Håstad's Separation of Constant-Depth Circuits Using Sipser Functions

Iddo Tzameret*

May 22, 2015

Abstract

This note contains a full proof of the exponential separation of depth-d circuits from depth-$(d+1)$ circuits due to Håstad [Has89]. The separating functions are the Sipser functions, denoted $f^{d+1,n}$. We use a simplified proof of Håstad's second switching lemma due to Neil Thapen [Tha09].

Key words and phrases: Circuit complexity, switching lemmas, random restrictions, lower bounds

1 Preliminaries

A *Boolean formula* is a tree with edges directed from the leaves toward the root and whose internal nodes are labeled with the unbounded fan-in connectives $\vee, \wedge, \neg$ (standing for OR, AND and NOT, respectively), and whose leaves are labeled with either $0, 1$ (standing for FALSE and TRUE, respectively) or Boolean variables denoted $x_1, x_2, \ldots$. The *depth* of a formula is the maximal nesting (*not* alternations) of $\wedge, \vee$ connectives. The $\neg$ connective does not increase the depth. Variables and their negation have depth 0. The *size* of a formula is the number of nodes in its underlying tree.

1.1 Chernoff's Bound

Let $X_1, \ldots, X_n$ be n mutually independent indicator random variables (that is, random variables taking $0, 1$ values with probability $\frac{1}{2}$ for each value), and let $\mu = \mathbb{E}\left(\sum_i^n X_i\right)$ be the expectation of their sum. Then, the (variant of the) Chernoff bound we shall use is the following (cf. [MU05]):

$$\mathbf{Pr}\left[\sum_{i=1}^{n} X_i < \mu/2\right] < e^{-\mu/8}$$

*Royal Holloway, University of London. email: *iddo.tzameret@gmail.com*

2 Switching Lemma for Blocks

In this section we follow Neil Thapen's proof of Håstad's second switching lemma [Tha09]. This is a simplification of the (already simplified) version of the original switching lemmas, provided by Razborov [Raz95] (and independently by Alan Woods). The idea governing this type of switching lemma is that we need to come up with a distribution on partial restrictions that will not collapse (with high probability) the Sipser functions (Definition 3.1) into a constant function. We will show that the switching lemma for blocks preserves the Sipser function in Section 4.

Definition 2.1 (Random restrictions $\mathcal{R}^+_{q,B}$ and $\mathcal{R}^-_{q,B}$) *Let $\bar{x}$ be a set of variables and let $0 < q < 1$. Partition the set of variables $\bar{x}$ into pairwise disjoint blocks $B_1, \ldots, B_r$. We define $\rho \in \mathcal{R}^+_q$ as $\rho : \{\overline{x}\} \to \{0, 1, *\}$ in the following way. Independently for every block B_j, $j = 1, ..., r$, do the following:*

(i) with probability q, let $s_j = *$, and otherwise $s_j = 0$;

(ii) for every variable $x_i \in B_j$ independently with probability q let $\rho(x_i) = s_j$, and otherwise $\rho(x_i) = 1$.

Define the random restriction $g(\rho)$ to be ρ such that for every block containing $*$'s (and possibly 1; but no 0), replace all $*$'s to 1's, except for the first $*$ (we assume some ordering, for the sake of definiteness).

The random restriction $\mathcal{R}^-_{q,B}$ is defined similarly, where 0 is flipped with 1.

Definition 2.2 (Query tree) *Let $X := \{x_{i_1}, \ldots, x_{i_k}\}$ be a set of variables (the ordering of the variables will not matter). The* query tree *for X is defined by induction on k as follows. If $k = 0$, then the tree is a single unlabeled node. If $k > 0$, the tree is rooted with a node labeled x_{i_1} from which two edges emanate: one edge labeled 0 and the other edge labeled 1. Each of these two edges leads to the decision tree for $x_{i_2}, \ldots, x_{i_k}$. Every branch starting from the root and ending at a leaf in the query tree determines a unique total truth assignment to the variables X (formed by the answers given to the queries along the branch).*

Definition 2.3 (Decision tree with blocks $\mathsf{btree}(F\restriction\rho)$) *Given a kDNF $F = \bigvee_{i\in I} C_i$ and a restriction ρ from either $\mathcal{R}^+_{q,B}$ or $\mathcal{R}^-_{q,B}$, the decision tree $\mathsf{btree}(F\restriction\rho)$ is defined as the binary tree, constructed by the following queries:*

1. *If for all $i \in I$, $C_i\restriction\rho \equiv 0$, then the tree consists of the single node* labeled 0.
2. *Otherwise, let C_1 be the first conjunct such that $C_1\restriction\rho \not\equiv 0$. Let $\rho^{-1}_{C_1}(*)$ denote the set of variables in C_1 assigned $*$ by ρ, and let $\mathsf{blocks}(\rho^{-1}_{C_1}(*))$ be the list of blocks in which these variables appear.*

   Consider the set of variables $\mathsf{first}^*[\mathsf{blocks}(\rho^{-1}_{C_1}(*))]$ of all the variables x_j that appear in some block in $\mathsf{blocks}(\rho^{-1}_{C_1}(*))$ such that $g(\rho)(x_j) = *$ (in each block that has $*$'s under ρ, there is only one [the first] such variable under $g(\rho)$).

