Lower Bounds for Approximations by Low Degree Polynomials Over Z_m

Noga Alon*
Tel Aviv University
noga@math.tau.ac.il

Richard Beigel†
Temple University
beigel@joda.cis.temple.edu

Abstract

We use a Ramsey-theoretic argument to obtain the first lower bounds for approximations over Z_m by nonlinear polynomials:

- *A degree-2 polynomial over Z_m (m odd) must differ from the parity function on at least a $1/2 - 1/2^{(\log n)^{\Omega(1)}}$ fraction of all points in the Boolean n-cube.*
- *A degree-$O(1)$ polynomial over Z_m (m odd) must differ from the parity function on at least a $1/2 - o(1)$ fraction of all points in the Boolean n-cube.*

These nonapproximability results imply the first known lower bounds on the top fanin of $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}_{O(1)}$ *circuits (*i.e.*, circuits with a single majority-gate at the output node,* MOD_m*-gates at the middle level, and constant-fanin AND-gates at the input level) that compute parity:*

- $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}_2$ *circuits that compute parity must have top fanin* $2^{(\log n)^{\Omega(1)}}$.
- *Parity cannot be computed by* $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}_{O(1)}$ *circuits with top fanin $O(1)$.*

Similar results hold for the MOD_q *function as well.*

1. Introduction

Obtaining lower bounds for constant-depth circuits involving MOD_m gates, when m is composite, is a famously difficulty problem. The prominence of this problem is mainly due to Smolensky's classic paper [14], which shows that constant-depth, polynomial-size circuits consisting of AND, OR, and MOD_{p^k} gates, where p is an odd prime, cannot compute parity.

Numerous research efforts have concentrated on depth-2 and depth-3 circuits with specific types of gates allowed at each level [8, 13, 6, 1, 9, 10, 4]. Such circuits can be surprisingly powerful. For example, quasipolynomial-size $\mathrm{SYM} \circ \mathrm{AND}_{\mathrm{polylog}\, n}$ contains all of ACC [15, 2].

Consider $\mathrm{MAJ} \circ \mathrm{MOD}_m$ circuits for m constant. Since $\mathrm{MAJ} \circ \mathrm{MOD}_m$ is contained in TC_2^0 (*i.e.*, depth-2 TC^0) [7], $\mathrm{MAJ} \circ \mathrm{MOD}_m$ circuits computing inner-product-mod-p require size $2^{\Omega(n)}$ [10]. Goldmann [4] has established the following sharper result: if q is divisible by a prime p that does not divide m, then $\mathrm{MAJ} \circ \mathrm{MOD}_m$ circuits computing MOD_q require size $2^{\Omega(n)}$. Since then, this lower bound has been generalized by Krause and Pudlak [9]: If q is divisible by a prime p that does not divide m, then $\mathrm{MAJ} \circ \mathrm{AND}_{O(1)} \circ \mathrm{MOD}_m$ circuits computing MOD_q require size $2^{\Omega(n)}$.

Goldmann and Hastad [8] have shown that $\mathrm{TC}_2^0 \circ \mathrm{AND}_{(\frac{1}{2}-\epsilon)\log n}$ circuits computing the generalized-inner-product function (GIP) require size $2^{\Omega(n)}$. Since $\mathrm{MAJ} \circ \mathrm{MOD}_m \subseteq \mathrm{TC}_2^0$ [7], it follows that $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}_{(\frac{1}{2}-\epsilon)\log n}$ circuits computing (GIP) require size $2^{\Omega(n)}$.

However, no lower bounds are known for computing parity with $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}_{O(1)}$ circuits, or even with $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}_2$ circuits. We will adopt the approach of [4]. Goldmann first proved a result of independent interest: MOD_m circuits cannot *approximate* MOD_q. Then he applied the "ϵ-discriminator" lemma of Hajnal, Maass, Pudlák, Szegedy, and Turán [7] to prove that small $\mathrm{MAJ} \circ \mathrm{MOD}_m$ circuits cannot *compute* MOD_q.

Using Ramsey theory, we will show that small $\mathrm{MOD}_m \circ \mathrm{AND}_{O(1)}$ circuits cannot approximate MOD_q. Our approximation bound suffices to provide lower bounds on $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}_{O(1)}$ circuits where the top fanin is small, but is not sharp enough to provide lower bounds for general $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}_{O(1)}$ circuits. It is still an open problem to prove lower bounds for that class of circuits.

