The Average Sensitivity of Bounded-Depth Formulas

Benjamin Rossman
University of Toronto[*]

July 25, 2016

Abstract

We show that unbounded fan-in boolean formulas of depth $d + 1$ and size s have average sensitivity $O(\frac{1}{d} \log s)^d$. In particular, this gives a tight $2^{\Omega(d(n^{1/d}-1))}$ lower bound on the size of depth $d+1$ formulas computing the PARITY function. These results strengthen the corresponding $2^{\Omega(n^{1/d})}$ and $O(\log s)^d$ bounds for circuits due to Håstad (1986) and Boppana (1997). Our proof technique studies a random process where the Switching Lemma is applied to formulas in an efficient manner.

1 Introduction

We consider boolean circuits with unbounded fan-in AND and OR gates and negations on inputs. *Formulas* are the class of tree-like circuits in which all gates have fan-out 1. *Size* of circuits (including formulas) is measured by the total number of gates. *Depth* is the maximum number of gates on an input-to-output path.

Lower bounds against bounded-depth circuits were first proved in the 1980s [1, 4, 9, 5], culminating in a tight size-depth tradeoff for circuits computing the PARITY function. The technique, based on random restrictions, applies more generally to boolean functions with high average sensitivity.

Theorem 1 (Håstad [5]). *Depth $d + 1$ circuits computing* PARITY *have size $2^{\Omega(n^{1/d})}$.*

Theorem 2 (Boppana [2]). *Depth $d + 1$ circuits of size s have average sensitivity $O(\log s)^d$.*

In this paper, we prove stronger versions of these results for bounded-depth formulas:

Theorem 3. *Depth $d + 1$ formulas computing* PARITY *have size $2^{\Omega(d(n^{1/d}-1))}$.*

Theorem 4. *Depth $d + 1$ formulas of size s have average sensitivity $O(\frac{1}{d} \log s)^d$.*

Theorems 3 and 4 directly strengthen Theorems 1 and 2 in light of the following

Fact 5. *Every depth $d+1$ circuit of size s is equivalent to a depth $d+1$ formula of size at most s^d.*

Theorems 1, 2, 3, 4 are asymptotically tight, since PARITY is computable by depth $d+1$ circuits (resp. formulas) of size $n2^{O(n^{1/d})}$ (resp. $2^{O(d(n^{1/d}-1))}$).

The main tool in the proof of Theorems 1 and 2 is Håstad's Switching Lemma [5]. The Switching Lemma states that every small-width CNF or DNF simplifies, with high probability under a random restriction, to a small-depth decision tree. This yields lower bounds against bounded-depth *circuits* via a straightforward depth-reduction argument. In this paper we show how the Switching Lemma can be applied more efficiently to bounded-depth *formulas*, though in a less straightforward manner.

In more detail: for independent uniformly distributed random $\sigma \in \{0,1\}^n$ ("assignment") and $\tau \in [0,1]^n$ ("timestamp"), we consider the family of restrictions $\{R_p^{\sigma,\tau}\}_{0 \le p \le 1}$ (i.e. functions $[n] \to \{0,1,*\}$ representing partial assignments to input variables $x_1, \ldots, x_n$) where $R_p^{\sigma,\tau}$ sets the variable x_i to σ_i if $\tau_i < p$ and leaves x_i unset if $\tau_i \ge p$. In the usual application of the Switching Lemma to circuits of depth $d+1$, all subcircuits of depth $k+1$ are hit with the restriction $R_{p_k}^{\sigma,\tau}$ for a fixed sequence $p_1 > \cdots > p_d$ (typically $p_k = n^{-k/(d+1)}$). In this paper we achieve sharper bounds against formulas by hitting each subformula Φ with the restriction $R_{\mathbf{q}(\Phi)}^{\sigma,\tau}$ where the parameter $\mathbf{q}(\Phi)$ ($= \mathbf{q}^{\sigma,\tau}(\Phi)$) is defined inductively, according to a random process indexed by subformulas of Φ. Our technical main theorem is a tail bound on $\mathbf{q}(\Phi)$, viewed as a random variable determined by σ and τ.

After preliminary definitions in §2, we state and prove our technical main theorem in §3 and §4. As a corollaries, we derive Theorem 3 in §5 and Theorem 4 in §6. In §7 we state a further corollary of our results on the relative power of formulas vs. circuits.

2 Preliminaries

$\mathbb{N} = \{0,1,2,\ldots\}$. $[n] = \{1,\ldots,n\}$. $\exp(\lambda) = e^\lambda$.

2.1 Formulas

A *formula* is a finite rooted tree whose leafs ("inputs") are labeled by literals (i.e. variables x_i or negated variables $\neg x_i$) and whose non-leafs ("gates") are labeled by AND or OR. (Gates have unbounded fan-in.) Every formula Φ computes a boolean function on the same set of variables.

The *size* of a formula Φ, denoted by $|\Phi|$, is the number of gates in Φ. (Note that every lower bound on *size* is also a lower bound on *leaf-size*, i.e., the number of leaves in a formula.) The *depth* of Φ is the maximum number of gates on an input-to-output path. Formulas of depth 0 are literals; formulas of depth 1 are clauses (i.e. an AND or OR of literals). We are often interested in formulas of depth ≥ 2 and speak of "depth $d+1$" where d is an arbitrary positive integer.