The tree will be a query tree for the variables $\mathsf{first}^*[\mathsf{blocks}(\rho_{C_1}^{-1}(*))]$. *Denote the terminal node in this query tree by* u.

Let π_1 *be the assignment which consists of the answers given to the queries of* $\mathsf{first}^*[\mathsf{blocks}(\rho_{C_1}^{-1}(*))]$ *and that gives value* 1*'s to all the other variables in* $\mathsf{blocks}(\rho_{C_1}^{-1}(*))$ *(that is,* π_1 *gives* 1 *to the variables in* $\mathsf{blocks}(\rho_{C_1}^{-1}(*))$ *excluding those variables in* $\mathsf{first}^*[\mathsf{blocks}(\rho_{C_1}^{-1}(*))]$*).* [1]

3. *For the leaf* u *above: if* $C_1 \restriction \rho\pi_1 = 1$, the label of u is 1*; otherwise,* u *is replaced by the tree induced by returning to step (1) with* $\rho \leftarrow \rho\pi_1$ *(that is,* ρ *is replaced by* $\rho\pi_1$*). The label on a leaf in* $\mathsf{btree}(F\restriction\rho)$ *is said to be the* output *of the tree, given the answers to the queries in the path leading to the leaf.*

Claim 2.1 *Given a tDNF, a restriction* ρ *and a partition of the variables into blocks* B_i*, the tree* $\mathsf{btree}(F\restriction\rho)$ *decides* $F\restriction g(\rho)$ *correctly, in the sense that for every assignment* v*,* $F\restriction g(\rho)v \equiv 1$ *iff* $\mathsf{btree}(F\restriction\rho)$ *outputs* 1 *when the queries are answered according to* v*.*

Proof of claim: The assignments π_1 decided along a branch are compatible with the assignments that are provided by $g(\rho)$: a query is done always on a $*$ variable in $g(\rho)$ (that is, the first $*$ variables x_j such that $\rho(x_j)$ [note that these variables stay $*$ variables under $g(\rho)$ by definition]). Further, all $*$ variables in ρ that are not the first ones in their blocks get the value 1 in π_1, and this is compatible with the definition of $g(\rho)$. Now, the values that $\mathsf{btree}(F\restriction\rho)$ outputs are decided (in parts (1) and (3), in Definition 2.3) by these π_1's and by $C_1\restriction\rho$'s, and so this is equivalent to deciding according to $C_i\restriction g(\rho)$, for some i. $\blacksquare_{\text{Claim}}$

Claim 2.2 *If the height of* $\mathsf{btree}(F\restriction\rho)$ *is at most* s*, then* $F\restriction g(\rho)$ *can be written both as an* s*CNF (namely, a conjunction of clauses of size* s*) and as an* s*DNF (namely, a disjunction of conjuncts of size at most* s*).*

Proof of claim: Each path in $\mathsf{btree}(F \restriction \rho)$ that terminates with a node labeled with 1 corresponds to a conjunct of literals. So, by Claim 2.1, taking the disjunction of all these conjuncts defines correctly $F\restriction g(\rho)$. This gives us an sDNF computing $\mathsf{btree}(F\restriction\rho)$. For the sCNF, we simply flip every leaf in the tree, obtaining a tree deciding $\neg F\restriction g(\rho)$. So we can write $\neg F\restriction g(\rho)$ as an sDNF, and by negating this sDNF (and transforming it into a negation normal form) we get an sCNF that computes $F\restriction g(\rho)$. $\blacksquare_{\text{Claim}}$

The following is the (second) switching lemma by Håstad [Has89], as simplified by Neil Thapen:

[1] Note that assigning 1's to all the variables in $\mathsf{blocks}(\rho_{C_1}^{-1}(*))$ excluding this variables in $\mathsf{first}^*[\mathsf{blocks}(\rho_{C_1}^{-1}(*))]$ is precisely what $g(\rho)$ does, since every block that is assigned $*$ cannot be assigned 0. Thus, π_1 provides (together with ρ) a *complete assignment* to the blocks in $\mathsf{blocks}(\rho_{C_1}^{-1}(*))$. Also note that the only variables to be queried are the $*$ variables from $F \restriction g(\rho)$ (that is, those unset in $F \restriction g(\rho)$). Furthermore, observe that if $C_i \restriction \rho \equiv 0$ we skip it, and if $C_i \restriction \rho \equiv 1$, by definition, we do not query any variable (because there are no $*$'s in it), and so we just arrive at Part (3) and output 1.

Theorem 2.3 (Second switching lemma; switching lemma for blocks) *Let F be a tDNF over the variables $X := \{x_1, \ldots, x_m\}$. Let B_i be a partition of the variables in X into B_i (pairwise disjoint) sets. Then,*

$$\mathbf{Pr}_{\rho \in \mathcal{R}^+_{q,B}} \left[\text{height} \left(\mathsf{btree}(F \restriction \rho) \right) \geq s \right] \leq (13qt)^s$$

The rest of this section is devoted to the proof of this theorem.

