DD2442 AUTUMN 2016: LECTURE 8

Today we want to prove what we just claimed last lecture, namely that

1. Using random restrictions of the right type, we can collapse d-DNF formulas to decision trees of small height with good probability (this is known as a SWITCHING LEMMA)

2. If we have a d-DNF resolution refutation of F $\pi = (H_1, H_2, \ldots, H_{L-1}, H_L = \bot)$ where all lines H_i can be represented by decision trees of small height, then F can be refuted in (standard) resolution in small width.

In future lectures, we will combine (1) and (2) with resolution <u>width</u> lower bounds to prove d-DNF resolution <u>length</u> lower bounds.

Let us start by proving (2), which is easier.
We need to refresh our memories about the concepts:

(a) Strong implication
(b) Strong soundness
(c) Strong representation

In what follows $G_1, \ldots, G_s, H$ are d-DNF formulas

$G_1, \ldots, G_s$ <u>imply</u> H if for any (total) truth value assignment α s.t. $\alpha(G_i) = 1 \ \forall i$ it also holds that $\alpha(H) = 1$

<u>Notation</u> $G_1 \wedge \cdots \wedge G_s \vDash H$

$G_1, \ldots, G_s$ <u>strongly imply</u> H if for any terms $t_i \in G_i$, $i \in [s]$ that are mutually consistent (i.e., $\bigwedge_{i=1}^{s} t_i$ satisfiable) there is a $t \in H$ s.t. $\bigwedge_{i=1}^{s} t_i \vDash t$.

A proof system operating with DNF formulas is <u>strongly sound</u> if whenever

$$\frac{G_1 \quad \cdots \quad G_s}{H}$$

is a derivation rule, it holds that $G_1, \ldots, G_s$ <u>strongly imply</u> H

As noted last time, d-DNF resolution is strongly sound, and this is all we need to know about the derivation rules today

DECISION TREE T for H

Binary tree; internal nodes labelled by queried variables; leaves by 0/1

Every leaf v defines partial assignment ρ_v
Every total assignment α leads to node v_α

T <u>represents</u> H if for any total α

$$\alpha(H) = \text{label}(v_\alpha)$$

Let $br_b(T) = \{ \rho_v \mid \overset{\text{leaf}}{v} \text{ labelled by } b \}$

T <u>strongly represents</u> H if

(a) $\forall \rho \in br_0(T) \quad \forall t \in H \quad t \restriction \rho = 0$

(b) $\forall \rho \in br_1(T) \quad \exists t \in H \quad t \restriction \rho = 1$

The <u>representation height</u> $h(H)$ of H is the minimum height of any decision tree strongly representing H.

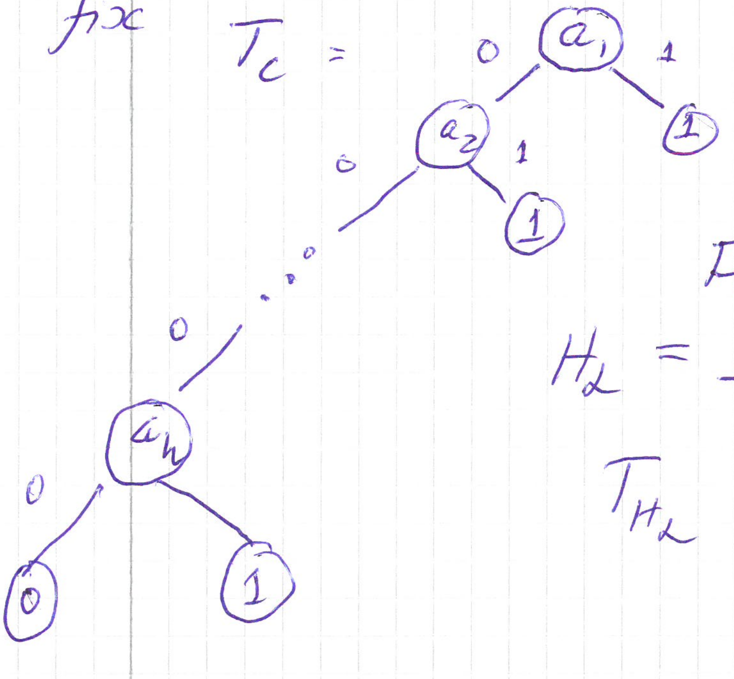
$H_\ell = \bot$

THEOREM 1 Let F be a h-CNF formula and suppose that $\pi = (H_1, H_2, \ldots, H_\ell)$ is a d-DNF resolution refutation of F such that for every line H_i it holds that $h(H_i) \le h$. Then there is a resolution refutation π' of F in width $W(\pi') \le 2h$.

