Polynomials that Sign Represent Parity and Descartes Rule of Signs

Saugata Basu*
School of Mathematics
Georgia Tech
saugata@math.gatech.edu

Nayantara Bhatnagar
College of Computing
Georgia Tech
nand@cc.gatech.edu

Parikshit Gopalan
College of Computing
Georgia Tech
parik@cc.gatech.edu

Richard J. Lipton†
College of Computing
Georgia Tech
rjl@cc.gatech.edu

Abstract

We study the sparsity of real polynomials that sign represent parity on n variables, each of which takes values from some finite subset A of integers. While the degree of such polynomials has been well studied [15, 1], relatively little is known about their sparsity. We study this problem using Descartes Rule of Signs, a classical result in algebra, relating the sparsity of a polynomial to its number of real roots.

We show that sign representing parity over $\{0, 1 \cdots, m-1\}^n$ with the degree in each variable at most $m-1$ requires sparsity at least m^n. We show a bound of $(m-1)^n$ for weak representations. We show that a tradeoff exists between sparsity and degree, by constructing a sign representation that has higher degree but lower sparsity. In some cases, the difference in sparsities is exponential. We show a lower bound of $n(m-2)+1$ on the sparsity of polynomials of any degree representing parity over $\{0, 1, \cdots, m-1\}^n$. We prove exact bounds on the sparsity of such polynomials for any two element subset A.

We show that for depth-two AND-OR-NOT circuits with a Threshold Gate at the top, the minimum circuit size for a function f equals the minimum sparsity of a polynomial sign representing f over a certain basis. We use this to give a simple proof that such circuits need size $(3/2)^n$ to compute parity, which improves on previous bounds [8]. We also show a tight lower bound of 2^n for the inner product function over $\{0,1\}^n \times \{0,1\}^n$.

The main technical tool used is Descartes rule of signs. Our bounds hold for various bases where Descartes sign rule is valid.

* Supported in part by an NSF Career Award 0133597 and a Sloan Foundation Fellowship

† Also with Telcordia. Supported in part by NSF CCR-0002299

1. Introduction

Let A be a subset of $\mathbb{Z}$ and let $X_1, \cdots X_n$ be variables taking values in A. Let $f : A^n \to \{0, 1\}$ be a 0-1 function on $X_1, \cdots, X_n$. A polynomial $P(X_1, \cdots, X_n)$ is said to sign represent f over A^n if

$$f(X_1, \cdots, X_n) = 0 \quad \Rightarrow \quad P(X_1, \cdots, X_n) > 0$$
$$f(X_1, \cdots, X_n) = 1 \quad \Rightarrow \quad P(X_1, \cdots, X_n) < 0$$

Such polynomials are also referred to as Polynomial Threshold Functions or Perceptrons.

Such representations have been well studied in computer science. They were by Minsky and Papert [15] in the context of computational learning theory. Many of the best known algorithms for learning classes of functions are obtained by showing that functions in that class can be represented by polynomials of low degree [11, 16]. They are extensively used in complexity theory to show bounds on small depth circuits. For instance, Aspnes *et al.* [1] study how well parity can be approximated by the sign of low degree polynomials. They use this to give a new proof that parity is not in AC_0. Beigel *et al.* [7] use certain polynomial constructions to prove closure properties of PP. Many related results are discussed in the survey by Beigel [5]. Finally, they are of interest as a restricted algebraic model of computation. This gives rise to some natural complexity measures, namely the minimum degree and sparsity needed to represent a function. Unlike in exact representations, polynomials that sign represent a given function are not unique, which makes the task of proving lower bounds non trivial. The survey by Saks [19] discusses many interesting questions related to algebra, geometry and combinatorics that arise from the study of sign representations.

