COMPLEX POLYNOMIALS AND CIRCUIT LOWER BOUNDS FOR MODULAR COUNTING

David A. Mix Barrington†
Department of Computer and Information Science
University of Massachusetts
Amherst, Massachusetts 01003
USA

Howard Straubing‡
Department of Computer Science
Boston College
Chestnut Hill, Massachusetts 02167
USA

Several recent papers [1,2,3,4] have considered the computational power of constant-depth unbounded fan-in circuits with AND, OR, and NOT gates, augmented with a limited number of MAJORITY gates. In particular Beigel, Reingold, and Spielman [3] have revived the notion of a *perceptron*, a term introduced by Minsky and Papert [6] to mean a majority-of-ANDs circuit. Such a circuit can be thought of as evaluating a polynomial in the input variables with integer coefficients, and taking its sign. The degree of the polynomial corresponds to the maximum fan-in of the AND gates. These devices can approximate AND-OR circuits of constant depth [1], and simulate them probabilistically [3], but cannot come close to computing the parity function of the inputs unless they have both polynomial degree and exponential size. Aspnes, Beigel, Furst and Rudich [1] and Beigel, Reingold and Spielman [4] consider a variant of the original perceptron, where the inputs to the single MAJORITY gate are constant-depth, unbounded fan-in AND-OR-NOT circuits. Following [4], we will call these "perceptrons" and view the original devices simply as polynomials. It is proved in [5] that a perceptron requires exponential size $(2^{n^{\Omega(1)}})$ to compute the parity function, and this is improved in [1] to show that any subexponential size perceptron must fail to compute parity on a constant fraction of the inputs. These lower bounds for perceptrons are used in [1,2,4] to get lower bounds for other types of circuits with a limited number of majority gates. In each case the circuit is converted into an equivalent perceptron. We know therefore that a constant-depth circuit with one majority gate in the middle [1], $o(\log n)$ majority gates [4], or even $n^{o(1)}$ majority gates [2] needs exponential size to compute parity. A primary tool in this work is to represent a function by a polynomial over the real numbers, whose variables are the inputs to the function. The lower bounds for the parity function make use of the fact that parity is in a sense embedded within the real numbers, as multiplication within the set $\{1, -1\}$. On the other hand, we would expect the other modular counting functions (the sum of the inputs modulo k for other constants k) to also be hard to compute with perceptrons.

† Research supported by NSF Grant CCR-8714714.
‡ Research supported by NSF Grant CCR-8902369.

Here we shall extend the previous lower bounds to circuits that compute the sum of the inputs modulo k for any $k > 1$. We do this by generalizing the real-valued "voting polynomials" of [1] to the complex domain, where it is possible to represent addition modulo k by multiplication of k^{th} roots of unity. Of course, with no obvious notion of the "sign" of a complex number we must redefine many of the concepts of [1]. We also make essential use of the rational approximation techniques of [4].

A few observations on complex polynomials.

Let $k > 1$. This will be fixed throughout the note. Let

$$\omega = e^{\frac{2\pi i}{k}},$$

and let

$$D = \{1, \omega, \ldots, \omega^{k-1}\}.$$

Observe that for $1 \leq j < k$,

$$\sum_{x \in D} x^j = 0.$$

We are going to consider polynomial functions in n variables restricted to D^n. We note the following facts: *(a)* Let

$$\mathbf{x} = (x_1, \ldots, x_n) \in D^n,$$

and let

$$\overline{\mathbf{x}} = (\overline{x_1}, \ldots, \overline{x_n}).$$

Then for any polynomial P,

$$\overline{P(\mathbf{x})} = P(\overline{\mathbf{x}}) = P(\overline{x_1}, \ldots, \overline{x_n}) = P(x_1^{k-1}, \ldots, x_n^{k-1}).$$

In particular, $\overline{P(\mathbf{x})}$ is equivalent to a polynomial of degree no more than $(k-1)$ times the degree of P.