*Address: Dept. of Mathematics, 69978 Tel Aviv, ISRAEL

†Address: Temple University, Dept. of Computer and Information Sciences, Wachman Hall Rm 304 (038-24), 1805 N Broad St, Philadelphia, PA 19122-6094. Research performed in part at Lehigh University, University of Illinois at Chicago, DIMACS, and NEC. Phone: (215)204-6975, fax: (215)204-5082, http://www.cis.temple.edu/∼beigel. Supported in part by the National Science Foundation under grants CCR-9796317, CCR-9996021, and CCR-0049019.

2. Definitions and Background

We first define the circuit classes that will occur in this article. Note that the size of a circuit is the number of edges it contains. In addition, unless otherwise indicated, all circuit classes allow gates of unbounded fanin and each input gate is labeled by an input variable, its negation, or a constant 0 or 1. MOD_m gates are defined by $\mathrm{MOD}_m(x_1, \ldots, x_n) = 1$ if and only if $\sum_{i=1}^n x_i \equiv 0 \pmod{m}$. For all of these classes, a subscript d denotes the subclass of circuits with depth exactly d.

MAJ denotes the class of polynomial-size circuits consisting of a single level of majority gates. The classes AND, OR and MOD_m are defined similarly. A subscript k used with AND and OR denotes the restriction to gates of fanin k.

Classes of circuits whose levels consist of various types of gates can be conveniently described as the composition of various classes of functions (see [11, 12]). If Γ and Λ are classes of Boolean functions (not necessarily circuit classes), then $\Gamma \circ \Lambda$ denotes the class of functions f of the form $f(x) = g(h(x))$, where $g \in \Gamma$, $h \in \Lambda$ and h has monotone increasing output length. For example, $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}$ is the class of functions computed by depth-3 polynomial-size circuits with majority gates at the output, MOD_m gates at level two, and AND gates at level one. In the case where Γ and Λ are both circuit classes, then we will refer to the corresponding circuits as $\Gamma \circ \Lambda$ circuits. Note that the condition on the output length of h guarantees that the complexity of g, which is measured relative to the output length of h, is related to the input length of f. This is important in some contexts but of no consequence for the circuit classes considered in this article.

We will often abuse notation and use the name of a class of polynomial-size circuits to denote the same class of circuits but with no implicit bound on the size. Then we will specify size restrictions explicitly, *e.g.*, "subexponential size $\mathrm{MOD}_2 \circ \mathrm{AND}_{O(1)}$ cannot compute MOD_3."

Correlation is defined as follows:

$$\begin{aligned} &\mathrm{Corr}(f, g) \\ &= \Pr[g(x) = 1 | f(x) = 1] - \Pr[g(x) = 1 | f(x) = 0] \end{aligned}$$

Upper bounds on correlation imply lower bounds on circuit size by (a special case of) the "ϵ-discriminator" lemma of Hajnal, Maass, Pudlák, Szegedy, and Turán, which states:

Lemma 1 (ϵ-discriminator [7]). *Let C be a circuit consisting of an unweighted threshold gate over subcircuits $c_1, \ldots, c_s$,* i.e.*,*

$$C(x) = 1 \textit{ iff } \sum_i c_i(x) \geq t$$

for some fixed natural number t. Then

$$s \cdot \max_i \mathrm{Corr}(C, c_i) \geq 1.$$

Mikael Goldmann has shown

Theorem 2 ([4]). *Let q be prime and let m be a number such that q does not divide m. Let p be a linear polynomial (not necessarily symmetric). Let*

$$g(x_1, \ldots, x_n) = \begin{cases} 1 & \textit{if } p(x_1, \ldots, x_n) \not\equiv 0 \pmod{m} \\ 0 & \textit{otherwise .} \end{cases}$$

Then $\mathrm{Corr}(\mathrm{MOD}_q, g) \leq 2^{-\Omega(n)}$.

A multivariate polynomial is called *symmetric* if it is invariant under permutations of its variables. Fred Green has shown (implicit in the proof of his Theorem 2.7)

Theorem 3 ([5]). *Let q be prime and let m be a number such that q does not divide m. Let p be a symmetric multivariate polynomial of degree $\log^{O(1)} n$. Let*

$$g(x_1, \ldots, x_n) = \begin{cases} 1 & \textit{if } p(x_1, \ldots, x_n) \not\equiv 0 \pmod{m} \\ 0 & \textit{otherwise} \end{cases}$$

Then $\mathrm{Corr}(\mathrm{MOD}_q, g) \leq 2^{-n^{\Omega(1)}}$.

Note: tight bounds for the case $q = 2$ are given by Cai, Green, and Thierauf [3].

Since $x_1, \ldots, x_n$ are Boolean variables we may assume, without loss of generality, that all polynomials over $x_1, \ldots, x_n$ are multilinear. This allows us to combine Ramsey theory with Green's result in order to prove lower bounds for the case of general polynomials (not necessarily symmetric).