1.1. Degree versus Sparsity

The degree of such representations is better understood than the sparsity for a couple of reasons. For many applications mentioned previously, the degree is the parameter of interest. Degree is less dependent than sparsity on the choice of A. It is invariant under linear transformations of A. Another tool which helps in studying the degree of symmetric functions is symmetrization, introduced in [15]. Let f be a symmetric function and let $P(X)$ be a polynomial that sign represents f. If $P(X)$ is symmetric, it is easier to show bounds on its degree. Otherwise,

$$Q(X) = \sum_{\sigma \in S_n} P(X_{\sigma(1)}, \cdots, X_{\sigma(n)})$$

is a symmetric polynomial of the same degree that sign represents f. Hence it suffices to show lower bounds for symmetric polynomials. Finally, suppose $A = \{a_1, \cdots a_m\}$. The polynomial $M_j = \prod_i (X_j - a_i)$ is identically zero over A for all j. Dividing a polynomial P by M_j does not change its value over A^n. Hence, it is sufficient to consider polynomials where the degree in each variable is at most $m - 1$. However it is easy to see that all the above operations could change the sparsity greatly, and hence they cannot in general be used to show bounds on sparsity.

Why is sparsity interesting and what can one hope to say about it? Sparsity of polynomials play an important role in real algebraic geometry. Unlike over algebraically closed fields, the number of real roots of a univariate real polynomial can be bounded in terms of the number of monomials appearing in it (independent of the degree). This is a consequence of a very old and classical result called Descartes Rule of Signs [4, 18]. Define the sparsity of a polynomial P to be the number of monomials that occur in it with non-zero coefficients. We will denote it by $sp(P)$.

Descartes Rule of Signs: *Let $P(X) \in \mathbb{R}[X]$ be a univariate polynomial. Then the number of positive real roots of P counted with multiplicities is bounded by the number of sign variations in the sequence of its non-zero coefficients written in order. In particular, the number of positive roots of P counted with multiplicity is bounded by $sp(P) - 1$.*

Descartes rule illustrates that for real univariate polynomials, sparsity is an important parameter controlling the number of real zeros. It forms the basis of many efficient algorithms for real root counting [4]. An important open problem in real algebraic geometry is to find proper analogues of Descartes rule for multivariate polynomials. The topological complexity (as measured by the Euler characteristics or the Betti numbers) of the real zeros of a multivariate real polynomial can still be bounded in terms of the sparsity of the polynomial independent of the degree [10, 3]. However, the known bounds are exponential in the sparsity and are believed to be nowhere near tight. A proper generalization of Descartes rule to multivariate polynomials is still elusive and remains a major open problem in real algebraic geometry (see [20] and [13] for interesting conjectures and counter-examples and [14] for results in special cases). One of our motivations is to understand the implications of Descartes rule for sign representations and to see if some non-trivial bounds on sparsity can be shown for multivariate polynomials in this simpler discrete setting. Parity is a natural candidate to begin with, since sign representing parity over $\{0, 1, \cdots, m-1\}$ is clearly related to Descartes rule in one variable.

Studying sparsity gives useful insights about sign representations. Unlike in exact representations, the polynomials that sign represent a function are not unique. A natural question to ask is what are the different *kinds* of polynomials that sign represent parity. Bounds on sparsity provide an answer to this question in some cases. For instance the polynomial $\prod_{i=1}^{n}(1 - 2X_i)$ sign represents parity over $A = \{0, 1\}$. We will show that in any multilinear sign representation, the sign of every coefficient must be the same as above. Another interesting question regarding sign representations is to study tradeoffs between parameters such as degree, sparsity and coefficient size for polynomials representing a function. Tradeoffs between coefficient size and degree are studied in [6]. We prove a degree versus sparsity tradeoff (over some bases) for parity. This is the first such tradeoff to our knowledge.

Lower bounds on the sparsity of sign representations can be used to prove circuit lower bounds. We show that for certain small depth circuits, circuit size is equal to the sparsity of sign representations of a function f over an appropriately chosen basis. This differs from most previous results relating sign representations to small depth circuits, in that the important parameter has been the degree [5].