2.2 Boolean functions and restrictions

A *restriction* is a function $\varrho : [n] \to \{0,1,*\}$, viewed as a partial assignment of boolean input variables $x_1, \ldots, x_n$ to 0, 1 or $*$ (meaning "unset"). For a boolean function $f : \{0,1\}^n \to \{0,1\}$, the restricted function $f{\restriction}\varrho : \{0,1\}^{\varrho^{-1}(*)} \to \{0,1\}$ is defined in the usual way. For $p \in [0,1]$, we write $\mathcal{R}_p$ for the distribution on restrictions ϱ where $\mathbb{P}[\varrho(i) = *] = p$ and $\mathbb{P}[\varrho(i) = 0] = \mathbb{P}[\varrho(i) = 1] = (1-p)/2$ independently for all $i \in [n]$.

2.3 Average sensitivity and decision-tree depth

The *average sensitivity* $\mathsf{as}(f)$ of a boolean function f is the expected number of input bits that, when flipped, change the output of f, starting with a random input assignment.

The *decision-tree depth* $\mathsf{D}(f)$ of f is the minimum depth of a decision tree which computes f; in particular, $\mathsf{D}(f) = 0$ iff f is constant. Two elementary facts which we will use later (see [2]): for every boolean function f,

(1) $$\mathsf{as}(f) \leq \mathsf{D}(f) \qquad \text{(i.e. average sensitivity is at most decision-tree depth)},$$

(2) $$\mathop{\mathbb{E}}_{\varrho \sim \mathcal{R}_p} [\, \mathsf{as}(f{\restriction}\varrho) \,] = p{\cdot}\mathsf{as}(f) \quad \text{for all } 0 \leq p \leq 1.$$

Håstad's Switching Lemma relates random restrictions and decision-tree depth. We give a somewhat nonstandard statement (the usual statement is in terms of width-k CNFs and width-ℓ DNFs).

Lemma 6 (Switching Lemma [5]). *Let $k, \ell \in \mathbb{N}$. Suppose f is the* AND *or* OR *of an arbitrary family $\{f_i\}$ of boolean functions with $\mathsf{D}(f_i) \leq k$ for all i. Then for all $0 \leq p \leq \frac{1}{2}$,*

$$\mathop{\mathbb{P}}_{\varrho \sim \mathcal{R}_p} [\, \mathsf{D}(f{\restriction}\varrho) \geq \ell \,] \leq (5pk)^\ell.$$

3   A random process associated with formulas

Definition 7. Let $\sigma \in \{0,1\}^n$ ("assignment") and $\tau \in [0,1]^n$ ("timestamp") be independent uniformly distributed random variables. For $0 \leq p \leq 1$, let $R_p^{\sigma,\tau} : [n] \to \{0,1,*\}$ be the restriction

$$R_p^{\sigma,\tau}(i) := \begin{cases} \sigma_i & \text{if } \tau_i > p, \\ * & \text{if } \tau_i \leq p. \end{cases}$$

We regard the family of restrictions $\{R_p^{\sigma,\tau}\}_{0 \leq p \leq 1}$ as a stochastic process where the parameter p represents a "time" which starts at 1 and decreases to 0. At the initial time $p = 1$, the assignment σ is fully masked (i.e. $R_1^{\sigma,\tau}$ is all $*$'s). As p decreases, the values of σ are gradually unmasked, until the final time $p = 0$ when σ is fully revealed (i.e. $R_0^{\sigma,\tau} = \sigma$). Of course, for any fixed p, $R_p^{\sigma,\tau}$ is simply a random restriction with distribution $\mathcal{R}_p$.

Definition 8 (Main Definition). For all formulas Φ, we define the "stopping time" $\mathbf{q}^{\sigma,\tau}(\Phi) \in [0,1]$ by the following induction:

- If Φ has depth 0 (i.e. Φ is a variable or negated variable), then $\mathbf{q}^{\sigma,\tau}(\Phi) := 1$.

- If Φ is $\text{AND}(\Psi_1, \ldots, \Psi_m)$ or $\text{OR}(\Psi_1, \ldots, \Psi_m)$, then

$$\mathbf{q}^{\sigma,\tau}(\Phi) := \frac{\mathbf{p}^{\sigma,\tau}(\Phi)}{14{\cdot}\mathbf{k}^{\sigma,\tau}(\Phi)}$$

where $\qquad \mathbf{p}^{\sigma,\tau}(\Phi) := \min_i \mathbf{q}^{\sigma,\tau}(\Psi_i), \qquad \mathbf{k}^{\sigma,\tau}(\Phi) := \max\{1, \max_i \mathsf{D}(\Psi_i{\restriction}R_{\mathbf{p}^{\sigma,\tau}(\Phi)}^{\sigma,\tau})\}.$

For the sake of readability, we will suppress σ and τ whenever possible and simply write $\mathbf{q}(\Phi)$, $\mathbf{p}(\Phi)$, $\mathbf{k}(\Phi)$. However, the reader should keep in mind that these random variables are determined, for all formulas Φ, by a single pair of σ of τ. (We will continue to write σ and τ when referring to restrictions $R_p^{\sigma,\tau}$.)