Let S be the set of restrictions ρ such that height(btree(F$\restriction\rho$)) > s. We show that the probability of $\rho \in S$, when ρ is chosen from $\mathcal{R}^+_{q,B}$, is at most $(13qt)^s$. We sometimes denote a probability of event like S by $|S|$.

Notation:

- Assume that π is the first (according to some ordering on the branches in the tree) path in $\mathsf{btree}(F\restriction\rho)$ of length $> s$. Suppose that $C_1, C_2, \ldots, C_k$, $\beta_1, \beta_2, \ldots, \beta_k$, $\pi_1, \pi_2, \ldots, \pi_k$ are the terms, lists of blocks, and assignments encountered along π. (Note that π_i is the complete assignment given by $g(\rho)$ to all blocks listed in β_i and by the answers to the queries of the first starred variables in these blocks. [In coding those π_i's via π'_i, we will only code the variables that were queried along the corresponding path.])

- Let σ_i, for $i = 1, \ldots, k$, be the assignments that assign 1 to all the starred variables that appear positive in $C_i \restriction \rho\pi_1 \ldots \pi_{i-1}$, and assign 0 to every other starred variable in every block $B_j \in \beta_i$. Let $\sigma := \sigma_1\sigma_2 \ldots \sigma_k$. Note that σ sets the same variables as $\pi_1\pi_2 \ldots \pi_k$ does.

  Comment 1 *Note that for every block $B_j \in \beta_i$,* only starred variables in B_j may be assigned 0 via σ_i*, and so if we know σ_i and β_i, then we know that every variable that is assigned 0 in σ_i is a starred variable in $B_j \in \beta_i$. Thus, we only need to know what are the starred variables that are assigned 1 under σ_i to recover fully all the starred variables in all $B_j \in \beta_i$. This is why we need γ_i's as follows:*

- Let γ_i, for $i = 1, \ldots, k$, be the set of variables assigned $*$ in $C_i\restriction\rho\pi_1 \ldots \pi_{i-1}$ that appear as positive literals (in $C_i\restriction\rho\pi_1 \ldots \pi_{i-1}$ – all of these variables appear in the blocks β_i).

Encoding.

- β'_i codes β_i. This is done by a list of $\leq 2t$ numbers, as follows. First, note that if we know C_i, then we only need to point on the positions of variables in that term, and this would determine the blocks. (We want to code ρ, but we want to avoid writing ρ explicitly, as otherwise the code would be too long [that is to say, its probability, or "density", would not decrease in our injective mapping, and so we would not get the desired bound on the probability].) Since we can indeed find out in the decoding process what are the C_i's, we can just list the positions of the variables in C_i that come from the blocks in β_i: for every block in $B \in \beta$ we list the position in C_i of the variable that comes from B; further, for each position entry we put a another number indicating whether it is the end of β_i or no. This amounts to $(2t)$ possible codewords for each β_i, and a total of $(2t)^s$ codewords for the whole concatenation of the β_i''s, denoted β'.

- π_i is coded by π_i', which is the string of answers to the queries along $C_i \restriction \rho$. Thus, coding $\pi_1\pi_2 \dots \pi_k$ amounts to 2^s possible codewords.

- γ_i is coded by a string γ_i' of t bits, indicating whether a variable in C_i is starred and appears positive in $C_i \restriction \rho\pi_1 \dots \pi_{i-1}$. Let $\gamma' := \gamma_1' \dots \gamma_k'$. There are 2^{ts} possible codewords for γ'.

Note we do not include C_i's in the coding.

Decoding. We define the mapping

$$\theta : \rho \longrightarrow (\rho\sigma, \beta', \pi_1' \dots \pi_k', \gamma')\,.$$

Claim 2.4 *The mapping θ is injective.*

Proof of claim: We show that given $(\rho\sigma, \beta', \pi_1' \dots \pi_k', \gamma')$ and knowing F and $F \restriction \rho\sigma$, we can deterministically recover the source $\rho = \theta^{-1}\left((\rho\sigma, \beta', \pi_1 \dots \pi_k, \gamma')\right)$.

1. We can find out what is C_1 as follows. Look for the first $C_i \restriction \rho\sigma \not\equiv 0$. If there is no such C_i then it means that there is no $C_i \restriction \rho \not\equiv 0$ (since, σ_i is consistent with C_i by definition). Thus, we get that $s = 0$ in contrast to the assumption. Otherwise, we claim that the C_i found is C_1. This is because, C_i cannot come before C_1 (since then it would mean that $C_i \restriction \sigma \not\equiv C_i \restriction \rho$, but then there must be starred variables in $C_i \restriction \rho$ which means that $C_i \restriction \rho \not\equiv 0$ and so this contradicts the assumption that C_1 is the first with this property). And also, C_i cannot come after C_1, since $C_1 \restriction \rho \not\equiv 0$ implies also $C_1 \restriction \rho\sigma \not\equiv 0$ (as σ is consistent with C_1 by definition).

2. Now that we know C_1, we can recover β_1 out of β': we can recover β_1' as it is encoded directly (recall that along each variable we put one bit indicating if it is the end of β_1'). Then β_1' just codes the position of variables within C_1 – and we can find the corresponding blocks of these variables.

