1 The Switching Lemma

A $k$-DNF (= disjunctive-normal-form formula of width $k$) is a depth-2 formula of the form $\text{OR}(C_1, \ldots, C_m)$ where each clause $C_i$ is an AND of $\leq k$ literals. A $k$-CNF (= conjunctive-normal-form formula of width $k$) is a depth-2 formula of the form $\text{AND}(C_1, \ldots, C_m)$ where each $C_i$ is an OR of $\leq k$ literals.

A decision tree of depth 0 is a constant (0 or 1). For $d \geq 1$, a decision tree of depth $\leq d$ is a triple $T = (x_i, T_0, T_1)$ where $x_i$ is a variable and $T_0$ and $T_1$ are decision trees of depth $\leq d - 1$. Decision trees compute boolean functions in the obvious way: if $T = (x_i, T_0, T_1)$, then $T(x) := T_{x_i}(x)$.

The decision-tree depth of a boolean function $f$, denoted $D(f)$, is the minimum depth of a decision tree that computes $f$. Note that $D(f) = 0$ iff $f$ is a constant, and $D(f) = 1$ iff $f$ is a literal. The function $f(a, b, c) = (a \land b) \lor (\neg a \land \neg c)$ has decision-tree depth 2. AND$_n$ and XOR$_n$ are examples of functions with the maximum possible decision-tree depth $n$.

It's easy to see that any function with decision-tree depth $k$ is equivalent to both a $k$-DNF and a $k$-CNF. (There is a weak converse to this fact: any function which can be expressed as both a $k$-DNF and an $\ell$-CNF has decision-tree depth at most $k\ell$. ) A corollary of this fact is that an OR (resp. AND) of arbitrarily many functions with decision-tree depth $k$ is equivalent to a $k$-DNF (resp. $k$-CNF).

Previously we studied the effect of the $p$-random restriction $R_p$ on DeMorgan formulas. $R_p$ also simplifies depth-$k$ decision trees, as well as $k$-DNF and $k$-CNF.

**Theorem 1** (Effect of $R_p$ on decision-tree depth). If $D(f) = k$, then

$$
P[\ D(f|R_p) \geq \ell \ ] \leq (2p)^{\ell} \binom{k}{\ell} = O(pk/\ell)^{\ell}
$$

for all $\ell \geq 1$.

*Proof.* Induction on $k$. Base case $k = 0$ is trivial, so assume $k \geq 1$ and $\ell \geq 1$. Let $T = (x_i, T_0, T_1)$
be a DT of depth \( k \). Then

\[
P[D(T|R_p) \geq \ell] = P[R_p(x_i) = * \text{ and } D(T|R_p) \geq t] \\
+ P[R_p(x_i) = 0 \text{ and } D(T|R_p) \geq \ell] + P[R_p(x_i) = 1 \text{ and } D(T|R_p) \geq \ell] \\
= pP[D(T_0|R_p) \geq \ell - 1 \text{ or } D(T_1|R_p) \geq \ell - 1] \\
+ \frac{1-p}{2}
\left( P[D(T_0|R_p) \geq \ell] + P[D(T_1|R_p) \geq \ell] \right) \\
\leq p\left( P[D(T_0|R_p) \geq \ell - 1] + P[D(T_1|R_p) \geq \ell - 1] \right) \\
+ \frac{1-p}{2}
\left( P[D(T_0|R_p) \geq \ell] + P[D(T_1|R_p) \geq \ell] \right) \\
\leq 2p(2p)^{\ell-1}\binom{k-1}{\ell-1} + (2p)^\ell\binom{k-1}{\ell} \\
= (2p)^\ell\left(\frac{k}{\ell}\right).
\]

\[ \square \]

Håstad’s Switching Lemma (1986) gives a similar bound for \( k \)-DNF and \( k \)-CNF formulas (i.e., OR’s or AND’s of depth-\( k \) decision trees). Instead of \( O(pk/\ell)^{\ell} \), we get a bound \( O(pk)^{\ell} \).

**Theorem 2** (Switching Lemma). If \( f \) is a \( k \)-DNF or \( k \)-CNF, then

\[
P[D(f|R_p) \geq \ell] \leq (5pk)^{\ell}.
\]


Fix \( k, \ell \geq 1 \) and \( p \in [0,1] \) and suppose \( f = \text{OR}(C_1, \ldots, C_m) \) where each clause \( C_j \) is an AND of \( \leq k \) literals. (In particular, we fix an ordering of clauses \( C_1, \ldots, C_m \).) Let \( \text{Vars}(C_j) \subseteq [n] \) denote the set of variables occurring in \( C_j \), that is, \( \text{Vars}(C_j) = \{i : x_i \text{ or } \pi_i \text{ occurs in } C_j\} \).