We view $\mathbf{q}(\Phi)$ as the stopping time for a stochastic process indexed by formulas Φ. For Φ of depth 0, $\mathbf{q}(\Phi)$ is the initial time 1 (when all variables are masked). For Φ of depth ≥ 1, $\mathbf{q}(\Phi)$ is defined in terms of two auxiliary parameters:

- $\mathbf{p}(\Phi)$ is the most advanced (i.e. minimum) stopping time $\mathbf{q}(\Psi)$ among children Ψ of Φ.

- $\mathbf{k}(\Phi)$ is the maximum decision-tree depth among children Ψ of Φ upon being hit with the restriction $R^{\sigma,\tau}_{\mathbf{p}(\Phi)}$. (For technical reasons, we set $\mathbf{k}(\Phi) = 1$ in the event that $D(\Psi{\upharpoonright}R^{\sigma,\tau}_{\mathbf{p}(\Phi)}) = 0$ for all Ψ.)

If Φ is an AND (resp. OR), then $\Phi{\upharpoonright}R^{\sigma,\tau}_{\mathbf{p}(\Phi)}$ is a $\mathbf{k}(\Phi)$-CNF (resp. DNF). The choice of definition $\mathbf{q}(\Phi) = \mathbf{p}(\Phi)/14{\cdot}\mathbf{k}(\Phi)$ allows us to apply the Switching Lemma to $\Phi{\upharpoonright}R^{\sigma,\tau}_{\mathbf{p}(\Phi)}$. This is made precise by the following lemma. (Since the dependence on σ and τ is crucial here, we use explicit notation: $\mathbf{q}^{\sigma,\tau}(\Phi)$, etc.)

Lemma 9. *Let Φ be a formula of depth ≥ 1 and let $q \in \mathrm{Supp}(\mathbf{q}^{\sigma,\tau}(\Phi))$ (i.e. $q = \mathbf{q}^{\sigma,\tau}(\Phi)$ for some $\sigma \in \{0,1\}^n$ and $\tau \in [0,1]^n$). Then for all $0 \leq \alpha \leq 1$ and $\ell \in \mathbb{N}$,*

$$\mathbb{P}_{\sigma,\tau}\left[\; D(\Phi{\upharpoonright}R^{\sigma,\tau}_{\alpha q}) \geq \ell \;\Big|\; \mathbf{q}^{\sigma,\tau}(\Phi) = q \;\right] \leq \left(\frac{\alpha}{e}\right)^{\ell}.$$

Proof. Fix Φ and q as in the hypothesis of the lemma. Since Φ has depth ≥ 1, it is the AND or OR of formulas Ψ_i. Let

$$I := \left\{ (p, \varrho, k) : \begin{array}{l} q = p/14k \text{ and there exist } \sigma \in \{0,1\}^n \text{ and } \tau \in [0,1]^n \\ \text{such that } \mathbf{p}^{\sigma,\tau}(\Phi) = p,\ R^{\sigma,\tau}_p = \varrho \text{ and } \mathbf{k}^{\sigma,\tau}(\Phi) = k \end{array} \right\}.$$

Note that I is nonempty and indexes a partition of the event $\{\mathbf{q}^{\sigma,\tau}(\Phi) = q\}$ into subevents $\{\mathbf{p}^{\sigma,\tau}(\Phi) = p,\ R^{\sigma,\tau}_p = \varrho \text{ and } \mathbf{k}^{\sigma,\tau}(\Phi) = k\}$.

To prove the lemma, consider any $(p, \varrho, k) \in I$. Conditioning on this subevent, we can view $R^{\sigma,\tau}_{\alpha q}$ as the composition of ϱ and an independent random restriction $\theta \sim \mathcal{R}_{\alpha/14k}$. Since $\Phi{\upharpoonright}\varrho$ is an AND or OR of functions $\Psi_i{\upharpoonright}\varrho$ of decision-tree depth $\leq k$, Lemma 6 implies

$$\mathbb{P}_{\sigma,\tau}\left[\; D(\Phi{\upharpoonright}R^{\sigma,\tau}_{\alpha q}) \geq \ell \;\Big|\; \mathbf{p}^{\sigma,\tau}(\Phi) = p,\ R^{\sigma,\tau}_p = \varrho \text{ and } \mathbf{k}^{\sigma,\tau}(\Phi) = k \;\right]$$

$$= \mathbb{P}_{\theta\sim\mathcal{R}_{\alpha/14k}}\left[\; D((\Phi{\upharpoonright}\varrho){\upharpoonright}\theta) \geq \ell \;\right] \leq \left(5\left(\frac{\alpha}{14k}\right)k\right)^{\ell} \leq \left(\frac{\alpha}{e}\right)^{\ell}. \qquad \square$$

4 Tail bound on $\mathbf{q}(\Phi)$

Our technical main theorem is a tail bound on the random variable $\mathbf{q}(\Phi)$ ($= \mathbf{q}^{\sigma,\tau}(\Phi)$) where the randomness is over independent uniform $\sigma \in \{0,1\}^n$ and $\tau \in [0,1]^n$. We state the result first with asymptotic notation.