3. Knowing C_1 and γ_1' we find out γ_1 (recall that γ_1' points to the positions of variables appearing positively in $C_1 \restriction \rho$). Therefore, by Comment 1, we can find out who are all the starred variables in the blocks in β_1. But these variables are precisely those variables that were assigned by σ_1, and so by considering $\rho\sigma$ and assigning these variables $*$ we can recover $\rho\rho_2 \dots \rho_k$.

4. This process can be iterated when $C_1 \longleftarrow C_2$, $\rho\sigma \longleftarrow \rho\sigma_2 \dots \sigma_k$, and similarly for $i = 3, \dots, k$, until we fully recover ρ.

■Claim

In what follows, $\mathbf{Pr}[\rho]$ denotes the probability that ρ was chosen from the distribution $\mathcal{R}^+_{q,B}$. Let us note the following facts (which stem directly from the definitsions):

Fact 1

1. *Let σ be an assignment $\sigma : B_i \rightarrow \{*, 0, 1\}$, that assigns a stars and c ones. Then,*

$$\mathbf{Pr}_{\rho \in \mathcal{R}^+_{q,B}}\left[\rho\restriction B_i = \sigma\right] = q^{a+1} \cdot (1-q)^c$$

(choose $s_i = *$ with probability q then choose s_i for a many times and then choose one for c times).

2. *Let σ be an assignment $\sigma : B_i \rightarrow \{*, 0, 1\}$, that assigns b zeros and c ones. Then,*

$$\mathbf{Pr}_{\rho \in \mathcal{R}^+_{q,B}}\left[\rho\restriction B_i = \sigma\right] = (1-q)^{b+1} \cdot (1-q)^c$$

(choose $s_i = 0$ with probability $1-q$ then choose s_i for b many times and then choose one for c times).

3. *Let σ_i's be a collection of assignments $\sigma_i : B_i \rightarrow \{*, 0, 1\}$, and let ρ be the (unique) assignment consistent with all σ_i's. Then,*

$$\mathbf{Pr}[\rho] = \mathbf{Pr}_{\rho' \in \mathcal{R}^+_{q,B}}\left[\bigwedge_i \rho'\restriction B_i = \sigma_i\right] \underset{\substack{\uparrow \\ \textit{By independence of the} \\ \textit{assignments to each block}}}{=} \prod_i \mathbf{Pr}_{\rho' \in \mathcal{R}^+_{q,B}}\left[\rho'\restriction B_i = \sigma_i\right] .$$

4. *Assume that $x_{i_1}, \dots, x_{i_m} \in B_i$ are assigned $*$ by σ'. If $\sigma : \{x_{i_1}, \dots, x_{i_m}\} \rightarrow \{0, 1\}$, where ℓ variables in σ assigned 1 (and $m - \ell$ are assigned 0), then, **in case** $m - \ell \geq 1$:*

$$\mathbf{Pr}_{\rho \in \mathcal{R}^+_{q,B}}\left[\rho\restriction B_i = \sigma'\right] = \mathbf{Pr}\left[\rho\restriction B_i = \sigma\right] \cdot \underbrace{\left(\frac{1}{q}\right)^{\ell}}_{\textit{unselect } \ell\ *\textit{'s}} \cdot \underbrace{(1-q)^{\ell}}_{\textit{instead select } \ell\ 1\textit{'s}} \cdot \underbrace{\frac{1}{q}}_{\substack{\textit{unselect } * \\ \textit{for block } B_i}} \cdot \underbrace{1-q}_{\substack{\textit{instead select } 0 \\ \textit{for block } B_i}} ;$$

*Otherwise, **in case** $m - \ell = 0$:*

$$\mathbf{Pr}\left[\rho\restriction B_i = \sigma'\right] = \mathbf{Pr}\left[\rho\restriction B_i = \sigma\right] \cdot \underbrace{\left(\frac{1}{q}\right)^{\ell}}_{\textit{unselect } \ell\ *\textit{'s}} \cdot \underbrace{(1-q)^{\ell}}_{\textit{instead select } \ell\ 1\textit{'s}} \cdot \underbrace{\frac{1}{q}}_{\substack{\textit{unselect } * \\ \textit{for block } B_i}} .$$

Thus, in both cases we have:

$$\mathbf{Pr}\left[\rho\restriction B_i = \sigma'\right] \geq \mathbf{Pr}\left[\rho\restriction B_i = \sigma\right] \cdot \left(\frac{1-q}{q}\right)^{\ell+1} .$$

From the last item in the Fact above we get:

Corollary 2.5 *Going from ρ to $\rho\sigma$ according to θ, where m is the number of starred variables in ρ assigned either* 0 *or* 1*, we have:*

$$\mathbf{Pr}\left[\rho\sigma\right] \geq \mathbf{Pr}\left[\rho\right] \cdot \left(\frac{1-q}{q}\right)^{m+s} .$$

The following concludes the proof of the theorem (the Second Switching Lemma):