For convenience, let us use resolution augmented with the <u>subsumption</u> or <u>weakening</u> rule saying that from a clause C we can derive $C \vee D$ for any D. All use of weakening can be eliminated from a resolution refutation without increasing the length (in fact, the length will decrease).

For every line H in π, fix a decision tree T_H of minimal height that strongly represents H

For a clause $C = a_1 \vee \cdots \vee a_h \in F$ we fix $T_C =$



For the final line $H_\ell = \bot$ we have

$T_{H_\ell} = 0$

For any partial assignment ρ, let us define

$$C_{\neg\rho} = \bigvee_{a \in \rho} \bar{a}$$

I.e., $C_{\neg\rho}$ is the unique maximal clause falsified by ρ.

By induction over $\pi = (H_1, H_2, \ldots, H_L = \bot)$ we will derive the set of clauses

$$\{ C_{\neg\rho} \mid \rho \in br_0(T_{H_i}) \}$$

in width at most $2h$.

<u>Base case</u> For a clause $C = a_1 \vee \cdots \vee a_h \in F$ we have $br_0(T_C) = \{ \bar{a}_1, \bar{a}_2, \ldots, \bar{a}_h \}$ and hence

$$\{ C_{\neg\rho} \mid \rho \in br_0(T_C) \} = \{ a_1 \vee a_2 \vee \cdots \vee a_h \}$$
$$= \{ C \} \qquad \text{OK.}$$

<u>After the final step</u> we will have derived

$$\{ C_{\neg\rho} \mid \rho \in br_0 (T_\bot) \} =$$
$$= \{ C_{\neg\rho} \mid \rho \in \emptyset \} = \bot$$

which is exactly what we want.
So if we can just do the inductive step, then we are done.

Suppose that H was derived from H_1 & H_2 (H derived from just H_1 is similar)

By our inductive hypothesis, we have derived all $C_{\neg\rho}$ for $\rho \in br_0(T_{H_i})$, $i = 1, 2$.

We want to derive $C_{\neg\rho}$ for all $\rho \in br_0(T_H)$.

To guide our inductive proof, we build a decision tree T for $H_1 \wedge H_2$ in the following way

① Take T_{H_1} and keep all internal nodes and 0-leaves

② For every 1-leaf v, say corresponding to ρ_v, replace this leaf by a (separate) copy of $T_{H_2} \upharpoonright \rho_v$

(See example picture on next page)

The height of T is clearly at most $2h$

CLAIM 2 The clauses $\{C_{\neg\rho} \mid \rho \in br_0(T)\}$ are derivable from $\{C_{\neg\rho} \mid \rho \in br_0(T_{H_1})\}$

$\cup \ \{C_{\neg\rho} \mid \rho \in br_0(T_{H_2})\}$

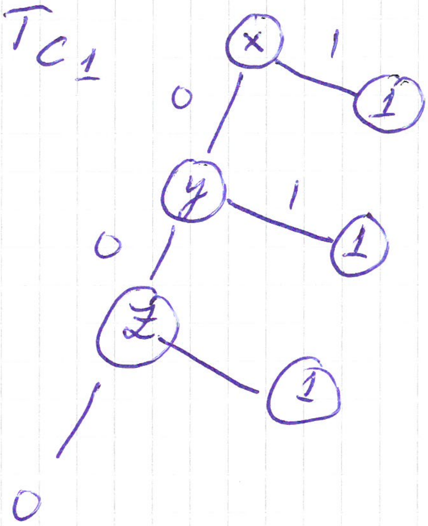
by weakening

<u>Proof</u> By construction every $\rho \in br_0(T)$ contains either some $\rho_1 \in br_0(T_{H_1})$ or some $\rho_2 \in br_0(T_{H_2})$, and for this ρ_i it then holds that $C_{\neg\rho_i} \subseteq C_{\neg\rho}$. ▨