3. Results

Previously nothing was known for general polynomials p, even of degree 2. We show

Theorem 4. *Let q be prime and let m be a number such that q does not divide m. Let p be a multivariate polynomial of degree 2. Let*

$$g(x_1, \ldots, x_n) = \begin{cases} 1 & \textit{if } p(x_1, \ldots, x_n) \not\equiv 0 \pmod{m} \\ 0 & \textit{otherwise} \end{cases}$$

Then $\mathrm{Corr}(\mathrm{MOD}_q, g) \leq 2^{-(\log n)^{\Omega(1)}}$.

Proof: Let p be a degree-2 multivariate polynomial in n variables. Without loss of generality, the coefficients of p belong to $\{0, \ldots, m-1\}$. Think of the polynomial p as a graph with one edge for each degree-2 monomial, and color each edge with the coefficient of that monomial. By Ramsey's theorem there exists a number $c \in \{0, \ldots, m-1\}$ and a set (which we denote after renumbering) $\{x_1, \ldots, x_{n'}\}$ consisting of $\Omega(\log n)$ variables such that every degree-2 monomial of p consisting only of variables in $\{x_1, \ldots, x_{n'}\}$ has coefficient c.

Obtain an n'-variate polynomial p' by assigning random Boolean values to $x_{n'+1}, \ldots, x_n$ in p. Then

$$\begin{aligned} p'(x_1, \ldots, x_{n'}) &= p''(x_1, \ldots, x_{n'}) + \sum_{1 \leq i < j \leq n'} c x_i x_j \\ &= p''(x_1, \ldots, x_{n'}) + c\binom{x_1 + \cdots + x_{n'}}{2} \end{aligned}$$

where p'' is linear. Color the variables $x_1, \ldots, x_{n'}$ according to their coefficients in p'', choose the most frequent color, and assign random Boolean values to the variables with any other color.

We are then left with a symmetric polynomial in $\Omega(\log n)$ variables. By Theorem 3, the resulting polynomial has correlation $2^{-(\log n)^{\Omega(1)}}$ with MOD_q. Therefore $\mathrm{Corr}(\mathrm{MOD}_q, g) = 2^{-(\log n)^{\Omega(1)}}$. ∎

We have the following interesting special case:

Corollary 5. *A degree-2 polynomial over Z_m (m odd) must differ from the parity function on at least a $1/2 - 1/2^{(\log n)^{\Omega(1)}}$ fraction of all points in the Boolean n-cube.*

The following is immediate from the ϵ-discriminator Lemma of [7] (our Lemma 1):

Corollary 6. $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}_2$ *circuits that compute* MOD_q *must have top fanin* $2^{(\log n)^{\Omega(1)}}$.

We now generalize Theorem 4:

Theorem 7. *Let q be prime and let m be a number such that q does not divide m. Let p be a multivariate polynomial of degree $O(1)$. Let*

$$g(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } p(x_1, \ldots, x_n) \not\equiv 0 \pmod{m} \\ 0 & \text{otherwise} \end{cases}$$

Then $\mathrm{Corr}(\mathrm{MOD}_q, g) = o(1)$.

Proof:

Induction hypothesis (induction on d): Let q be prime and let m be a number such that q does not divide m. Let p be a multivariate polynomial of degree d in n variables. Let s by a symmetric multivariate polynomial of degree $O(1)$. Let

$$\begin{aligned} &g(x_1, \ldots, x_n) \\ &= \begin{cases} 1 & \text{if } p(x_1, \ldots, x_n) + s(x_1, \ldots, x_n) \not\equiv 0 \pmod{m} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Then $\mathrm{Corr}(\mathrm{MOD}_q, g) = o(1)$.

Base case: $d = 0$: Then p is constant, so $p + s$ is symmetric. In this case, the claim follows from Theorem 3.

Inductive case: Assume that the induction hypothesis is true for $d - 1$. Let p be a degree-d multivariate polynomial in n variables. Without loss of generality, the coefficients of p belong to $\{0, \ldots, m-1\}$. Think of the polynomial p as a hypergraph with one hyperedge for each monomial, and color each hyperedge with the coefficient of that monomial. By Ramsey's theorem for hypergraphs there exists a number $c \in \{0, \ldots, m-1\}$ and a set (which we denote after renumbering) $\{x_1, \ldots, x_{n'}\}$ consisting of $\omega(1)$ variables such that every degree-d monomial of p consisting only of variables in $\{x_1, \ldots, x_{n'}\}$ has coefficient c.