Lemma 2.6

$$\mathbf{Pr}_{\rho\in\mathcal{R}^+_{q,B}}\left[\rho \in S\right] < (13qt)^s .$$

Proof: Fix some β', π' and γ'. Define θ_1 to be the projection of θ on the first coordinate, when θ is restricted to β', π', γ', that is, $\theta_1 : \rho \longrightarrow \rho\sigma$. Assume that m many starred variables in ρ assigned 0 or 1 in σ. Since θ is injective, so does θ_1. Let $S_{\beta',\pi'\gamma'}$ be the *source* of θ_1. By the injectivity of θ_1 we get that each $\rho \in \theta_1$ is mapped to a distinct $\rho\sigma$ (see Figure 1 for an illustration of this). Therefore, we have (we write $\mathbf{Pr}\left[\rho\right]$ to denote $\mathbf{Pr}_{\rho\in\mathcal{R}^+_{q,B}}\left[\rho\right]$):

$$\begin{aligned}
\mathbf{Pr}\left[\rho \in S_{\beta',\pi'\gamma'}\right] &= \sum_{\rho\in S_{\beta',\pi'\gamma'}} \mathbf{Pr}\left[\rho\right] \\
\text{(by Corollary 2.5)} \qquad &\leq \sum_{\rho\in S_{\beta',\pi'\gamma'}} \mathbf{Pr}\left[\theta_1(\rho)\right] \cdot \left(\frac{q}{1-q}\right)^{m+s} \\
&= \left(\frac{q}{1-q}\right)^{m+s} \cdot \underbrace{\sum_{\rho\in S_{\beta',\pi'\gamma'}} \mathbf{Pr}\left[\theta_1(\rho)\right]}_{\substack{\text{by injectivity of } \theta_1 \text{ (Claim 2.4)} \\ \text{the sum is } \leq 1}} ,
\end{aligned}$$

and thus,

$$\mathbf{Pr}\left[\rho \in S_{\beta',\pi'\gamma'}\right] \leq \left(\frac{q}{1-q}\right)^{m+s} . \qquad (1)$$

Equation (1) holds for every β', π', γ', such that m many starred variables in ρ assigned 0 or 1 in σ. The maximal value for such m is ts. Thus, we can compute the probability

$\mathbf{Pr}\left[S_{\beta',\pi'}\right]$ (running over all $|\gamma'| = 0, \ldots, ts$), as follows (assuming that $q < \frac{1}{2t}$):

$$\begin{aligned}
\mathbf{Pr}\left[\rho \in S_{\beta',\pi'}\right] = \sum_{\gamma'} \mathbf{Pr}\left[\rho \in S_{\beta',\pi',\gamma'}\right] &\le \sum_{m=0}^{ts} \sum_{|\gamma'|=m} \mathbf{Pr}\left[\rho \in S_{\beta',\pi',\gamma'}\right] \\
&\le \sum_{m=0}^{ts} \binom{ts}{m} \cdot \left(\frac{q}{1-q}\right)^{m} \cdot \left(\frac{q}{1-q}\right)^{s} \\
&\underset{\substack{\uparrow \\ \text{by binomial expansion}}}{=} \left(1 + \frac{q}{1-q}\right)^{ts} \cdot \left(\frac{q}{1-q}\right)^{s} \\
&\le e^{\frac{qts}{1-q}} \cdot \left(\frac{q}{1-q}\right)^{s} \\
&\underset{\substack{\uparrow \\ \text{Since } q < \frac{1}{2t}}}{\le} \left(\frac{3q}{1-q}\right)^{s}
\end{aligned}$$

If we now consider the sources of S (without restriction to some fixed β', π', γ''s), then we can sum up everything by the inequalities above:

$$\mathbf{Pr}\left[\rho \in S\right] \le (2t)^s 2^s \left(\frac{3q}{1-q}\right)^{s} = \left(\frac{12qt}{1-q}\right)^{s}$$

giving the desired result. ■

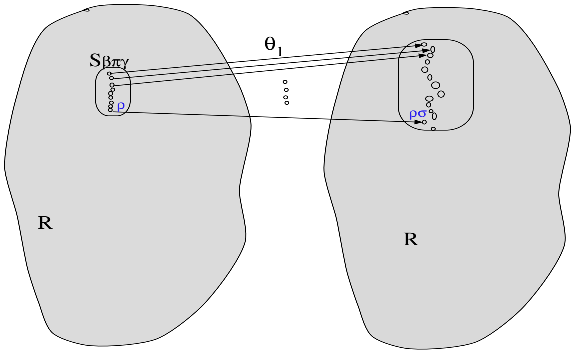

Figure 1: The θ_1 mapping. The mapping θ_1 maps (injectively) an assignment ρ to an extension of it $\rho\sigma$ *which has a bigger probability in* $\mathcal{R}^{+}_{q,B}$ (abbreviated by R in the figure). Thus, θ_1 maps (the event that ρ is taken from) $S_{\beta',\pi'\gamma'}$ into an event with a bigger probability.