Theorem 10. *For every depth $d + 1$ formula Φ and $0 < \lambda \leq 1$,*

$$\mathbb{P}\left[\; \mathbf{q}(\Phi) \leq \lambda \;\right] \leq \frac{|\Phi|}{\exp(\Omega(d\lambda^{-1/d}) - O(d))}.$$

4

In order to have a useable induction hypothesis, we restate Theorem 10 with explicit constants:

Theorem 10 (more precisely). *For every depth $d+1$ formula Φ and $\ell > 0$,*

$$\mathbb{P}\left[\ \mathbf{q}(\Phi) \leq \frac{1}{14^{d+1}\ell}\ \right] \leq |\Phi|\frac{C^d}{\exp(\mathrm{e}^{-2}d\ell^{1/d})}$$

where $C = 1 + \sum_{i=0}^{\infty}\left(\dfrac{1}{\exp(\mathrm{e}^{i-1} - (i+1)\mathrm{e}^{-2})} + \sum_{j=0}^{\infty}\dfrac{1}{\exp((j+1)\mathrm{e}^{i-1} - (i+j+2)\mathrm{e}^{-2})}\right) \approx 7.83.$

Proof. We first note that the theorem is trivial if $\ell < \mathrm{e}^d$ (as the RHS is $> (C/\exp(\mathrm{e}^{-1}))^d > 1$ since $C > \exp(\mathrm{e}^{-1})$). Therefore, we assume that $\ell \geq \mathrm{e}^d$. We argue by induction on d.

Consider the base case $d = 1$ where Φ is a depth 2 formula. Note that $\mathbf{q}(\Psi) = 1/14$ for each depth 1 subformula Ψ of Φ; hence $\mathbf{p}(\Phi) = 1/14$. Also, each Ψ is the AND or OR of decision-trees of depth 1; so by Lemma 6,

$$\mathop{\mathbb{P}}_{\sigma,\tau}\left[\ \mathsf{D}(\Psi{\upharpoonright}R^{\sigma,\tau}_{1/14}) \geq \ell\ \right] = \mathop{\mathbb{P}}_{\varrho\sim\mathcal{R}_{1/14}}\left[\ \mathsf{D}(\Psi{\upharpoonright}\varrho) \geq \ell\ \right] \leq \left(\frac{1}{\mathrm{e}}\right)^{\ell}.$$

Since $\mathbf{q}(\Phi) = \mathbf{p}(\Phi)/14{\cdot}\mathbf{k}(\Phi) = 1/14^2{\cdot}\mathbf{k}(\Phi)$, we have

$$\mathbb{P}\left[\ \mathbf{q}(\Phi) \leq \frac{1}{14^2\ell}\ \right] = \mathbb{P}\left[\ \mathbf{k}(\Phi) \geq \ell\ \right] = \mathbb{P}\left[\ \bigvee_{\Psi}\mathsf{D}(\Psi{\upharpoonright}R^{\sigma,\tau}_{\mathbf{p}(\Phi)}) \geq \ell\ \right]$$

$$\leq \sum_{\Psi}\mathbb{P}\left[\ \mathsf{D}(\Psi{\upharpoonright}R^{\sigma,\tau}_{1/14}) \geq \ell\ \right]$$

$$\leq |\Phi|\frac{1}{\exp(\ell)} < |\Phi|\frac{C^d}{\exp(\mathrm{e}^{-2}d\ell^{1/d})}.$$

For the induction step, let $d \geq 2$ and assume the theorem holds for $d-1$. Let Φ be a formula of depth $d+1$. Let Ψ range over depth-d subformulas of Φ. In particular, we have $|\Phi| = 1 + \sum_{\Psi}|\Psi|$.

We will define a family of events denoted $\mathcal{A}$ and $\mathcal{B}_i$ $(i \in \mathbb{N})$ and $\mathcal{C}_{i,j}$ $(i,j \in \mathbb{N})$ and show that the union of these events covers the event $\{\mathbf{q}(\Phi) \leq \frac{1}{14^{d+1}\ell}\}$. We will then bound the probability of each of these events and show that the (infinite) sum of these probabilities is at most $|\Phi|\frac{C^d}{\exp(\mathrm{e}^{-2}d\ell^{1/d})}$.

For all $i \in \mathbb{N}$, define k_i and α_i by

$$k_i := \mathrm{e}^{i-1}\ell^{1/d}, \qquad \alpha_i := \frac{k_i}{14^d\ell}\ \left(= \frac{1}{14^d\mathrm{e}^{1-i}\ell^{(d-1)/d}}\right).$$

Events $\mathcal{A}$ and $\mathcal{B}_i$ and $\mathcal{C}_{i,j}$ $(i,j \in \mathbb{N})$ are defined as follows:

$$\mathcal{A} \overset{\text{def}}{\iff} \left(\mathbf{p}(\Phi) \leq \alpha_0\right),$$

$$\mathcal{B}_i \overset{\text{def}}{\iff} \bigvee_{\Psi}\left(\mathbf{q}(\Psi) \leq \alpha_{i+1}\right) \wedge \left(\mathsf{D}(\Psi{\upharpoonright}R^{\sigma,\tau}_{\mathbf{q}(\Psi)}) \geq k_i\right),$$

$$\mathcal{C}_{i,j} \overset{\text{def}}{\iff} \bigvee_{\Psi}\left(\alpha_{i+j+1} < \mathbf{q}(\Psi) \leq \alpha_{i+j+2}\right) \wedge \left(\mathsf{D}(\Psi{\upharpoonright}R^{\sigma,\tau}_{\alpha_{i+1}}) \geq k_i\right).$$

5

<u>Claim:</u> If $\mathbf{q}(\Phi) \leq \dfrac{1}{14^{d+1}\ell}$, then $\mathcal{A} \vee \bigvee\limits_{i=0}^{\infty}\left(\mathcal{B}_i \vee \bigvee\limits_{j=0}^{\infty}\mathcal{C}_{i,j}\right)$.