For every restriction \( \rho : [n] \rightarrow \{0,1,*\} \), we define a decision tree \( T(f,\rho) \) called the “canonical decision tree of \( f|\rho \)”. This is defined as follows. If \( \rho \) fixes every clause to 0, then \( T(f,\rho) \) outputs 0. Otherwise, let \( C_j \) be the first clause not fixed to 0 by \( \rho \) and proceed as follows:

- If \( C_j \) is fixed to 1 by \( \rho \) (i.e. every literal is set to 1), then \( T(f,\rho) \) outputs 1.
- If \( C_j \) is not fixed to 1 by \( \rho \) (i.e. no literal is set to 0 and at least one literal has value *), then \( T(f,\rho) \) queries all free variables in \( C_j \) and proceeds as the decision tree \( T(f,\rho\pi) \) where
  - \( \pi \in \{0,1\}^{\text{Vars}(C_j) \cap \text{Stars}(\rho)} \) is the assignment to the queried variables of \( C_j \),
  - \( \rho\pi \in \{0,1,*\}^n \) is the combined restriction with \( (\rho\pi)_i = \begin{cases} \pi_i & \text{if } i \in \text{Vars}(C_j) \cap \text{Stars}(\rho), \\ \rho_i & \text{otherwise}. \end{cases} \)
Clearly the depth of $T(f, R_p)$ is an upper bound on $D(f| R_p)$. Therefore, it suffices to show

\[(1) \quad P[ \text{depth}(T(f, R_p)) \geq \ell ] \leq (16pk)^\ell \]

Let’s name this bad event

$$ \text{BAD} \stackrel{\text{def}}{=} \{ \rho : \text{depth}(T(f, \rho)) \geq \ell \}. $$

To prove (1), we will associate each $\rho \in \text{BAD}$ with a restriction $\hat{\rho}$ (not necessarily in BAD) such that

(i) $|\text{Stars}(\hat{\rho})| = |\text{Stars}(\rho)| - \ell$,

(ii) the function $\rho \mapsto \hat{\rho}$ is at most $(4k)^\ell$-to-1,

that is, for every restriction $\sigma$, we have $\#\{\rho \in \text{BAD} : \hat{\rho} = \sigma\} \leq (4k)^\ell$.

Note that property (i) implies $P[ R_p = \rho ] = (\frac{2p}{1-p})^\ell P[ R_p = \hat{\rho} ]$. (This follows from the observation that $P[ R_p = \sigma ] = p^{|\text{Stars}(\sigma)|} (\frac{1-p}{2})^{|\text{Nonstars}(\sigma)|}$ for all restrictions $\sigma$.) Without loss of generality, we may assume that $p \leq 1/2$ (since the Theorem is trivial if $p \leq 1/16$). Therefore, we have

\[(2) \quad P[ R_p = \rho ] \leq (4p)^\ell P[ R_p = \hat{\rho} ]. \]

Assuming we have a function $\rho \mapsto \hat{\rho}$ satisfying (i) and (ii), we obtain inequality (1) as follows:

$$ P[ R_p \in \text{BAD} ] = \sum_{\rho \in \text{BAD}} P[ R_p = \rho ] $$

$$ \leq (4p)^\ell \sum_{\rho \in \text{BAD}} P[ R_p = \hat{\rho} ] \quad \text{(by (2))} $$

$$ = (4p)^\ell \sum_{\sigma : [n] \rightarrow \{0,1,\ast\}} P[ R_p = \sigma ] \cdot \#\{\rho \in \text{BAD} : \hat{\rho} = \sigma\} $$

$$ \leq (16pk)^\ell \sum_{\sigma : [n] \rightarrow \{0,1,\ast\}} P[ R_p = \sigma ] \quad \text{(by (ii))} $$

$$ = (16pk)^\ell. $$

**Definition of $\hat{\rho}$**. It remains to define the function $\rho \mapsto \hat{\rho}$ and show that it satisfies (i) and (ii). Consider any $\rho \in \text{BAD}$. By definition, the decision tree $T(f, \rho)$ contains a path of length $\geq \ell$. Fix any such “long path” in $T(f, \rho)$. Let $Q \subseteq [n]$, $|Q| = \ell$, consist of the first $\ell$ variables queries on this path, and let $\pi : Q \rightarrow \{0,1\}$ be the corresponding assignment of these variables.