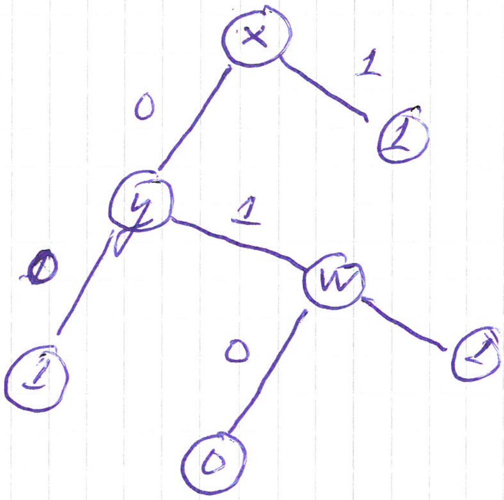
T4

$$C_1 = x \vee y \vee z$$
$$C_2 = x \vee \bar{y} \vee w$$

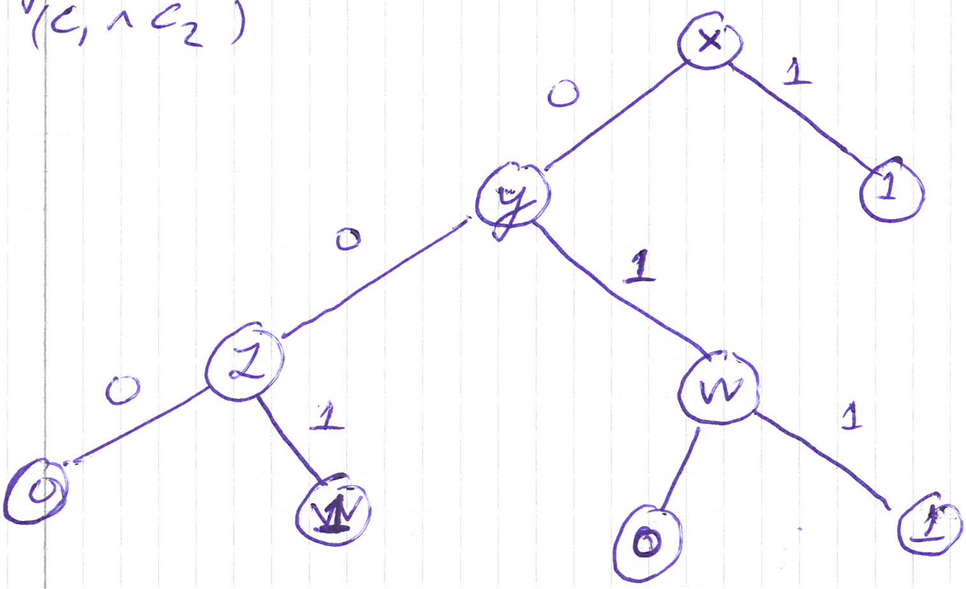
T_{C_1}



T_{C_2}



$T_{(C_1 \wedge C_2)}$



LEMMA 3 With notation as above, fix any $\sigma \in Gr_0(T_H)$. Then for all $\rho \in Gr_0(T) \cup Gr_1(T)$ it holds that if ρ and σ are consistent, then $\rho \in Gr_0(T)$.

<u>Proof</u> Suppose that $\rho \in Gr_1(T)$. We claim that there exists a term $t \in H$ such that ρ satisfies t (i.e., $t \restriction \rho = 1$). By construction of T there are ~~some~~ $\rho_1 \in Gr_1(T_{H_1})$ and $\rho_2 \in Gr_1(T_{H_2})$ such that $\rho_1 \cup \rho_2 = \rho$.

Since T_1 & T_2 strongly represent H_1 & H_2, there are $t_1 \in H_1$ and $t_2 \in H_2$ such that $\rho = \rho_1 \cup \rho_2$ satisfy $t_1 \wedge t_2$. By the strong soundness of d-DNF resolution, there exists a term $t \in H$ such that $t \restriction \rho = 1$.

Suppose now that $\sigma \in Gr_0(T_H)$. Since T_H strongly represents H, σ falsifies all terms $t \in H$. Since $\rho \in Gr_1(T)$ satisfies some term in H, σ and ρ cannot be consistent. $\square$

Why does this lemma helps us ?

