{i,j}}$ $\quad i \in [m]$

Hole axioms $\boxed{H_j^{i,i'} \doteq \overline{x}_{i,j} \vee \overline{x}_{i',j}}$ $\quad i < i' \in [m]$, $j \in [n]$

PIGEONHOLE PRINCIPLE FORMULA

$$\boxed{PHP_n^m \doteq \bigwedge_{i=1}^{m} P^i \bigwedge_{i'=2}^{m} \bigwedge_{i=1}^{i'-1} \bigwedge_{j=1}^{n} H_j^{i,i'}}$$

Can also add

Functionality axioms $\boxed{F_{j,j'}^{i} \doteq \overline{x}_{i,j} \vee \overline{x}_{i,j'}}$ $\quad i \in [m]$, $j < j' \in [n]$

Onto axioms $\boxed{S_j \doteq \bigvee_{i=1}^{m} x_{i,j}}$ $\quad j \in [n]$

Today focus on "vanilla PHP formula"

PHP formulas intensely studied

Entire survey [Razborov '02]

(not fully up-to-date, but very readable)

Different flavours of PHP in different parameter regimes $m > n$ studied for different proof systems & measures

Today focus on $\boxed{\text{resolution size}}$

[Haken '85] Lower bound $\exp(\Omega(n))$

[Buss, Pitassi '97] Upper bound $\exp(O(n))$

For $m = \exp(\Omega(\sqrt{n \log n}))$ upper bound $\exp(O(\sqrt{n \log n}))$

For any $m > n$, sequence of works [Raz '04, Razborov '01, '03, '04] establishing lower bound $\exp(\Omega(\sqrt[3]{n}))$

WHAT IS THE RIGHT BOUND?

Can also study

- graph PHP — each pigeon has restricted set of holes as specified by bipartite graph
- other encodings like binary PHP formulas (hole for pigeon encoded as $\lceil \log(n+1) \rceil$ bits

$\boxed{\text{"Weak pigeonhole principle"}}$: m "large", say $m = \Omega(n^2)$

Hardness open for polynomial calculus

Known proof techniques break

Today we will prove

THEOREM 1 [Haken '85]
Resolution refutations of PHP_n^{n+1} require resolution length $\exp(\Omega(n))$

In terms of formula size $N = \Theta(n^3)$ get lower bound $\exp(\Omega(\sqrt[3]{N}))$

For any CNF formula of size N
$\exists$ resolution refutations of length $\exp(O(N))$
We will see tight lower bounds in later lectures

Follow exposition in [Pudlák '00]
Model resolution as Prosecutor-Defendant game played on unsatisfiable CNF formula F

Prosecutor asks about values of variables, or forgets variable values
Defendant answers to questions
Prosecutor has record = partial truth value assignment
Game of full information: Defendant knows current record. Doesn't need to answer consistently if asked about forgotten variable.
Prosecutor wins when record falsifies axiom clause in F

Formally, game played in rounds.
Start with record $R_{init} = \emptyset$
In each round, depending on R
Prosecutor does one of the following

$\boxed{\text{ASK}}$ Ask about variable not in R
Defendant answers $x = b \in \{0,1\}$
New record $R \cup \{x = b\}$

$\boxed{\text{FORGET}}$ Prosecutor shrinks R to $R' \subsetneq R$

Winning position for Prosecutor:
R falsifies clause in F

Prosecutor strategy: Collection of records and associated moves that lead to winning position regardless of how Defendant plays

Complexity measures for strategy

- \# records
- max size of record
- max \# rounds of play

LEMMA 2 Let F be an unsatisfiable CNF formula.

(a) If there is a resolution refutation π in length L, width w, and depth d, then there is a prosecutor strategy with

$\leq 3L$ records

every record has size $\leq w+1$

rounds required is $\leq 2d$

(b) If there is a Prosecutor strategy of size S with records of size $\leq k$ requiring at most r rounds of play, then there is a resolution refutation in length $\leq S$, width $\leq k$, and depth $\leq r$.

Proof sketch We will only need (a), and leave (b) as exercise

Look at DAG G_π representing π

Maintain record falsifying current clause. Start with $\emptyset$ falsifying $\bot$