(b) Since $x^k = 1$ for $x \in D$, any polynomial is equivalent to one in which every monomial has the form

$$x_1^{i_1} \cdots x_n^{i_n},$$

where $0 \leq i_j < k$ for $j = 1, \ldots, n$. We will henceforth suppose that every polynomial is in this form. In particular every polynomial has degree no greater than $(k-1)n$. The *weight* of the monomial is the number of nonzero i_j. Observe that if m is a monomial of weight w and degree d, then $\overline{m}$ has weight w and degree $kw - d$.

(c) The polynomial

$$M_k(\mathbf{x}) = x_1^{k-1} \cdots x_n^{k-1}$$

maps $(\omega^{i_1}, \ldots, \omega^{i_n})$ to $\omega^{-(i_1 + \cdots + i_n) \bmod k}$.

(d) If $z_1, z_2 \in \mathbf{C}$, then $Re(z_1\overline{z_2})$ is the inner product of z_1 and z_2 considered as vectors in $\mathbf{R}^2$. Thus $Re(z_1\overline{z_2}) > 0$ if and only if the angle between z_1 and z_2 is less than $\frac{\pi}{2}$.

(e) Consider a polynomial P that vanishes on D^n. We can write it as a polynomial of degree $k-1$ in a single variable, with coefficients that are polynomials in $n-1$ variables. That is,

$$P(x_1, \ldots, x_n) = p_0(\mathbf{x}) + p_1(\mathbf{x})x_n + \cdots + p_{k-1}(\mathbf{x})x_n^{k-1},$$

where $\mathbf{x} = (x_1, \ldots, x_{n-1})$. Thus for each $\mathbf{x} \in D^{n-1}$ this polynomial vanishes on D, and thus each p_i vanishes on D^{n-1}. It follows by induction that P is the zero polynomial.

(f) A polynomial P *weakly represents* a function $f : D^n \to \mathbf{C}$ if P is not the zero polynomial, and for all $\mathbf{x} \in D^n$ such that $P(\mathbf{x}) \neq 0$, $Re(P(\mathbf{x})\overline{f(\mathbf{x})}) > 0$. The *weak degree* of f, denoted $d_w(f)$, is the degree of the smallest-degree polynomial that weakly represents f.

Weak degree of the product function M_k.

This is a generalization of Lemma 3 of [1]:

Lemma 1. $d_w(M_k) = n(k-1)$.

Proof. We introduce the following inner product on the space of functions from D^n to $\mathbf{C}$:

$$\langle f, g \rangle = \sum_{\mathbf{x} \in D^n} f(\mathbf{x})\overline{g(\mathbf{x})}.$$

It follows that if P is a polynomial weakly representing g, then $Re(\langle P, g \rangle) > 0$. On the other hand, suppose that the degree of P is less than $n(k-1)$, and that P weakly represents M_k. Then P is a sum of monomials in which at least one variable appears to a power less than $k-1$. Thus for each such monomial m, we can factor

$$\sum_{x_i \in D} x_i^j,$$

for some $1 \leq i \leq n$, $1 \leq j < k$, out of $\langle m, M_k \rangle$. Thus $\langle m, M_k \rangle = 0$, and consequently $\langle P, M_k \rangle = 0$, a contradiction. ∎

Killing a large set with a real-valued polynomial of small degree

Here we generalize Lemma 1 of [1].

Lemma 2. There is a positive constant c (depending on k) with the following property. Let $S \subseteq D^n$, with $|S| < c \cdot k^n$. Then for sufficiently large n there is a nonzero real-valued polynomial q with degree no more than

$$n \cdot (k-1) - \sqrt{n}$$

such that $q(\mathbf{x}) = 0$ for $\mathbf{x} \in S$ and $q(\mathbf{x}) \geq 0$ for $x \in D^n \backslash S$.