Recall that we have $\{C_{\neg\rho} \mid \rho \in br_0(T)\}$ by our inductive hypothesis.
We want to derive $C_{\neg\sigma}$ for $\sigma \in br_0(T_H)$.
Let us consider what happens when we walk in T according to σ. A priori 3 cases

(1) Hit a leaf v labelled 1. Let ρ_v corresponding assignment. Then $\rho_v \in br_1(T)$ and ρ_v and σ consistent.
This doesn't happen — the lemma we just proved rules this out

(2) Hit a leaf v labelled 0. Then $\rho_v \in br_0(T)$ and $\rho_v \subseteq \sigma$.
But then $C_{\neg\sigma} \supseteq C_{\neg\rho_v} \in \{C_{\neg\rho} \mid \rho \in br_0(T)\}$
which is a clause we already have, so $C_{\neg\sigma}$ can be derived by weakening

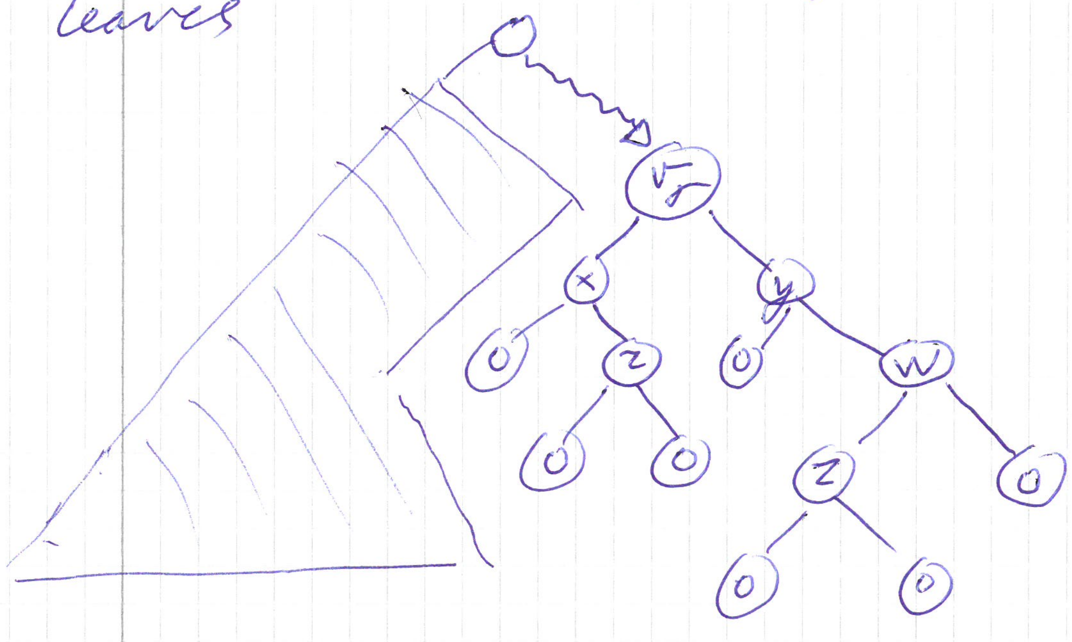
(3) Walk some way down in T but don't hit a leaf. Then look at all $\rho \in br_0(T)$, $\rho \supseteq \sigma$, corresponding to leaves in T below where walk according to σ stopped. We claim that we can resolve these clauses $C_{\neg\rho}$ to derive $C_{\neg\sigma}$. If we can do this, then we are done with the proof of the theorem.

But let us think a little bit more...

CLAIM 4 Let v_σ be the node reached in T when walking according to $\sigma \in Gr_0(T_H)$. Then there is no path to a 1-labelled leaf in T going through v_σ.