By definition of $T(f, \rho)$, there exists a partition $Q = Q_1 \uplus \cdots \uplus Q_t$ and clauses $C_{j_1}, \ldots, C_{j_t}$ ($1 \leq j_1 < \cdots < j_t \leq m$) where $C_{j_i}$ is responsible for queries $Q_i$ in the process defining $T(f, \rho)$. Let $\pi_i : Q_i \rightarrow \{0,1\}$ denote the corresponding sub-assignment of $\pi$. In addition:

- let $a_i \in \{0,1\}^k$ be the characteristic function of $Q_i$ among variables of $C_{j_i}$,
• let $b_i \in \{0, 1\}^{Q_i}$ encode $\pi_i$ (under the order in which variables occur in $C_{j_i}$),
• let $\tilde{\pi}_i : Q_i \rightarrow \{0, 1\}$ be the unique assignment to $Q_i$ such that $C_i|\rho\pi_1 \cdots \pi_{i-1}\tilde{\pi}_i \neq 0$.

Finally, we define $\tilde{\rho}$ by

$$\tilde{\rho} \overset{\text{def}}{=} \rho\tilde{\pi}_1 \cdots \tilde{\pi}_t.$$

Property (i) clearly holds, since $\tilde{\rho}$ fills in exactly $\ell$ stars of $\rho$. As for property (ii), we establish that $\rho \mapsto \tilde{\rho}$ is at most $(4k)^\ell$-to-1 over BAD by showing:

(ii-a) the function $\rho \mapsto (\tilde{\rho}, a, b)$ is 1-to-1 over BAD,

(ii-b) the pair $(a, b)$ (i.e. the string $(a_1, \ldots, a_t, b_1, \ldots, b_t)$) takes at most $(4k)^\ell$ possible values over $\rho \in \text{BAD}$.

To see that (ii-a) holds, we describe a procedure for inverting $\rho \mapsto (\tilde{\rho}, a, b)$ over BAD. Given $(\tilde{\rho}, a, b)$:

• Note that $C_{j_1}$ is the first clause of $f$ with the property that $C_{j_1}|\tilde{\rho} \neq 0$. Therefore, $\tilde{\rho}$ gives knowledge of $C_{j_1}$, and $a_1, b_1$ then give knowledge of $Q_1, \pi_1$. This allows us to determine $\rho\pi_1\tilde{\pi}_2 \cdots \tilde{\pi}_t$.

• Next (if $|Q_1| < \ell$), note that $C_{j_2}$ the first clause of $f$ with the property that $C_{j_2}|\rho\pi_1\tilde{\pi}_2 \cdots \tilde{\pi}_t \neq 0$. Via $a_2, b_2$, we now have knowledge of $Q_2, \pi_2$. This allows us to determine $\rho\pi_1\pi_2\tilde{\pi}_3 \cdots \tilde{\pi}_t$.

• This process continues until we have learned $Q_1, \ldots, Q_t, \pi_1, \ldots, \pi_t$ and $\rho\pi_1 \cdots \pi_t$, as which point we know $\rho$.

Finally, to show (ii-b), we note that each $(a_1, \ldots, a_t)$ is an element of $\{0, 1\}^k$ where $|a_1|, \ldots, |a_t| \geq 1$ and $|a_1| + \cdots + |a_t| = \ell$. The number of such sequences is at most $(2k)^\ell$. The possibilities for $(b_1, \ldots, b_t)$, given each $(a_1, \ldots, a_t)$, contribute another $2^\ell$ factor. \hfill \Box

2 Lower bounds for XOR$_n$

Using the Switching Lemma, we able to prove tight lower bounds for the depth $d + 1$ circuit size (as well as the depth $d + 1$ formulas size) of XOR$_n$.

Theorem 3. Let $C$ be an AC$^0$ circuit of depth $d + 1$ and size $S$. Let $p = \frac{1}{10(20 \log S)^d}$. Then

$$\Pr\left[ D(C|R_p) \geq \ell \right] \leq \frac{1}{2^\ell} + \frac{1}{S}.$$
Proof. Let $p_1 = 1/10$ and let $\rho_1$ be a $p_1$-random restriction over the variables of $C$. Note that each bottom-level gate $g$ of $C$ is an AND or OR of literals, hence a 1-CNF or 1-DNF. Therefore, by the Switching Lemma, $\mathbb{P}[D(g|\rho_1) > 2 \log S] \leq (5p_1)^{2\log S} \leq 1/S^2$.