Proof. We will first show that a constant fraction of the k^n possible monomials have the property that both their degree and that of their conjugate are at most $\frac{1}{2}(n(k-1) - \sqrt{n})$. Consider the uniform probability distribution on these monomials. The weight of a randomly selected monomial is the sum of n independent random variables each taking the values 0 and 1 with probabilities $\frac{1}{k}$ and $\frac{k-1}{k}$, respectively. Thus the average weight of

a randomly selected monomial is $n(k-1)/k$, and the standard deviation is $\sigma = (\frac{\sqrt{k-1}}{k})\sqrt{n}$. Let $\mathcal{N}$ denote the normal cumulative distribution function. By the Central Limit Theorem, for large values of n, the probability that a randomly chosen monomial has weight between $n(k-1)/k - 2\sqrt{n}/k$ and $n(k-1)/k - 3\sqrt{n}/k$ is very close to $\mathcal{N}(-2\sqrt{n}/k\sigma) - \mathcal{N}(-3\sqrt{n}/k\sigma)$, a positive constant depending only on k. For a particular weight $w = n(k-1)/k - d\sqrt{n}/k$ in this range, the degree of a randomly chosen monomial of this weight is the sum of w independent random variables, each with value uniformly chosen from the set $\{1, \ldots, k-1\}$. The mean degree is thus $\mu_w = kw/2 = n(k-1)/2 - d\sqrt{n}/2$, and the standard deviation is

$$\sigma_w = \sqrt{w} \cdot \sqrt{\frac{k^2 - 2k}{12}} = \Theta(\sqrt{n}).$$

Recall that for any single monomial m of weight w, the sum of the degrees of m and of $\overline{m}$ is $kw = n(k-1) - d\sqrt{n}$. As d is at least 2, if the degree of m is within $\sqrt{n}/2$ of the mean μ_w then both it and the degree of $\overline{m}$ will be no more than $n(k-1)/2 - \sqrt{n}/2$, as desired. This happens with probability approximately $\mathcal{N}(2\sigma_w/\sqrt{n}) - \mathcal{N}(-2\sigma_w/\sqrt{n})$, which is asymptotically a positive constant for each w and bounded below by the positive constant for the minimum weight $w = n(k-1)/k - 3\sqrt{n}/k$. Thus the overall probability that both the degree and conjugate degree constraints are satisfied is bounded below by a positive constant c.

Now suppose $|S| < c \cdot k^n$, as in the hypothesis. Let us consider a polynomial p that is a linear combination of monomials satisfying the weight and degree constraints given above. Setting $p(\mathbf{x}) = 0$ for all $\mathbf{x} \in S$ gives a system of $|S|$ equations in at least $c \cdot k^n$ unknowns, and thus has a nontrivial solution over $\mathbf{C}$. Let us now set $q = p\overline{p}$. q is real and nonnegative on D^n and zero on S. The degree of q is bounded above by the sum of the degrees of p and $\overline{p}$, which is no more than

$$n \cdot (k-1) - \sqrt{n},$$

completing the proof.∎

Connection with bounded-depth circuits.

We quote without proof a theorem from [1]. The OR function mentioned in the statement maps $\{0,1\}^n$ to $\{0,1\}$. The polynomial obtained has integer values, and gives the value $OR(x_1, \ldots, x_n)$ on all but a small fraction of the 2^n inputs:

Lemma 3. For any $\epsilon > 0$ and any distribution of the inputs there exists a degree $O(\log(1/\epsilon)\log n)$ polynomial with integer coefficients that computes $OR(x_1, \ldots, x_n)$ with probability at least $1 - \epsilon$. ∎

We will consider unbounded fan-in boolean circuits with AND, OR and NOT gates and depth d fixed independently of the number n of inputs. These circuits ordinarily compute boolean functions from $\{0,1\}^n$ into $\{0,1\}$, but we can create circuits that compute functions from D^n into $\{0,1\}$: Such a circuit has $(k-1)n$ binary inputs, arranged in n groups of $k-1$ bits. Each input $\omega^i \in D$ is encoded by the bit string $1^i 0^{k-1-i}$.

Lemma 4. Let $\epsilon > 0$. Given a circuit as described above, with depth d and size s, there is a polynomial F of n variables with degree $O((\log(s/\epsilon)\log s)^d)$ such that F computes the function from D^n into $\{0,1\}$ computed by the circuit with no more than $k^n\epsilon$ errors.