<u>Proof of claim:</u> Assume $\mathbf{q}(\Phi) \leq 1/14^{d+1}\ell$ and further assume that $\mathcal{A}$ does not hold. Clearly there exists a unique $i \in \mathbb{N}$ such that $\alpha_i < \mathbf{p}(\Phi) \leq \alpha_{i+1}$ (since α_i is eventually > 1). Since $\mathbf{q}(\Phi) = \mathbf{p}(\Phi)/14 \cdot \mathbf{k}(\Phi)$, we have $\mathbf{k}(\Phi) > \alpha_i 14^d \ell = k_i$. Note that $k_i \geq k_0 = \mathrm{e}^{-1}\ell^{1/d} \geq 1$ (using the assumption that $\ell \geq \mathrm{e}^d$). Since $\mathbf{k}(\Phi) = \max\{1, \max_\Psi \mathsf{D}(\Psi{\restriction}R^{\sigma,\tau}_{\mathbf{p}(\Phi)})\}$, it follows that there exists a Ψ such that $\mathsf{D}(\Psi{\restriction}R^{\sigma,\tau}_{\mathbf{p}(\Phi)}) \geq k_i$.

Fix an arbitrary choice of Ψ such that $\mathsf{D}(\Psi{\restriction}R_{\mathbf{p}(\Phi)}) \geq k_i$. There are two cases to consider: either $\mathbf{q}(\Psi) \leq \alpha_{i+1}$ or $\alpha_{i+j+1} < \mathbf{q}(\Psi) \leq \alpha_{i+j+2}$ for some $j \in \mathbb{N}$.

- Assume $\mathbf{q}(\Psi) \leq \alpha_{i+1}$. In this case, we have $\mathsf{D}(\Psi{\restriction}R_{\mathbf{p}(\Phi)}) \leq \mathsf{D}(\Psi{\restriction}R_{\mathbf{q}(\Psi)})$ since $\mathbf{p}(\Phi) \leq \mathbf{q}(\Psi)$. Therefore, $\mathsf{D}(\Psi{\restriction}R^{\sigma,\tau}_{\mathbf{q}(\Psi)}) \geq k_i$. We conclude that $\mathcal{B}_i$ holds.

- Assume $\alpha_{i+j+1} < \mathbf{q}(\Psi) \leq \alpha_{i+j+2}$ for some $j \in \mathbb{N}$. We have $\mathsf{D}(\Psi{\restriction}R_{\mathbf{p}(\Phi)}) \leq \mathsf{D}(\Psi{\restriction}R_{\alpha_{i+1}})$ since $\mathbf{p}(\Phi) \leq \alpha_{i+1}$. Therefore, $\mathsf{D}(\Psi{\restriction}R^{\sigma,\tau}_{\alpha_{i+1}}) \geq k_i$. We conclude that $\mathcal{C}_{i,j}$ holds.

This concludes the proof of the claim.

To complete the proof of the theorem, we will bound the probabilities of events $\mathcal{A}$, $\mathcal{B}_i$ and $\mathcal{C}_{i,j}$ and take a union bound. We ignore the fact that all but finitely many of these events have zero probability, since $\mathbb{P}[\,\mathcal{B}_i\,] = 0$ (resp. $\mathbb{P}[\,\mathcal{C}_{i,j}\,] = 0$) for all $\alpha_i > 1$ (resp. $\alpha_{i+j+1} > 1$). Instead, we show that $\mathbb{P}[\,\mathcal{B}_i\,]$ is exponentially decreasing in i, while $Pr[\,\mathcal{C}_{i,j}\,]$ is exponentially decreasing in j and doubly exponentially decreasing in i.

We first bound the probability of $\mathcal{A}$:

$$
\begin{aligned}
\mathbb{P}[\,\mathcal{A}\,] = \mathbb{P}\left[\ \bigvee_{\Psi} \mathbf{q}(\Psi) \leq \frac{1}{14^d \mathrm{e}\ell^{(d-1)/d}}\ \right] &\leq \sum_{\Psi} \mathbb{P}\left[\ \mathbf{q}(\Psi) \leq \frac{1}{14^d \mathrm{e}\ell^{(d-1)/d}}\ \right]\\
&\leq |\Phi|\frac{C^{d-1}}{\exp(\mathrm{e}^{-2}(d-1)\mathrm{e}^{1/(d-1)}\ell^{1/d})} \quad \text{(induction hypothesis)}\\
&\leq |\Phi|\frac{C^{d-1}}{\exp(\mathrm{e}^{-2}d\ell^{1/d})} \quad \text{(using } \mathrm{e}^{1/(d-1)} \geq \tfrac{d}{d-1}\text{).}
\end{aligned}
$$