For $i \in \{2, \ldots, d+1\}$, let $p_i = p_{i-1}/20 \log S$ and let $\rho_i$ be a $p_i$-random restriction over the stars of $\rho_{i-1}$. For each gate $g = \text{AND/OR}(g_1, \ldots, g_m)$ of depth $i \leq d$, if we condition on $D(g_j|\rho_1 \ldots \rho_{i-1}) \leq 2 \log S$ for all $j \in [m]$ (in which case $g$ is a 2-$\log S$-CNF/DNF), then by the Switching Lemma $D(g|\rho_1 \ldots \rho_i) \leq 2 \log S$ except with probability $(5p_i \cdot 2 \log S)^{2\log S} = 2^{-2\log S} = 1/S^2$.

It follows that, except with probability $1/S$, we have $D(g|\rho_1 \ldots \rho_d) \leq 2 \log S$ for all gates $g$ below the output gate of $C$. If we condition on this event, then by the Switching Lemma $D(C|\rho_1 \ldots \rho_{d+1}) \leq \ell$ except with probability $(5p_{d+1} \cdot 2 \log S)^{\ell} = 2^{-\ell}$. The proof is completed by noting that $\rho_1 \ldots \rho_{d+1}$ in aggregate is a $p_1 \cdot \ldots \cdot p_{d+1}$-random restriction and that $p_1 \cdot \ldots \cdot p_{d+1} = 1/(20 \log S)^d$.

**Corollary 4.** $C_{d+1}(\text{XOR}_n) = 2^{\Omega(n^{1/d})}$

**Proof.** Let $S = C_{d+1}(\text{XOR}_n)$ and let $p = \frac{1}{10^{d+1}(2 \log S)^d}$. We have

$$\mathbb{P}[D(\text{XOR}_n|R_p) \geq 1] \leq \frac{1}{2} + \frac{1}{S}.$$ 

Assuming $S \geq 4$ (without loss of generality), it follows that

$$\mathbb{P}[D(\text{XOR}_n|R_p) = 0] \geq \frac{1}{4}.$$ 

Since $\mathbb{P}[D(\text{XOR}_n|R_p) = 0] = \mathbb{P}[\ Bin(n, p) = 0]$, it follows that $p = O(1/n)$ and hence

$$\frac{1}{10^{d+1}(2 \log S)^d} = \Omega(n).$$

We conclude that $S = 2^{\Omega(n^{1/d})}$.

**Exercise.** Show $C_d(\text{MAJ}_n) = 2^{\Omega(n^{1/d})}$ by reduction to XOR$_n$.

3 Lower Bounds for $\text{AC}^0[p]$ by the Polynomial Method (Razborov’87, Smolensky’87)

We work over the field $\mathbb{F}_p$ for an arbitrary prime $p$.

Recall that $\text{AC}^0[p]$ circuits and formulas have inputs labeled by literals and unbounded fan-in AND, OR, $\text{MOD}_p$ gates where $\text{MOD}_p(x_1, \ldots, x_n) = 1 \iff x_1 + \cdots + x_n = 0 \text{ mod } p$. 

5
**Definition 5.** Let $A \in \mathbb{F}_p[x_1, \ldots, x_n]$ be a random polynomial (i.e. a random variable over $\mathbb{F}_p[x_1, \ldots, x_n]$).

The degree of a random polynomial $A \in \mathbb{F}_p[x_1, \ldots, x_n]$ is the maximum degree of a polynomial in the support of $A$.

The $\varepsilon$-approximate degree of $f : \{0,1\}^n \to \{0,1\}$, denoted $\deg_\varepsilon(f)$, is the minimum degree of a random polynomial $A \in \mathbb{F}_p[x_1, \ldots, x_n]$ such that $\Pr[A \neq A(x)] \leq \varepsilon$ for every $x \in \{0,1\}^n$.

**Lemma 6.** There exists a non-random polynomial $a \in \mathbb{F}_p[x_1, \ldots, x_n]$ of depth $\deg_\varepsilon(f)$ such that $\Pr_{x \in \{0,1\}^n}[a(x) \neq f(x)] \leq \varepsilon$.

**Proof.** Let $A$ be an $\varepsilon$-approximating polynomial for $f$. By Markov’s inequality

$$\Pr_{x \in \{0,1\}^n}[A(x) \neq f(x)] > \varepsilon \leq \frac{\mathbb{E}_{x \in \{0,1\}^n}[A(x) \neq f(x)]}{\varepsilon} < 1.$$

Therefore, there exists $a \in \text{Supp}(A)$ such that $\Pr_{x \in \{0,1\}^n}[a(x) \neq f(x)] \leq \varepsilon$. \hfill \Box