3 The Sipser Functions

Definition 3.1 (Sipser function $f^{d,n}$) *A Sipser function $f^{d,n}$ is a d-layered formula over the variables $x_{i_1,\ldots,i_d}$, for $(i_1 \ldots i_d) \in [n]^d$, denoted X, whose form is:*

$$\bigwedge_{i_1=1}^{\sqrt{\frac{n}{\log n}}} \bigvee_{i_2=1}^{n} \cdots \bigwedge_{i_d=1}^{\sqrt{\frac{1}{2}dn\log n}} x_{i_1,\ldots,i_d}\,, \quad \textit{if } d \textit{ is odd, and}$$

$$\bigwedge_{i_1=1}^{\sqrt{\frac{n}{\log n}}} \bigvee_{i_2=1}^{n} \bigwedge_{i_3=1}^{n} \cdots \bigvee_{i_d=1}^{\sqrt{\frac{1}{2}dn\log n}} x_{i_1,\ldots,i_d}\,, \quad \textit{if } d \textit{ is even}.$$

Note that every variable appears once in (the bottom level of) a Sipser function. The number of variables in $f^{d,n}$, is

$$m = \sqrt{\frac{n}{\log n}} \cdot n^{d-2} \cdot \sqrt{\frac{1}{2}dn\log n} = n^{d-1} \cdot \sqrt{d/2}\,.$$

4 Function Preserving

Definition 4.1 *A $\bigwedge_d^{S,t}$ formula is a formula of depth at most $d+1$, and if it is of depth exactly $d+1$ then the uppermost connective is $\wedge$, and such that the following hold:*

(i) The bottom level connectives are of fan-in at most t;

(ii) The total number of connectives above the bottom level (that is, the connectives excluding the bottom level) is at most S.

Theorem 4.1 (Function preserving) *Let C be an $\bigwedge_d^{S,t}$ circuit computing $f^{d,n}$ and assume that B_i is a partition of the variables in C that corresponds to the bottom level variables in $f^{d,n}$. Let $q = \frac{1}{\sqrt{\frac{n}{2d\log n}}}$ and suppose that $\rho \in \mathcal{R}^+_{q,B}$. Then, with probability $\geq 2/3$, $C\restriction g(\rho)$ contains a copy of a circuit that computes the function $f^{d-1,n}$ (that is, by setting some [possibly none] of the variables of $C\restriction g(\rho)$ we get a circuit computing $f^{d-1,n}$).*

Proof of Theorem 4.1. We consider the defining circuit of $f^{d,n}$. We apply to this circuit a random restriction and we claim that the circuit would contain with high probability a Sipser function of depth $d-1$. Thus, for C that computes $f^{d,n}$ (which may be completely different from the defining circuit of $f^{d,n}$) it must be that $C\restriction g(\rho)$ *will also compute the Sipser function of depth* $d-1$ (note that if two different circuits compute the same function then restricting them to the same partial assignment will result in two circuits that compute the same function).

We consider the case where d is odd (the other case where d is even is similar). Thus, in $f^{d,n}$ the $d-1$ level is an OR of ANDs, that is, a DNF with bottom fan-in $\sqrt{nd\log n \cdot \frac{1}{2}}$.

Claim 4.2 *The probability that an AND gate C_i in the bottom level of $f^{d,n}$ is not assigned s_i under $g(\rho)$ is:*[2]

$$(1-q)^{|B_i|} \le e^{-d\log n}\,.$$

Proof of claim: Note that the event that a conjunct C_i is not assigned an s_i is the same event that C_i is assigned 1 which is the same event that *every variable in* C_i is assigned a 1. This happens with probability $(1-q)^{|B_i|}$. We have:

$$(1-q)^{|B_i|} = (1-\frac{1}{\sqrt{\frac{n}{2d\log n}}})^{\sqrt{\frac{1}{2}dn\log n}}\,.$$

Since $(1-\frac{1}{x})^y = (1-\frac{1}{x})^{x\cdot\frac{y}{x}} \le e^{-y/x}$, then $(1-q)^{|B_i|} \le e^{-\sqrt{\frac{1}{2}dn\log n}\cdot\sqrt{\frac{2d\log n}{n}}} = e^{-\sqrt{dn\log n\cdot\frac{d\log n}{n}}} = e^{-d\log n}$. $\blacksquare_{\text{Claim}}$

Corollary 4.3 *The probability that* all *blocks B_i in the bottom level (there are approximately n^{d-1} such blocks) are assigned s_i is at least* $5/6$.

Proof: By the previous claim, the probability that a single block B_i is *not assigned s_i* is $\le e^{-d\log n}$, and so by the union bound we get that the probability that *all* (at most) n^{d-1} blocks B_i are not assigned s_i is $\le e^{-d\log n}\cdot n^{d-1} \le n^{-d}\cdot n^{d-1} < 1/6$. [3] $\blacksquare$

Lemma 4.4 *Let Υ be the event that every block B_i in the bottom level is assigned s_i under $g(\rho)$. Then, conditioned on the event Υ, with probability at least* $5/6$ all *OR gates at the $d-2$ level are assigned at least $\sqrt{\frac{1}{2}(d-1)n\log n}$ inputs that are $*$ (that is, each of the n many OR gates have at least that much [unassigned] input variables).*

Proof: We need to show the probability of the event that for *every OR* (out of the $< n^{d-2}$ such gates) in the second level from below, there are at least $\sqrt{\frac{1}{2}dn\log n}$ variables (that is, $*$). First, we determine the probability that for some *fixed* OR gate there are *less than* $\sqrt{\frac{1}{2}dn\log n}$ variables (that is, $*$). Then, we use the union bound to determine the probability that none of these "bad" events happen (which is the probability we need to determine).