Proof. Consider the probability distribution on the $2^{(k-1)n}$ binary inputs that assigns probability k^{-n} to each encoding of an element of D^n and 0 to all other inputs. We replace each AND gate by an OR gate whose inputs and outputs are negated, and consider the distribution of the inputs to each OR gate induced by the above distribution. By Lemma 3, for each gate there is a polynomial in $x_1, \ldots, x_{(k-1)n}$ with integer coefficients and degree $O(\log(s/\epsilon)\log s)$ that computes the OR with probability at least $1 - \epsilon/s$. By taking the composition of these polynomials we obtain a polynomial p of degree $O((\log(s/\epsilon)\log s)^d)$ that computes the boolean function computed by the circuit with probability at least $1 - \epsilon$. Thus p makes at most $k^n\epsilon$ errors on the binary encodings of elements of D^n. By interpolation we can construct $k-1$ one-variable polynomials $u_1, \ldots, u_{k-1}$ each of degree $k-1$ such that $u_j(\omega^r)$ is the j^{th} bit of the binary encoding of ω^r. Thus

$$F(\mathbf{x}) = f(u_1(x_1), \ldots, u_{k-1}(x_1), \ldots, u_1(x_n), \ldots, u_{k-1}(x_n))$$

is a polynomial whose degree is at most $k-1$ times that of f. F computes the function from D^n into $\{0,1\}$ computed by the circuit with at most $k^n\epsilon$ errors. ∎

We now proceed to our principal result. Suppose we have a family of circuits of depth d such that the n^{th} circuit in the family has n inputs, a single MAJORITY gate at the output, and gives the answer 1 if and only if the number of input bits that are on is divisible by k. We will show that the size of the circuits must grow exponentially in n. Let us choose a large value of n (how large will be determined later) and consider the circuit in the family with $(k-1)(n+1)$ binary inputs. By fixing j of the last $k-1$ input bits to 1 and the remaining $k-1-j$ of these bits to 0, we obtain for $j = 0, \ldots, k-1$ a circuit $\mathcal{C}_j$ that computes the function $g_j : D^n \to \{0,1\}$ given by

$$g_j(\omega^{i_1}, \ldots, \omega^{i_n}) = 1$$

if and only if

$$-(i_1 + \cdots + i_n) \bmod k = j.$$

We can assume that the MAJORITY gate has an odd number $(2r+1)$ of inputs. Each of these inputs is the output of an unbounded fan-in circuit of AND, OR and NOT gates whose size is less than s and depth less than d. Thus by Lemma 4 we can approximate these subcircuits by polynomials

$$f_1^{(j)}, \ldots, f_{2r+1}^{(j)},$$

each of which has degree $O((\log(\frac{s^2}{\epsilon})\log s)^{d-1})$ and makes at most $k^n \frac{\epsilon}{k(2r+1)}$ errors. Let

$$f^{(j)} = \sum_{i=1}^{2r+1} f_i^{(j)} - r.$$

Then $f^{(j)}$ has degree $O((\log(\frac{s^2}{\epsilon})\log s)^{d-1})$, and for all but $k^n\frac{\epsilon}{k}$ values of

$$\mathbf{x} = (\omega^{i_1}, \ldots, \omega^{i_n}) \in D^n,$$

$f^{(j)}(\mathbf{x}) > 0$ if and only if

$$-(i_1 + \cdots + i_n) \bmod k = j.$$

We shall now use the following fact from [4], assuring the existence of rational functions that closely approximate the sign function:

Lemma 5. Let $\delta > 0$. There is a rational function of one variable with real coefficients $S(y) = Q(y)/R(y)$ such that the degrees of Q and R are $O(\log r)$, and for each nonzero integer $y \in [-r, r+1]$,

$$S(y) = \beta(y)y/|y|,$$

where $1 \leq \beta(y) \leq 1 + \delta$. ∎

Now let

$$P(\mathbf{x}) = \sum_{j=0}^{k-1} (S(f^{(j)}(\mathbf{x})) + 1)\omega^j,$$

where S is chosen as in Lemma 5, with $\delta = 1/k$. We claim that $Re(P(\mathbf{x})\overline{M_k(\mathbf{x})}) > 0$ for all but $k^n\epsilon$ points $\mathbf{x} \in D^n$: The union of the error sets of the $f^{(j)}$ has cardinality less than $k^n\epsilon$. For a point $\mathbf{x}$ outside this union,