We next bound the probability of $\mathcal{B}_i$:

$$\mathbb{P}[\mathcal{B}_i] = \mathbb{P}\left[\bigvee_{\Psi}\left(\mathbf{q}(\Psi) \leq \alpha_{i+1}\right) \wedge \left(\mathsf{D}(\Psi{\restriction}R^{\sigma,\tau}_{\mathbf{q}(\Psi)}) \geq k_i\right)\right]$$

$$\leq \sum_{\Psi}\mathbb{P}\left[\mathbf{q}(\Psi) \leq \alpha_{i+1}\right]\mathbb{P}\left[\mathsf{D}(\Psi{\restriction}R^{\sigma,\tau}_{\mathbf{q}(\Psi)}) \geq k_i \mid \mathbf{q}(\Psi) \leq \alpha_{i+1}\right]$$

$$\leq \left(\frac{1}{\mathrm{e}}\right)^{k_i}\sum_{\Psi}\mathbb{P}\left[\mathbf{q}(\Psi) \leq \alpha_{i+1}\right] \qquad\qquad \text{(Lemma 9)}$$

$$= \frac{1}{\exp(\mathrm{e}^{i-1}\ell^{1/d})}\sum_{\Psi}\mathbb{P}\left[\mathbf{q}(\Psi) \leq \frac{1}{14^d\mathrm{e}^{-i}\ell^{(d-1)/d}}\right]$$

$$\leq \frac{1}{\exp(\mathrm{e}^{i-1}\ell^{1/d})}|\Phi|\frac{C^{d-1}}{\exp(\mathrm{e}^{-2}(d-1)\mathrm{e}^{-i/(d-1)}\ell^{1/d})} \qquad \text{(induction hypothesis)}$$

$$\leq \frac{1}{\exp(\mathrm{e}^{i-1}\ell^{1/d})}|\Phi|\frac{C^{d-1}}{\exp(\mathrm{e}^{-2}(d-1)\ell^{1/d} - i\mathrm{e}^{-2}\ell^{1/d})} \qquad (\mathrm{e}^{-i/(d-1)} \geq 1 - \tfrac{i}{d-1})$$

$$= \frac{1}{\exp((\mathrm{e}^{i-1} - (i+1)\mathrm{e}^{-2})\ell^{1/d})}|\Phi|\frac{C^{d-1}}{\exp(\mathrm{e}^{-2}d\ell^{1/d})}$$

$$\leq \frac{1}{\exp(\mathrm{e}^{i-1} - (i+1)\mathrm{e}^{-2})}|\Phi|\frac{C^{d-1}}{\exp(\mathrm{e}^{-2}d\ell^{1/d})}.$$

The last inequality uses the assumption $\ell^{1/d} \geq 1$ as well as the nonnegativity of $\mathrm{e}^{i-1} - (i+1)\mathrm{e}^{-2}$ for all $i \in \mathbb{N}$.

Finally, we bound the probability of $\mathcal{C}_{i,j}$:

$$\mathbb{P}[\mathcal{C}_{i,j}] = \mathbb{P}\left[\bigvee_{\Psi}\left(\alpha_{i+j+1} < \mathbf{q}(\Psi) \leq \alpha_{i+j+2}\right) \wedge \left(\mathsf{D}(\Psi{\restriction}R^{\sigma,\tau}_{\alpha_{i+1}}) \geq k_i\right)\right]$$

$$\leq \sum_{\Psi}\mathbb{P}\left[\mathbf{q}(\Psi) \leq \alpha_{i+j+2}\right]\mathbb{P}\left[\mathsf{D}(\Psi{\restriction}R^{\sigma,\tau}_{\alpha_{i+1}}) \geq k_i \mid \alpha_{i+j+1} < \mathbf{q}(\Psi) \leq \alpha_{i+j+2}\right]$$

$$\leq \left(\frac{\alpha_{i+1}/\alpha_{i+j+1}}{\mathrm{e}}\right)^{k_i}\sum_{\Psi}\mathbb{P}\left[\mathbf{q}(\Psi) \leq \alpha_{i+j+2}\right] \qquad\qquad \text{(Lemma 9)}$$

$$= \frac{1}{\exp((j+1)\mathrm{e}^{i-1}\ell^{1/d})}\sum_{\Psi}\mathbb{P}\left[\mathbf{q}(\Psi) \leq \frac{1}{14^d\mathrm{e}^{-(i+j+1)}\ell^{(d-1)/d}}\right]$$

$$\leq \frac{1}{\exp((j+1)\mathrm{e}^{i-1}\ell^{1/d})}|\Phi|\frac{C^{d-1}}{\exp(\mathrm{e}^{-2}(d-1)\mathrm{e}^{-(i+j+1)/(d-1)}\ell^{1/d})} \qquad \text{(ind. hyp.)}$$

$$\leq \frac{1}{\exp((j+1)\mathrm{e}^{i-1}\ell^{1/d})}|\Phi|\frac{C^{d-1}}{\exp(\mathrm{e}^{-2}(d-1)\ell^{1/d} - (i+j+1)\mathrm{e}^{-2}\ell^{1/d})}$$

$$= \frac{1}{\exp(((j+1)\mathrm{e}^{i-1} - (i+j+2)\mathrm{e}^{-2})\ell^{1/d})}|\Phi|\frac{C^{d-1}}{\exp(\mathrm{e}^{-2}d\ell^{1/d})}$$

$$\leq \frac{1}{\exp((j+1)\mathrm{e}^{i-1} - (i+j+2)\mathrm{e}^{-2})}|\Phi|\frac{C^{d-1}}{\exp(\mathrm{e}^{-2}d\ell^{1/d})}.$$

The last inequality uses the assumption $\ell^{1/d} \geq 1$ and the nonnegativity of $(j+1)\mathrm{e}^{i-1} - (i+j+2)\mathrm{e}^{-2}$ for all $i, j \in \mathbb{N}$.