Proof Let w be such a leaf and consider $\rho_w \in Gr_1(T)$. By construction $\sigma \subseteq \rho_w$ and so σ and ρ_w are consistent. But that contradicts the lemma $\blacksquare$

So, pictorially v_σ is the root of a subtree containing only 0-labelled leaves



We have all clauses $C_{\neg\rho_w}$ for the leaves w in this subtree. It is intuitively kind of clear that we should just systematically resolve away the variables in the subtree bottom-up to derive $C_{\neg\sigma}$. For any ρ, let v_ρ be node in T reached when walking according to ρ

Formally, we prove the following by induction

LEMMA 5 With notation as above, let $\sigma \in \mathcal{b\sigma}_0(T_H)$ and let $v_\sigma \in T$ be the node reached when walking from the root of T according to σ. Let v_ρ be any node in the subtree rooted at v_σ, i.e., such that $\rho \supseteq \sigma$. Then $C_{\neg\rho}$ is derivable in width at most $2h$ from $\mathcal{C} = \{C_{\neg\rho} \mid \rho \in \mathcal{b\sigma}_0(T), \rho \supseteq \sigma\}$.

__Proof__ All clauses we will see will have width $\leq$ depth of $T \leq 2h$, so let us not worry about that.

__Base case__ $\rho \in \mathcal{b\sigma}_0(T)$ ρ corresponds to leaf. Then we know

Then $C_{\neg\rho} \in \mathcal{C}$ and there is nothing to prove

__Inductive step__ $\rho \notin \mathcal{b\sigma}_0(T)$, ρ does not correspond to leaf, so there are nodes in T labelled by $\rho \cup \{x\}$ & $\rho \cup \{\bar{x}\}$ for the variable x queried at v_ρ

By induction, can derive

$$C_{\neg(\rho\cup\{x\})} = C_{\neg\rho} \vee \bar{x} \;\&\; C_{\neg(\rho\cup\{\bar{x}\})} = C_{\neg\rho} \vee x$$

Resolve to get $C_{\neg\rho}$. Done.

Note that width of clauses only shrink as we work our way up the tree.

Now we have established Thm 1.

SWITCHING LEMMA

Q1 When can an OR of ANDs be represented as an AND of ORs?

A1 Always. Both CNFs and DNFs can express any Boolean function.

Q2 When can the conversion from DNF to CNF (and vice versa) be efficient?

A2 Not always.

Consider 2-DNF

$$(x_1 \wedge x_2) \vee (x_3 \wedge x_4) \vee \cdots \vee (x_{2m-1} \wedge x_{2m})$$

CNF becomes

$$\begin{aligned} & (x_1 \vee x_3 \vee \cdots \vee x_{2m-3} \vee x_{2m-1}) \\ \wedge & (x_1 \vee x_3 \vee \cdots \vee x_{2m-3} \vee x_{2m}) \\ \wedge & (x_1 \vee x_3 \vee \cdots \vee x_{2m-2} \vee x_{2m-1}) \\ \wedge & (x_1 \vee x_3 \vee \cdots \vee x_{2m-2} \vee x_{2m}) \end{aligned}$$

et cetera.

Switching lemmas say that when we first hit DNF with suitable random restriction, then conversion likely to be efficient.

In our case, want to convert DNF H to shallow decision tree (and a decision tree of height h can always be represented as a h-CNF formula).

We stated last time:

THEOREM 12 [SBI '04]
Let $d, s, t \in \mathbb{N}^+$; let $\gamma, \delta \in (0,1]$; and let $\mathcal{D}$ be a distribution on partial truth value assignments such that for every d-DNF formula G it holds that

$$\Pr_{\rho \sim \mathcal{D}}\left[G\restriction_\rho \neq 1 \right] \leq t \cdot 2^{-\delta(cov(G))^\gamma}.$$

Then for every d-CNF formula H it holds that

$$\Pr_{\rho \sim \mathcal{D}}\left[h(H\restriction_\rho) > 2s \right] \leq dt \cdot 2^{-\delta' s^{\gamma'}}$$

for $\delta' = 2(\delta/4)^d$ and $\gamma' = \gamma^d$.

For H d-DNF formula $S \subseteq Vars(H)$ is a <u>cover</u> of H if every term $t \in \mathcal{H}$ contains a variable from S. The <u>covering number</u> $cov(\mathcal{H})$ of $\mathcal{H}$ is the minimum size of any cover of $\mathcal{H}$.

Will actually prove ~~more~~ general theorem from which Thm 1 follows as corollary. Take a deep breath...