Clause C derived by resolving $C_1 \vee x$ and $C_2 \vee \overline{x}$

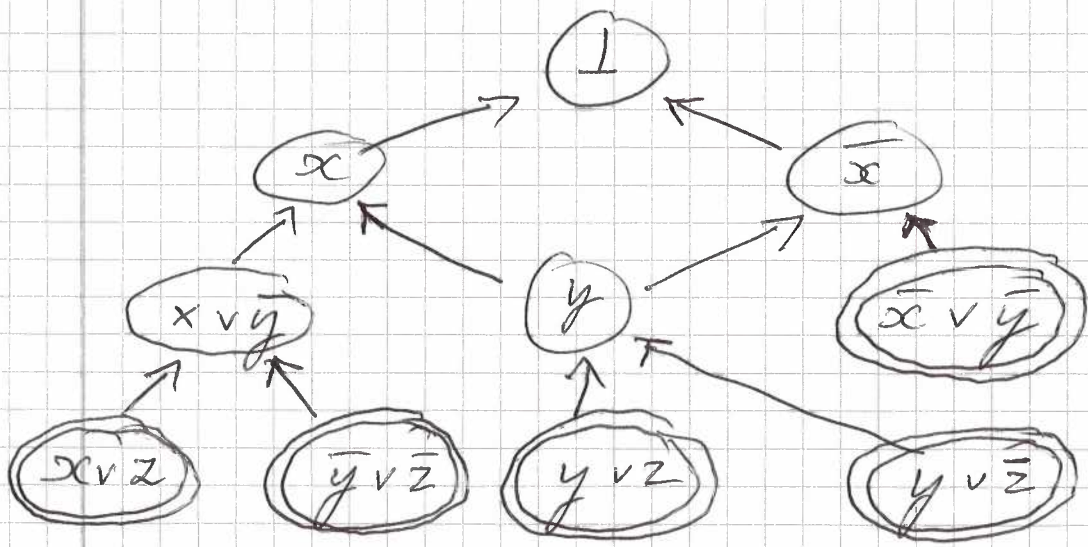
Prosecutor asks about x

Moves to falsified clause

Forgets to get minimal falsifying assignment. $\square$

Example

$$F = (x \vee z) \wedge (\bar{x} \vee \bar{y}) \wedge (y \vee z) \wedge (y \vee \bar{z}) \wedge (\bar{y} \vee \bar{z})$$



Note

Tree-like resolution $\iff$ no Forget moves

It follows from Lemma 2 that any unsatisfiable CNF formula F over n variables has tree-like resolution refutation in

- size $\exp(O(n))$
- depth $O(n)$

In view of Lemma 2, sufficient to prove that there is no small Prosecutor strategy for PHP_n^{n+1}

LEMMA 3

There is a $\delta > 0$ such that for large enough $n \in \mathbb{N}^+$ any Prosecutor strategy requires $\geq 2^{\delta n}$ records

Prove this by exhibiting Defendant strategy that forces Prosecutor to have many records.

Use randomization

DEFENDANT STRATEGY

Choose $n/4$ pigeons and assign to $n/4$ holes uniformly at random (assume for simplicity $4 \mid n$)

Let $\mathcal{M}_{init}$ be this partial matching

For any partial matching $\mathcal{M}$, can identify it with partial truth value assignment $\rho_{\mathcal{M}}$ by

$$\rho_{\mathcal{M}}(x_{i,j}) = \begin{cases} 1 & \text{if } (i,j) \in \mathcal{M} \\ 0 & \text{if } (i,j') \in \mathcal{M} \text{ for } j' \neq j \\ * & \text{otherwise} \end{cases}$$

$*$ means unassigned

From now on, identify matchings $\mathcal{M}$ and assignments $\rho_{\mathcal{M}}$

Defendant always maintains $\boxed{\mathcal{M} \supseteq \mathcal{M}_{init}}$
and $\rho_{\mathcal{M}} \supseteq \rho_{init} = \rho_{\mathcal{M}_{init}}$

Hole j $\boxed{\text{ASSIGNED}}$ to pigeon i if $x_{i,j} = 1$ on record

Hole j $\boxed{\text{PROHIBITED}}$ for pigeon i if Prosecutor has $x_{i,j} = 0$ on record

How Defendant answers question about $x_{i,j}$

(1) If $i \in \mathrm{Dom}(\mathcal{M})$
answer $\rho_{\mathcal{M}}(x_{i,j})$

(2) If $i \notin \mathrm{Dom}(\mathcal{M})$
answer $x_{i,j} = 0$

Then look at # prohibited holes for pigeon i

If $\geq n/2$, choose smallest j^* that is not prohibited and is consistent with $\mathcal{M}$ and extend $\mathcal{M}$ to $\mathcal{M} \cup \{i \mapsto j^*\}$

If this is not possible, then give up

If Prosecutor forgets

If pigeon i ~~now~~ $\in \mathrm{Dom}(\mathcal{M})$ is not assigned to hole and has less than $n/2$ prohibited holes then REMOVE i from $\mathcal{M}$.

