PARITY $\notin$ AC^0 from DNF Sparsification

Benjamin Rossman
University of Toronto

Abstract

We present a simple alternative proof of PARITY $\notin$ AC^0 using the DNF Sparsification Theorem of Gopalan, Meka and Reingold [1].

1. Introduction

PARITY $\notin$ AC^0 is a classic theorem in circuit complexity, which states that the n-variable parity function is not computable by unbounded fan-in Boolean circuits of $\text{poly}(n)$ size and constant depth [2, 3]. In this note, we present a simple alternative proof of this result using the DNF Sparsification Theorem from a paper of Gopalan, Meka and Reingold [1]. Recall that a *k-DNF* is a OR-of-ANDs formula with bottom fan-in at most k.

Theorem 1 (DNF Sparsification Theorem [1])**.** *Every k-DNF $f : \{0,1\}^n \to \{0,1\}$ has an ε-sandwiching approximation by k-DNFs with at most $(k\log(1/\varepsilon))^{O(k)}$ terms. That is, there exist k-DNFs f_ℓ, f_u with at most $(k\log(1/\varepsilon))^{O(k)}$ terms such that $f_\ell \leq f \leq f_u$ and*

$$\mathop{\mathbb{P}}_{x\in\{0,1\}^n}[\ f_\ell(x) \neq f_u(x)\ ] \leq \varepsilon.$$

The paper [1] includes two applications of Theorem 1: a derandomized switching lemma and a faster deterministic algorithm for approximately counting the number of satisfying solutions to a DNF. Our alternative proof of PARITY $\notin$ AC^0 is similar to (but more elementary than) the derandomized switching lemma of [1]. Precisely, we show that Theorem 1 implies the following lower bound:

Theorem 2 (Main result of this note)**.** *Depth-d circuits computing the n-variable parity function require $\exp(\Omega(\frac{\log^2 n}{d^2 \log\log n}))$ gates.*

Quantitatively, this is weaker than the best known (asymptotically tight) $\exp(\Omega(n^{1/(d-1)}))$ lower bound proved by Håstad using his Switching Lemma [4]. On the other hand, the proof of Theorem 2 uses random restrictions in a more simplistic way (with the potential advantage of being easier to extend to non-uniform input distributions).

The basic idea of the proof behind is to work with juntas instead of decision trees. Recall that an *m-junta* is a Boolean function that depends on at most m variables. To prove Theorem 2, we do not require the full power of Theorem 1. Rather, we only need the following Junta Approximation Theorem, which is an immediate corollary of Theorem 1.

Theorem 3 (Junta Approximation Theorem, Corollary 1.3 in [1])**.** *Every k-DNF is ε-close to an m-junta where $m = (k\log(1/\varepsilon))^{O(k)}$.*

To see that Theorem 3 follows from Theorem 1, simply note that the k-DNF f_ℓ is both ε-close to f and an kt-junta where $t = (k\log(1/\varepsilon))^{O(k)}$ is the number of terms in f_ℓ. Indeed, the same observation applies to f_u; thus, every k-DNF has an ε-sandwiching approximation by m-juntas (however, we do not need this stronger fact).

Our proof of Theorem 2 combines the above Junta Approximation Theorem with the following "Junta Switching Lemma" (which is really just a trivial observation). We denote by $\mathbf{R}_p$ the random restriction $[n] \to \{0, 1, *\}$ which maps each coordinate $i \in [n]$ to $*, 0, 1$ independently with probability $p, \frac{1-p}{2}, \frac{1-p}{2}$ respectively.

Lemma 4 ("Junta Switching Lemma")**.** *Every m-junta f satisfies*

$$\mathbb{P}[\ f{\restriction}\mathbf{R}_p \textit{ is not a } k\textit{-junta}\ ] \leq (pm)^k.$$

The proof is simply the observation that $\mathbb{P}[\ f{\restriction}\mathbf{R}_p$ is not a k-junta $] \leq \mathbb{P}[\ \mathrm{Bin}(m,p) > k\ ]$, which is at most $(pm)^k$. Note the superficial similarity between Lemma 4 and

Theorem 5 (Håstad's Switching Lemma [4])**.** *Every m-DNF f satisfies*

$$\mathbb{P}[\ f{\restriction}\mathbf{R}_p \textit{ is not a } k\textit{-CNF}\ ] \leq (5pm)^k.$$

To prove our lower bound for depth-d circuits computing PARITY_n (Theorem 2), we repeatedly apply Theorem 3 followed by Lemma 4 to each layer of the circuit in a sequence of $d-1$ steps (with $p = n^{-1/2d}$ and an appropriate choice of parameters k, m). This is similar to the depth reduction argument that yields a tight $\exp(n^{1/(d-1)})$ lower bound via Håstad's Switching Lemma. The fundamental difference between the two lower bounds is that the real work in Theorem 2 is done by Theorem 1 (the DNF Sparsification Theorem), which is based on a deterministic sunflower-plucking argument along the lines of Razborov's lower bounds for monotone circuits computing the $\mathrm{CLIQUE}_{k,n}$ function [5].