THEOREM 2 [SBI '04]
Let $d \in \mathbb{N}^+$; let $s_i \in \mathbb{N}^+$, $i \in [0, d-1]$, and $p_i \in (0,1)$, $i \in [1, d]$, ~~and~~ be sequences of positive numbers, and let $\mathcal{D}$ be a distribution on partial assignments such that for every $i \leq d$ and every i-DNF formula G it holds that if $cov(G) > s_{i-1}$, then

$$\Pr_{\rho \sim \mathcal{D}}\left[G\restriction_\rho \neq 1 \right] \leq p_i$$

Then for every d-DNF formula $\mathcal{H}$ it holds that

$$\Pr_{\rho \sim \mathcal{D}}\left[h(\mathcal{H}\restriction_\rho) > \sum_{i=0}^{d-1} s_i \right] \leq \sum_{i=1}^{d} p_i \cdot 2^{\sum_{j=i}^{d-1} s_j}$$

Wait... Aren't there minus signs missing somewhere? $\sum_{j=i}^{d-1} s_j > 1$, so $2^{\sum_{j=i}^{d-1} s_i}$ is kind of large, no?

Yes, so the p_i had better be small...

For what remains of today's lecture we will prove this theorem. The proof is by induction on d, the size of the terms.

<u>Base case</u> [$d=1$, i.e., H is a clause]

(b) If $\underline{cov(H) \leq s_0}$, this means that the width of the clause H is $W(H) \leq s_0$, or expressed differently that H contains at most s_0 variables

Build decision tree T of height $\leq s_0$ that queries <u>all</u> variables in H. Since all $\rho \in Br_0(T) \cup Br_1(T)$ assign all variables in H, T obviously strongly represents H.

(a) If $\underline{cov(H) > s_0}$, then by assumption $Pr_{\rho \in \mathcal{D}}[H\restriction_\rho \neq 1] \leq p_1$, and if $H\restriction_\rho = 1$ then the decision tree 1 of height 0 will do

So

$$\Pr_{g \sim D}\left[h(H \upharpoonright g) > s_0 \right] \leq$$

$$\Pr_g \left[h(H \upharpoonright g) \neq 1 \right] \leq$$

$$p_1 = p_1 \, 2^{\sum_{j=1}^{0} s_j} = p_1 \cdot 2^0 .$$

Inductive step

Theorem holds for d-DNF formulas

Let H $(d+1)$-DNF formula.

Let $s_0, \dots, s_d, p_1, \dots, p_{d+1}$ be sequences as in the statement of the theorem.

Case (a) $\quad cov(H) > s_d$ $\quad$ Similar to base case:

$$\Pr_{g \sim D}\left[h(H \upharpoonright g) > \sum_{i=0}^{d} s_i \right] \leq$$

$$\Pr_{g \sim D}\left[H \upharpoonright g \neq 1 \right] \leq p_{d+1} \leq \sum_{i=1}^{d+1} p_i \cdot 2^{\sum_{j=i}^{d} s_j} .$$

Case (b) $\quad cov(H) \leq s_d$

Let $S \subseteq Vars(H)$ cover of size $|S| \leq s_d$

For any restriction $\tau : S \to \{0,1\}$

$H \upharpoonright \tau$ is a d-DNF formula (since τ hits at least one literal in every term $t \in H$).

What to prove that $H\restriction_\rho$ is very likely to have shallow decision tree.

By induction, this is true for $(H\restriction_\tau)\restriction_\rho$

First sample $\rho \sim D$.
For any τ consistent with ρ we have $(H\restriction_\tau)\restriction_\rho = (H\restriction_\rho)\restriction_\tau$ and hence by the induction hypothesis:

$$\Pr\left[h((H\restriction_\tau)\restriction_\rho) > \sum_{i=0}^{d-1} s_i\right] \leq \sum_{i=1}^{d} p_i \cdot 2^{\sum_{j=i}^{d-1} s_j} \quad (1)$$

There are at most $2^{|S|} \leq 2^{sd}$ restrictions $\tau: S \to \{0,1\}$ that are consistent with ρ.
Hence, by the union bound applied on (1) we get

$$\Pr_\rho\left[\exists \tau \text{ consistent with } \rho \text{ s.t. } h((H\restriction_\rho)\restriction_\tau) > \sum_{i=0}^{d-1} s_i\right] \leq$$

$$\leq 2^{sd}\left(\sum_{i=1}^{d} p_i 2^{\sum_{j=1}^{d-1} s_j}\right)$$

$$\Leftarrow \sum_{i=1}^{d} p_i \cdot 2^{\sum_{j=1}^{d} s_j} < \sum_{i=1}^{d+1} p_i \cdot 2^{\sum_{j=i}^{d} s_j} \quad (2)$$

Note that (2) is exactly the probability bound that we are shooting for.