Claim 4.5 *For any given OR gate, with probability at most $e^{-\sqrt{\frac{1}{32}dn\log n}}$ there are* less than *$\sqrt{\frac{1}{2}(d-1)n\log n}$ variables (that is, $*$) coming into this OR gate.*

[2] In this case, all literals in C_i are assigned 1; and hence, also the OR gate that leads to C_i is 1 under the restriction.

[3] It is unclear why [Has89] multiplies in n^d, and then gets the 1/6 bound. I believe that the probability is polynomially small, and not just constant, because the number of blocks is smaller than n^d. However, the former is sufficient for the proof, of course.

Proof of claim: By Chernoff's bound. Let X be the sum of n independent random variables with success probability $q = \frac{1}{\sqrt{\frac{n}{2d\log n}}}$. The expectation of X is

$$\mu = \mathbb{E}(X) = \frac{n}{\sqrt{\frac{n}{2d\log n}}} = \sqrt{2dn\log n},$$

and so

$$\sqrt{\frac{1}{2}dn\log n} = \sqrt{\frac{2}{4}dn\log n} = \frac{\sqrt{2dn\log n}}{2} = \frac{\mu}{2}.$$

Thus, by Chernoff's bound:

$$\mathbf{Pr}\left[X < \sqrt{\frac{1}{2}dn\log n}\right] < e^{-\frac{\mu}{8}} = e^{-\sqrt{\frac{1}{32}dn\log n}}.$$

Since $\mathbf{Pr}\left[X < \sqrt{\frac{1}{2}(d-1)n\log n}\right] < \mathbf{Pr}\left[X < \sqrt{\frac{1}{2}dn\log n}\right]$ we conclude the claim. $\blacksquare_{\text{Claim}}$

Now, by the union bound, and using the previous claim, the probability that the event in the statement does not hold is at most $n^{d-2} \cdot e^{-\sqrt{\frac{1}{32}dn\log n}} = o(1)$. Thus, for sufficiently large n, the event in the statement holds with probability at least $5/6$.

$\blacksquare$

The following claim is sufficient to complete the proof of Theorem 4.1:

Claim 4.6 *With probability at least* $2/3$, every *OR gate in the level* $d-2$ *in* $f^{d,n}$ *is (logically equivalent) to an OR of* $\geq \sqrt{\frac{1}{2}(d-1)n\log n}$ *variables (where each OR has distinct and disjoint sets of input variables), under* $g(\rho)$*, for sufficiently large* n*.*

Proof of claim: By Corollary 4.3, with probability $\geq 5/6$ every block B_i in the bottom level is assigned s_i under $g(\rho)$. Conditioned on this event, by Lemma 4.4, with probability $\geq 5/6$ every OR gate in level $d-2$ has $\geq \sqrt{\frac{1}{2}(d-1)n\log n}$ input variables (for sufficiently large n). Thus (since $\mathbf{Pr}[A \;\&\; B] = \mathbf{Pr}[A|B] \cdot \mathbf{Pr}[B]$), the probability of the event in the statement is $\geq (\frac{5}{6})^2 > 2/3$, for sufficiently large n. $\blacksquare_{\text{Claim}}$

5 The Lower Bound

Theorem 5.1 (Main) *Let* $c_d = \frac{1}{27\sqrt{2d}}$ *and let* $t \leq c_d\sqrt{\frac{n}{\log n}}$*. Then, for every* $\bigvee_d^{S,t}$ *formula computing* $f^{d,n}$*,* $S \geq 2^{c_d\sqrt{\frac{n}{\log n}}}$*.*

Proof: By induction on d.
Base case: $d = 2$.

Claim 5.2 *For any S, there is no $\bigvee_1^{S,t}$ (that is, a tDNF) formula and no $\bigwedge_1^{S,t}$ (that is, tCNF) formula computing $f^{2,n}$, for $t \le c_d\sqrt{\frac{n}{\log n}}$, if $c_d < 1$. (Note that a tDNF has only one node in its second level, and so the size measure S has no effect here.)*

Proof of claim: We need to show that there is no tDNF with $t \le c_d\sqrt{n/\log n}$ that computes the function $f^{2,n}$, that is, $\bigwedge_{i_1=1}^{\sqrt{\frac{n}{\log n}}}\bigvee_{i_2=1}^{\sqrt{n\log n}} x_{i_1,i_2}$ (when $c_d < 1$). An assignment α satisfies $f^{2,n}$ iff for all $1 \le i_1 \le \sqrt{n/\log n}$ there exists $1 \le i_2 \le \sqrt{n\log n}$ such that x_{i_1,i_2} is assigned 1 under α. But this means that every minterm of $f^{2,n}$ is of size $\ge \sqrt{n/\log n} > c_d\sqrt{n/\log n} = t$ (because $c_d < 1$).