1.2. Our Results

We study the following problem: *Let $A = \{a_1, \cdots a_m\}$ be a set of integers of cardinality m. Let $X_i \in A$ for $i \in [n]$. Define the parity function $Par(X_1, \cdots X_n)$ to be 0 if the sum of the X_i is even and 1 otherwise. What is the sparsity of polynomials that sign represent parity over A^n?*

We start by considering $A = \{0, 1\}$. We reprove the result of Minsky and Papert [15] that any polynomial that sign represents parity has degree n and sparsity 2^n. We generalize this to show that representing parity over $\{0, 1, \cdots, m-1\}^n$ with the degree in each variable at most $m - 1$ requires sparsity m^n. Moreover, we show that the sign of each coefficient is fixed.

Weak representations can have exponentially smaller sparsity than strong representations. For instance, the polynomial $\prod_i X_i$ weakly represents parity over $\{0,1\}^n$. We show a bound of $(m-1)^n$ on the sparsity of weak representations over $\{0,1,\cdots m-1\}^n$, with the degree in each variable at most $m-1$. In fact for $\{1,\cdots,m\}^n$, we show a lower bound of m^n even for weak representations.

These results show that low degree representations must have high sparsity. We show that a tradeoff exists by a simple construction of small sparsity but higher degree. For $A=\{1,2\}$ this gives sparsity $n+1$ with degree n^2. For $A=\{0,1,\cdots m-1\}$, our construction gives a weak representation of sparsity $o(m^n)$. We show that for certain sets A of size m, the sparsity can be as low as $n(m-1)+1$.

We then show a lower bound of $n(m-1)+1$ on the sparsity for polynomials of any degree representing parity over $\{1,\cdots,m\}^n$. This allows us to prove tight upper and lower bounds for any 2 element set A. For large values of m there is a large gap between this lower bound and our upper bounds. We are unable to close this gap.

We study depth-two AND-OR-NOT circuits with a threshold gate at the top. We show that the minimum circuit size for a function f in this model equals the minimum sparsity of a polynomial sign representing f over a certain generating set of polynomials. Using this we give a simple proof that any such circuit for parity has size at least $(\frac{3}{2})^n$. The best bound known previously was $(\frac{4}{3})^{\frac{n}{2}}$. We also show a lower bound of 2^n for computing the inner product function over $\{0,1\}^n \times \{0,1\}^n$ which is tight.

We use the self-reducibility of parity to prove that polynomials representing parity over A^n must have a certain structure: they must be built from polynomials which represent parity in fewer variables over A. For instance, if $P(X_1,\cdots,X_n)$ sign represents parity over $\{0,1,\cdots,m-1\}^n$, and if the degree in X_n is bounded by $m-1$, then we show that

$$P(X_1,\cdots,X_n)=\sum_{i=0}^{m-1} X_n^i Q_i(X_1,\cdots,X_{n-1})$$

where $(-1)^i Q_i$ represents parity on $n-1$ variables. We make extensive use of Descartes rule of signs which relates the roots of a univariate polynomial to its sparsity and the signs of its coefficients. In generalizing results for $\{0,1\}^n$ to $\{0,1,\cdots,m-1\}^n$, we use properties of the generalized Vandermonde matrices and their inverses, which are in fact consequences of Descartes rule. One of the main difficulties in studying sparsity is that a change of basis can affect sparsity greatly. Our techniques overcome this obstacle and hold for many different bases where Descartes rule is valid.

1.3. Related Work

Minsky and Papert prove that representing parity over $\{0,1\}$ inputs needs degree n and sparsity 2^n [15]. Krause and Pudlak [12] show that there is a Boolean function f that has exponential sparsity in the $\{-1,1\}$ basis but polynomial sparsity in the $\{0,1\}$ basis. O'Donnell and Servedio [17] study various extremal properties of such representations. They show that over the $\{-1,1\}$ basis, every Boolean function has sparsity at most $(1-\frac{1}{O(n)})2^n$. Many other results regarding sparsity of random Boolean functions in the $\{-1,+1\}$ basis can be found in [17, 19].