We finish the proof by taking a union bound:

$$\mathbb{P}\left[\ \mathbf{q}(\Phi) \le \frac{1}{14^{d+1}\ell}\ \right] \le \mathbb{P}[\mathcal{A}] + \sum_{i=0}^{\infty}\left(\ \mathbb{P}[\mathcal{B}_i] + \sum_{j=0}^{\infty}\mathbb{P}[\mathcal{C}_{i,j}]\right) \le |\Phi|\frac{C^d}{\exp(\mathrm{e}^{-2}d\ell^{1/d})}. \qquad \square$$

5   PARITY

We use the results of the last section to prove our lower bound for the PARITY function.

Theorem 3 (restated). *Depth $d+1$ formulas computing* PARITY *require size* $\exp(\Omega(d(n^{1/d}-1)))$.

Proof. Suppose Φ is a depth $d+1$ formula computing PARITY. Then

$$\mathop{\mathbb{P}}_{\varrho \sim \mathcal{R}_{1/n}}\left[\ \Phi{\upharpoonright}\varrho \text{ is non-constant}\ \right] = 1 - \left(1 - \frac{1}{n}\right)^n > 1 - \frac{1}{\mathrm{e}}.$$

On the other hand, by Theorem 10 and Lemma 9,

$$\mathop{\mathbb{P}}_{\varrho \sim \mathcal{R}_{1/n}}\left[\ \Phi{\upharpoonright}\varrho \text{ is non-constant}\ \right] = \mathop{\mathbb{P}}_{\sigma,\tau}\left[\ \mathsf{D}(\Phi{\upharpoonright}R^{\sigma,\tau}_{1/n}) \ge 1\ \right]$$

$$\le \mathbb{P}\left[\ \mathsf{D}(\Phi{\upharpoonright}R^{\sigma,\tau}_{\max\{1/n,\mathbf{q}(\Phi)\}}) \ge 1\ \right]$$

$$\le \mathbb{P}\left[\ \mathbf{q}(\Phi) \le 1/n\ \right] + \mathbb{P}\left[\ \mathsf{D}(\Phi{\upharpoonright}R^{\sigma,\tau}_{\mathbf{q}(\Phi)}) \ge 1\ \right]$$

$$\le \frac{|\Phi|}{\exp(\Omega(dn^{1/d}) - O(d))} + \frac{1}{\mathrm{e}}.$$

Therefore,

$$|\Phi| \ge \left(1 - \frac{2}{\mathrm{e}}\right)\exp\left(\Omega(dn^{1/d}) - O(d)\right).$$

It follows that there exist universal constants $c_0, c_1 > 0$ (determined by the constants in the $\Omega(\cdot)$ and $O(\cdot)$) such that $|\Phi| \ge \exp(c_0 d(n^{1/d}-1))$ in the regime $d \le c_1 \ln n$.

In the regime $d > c_1 \ln n$, we have $d(n^{1/d}-1) = \Theta(\ln n)$, more precisely,

$$\ln n < d(n^{1/d} - 1) < c_1(\mathrm{e}^{c_1} - 1)\ln n.$$

Note that $d(n^{1/d}-1)$ is decreasing in d and $\lim_{d\to\infty} d(n^{1/d}-1) = \ln n$. In order to claim a tight lower bound of $\exp(\Omega(d(n^{1/d}-1)))$ for all d and n (in particular when $d = \omega(\log n)$), we require an $n^{\Omega(1)}$ lower bound on the (gate-)size of unbounded-depth formulas computing PARITY. A lower bound of n follows from Khrapchenko's classic n^2 leaf-size lower bound [6], which completes our proof. (While Khrapchenko's bound suffices our purposes, we remark that an asymptotically tight $\Omega(n^2)$ formula (gate-)size lower bound for PARITY was shown by Childs, Kimmel and Kothari [3] using quantum techniques.) $\qquad \square$

6   Average Sensitivity

Theorem 4 (restated). *Depth $d+1$ formulas of size s have average sensitivity $O(\frac{1}{d}\ln s)^d$.*

Proof. Let Φ be a formula of depth $d+1$ and size s (recall that size is the number of gates). Assume $\mathsf{as}(\Phi) \geq 1$, since otherwise the theorem is trivial. We further assume that Φ has bottom fan-in $\leq s$; otherwise it is easily shown that $\mathsf{as}(\Phi) = O(\mathsf{as}(\Phi'))$ where Φ' is obtained from Φ by replacing every bottom AND (resp. OR) gate with fan-in $> s$ with 0 (resp. 1). In particular, Φ has leaf-size $\leq s^2$, so it depends on $\leq s^2$ distinct variables.