2. Proof of Theorem 2

Let $c \geq 1$ be the universal constant in Theorem 3 such that every k-DNF is ε-close to a $(k\log(1/\varepsilon))^{ck}$-junta. Suppose $\mathcal{C}$ is a depth-d circuit on n variables with $s = n^\ell$ gates where

$$\ell \leq \frac{\log n}{96cd^2 \log\log n}.$$

We will show that $\mathcal{C}$ does not compute PARITY_n, thereby proving Theorem 2.

Set parameters k and m as follows:

$$k := 8d\ell, \qquad m := (2k\ell\log n)^{ck}.$$

With this choice of parameters, Theorem 3 implies

Lemma 6. *Every* OR *or* AND *of (arbitrarily many) k-juntas is $1/s^2$-close to an m-junta.*

Proof. Note that k-DNFs (resp. k-CNFs) are the same things as ORs (resp. ANDs) of k-juntas. The statement for "ORs" is precisely Theorem 3 with the given parameters. The statement for "ANDs" follows by duality (taking negations). □

Recall that a *restriction* is a function $\varrho : \{1, \dots, n\} \to \{0, 1, *\}$ which maps each variable (i.e. coordinate of $\{0,1\}^n$) to 0, 1 or $*$ (meaning "unrestricted"). For $f : \{0,1\}^n \to \{0,1\}$, we write $f{\restriction}\varrho$ for the restricted function $\{0,1\}^{|\varrho^{-1}(*)|} \to \{0,1\}$. For $p \in [0,1]$, let $\mathbf{R}_p$ be the *random restriction* that maps each variable independently to $*$ with probability p and to 0 and 1 each with probability $\frac{1-p}{2}$.

Lemma 7. *Every m-junta f satisfies $\mathbb{P}[\ f{\restriction}\mathbf{R}_{n^{-1/2d}}$ is <u>not</u> a k-junta $] \le 1/s^2$.*

Proof. By Lemma 4, this probability is at most $(n^{-1/2d}m)^k$. Plugging in parameters, we have

$$\left(\frac{m}{n^{1/2d}}\right)^k \le \frac{(2k\ell \log n)^{ck^2}}{n^{k/2d}} = \left(\frac{(16d\ell^2 \log n)^{16cd^2\ell}}{n}\right)^{4\ell} \le \left(\frac{((\log n)^3)^{\log n/6 \log\log n}}{n}\right)^{4\ell} = \frac{1}{s^2}. \qquad \square$$

The next lemma applies Lemmas 6 and 7 iteratively to the gates of $\mathcal{C}$.

Lemma 8. *For all $\varepsilon > 0$, $\mathbb{P}\left[\ \mathcal{C}{\restriction}\mathbf{R}_{1/\sqrt{n}}$ is <u>not</u> ε-close to a k-junta $\right] \le \dfrac{2}{\varepsilon s}$.*

Before giving the proof of Lemma 8, we show how to obtain Theorem 2 as a consequence. Since $k < \log n$, we have (assuming $s = \omega_{n\to\infty}(1)$, which is clearly acceptable)

$$\mathbb{P}\left[\ \mathcal{C}{\restriction}\mathbf{R}_{1/\sqrt{n}} \text{ is } 1/4\text{-close to a } \log n\text{-junta } \right] \ge 1 - o_{n\to\infty}(1).$$

On the other hand,

$$\begin{aligned}\mathbb{P}\left[\ \mathrm{PARITY}_n{\restriction}\mathbf{R}_{1/\sqrt{n}} \text{ is } 1/4\text{-close to a } \log n\text{-junta } \right] &= \mathbb{P}\left[\ \mathrm{Bin}(n, 1/\sqrt{n}) \le \log n\ \right] \\ &\le \exp(n^{-\Omega(1)}).\end{aligned}$$

These two inequalities imply that $\mathcal{C}$ does not compute PARITY_n (for sufficiently large n), since otherwise we get a contradiction $1 - o_{n\to\infty}(1) \le \exp(n^{-\Omega(1)})$. It remains to prove Lemma 8 using Lemmas 6 and 7.