Similarly, we can show that there is no tCNF with $t \le c_d\sqrt{n/\log n}$ that computes $f^{2,n}$.
$\blacksquare_{\text{Claim}}$

Induction step: Assume, by a way of contradiction that there exists a $\bigvee_d^{S,t}$ circuit computing $f^{d,n}$, where $t \le c_d\sqrt{n/\log n}$ and $S < 2^{c_d\sqrt{n/\log n}}$. Assume that in level $d-1$ there are tDNFs (the case for tCNFs is similar). Apply a random restriction from $\mathcal{R}^+_{q,n}$, with $q = \frac{1}{\sqrt{\frac{n}{2d\log n}}}$ and with each block B_i corresponding to variables in a bottom AND gate. Let $s = c_{d-1}\sqrt{\frac{n}{\log n}}$. By the switching lemma (Theorem 2.3), for each bottom tDNF denoted F, with probability $\le (13qt)^s$ we cannot write $F\!\restriction g(\rho)$ as an sCNF. Thus, by the union bound and since by assumption the number of gates in level $d-1$ is at most S, with probability $\le S\cdot(13qt)^s$ there exists (at least one) bottom tDNF, F, such that $F\!\restriction g(\rho)$ cannot be written as an sCNF. In other words, with probability at least

$$\begin{aligned}
1 - S\cdot(13qt)^s &> 1 - 2^{c_d\sqrt{\frac{n}{\log n}}}\cdot\left(13\cdot\frac{1}{\sqrt{\frac{n}{2d\log n}}}\cdot c_d\sqrt{\frac{n}{\log n}}\right)^{c_{d-1}\sqrt{\frac{n}{\log n}}}\\
&= 1 - 2^{c_d\sqrt{\frac{n}{\log n}}}\cdot\left(13\cdot\sqrt{2d}\cdot c_d\right)^{c_{d-1}\sqrt{\frac{n}{\log n}}}\\
&= 1 - 2^{\frac{1}{27\sqrt{2d}}\cdot\sqrt{\frac{n}{\log n}}}\cdot\left(13\cdot\sqrt{2d}\cdot\frac{1}{27\sqrt{2d}}\right)^{\frac{1}{27\sqrt{2(d-1)}}\sqrt{\frac{n}{\log n}}}\\
&\underset{\substack{\uparrow\\ \frac{1}{27\sqrt{2d}}<\frac{1}{27\sqrt{2(d-1)}}}}{>} 1 - \left(\frac{26}{27}\right)^{\frac{1}{27\sqrt{2d}}\cdot\sqrt{\frac{n}{\log n}}}\\
&\underset{\substack{\uparrow\\ \text{for } n \text{ big enough}}}{>} \frac{1}{6}
\end{aligned}$$

we can transform all the bottom tDNFs into sCNFs, and then merge the AND gates in the $d-2$ and $d-1$ levels into one level. This yields a $\bigvee_{d-1}^{S,t}$ circuit C , where $S < 2^{c_d\sqrt{\frac{n}{\log n}}}$ and $t \le c_d\frac{n}{\log n}$. But, by Theorem 4.1, after applying the random restriction $g(\rho)$ on the circuit, with probability 5/6 the circuit contains the Sipser function $f^{d-1,n}$. Since $S < 2^{c_d\sqrt{\frac{n}{\log n}}} < 2^{c_{d-1}\sqrt{\frac{n}{\log n}}}$ and $t \leq c_d\sqrt{\frac{n}{\log n}} < c_{d-1}\sqrt{\frac{n}{\log n}}$, we arrive at a contradiction to induction hypothesis. We thus conclude that there is no $\bigvee_d^{S,t}$ circuit computing $f^{d,n}$, where $t \leq c_d\sqrt{n/\log n}$ and $S < 2^{c_d\sqrt{n/\log n}}$. ■

Acknowledgments

I wish to thank Wang Zhengyu for referring my attention to several typos and inaccuracies in this note.

A Parameters

number of variables in $f^{d,n}$	$n^{d-1} \cdot \sqrt{d/2}$	(2)
top fan-in for $f^{d,n}$	$\sqrt{\frac{n}{\log n}}$	(3)
bottom fan-in for $f^{d,n}$	$\sqrt{\frac{1}{2}dn\log n}$	(4)
number of blocks B_i in $f^{d,n}$	$n^{d-2} \cdot \sqrt{\frac{n}{\log n}}$	(5)
probability of choosing $s_i = *$	$q = \frac{1}{\sqrt{\frac{n}{2d\log n}}}$	(6)
probability of choosing $x_j = s_i$, for $x_j \in B_i$	″	(7)
lower bound for bottom fan-in in depth-d circuits	$c\sqrt{\frac{n}{\log n}}$	(8)
lower bound for gates in depth ≥ 2 in depth-d circuits	$2^{c\sqrt{\frac{n}{\log n}}}$	(9)
	$c = \frac{1}{27\sqrt{2d}}$	(10)
tDNF switchs to sCNF with probability $\leq$	$(13qt)^s$	(11)

References

[Has89] Johan Haståd. *Advances in Computer Research*, volume 5, chapter Almost optimal lower bounds for small depth circuits, pages 143–170. JAI Press, 1989. (document), 2, 3

[MU05] Michael Mitzenmacher and Eli Upfal. *Probability and Computing – Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005. 1.1

[Raz95] Alexander A. Razborov. Unprovability of lower bounds on circuit size in certain fragments of bounded arithmetic. *Izv. Ross. Akad. Nauk Ser. Mat.*, 59(1):201–224, 1995. 2

[Tha09] Neil Thapen. Note on switching lemmas. *manuscript*, Feb. 2009. (document), 2