Say that pigeon i is $\boxed{\text{THOROUGHLY EXAMINED}}$ in Prosecutor record R if R contains either

(a) $x_{i,j} = 1$, i.e., pigeon i assigned to hole; or

(b) $x_{i,j} = 0$ for $\geq n/2$ j's, i.e., at least $n/2$ forbidden holes

LEMMA 4

Before Prosecutor wins, there will be a record with $\geq n/4$ thoroughly examined pigeons.

Let us call such a record $\boxed{\text{INFORMATIVE}}$

Proof As long as Defendant follows strategy cannot lose. Hence update of $\mathcal{M}$ in (2) must fail

I.e., some pigeon i^* has reached $n/2$ prohibited holes, but there is no available hole j^*.

This means $|\mathrm{Dom}(\mathcal{M})| \geq n/2$

so $|\mathrm{Dom}(\mathcal{M}) \setminus \mathrm{Dom}(\mathcal{M}_{init})| \geq n/4$

But $i \in \mathrm{Dom}(\mathcal{M}) \setminus \mathrm{Dom}(\mathcal{M}_{init})$ only if (a) or (b) above holds $\square$

Want to prove that Prosecutor strategy must contain $\geq 2^{\delta n}$ distinct informative records

Prove that for fixed informative record R that probability of reaching R is exponentially small

In fact, prove something slightly stronger

$$\boxed{\Pr_{\mathcal{M}_{init}}[R \text{ consistent with } \mathcal{M}_{init}] \leq 2^{-\delta n}}$$

But we know that for every $\mathcal{M}$ the play reaches some informative record R with 100% probability. So

$$1 = \Pr_{\mathcal{M}}[\exists \text{ informative } R \text{ consistent with } \mathcal{M}]$$

$$\leq \sum_{\substack{\text{informative} \\ R}} \Pr_{\mathcal{M}}[R \text{ consistent with } \mathcal{M}]$$

$$\leq 2^{-\delta n} \cdot (\# \text{ informative } R)$$

and there must be $\geq 2^{\delta n}$ informative records

Fix some informative record R

Let $\boxed{I_R = \{\text{thoroughly investigated pigeons in } R\}}$

$$|I_R| \geq n/4$$

For random M_{init} the expected size of the intersection is

$$\mathop{E}_{M}\left[\,|I_R \cap Dom(M_{init})|\,\right] \geq n/16$$

by linearity of expectation.

By concentration of measure, we have

$$\boxed{|I_R \cap Dom(M_{init})| \geq n/32}$$

except with exponentially small probability

Intuitive argument:

Pick pigeons in $M = M_{init}$ one by one
Every time chance of picking pigeon in I_R is $\approx 1/4$.
This experiment is performed $n/4$ times
Roughly flipping coins with probability $1/4$ of heads. Actual # heads will be sharply concentrated around the expected number

Formal argument

Trim I_R to size exactly $n/4$. Then

$$\Pr_{\mathcal{M}}\left[|I_R \cap \mathrm{Dom}(\mathcal{M})| \leq n/32 \right] =$$

$$= \frac{\sum_{i=0}^{n/32} \binom{n/4}{i} \binom{n+1-n/4}{n/4-i}}{\binom{n+1}{n/4}} \leq$$

$$\leq \quad \ldots \text{ calculations } \ldots \quad \leq$$

$$\leq 2^{-\delta' n}$$

for some $\delta' > 0$

Suppose $|I_R \cap \mathcal{M}| > n/32$

For every $i \in I_R$ it holds that R either
(a) specifies a hole j^*; or
(b) prohibits $n/2$ holes j

To be consistent, $\mathcal{M}$ must comply with these restrictions

Choose pigeons in $I_R \cap \mathcal{M}$ one by one

For $(i+1)$ st pigeon

(a) Probability $\leq \frac{1}{n-i}$ to hit exactly right hole

(b) Probability $\leq \frac{n/2}{n-i}$ to avoid prohibited hole

$i \leq n/32 \Rightarrow$ Probabilities in (a) & (b) $\boxed{\ll 2/3}$

Probability of choosing all pigeons in $I_R \cap Dom(M)$ consistently with R if intersection $|I_R \cap Dom(M)| \geq n/32$

$$\leq \left(\frac{2}{3}\right)^{n/32} < 2^{\delta'' n}$$

for some $\delta'' > 0$.

Putting everything together

$$\Pr_M\left[\text{Information } R \text{ and } M \text{ consistent}\right] \leq$$

$$\leq \Pr_M\left[|I_R \cap Dom(M)| \text{ small}\right] +$$

$$\Pr_M\left[|I_R \cap Dom(M)| \text{ large but } M \text{ consistent with } R\right]$$

$$\leq 2^{-\delta' n} + 2^{-\delta'' n}$$

$$\leq 2^{-\delta n}$$

for some $\delta > 0$, QED $\boxed{}$

This establishes Lemma 3, and Theorem 1 with the PHP lower bound follows