$$Re(P(\mathbf{x})\overline{M_k(\mathbf{x})}) = Re(c_0 + c_1\omega + \cdots c_{k-1}\omega^{k-1}),$$

where $c_0 \geq 2$ and, for $j > 0$, $c_j \geq -1/k$, which proves the claim. We can write P as a rational function whose denominator Z is real-valued on D^n. Thus, since $r < s$, $Y = PZ^2$ is a polynomial of degree

$$\Delta = O((\log(\frac{s^2}{\epsilon})\log s)^d) = O((\log s)^{2d})$$

such that

$$|\{\mathbf{x} \in D^n : Re(Y(\mathbf{x})\overline{M_k(\mathbf{x})}) \leq 0\}| \leq k^n\epsilon.$$

We now choose ϵ to be less than the value c given in Lemma 2, and take q to be the polynomial found in that lemma. Then qY weakly represents M_k, so by Lemma 1, $\Delta = \Omega(\sqrt{n})$, and thus

$$\log s = \Omega(n^{\frac{1}{4d}}).$$

We have proved:

Theorem 6. Let $k > 1$. Consider a family of unbounded fan-in circuits with AND, OR and NOT gates and with a single MAJORITY gate at the output. Suppose that the depth

of the circuits is a constant d, and that the n^{th} circuit in the family determines whether the sum of the n input bits is divisible by k. Then the size of the n^{th} circuit is

$$2^{\Omega(n^{\frac{1}{4d}})}.$$

∎

Beigel, Reingold and Spielman [4] show how to simulate an unbounded fan-in depth d threshold circuit (that is, a circuit with AND, OR, NOT and MAJORITY gates) by a depth $d+4$ circuit with a single MAJORITY gate at the output. The increase in size is small enough so that subexponential size ($2^{n^{o(1)}}$) can be preserved if the original circuit has $o(\log n)$ MAJORITY gates. Beigel [2] gives an improved construction of the same sort. His can eliminate $n^{o(1)}$ majority gates while preserving subexponential size and constant depth. This fact together with Theorem 6 implies

Theorem 7. Let $k > 1$. A constant-depth family of threshold circuits of size $2^{n^{o(1)}}$, with $n^{o(1)}$ MAJORITY gates, cannot determine whether the sum of the input bits is divisible by k.∎

An open problem.

Smolensky [7] has shown that if p is prime, a constant-depth, unbounded fan-in family of circuits with AND, OR, NOT and MOD p gates and size $2^{n^{o(1)}}$ cannot determine whether the sum of the input bits is divisible by q, unless q is a power of p. Naturally we conjecture that the same result holds even if we allow a majority gate at the output, or even $n^{o(1)}$ MAJORITY gates anywhere in the circuit. Smolensky represents the circuit's behavior by a polynomial over a field of characteristic p, while our techniques rely on polynomials over a field of characteristic 0. We wonder if some combination of these methods could be applied to circuits with both MOD p gates and MAJORITY gates.

Acknowledgements

We would like to thank Richard Beigel and Zhi-Li Zhang for various helpful discussions on this work.

References.

1. J. Aspnes, R. Beigel, M. Furst, and S. Rudich, The expressive power of voting polynomials, *Proc. 23rd ACM STOC,* 402-409, 1991.

2. R. Beigel, personal communication and presentation at Jan. 1991 DIMACS workshop.

3. R. Beigel, N. Reingold, and D. Spielman, The perceptron strikes back, *Proc. 6th Structure in Complexity Theory*, to appear, 1991. Also TR-813 (Sept. 1990), Yale Dept. of Computer Science.

4. R. Beigel, N. Reingold, and D. Spielman, PP is closed under intersection, *Proc. 23rd ACM STOC,* 1-9, 1991.

5. F. Green, An oracle separating $\oplus P$ from PP^{PH}, *Proc. 5th Structure in Complexity Theory*, 1990, 295-298.

6. M. L. Minsky and S. A. Papert, *Perceptrons* (Cambridge, MA, MIT Press, 1988). (Expanded edition, original edition was in 1968.)

7. R. Smolensky, Algebraic methods in the theory of lower bounds for Boolean circuit complexity, *Proc. 19th ACM STOC),* 77-82, 1987.