**Lemma 7.** Suppose $f(x) = g(h_1(x), \ldots, h_m(x))$. Then for all $\delta, \varepsilon_1, \ldots, \varepsilon_m$,

$$\deg_{\delta + \varepsilon_1 + \ldots + \varepsilon_m}(f) \leq \deg_\delta(g) \cdot \max_i \deg_{\varepsilon_i}(h_i).$$

**Proof.** Let $A_g \in \mathbb{F}_p[y_1, \ldots, y_m]$ be a $\delta$-approx random poly for $g$ and let $A_{h_i} \in \mathbb{F}_p[x_1, \ldots, x_n]$ be $\varepsilon_i$-approx random polys for $h_i$. Let $A_f(x) := A_g(A_{h_1}(x), \ldots, A_{h_m}(x))$. Then $\deg(A_f) = \deg(A_g) + \max_i \deg(A_{h_i})$. And

$$\Pr_{A_f}[A_f(x) \neq f(x)] \leq \Pr_{A_g,A_{h_1},\ldots,A_{h_m}}\left[ \bigvee_i \left( A_{h_i}(x) \neq h_i(x) \right) \lor A_g(x) \neq g(x) \right] \leq \delta + \sum_i \varepsilon_i. \hfill \Box$$

We use this lemma together with bounds on $\text{MOD}_{p,n}$ and $\text{OR}_n$ and $\text{AND}_n$ to obtain bounds on $\deg_\varepsilon$ for $\text{AC}^0[p]$ circuits and formulas.

**Lemma 8.** For all $\varepsilon$ and $n$, we have $\deg_\varepsilon(\text{MOD}_{p,n}) \leq p - 1$.

**Note:** This bound does not depend on $\varepsilon$ or $n$.

**Proof.** For all $x \in \{0,1\}^n$, we have $\text{MOD}_p(x_1, \ldots, x_n) = (x_1 + \cdots + x_n)^{p-1}$ by Fermat’s Little Theorem. Therefore, $\deg_\varepsilon(\text{MOD}_{p,n}) \leq \deg_0(\text{MOD}_{p,n}) \leq p - 1$. \hfill \Box

**Lemma 9.** $\deg_\varepsilon(\text{OR}_n) \leq p(\log_p(1/\varepsilon) + 1)$

**Note:** Again, bound does not depend on the fan-in $n$. \

6
Proof. Fix any \( x \in \{0,1\}^n \). For random \( \lambda \in \mathbb{F}_p^n \), we have

\[
\mathbb{P}_{\lambda} \left[ \text{OR}(x) \neq (\lambda_1 x_1 + \cdots + \lambda_n x_n)^{p-1} \right] = \begin{cases} 0 & \text{if } x = (0, \ldots, 0), \\ 1/p & \text{if } x \neq (0, \ldots, 0). \end{cases}
\]

Therefore, for independent random \( \lambda^{(1)}, \ldots, \lambda^{(t)} \in \mathbb{F}_p^n \),

\[
\mathbb{P}_{\lambda^{(1)}, \ldots, \lambda^{(t)}} \left[ \text{OR}(x) \neq 1 - \prod_{i=1}^{t} \left( 1 - (\lambda^{(i)}_1 x_1 + \cdots + \lambda^{(i)}_n x_n)^{p-1} \right) \right] \leq 1/p^t.
\]

Thus, \( \text{OR}(x) \) is approximated with error \( 1/p^t \) on every \( x \in \{0,1\}^n \) by a random polynomial of degree \( t(p-1) \).

For error \( \varepsilon \), we take \( t = \lceil \log_p (1/\varepsilon) \rceil \) and get degree \( pt \leq (\log_p (1/\varepsilon) + 1) \).

Corollary 10. \( \deg_{\varepsilon}(\text{AND}_n) \leq p(\log_p (1/\varepsilon) + 1) \)

Theorem 11. If \( C \) is an \( \mathsf{AC}^0[p] \) circuit of depth \( d \) and size \( S \), then \( \deg_{1/4}(C) \leq O(p \log_p (S))^d \).

Proof. Replace each \( \text{AND}/\text{OR}/\text{MOD}_p \) gate \( g : \{0,1\}^n \to \{0,1\} \) with an \( 1/4S \)-approximating polynomial \( A_g \in \mathbb{F}_p[y_1, \ldots, y_m] \) of degree \( O(p \log_p (S/4)) = O(p \log_p (S)) \). The resulting random polynomial has degree \( O(p \log_p (S))^d \) and approximates \( C \) with error at most \( S \cdot (1/4S) = 1/4 \).

Next lecture we will use this theorem to show:

Theorem 12. Depth-\( d \) \( \mathsf{AC}^0[3] \) circuits for XOR_\( n \) require size \( 2^{\Omega(n^{1/2d})} \).