Proof of Lemma 8. For $i \in \{0, \dots, d\}$, let G_i be the set of depth-i subcircuits in $\mathcal{C}$. Thus, G_0 is the set of inputs; $G_d = \{\mathcal{C}\}$; and for $i \in \{1, \dots, d-1\}$, G_i corresponds to the set of gates at depth i. Note that $s = n^\ell = |G_1 \cup \cdots \cup G_d|$.

We define a joint distribution $(\varrho_1, \dots, \varrho_d, x)$, where $\varrho_1, \dots, \varrho_d$ are restrictions and x is an element of $\{0,1\}^n$, as follows:

- Sample $\varrho_1 : [n] \to \{0, 1, *\}$ with $*$-probability $n^{-1/2d}$.
- For $i \in \{2, \dots, d\}$, given $\varrho_1, \dots, \varrho_{i-1}$, sample $\varrho_i : \varrho_{i-1}^{-1}(*) \to \{0, 1, *\}$ with $*$-probability $n^{-1/2d}$. Note that the composed restriction $\varrho_1 \cdots \varrho_d : [n] \to \{0, 1, *\}$ has distribution $\mathbf{R}_{1/\sqrt{n}}$.
- Let x be a uniform random element of $\{0,1\}^n$ conditioned on being consistent with $\varrho_1 \cdots \varrho_d$. (That is, generate x as $\varrho_1 \cdots \varrho_d \sigma$ for uniform random $\sigma : \varrho_d^{-1}(*) \to \{0,1\}$.)

By induction on $i \in \{0, \dots, d\}$, we associated with each $g \in G_i$ a random k-junta denoted $\widetilde{g}$, which is determined by the first i random restrictions $\varrho_1, \dots, \varrho_i$. In the base case $i = 0$, for each input $g \in G_0$, let $\widetilde{g} := g$ and note that this is a k-junta (since $k \ge 1$). For $i \in \{1, \dots, d\}$ and $g \in G_i$, we define $\widetilde{g}$ as follows (via intermediate functions $\breve{g}$ and $\widehat{g}$ and an indicator random variable $\mathrm{Bad}(g)$). Suppose $g = \mathrm{OR/AND}(f_1, \dots, f_t)$ where $f_1, \dots, f_t \in G_0 \cup \cdots \cup G_{i-1}$.

- Let $\breve{g} := \mathrm{OR/AND}(\widetilde{f}_1, \dots, \widetilde{f}_t)$. Note that $\breve{g}$ is a k-DNF/CNF, since each $\widetilde{f}_i$ is a k-junta. Moreover, $\breve{g}$ is determined by the first $i-1$ restrictions $\varrho_1, \dots, \varrho_{i-1}$ and only depends on variables in $\varrho_{i-1}^{-1}(*)$, since the same is true of each $\widetilde{f}_i$.

- By Lemma 6, $\breve{g}$ is $1/s^2$-close to an m-junta. Let $\widehat{g}$ be a canonical m-junta that is $1/s^2$-close to $\breve{g}$ (say, the lexicographically minimal one). Thus, $\widehat{g}$ satisfies $\mathbb{P}_{\varrho_1,\dots,\varrho_{i-1}}[\, \mathbb{P}_x[\, \widehat{g}(x) = \breve{g}(x)\,] \le 1/s^2\,] = 1$, hence

$$\mathbb{P}[\, \widehat{g}(x) = \breve{g}(x)\,] \le 1/s^2. \tag{1}$$

  (Here we have used the fact that the joint distribution $(\varrho_1, \dots, \varrho_{i-1}, x)$ is equivalent to the joint distribution $(\varrho_1, \dots, \varrho_{i-1}, y)$ where y is uniform random in $\{0,1\}^n$ condition on being consistent with $\varrho_1 \cdots \varrho_{i-1}$.)

- Let $\mathrm{Bad}(g) \in \{0,1\}$ be the indicator variable for the event $\{\widehat{g}\restriction \varrho_1 \cdots \varrho_i$ is not a k-junta$\}$. By Lemma 7, we have $\mathbb{P}_{\varrho_1,\dots,\varrho_{i-1}}[\, \mathbb{P}_{\varrho_i}[\, \mathrm{Bad}(g)\,] \le 1/s^2\,] = 1$, hence

$$\mathbb{P}[\, \mathrm{Bad}(g)\,] \le 1/s^2. \tag{2}$$

- Finally, let

$$\widetilde{g} := \begin{cases} 0 \text{ (i.e. the identically zero function)} & \text{if } \mathrm{Bad}(g), \\ \widehat{g}\restriction \varrho_1 \cdots \varrho_i & \text{otherwise.} \end{cases}$$

  Note that $\widetilde{g}$ is a k-junta on variables $\varrho_i^{-1}(*)$ (with probability 1).