Hence, if there is a τ s.t. $h((H\restriction_\rho)\restriction_\tau)$ is too large, we can afford to just give up.

But if for all τ consistent with ρ it holds that

$$h\left((H\restriction_\rho)\restriction_\tau\right) \leq \sum_{i=0}^{d-1} s_i \qquad (3)$$

we proceed to construct a decision tree strongly representing $H\restriction_\rho$

First query all variables in S left unset by ρ.
Formally let $\beta = \tau \setminus \rho$ (where again we think of ρ = {literals set to true by ρ})
Build complete binary tree (if necessary) for all assignments $\beta : (\mathrm{dom}(\tau) \setminus \mathrm{dom}(\rho)) \to \{0,1\}$
Let this be the "upper part" of T
At each leaf, the residual formula is

$$(H\restriction_\rho)\restriction_\beta = (H\restriction_\rho)\restriction_\tau$$

at the corresponding leaf

For each β, plug in a decision tree of minimum height that strongly represents $(H\restriction_\rho)\restriction_\tau$, which has height

$$\leq \sum_{i=0}^{d-1} s_i \qquad \text{by (3)}$$

Let this be the "lower part" of the tree. S VIII
Let T be the tree constructed in this way. We claim two properties of T.

(1) T has height $\leq \sum_{i=0}^{d} s_i$.

(2) T strongly represents $H\restriction_g$.

Note that if we can establish these two claims, then the inductive step is finished and Thm 2 follows by the induction principle.

Claim 1: This is the easy part

The upper part has height

$$\begin{aligned} |dom(\beta)| &= |dom(\tau) \setminus dom(g)| \\ &\leq |dom(\tau)| \\ &\leq |S| \\ &\leq s_d \end{aligned} \qquad (4)$$

As already noted, the lower part has height at most

$$\sum_{i=0}^{d-1} s_i \qquad (5)$$

Just sum up (4) & (5). Done.

S IX

Let $\psi \in br_0(T) \cup br_1(T)$

Can write

$$\psi = \beta \cup \sigma$$

for

$$\beta : S \setminus dom(\rho) \to \{0,1\}$$

and

$$\sigma \in br_0(T(\beta)) \cup br_1(T(\beta))$$

where $T(\beta)$ is a tree that strongly represents $(H\restriction\rho)\restriction\beta$

Need to find term in $H\restriction\rho$ fixed to true by ψ

Suppose $\psi \in br_1(T)$. Then there is a term $t' \in (H\restriction\rho)\restriction\beta$ such that

$$t'\restriction\sigma = 1.$$

This t' comes from $t \in H$ such that

$$t' = t\restriction(\rho\cup\beta) = (t\restriction\rho)\restriction\beta$$

for $t\restriction\rho \in H\restriction\rho$. For this term

$$(t\restriction\rho)\restriction\psi = (t\restriction\rho)\restriction(\beta\cup\sigma) \equiv 1$$

Suppose $\psi \in \mathfrak{S}_0(T)$.

As before, write $\psi = \beta \cup \sigma$.

We get two cases:

(a) $(H\restriction_g)\restriction_\beta = 0$

(b) $(H\restriction_g)\restriction_\beta \neq 0$

In case (a) β already falsifies every term in $H\restriction_g$, and extending β to $\psi = \beta \cup \sigma$ does not change this.

In case (b), since subtree underneath β strongly represents $(H\restriction_g)\restriction_\beta$ and $\sigma \in \mathfrak{S}_0(T(\beta))$ for every term $t'' \in (H\restriction_g)\restriction_\beta$ it holds that $t''\restriction_\sigma = 0$

Thus, for every $t' \in H\restriction_g$ it holds that

$$t'\restriction_{(\beta\cup\sigma)} = t'\restriction_\psi = 0,$$

either because t' is inconsistent with β, or if not, then with σ.

This shows that T strongly represents $H\restriction_g$, and the theorem follows ▨