Obtain an n'-variate polynomial p' by assigning random Boolean values to $x_{n'+1}, \ldots, x_n$ in p. Then

$$\begin{aligned} &p'(x_1, \ldots, x_{n'}) \\ &\quad= p''(x_1, \ldots, x_{n'}) + \sum_{1 \leq i_1 < \cdots < i_d \leq n'} c x_{i_1} \cdots x_{i_d} \\ &\quad= p''(x_1, \ldots, x_{n'}) + c\binom{x_1 + \cdots + x_{n'}}{d} \end{aligned}$$

where p'' is a degree-$(d-1)$ polynomial. Since $c\binom{x_1+\cdots+x_{n'}}{d}$ is symmetric, we are now done by the inductive hypothesis. ∎

Note: If p has degree d, then in fact, $\mathrm{Corr}(\mathrm{MOD}_q, g) \leq 1/f^{-1}(n)$, where f is roughly the $(d+2)$nd level of Ackerman's function.

We have the following interesting special case:

Corollary 8. *A degree-$O(1)$ polynomial over Z_m (m odd) must differ from the parity function on at least a $1/2 - o(1)$ fraction of all points in the Boolean n-cube.*

The following is immediate from the ϵ-discriminator of [7]:

Corollary 9. *The* MOD_q *function cannot be computed by* $\mathrm{MAJ} \circ \mathrm{MOD}_m \circ \mathrm{AND}_{O(1)}$ *circuits with top fanin $O(1)$.*

4. Open Problems

- If the degree of g is fixed, is $\mathrm{Corr}(\mathrm{MOD}_q, g)$ maximized when g is symmetric? This would imply that in general $\mathrm{Corr}(\mathrm{MOD}_q, g) = 2^{-n^{\Omega(1)}}$.
- Is $\mathrm{Corr}(\mathrm{MOD}_q, g) = 2^{-\Omega(n)}$? $2^{-n^{\Omega(1)}}$? $n^{-\omega(1)}$?

Acknowledgments

We thank the organizers of LATIN '98 for providing a warm and stimulating research environment, Pavel Pudlák and Bill Gasarch for helpful discussions, and Alexis Maciel for proofreading. We would also like to thank the Complexity 2001 program committee for helpful suggestions.

References

[1] R. Beigel and A. Maciel. Upper and lower bounds for some depth-3 circuit classes. In *Proc. 12th Ann. IEEE Conf. Comput. Complexity Theory*, 1997.

[2] R. Beigel and J. Tarui. On ACC. *Comput. Complexity*, 4:350–366, 1994. Special issue devoted to the 4th Annual McGill Workshop on Complexity Theory.

[3] J. Cai, F. Green, and T. Thierauf. On the correlation of symmetric functions. *Math. Systems Theory*, 29, 1996.

[4] M. Goldmann. A note on the power of majority gates and modular gates. *Inform. Process. Lett.*, 53:321–327, 1995.

[5] F. Green. Exponential sums and circuits with a single threshold gate and mod-gates. *Theory Comput. Systems*, 32:453–566, 1999.

[6] V. Grolmusz. A weight–size trade–off for circuits with MOD m gates. In *Proc. 26th Ann. ACM Symp. Theory Comput.*, pages 68–74, 1994.

[7] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *J. Comput. System Sci.*, 46(2):129–154, 1993.

[8] J. Håstad and M. Goldmann. On the power of small-depth threshold circuits. *Comput. Complexity*, 1:113–129, 1991.

[9] M. Krause and P. Pudlák. On the computational power of depth 2 circuits with threshold and modulo gates. *Theor. Computer Sci.*, 174:137–156, 1997.

[10] M. Krause and S. Waack. Variation ranks of communication matrices and lower bounds for depth two circuits having symmetric gates with unbounded fan-in. In *Proc. 32th Ann. IEEE Symp. Found. Computer Sci.*, pages 777–782, 1991.

[11] A. Maciel. *Threshold Circuits of Small Majority-Depth*. PhD thesis, School of Computer Science, McGill University, Montréal, Québec, Canada, 1995.

[12] A. Maciel and D. Thérien. Threshold circuits of small majority-depth. Technical Report SOCS–95.5, School of Computer Science, McGill University, Montréal, Québec, Canada, 1995. To appear in *Inform. and Comput.*

[13] A. Razborov and A. Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth 3 threshold circuits with AND gates at the bottom. *Inform. Process. Lett.*, 45, 1993.

[14] R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proc. 19th Ann. ACM Symp. Theory Comput.*, pages 77–82, 1987.

[15] A. C.-C. Yao. On ACC and threshold circuits. In *Proc. 31th Ann. IEEE Symp. Found. Computer Sci.*, pages 619–627, 1990.