Having defined random k-juntas $\widetilde{g}$ for all gates g, we complete the proof of Lemma 8 as follows:

$$\begin{aligned}
&\mathbb{P}\left[\, \mathcal{C}\restriction \mathbf{R}_{1/\sqrt{n}} \text{ is not } \varepsilon\text{-close to a } k\text{-junta}\,\right] \\
&\qquad \le \mathbb{P}\left[\, \mathcal{C}\restriction \varrho_1 \cdots \varrho_d \text{ is not } \varepsilon\text{-close to } \widetilde{\mathcal{C}}\,\right] \\
&\qquad = \mathop{\mathbb{P}}_{\varrho_1,\dots,\varrho_d}\left[\, \mathop{\mathbb{P}}_{x}\left[\, \mathcal{C}(x) \ne \widetilde{\mathcal{C}}(x)\,\right] > \varepsilon\,\right] \quad \text{(recall } x \text{ is uniform consistent with } \varrho_1 \cdots \varrho_d\text{)} \\
\text{(Markov's ineq.)} \quad &\qquad \le \frac{1}{\varepsilon}\,\mathbb{P}\Big[\, \underbrace{\mathcal{C}(x) \ne \widetilde{\mathcal{C}}(x)}_{\Rightarrow\, \exists \text{ minimal gate } g \text{ with } g(x) \ne \widetilde{g}(x)} \,\Big] \\
&\qquad = \frac{1}{\varepsilon}\,\mathbb{P}\left[\, \bigvee_{g \in G_1 \cup \cdots \cup G_d} \Bigg(\underbrace{\Big(\widetilde{g}(x) \ne g(x)\Big)}_{\Rightarrow\, \mathrm{Bad}(g) \vee (\widehat{g}(x) \ne g(x))} \wedge \underbrace{\bigwedge_{\text{children } f \text{ of } g} \Big(\widetilde{f}(x) = f(x)\Big)}_{\Rightarrow\, \breve{g}(x) = g(x)} \Bigg) \,\right] \\
&\qquad \le \frac{1}{\varepsilon}\,\mathbb{P}\left[\, \bigvee_{g \in G_1 \cup \cdots \cup G_d} \mathrm{Bad}(g) \vee \Big(\widehat{g}(x) \ne \breve{g}(x)\Big) \right] \\
&\qquad \le \frac{1}{\varepsilon} \sum_{g \in G_1 \cup \cdots \cup G_d} \mathbb{P}\left[\, \mathrm{Bad}(g)\,\right] + \mathbb{P}\left[\, \widehat{g}(x) \ne \breve{g}(x)\,\right] \\
\text{(by (1) and (2))} \quad &\qquad \le \frac{1}{\varepsilon} \sum_{g \in G_1 \cup \cdots \cup G_d} \frac{2}{s^2} \\
&\qquad = \frac{2}{\varepsilon s}.
\end{aligned}$$

□

3. Final Remarks

A different junta approximation theorem (incomparable to Theorem 3) follows from known results in Boolean analysis.

Theorem 9 (Corollary 9.30 of [6]). *Every k-DNF is ε-close to an $(1/\varepsilon)^{O(k)}$-junta.*

(Note: This is better than the $2^{-O(k/\varepsilon)}$ bound that follows from Friedgut's Junta Theorem [7] for Boolean functions with average sensitivity k.) It is an open question, highlighted in [1], whether every k-DNF is ε-close to an $(\log(1/\varepsilon))^{O(k)}$-junta.

Finally, it would be interesting to find an application of the lower bound proof presented in this note. Perhaps the lower bound technique via DNF sparsification can be applied to input distributions other than the uniform distribution, where extensions of Håstad's Switching Lemma may be hard to prove.

References

[1] P. Gopalan, R. Meka, O. Reingold, DNF sparsification and a faster deterministic counting algorithm, Computational Complexity 22 (2) (2013) 275–310.

[2] M. Ajtai, σ_1^1 formulae on finite structures, Annals of Pure and Applied Logic 24 (1983) 1–48.

[3] M. L. Furst, J. B. Saxe, M. Sipser, Parity, circuits, and the polynomial-time hierarchy, Mathematical Systems Theory 17 (1984) 13–27.

[4] J. Håstad, Almost optimal lower bounds for small depth circuits, in: 18th Annual ACM Symposium on Theory of Computing, 1986, pp. 6–20.

[5] A. A. Razborov, Lower bounds on the monotone complexity of some boolean functions, in: Dokl. Akad. Nauk SSSR, Vol. 281, 1985, pp. 798–801.

[6] R. O'Donnell, Analysis of boolean functions, Cambridge University Press, 2014.

[7] E. Friedgut, Boolean functions with low average sensitivity depend on few coordinates, Combinatorica 18 (1) (1998) 27–35.