Letting $p = 1/\mathsf{as}(\Phi)$ and using facts (1) and (2), we have

$$1 = p{\cdot}\mathsf{as}(\Phi) = \mathop{\mathbb{E}}_{\varrho \sim \mathcal{R}_p} \big[\, \mathsf{as}(\Phi{\restriction}\varrho) \,\big] \leq \mathop{\mathbb{E}}_{\sigma,\tau} \big[\, \mathsf{D}(\Phi{\restriction}R_p^{\sigma,\tau}) \,\big] = \sum_{k=1}^{s^2} \mathop{\mathbb{P}}_{\sigma,\tau} \big[\, \mathsf{D}(\Phi{\restriction}R_p^{\sigma,\tau}) \geq k \,\big].$$

For all $k \in \mathbb{N}$, by Theorem 10 and Lemma 9,

$$\mathop{\mathbb{P}}_{\sigma,\tau} \big[\, \mathsf{D}(\Phi{\restriction}R_p^{\sigma,\tau}) \geq k \,\big] \leq \mathop{\mathbb{P}}_{\sigma,\tau} \big[\, \mathsf{D}(\Phi{\restriction}R_{\max\{p,\mathbf{q}(\Phi)\}}^{\sigma,\tau}) \geq k \,\big]$$
$$\leq \mathbb{P}\big[\, \mathbf{q}(\Phi) \leq p \,\big] + \mathbb{P}\big[\, \mathsf{D}(\Phi{\restriction}R_{\mathbf{q}(\Phi)}^{\sigma,\tau}) \geq k \,\big]$$
$$\leq \frac{s}{\exp(\Omega(d{\cdot}\mathsf{as}(\Phi)^{1/d}) - O(d))} + \frac{1}{\mathrm{e}^k}.$$

Combining these inequalities, we have

$$\exp(\Omega(d{\cdot}\mathsf{as}(\Phi)^{1/d}) - O(d)) \leq \frac{s^3}{1 - \sum_{k=1}^{\infty} \mathrm{e}^{-k}} = \frac{1 - \mathrm{e}^{-1}}{1 - 2\mathrm{e}^{-1}} s^3 = O(s^3).$$

It follows that $\Omega(d{\cdot}\mathsf{as}(\Phi)^{1/d}) \leq 3\ln s + O(d)$ and therefore $\mathsf{as}(\Phi) = O(\frac{1}{d}\ln s)^d$. $\qquad\square$

7 Formulas vs. Circuits

Our lower bound for PARITY (Theorem 3) implies a separation between the power of depth $d + 1$ formulas vs. circuits. We write {poly-size depth $d + 1$ circuits/formulas} for the non-uniform complexity class of languages computable by $n^{O(1)}$-size depth $d+1$ circuits/formulas where $d(n)$ is an arbitrary function of n.

Corollary 11. *For all $d(n) = o(\log n)$ with $\lim_{n\to\infty} d(n) = \infty$,*

(3) $\qquad\qquad$ *{poly-size depth $d + 1$ formulas} $\neq$ {poly-size depth $d + 1$ circuits}.*

Moreover, for all $d \leq C\frac{\log n}{\log\log n}$ (for some universal constant $C > 0$),

(4) $\qquad\qquad$ *{poly-size depth $d + 1$ circuits} $\not\subseteq$ {$n^{o(d)}$-size depth $d + 1$ formulas}.*

Separation (3) may be regarded as the depth $d + 1$ analogue of the conjectured separation {poly-size formulas} $\neq$ {poly-size circuits}, also known as $\mathsf{NC}^1 \neq \mathsf{P}/\mathsf{poly}$. By Spira's theorem [8], every poly-size formula is equivalent to a poly-size formula of depth $O(\log n)$; thus, extending (3) from depth $o(\log n)$ to depth $O(\log n)$ would imply $\mathsf{NC}^1 \neq \mathsf{P}/\mathsf{poly}$ (in fact $\mathsf{NC}^1 \neq \mathsf{AC}^1$).

For the smaller range of $d \leq c\frac{\log n}{\log\log n}$, we get the stronger separation (4). In light of Fact 5, this is the strongest possible separation between formulas and circuits of the same depth.

We remark that until recently not even the weak separation (3) was known to hold for any super-constant $d \not\leq O(1)$. The first progress on this question was made in [7], where (4) was shown

to hold for all $d \leq \log \log \log n$ via a lower bound for DISTANCE-$\log \log n$ ST-CONNECTIVITY. In fact, the lower bound of [7] implies a much stronger result: for all $d \leq \log \log \log n$,

$$(5) \qquad \{\text{poly-size depth } d+1 \text{ circuits}\} \nsubseteq \{n^{o(d)}\text{-size depth } \tfrac{\log n}{(\log \log n)^3} \text{ formulas}\}.$$

It remains an open problem to push separation (5) to greater depths.

Acknowledgements

References

[1] Miklós Ajtai. Σ_1^1 formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[2] Ravi B. Boppana. The average sensitivity of bounded-depth circuits. *Information Processing Letters*, 63(5):257–261, 1997.

[3] Andrew M. Childs, Shelby Kimmel, and Robin Kothari. The quantum query complexity of read-many formulas. In *European Symposium on Algorithms*, pages 337–348, 2012.

[4] Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17:13–27, 1984.

[5] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *18th Annual ACM Symposium on Theory of Computing*, pages 6–20, 1986.

[6] V.M. Khrapchenko. Complexity of the realization of a linear function in the case of Π-circuits. Math. Notes Acad. Sciences, 9:21–23, 1971.

[7] Benjamin Rossman. Formulas vs. circuits for small distance connectivity. In *46th Annual ACM Symposium on Theory of Computing*, pages 203–212, 2014.

[8] P.M. Spira. On time-hardware complexity tradeoffs for Boolean functions. In *4th Hawaii Symposium on System Sciences*, pages 525–527, 1971.

[9] Andrew C.C. Yao. Separating the polynomial-time hierarchy by oracles. In *26th Annual IEEE Symposium on Foundations of Computer Science*, pages 1–10, 1985.