Constant depth AND-OR-NOT circuits with a threshold gate at the top have been studied previously (see [8] and the references therein). Goldmann shows that any such circuit with constant depth for computing parity needs exponential size. There it is shown using random restrictions that a depth-two circuit computing parity needs size $\frac{1}{4}(\frac{4}{3})^{\frac{n}{2}}$, this is attributed to Hofmeister, Krause and Pudlak.

2. Preliminaries

We use the vector notation $X_{[n]}$ to denote $X_1,\cdots,X_n$ and $a_{[n]}$ to denote $a_1,\cdots,a_n$. If $P(X_{[n]})$ is a polynomial in $X_1,\cdots,X_n$, we use $P(X_{[n-1]},c)$ to denote the polynomial in $X_1,\cdots,X_{n-1}$ obtained by substituting $X_n=c$ in $P(X_{[n]})$. For $c\in\mathbb{R}$, the sign of c denoted $sgn(c)$ is is $+1,-1$ or 0 depending on whether c is positive, negative or 0.

For a subset A of the integers, we define the Parity function $Par(X_{[n]})$ on A^n as follows: for $a_{[n]}\in A^n$, $Par(a_{[n]})=0$ if $\sum_i a_i$ is even and $Par(a_{[n]})=1$ if $\sum_i a_i$ is odd.

The degree of a polynomial $P(X_{[n]})$ is denoted by $Deg(P)$. It is the maximum of $\sum d_i$ over all monomials $\prod_i X_i^{d_i}$ that occur in $P(X_{[n]})$ with non-zero support. We will also use the degree in each variable. In addition to the overall sparsity $sp(P)$, we also define the sparsity in a particular variable X_i to be the number of distinct powers of X_i that occur in $P(X_{[n]})$. Note that this is different from the number of monomials of $P(X)$ in which X_i appears. Clearly if $P(X_{[n]})$ sign represents f, then $-P(X_{[n]})$ sign represents $\overline{f}$.

Definition 2.1 *[1] A polynomial $P(X_{[n]})$ weakly sign represents f over A^n if on every point $a_{[n]}\in A^n$*

$$\begin{array}{lcl} f(a_{[n]})=0 & \Rightarrow & P(a_{[n]})\geq 0 \\ f(a_{[n]})=1 & \Rightarrow & P(a_{[n]})\leq 0 \end{array}$$

and further f is not zero at every point in A^n.

If a polynomial does not output 0 anywhere in A^n, we call it a strong representation. A sign representation by default will mean a strong representation. The next lemma states that the set of polynomials that sign represent f is closed under positive linear combinations and will be used repeatedly.

Lemma 2.2 *For $i\in[k]$, let $P_i(X_{[n]})$ be polynomials in $\mathbb{R}[X_{[n]}]$ that weakly sign represent f and let c_i be positive* *reals. Let*

$$Q(X_{[n]}) = \sum_{i\in[k]} c_i P_i(X_{[n]})$$

Then $Q(X_{[n]})$ weakly sign represents f.

Proof: Let $a_{[n]} \in A^n$. Suppose $f(a_{[n]}) = 0$. Then

$$\forall i, P_i(a_{[n]}) \geq 0 \quad \Rightarrow \quad Q(a_{[n]}) = \sum_{i=1}^{k} c_i P_i(a_{[n]}) \geq 0$$

Similarly if $f(a_{[n]}) = 1, Q(a_{[n]}) \leq 0$. Further if for some index i, $P_i(a_{[n]}) \neq 0$ then $Q_i(a_{[n]}) \neq 0$. Since the polynomials $P_i(X_{[n]})$ are not identically zero over A^n, $Q(X_{[n]})$ is also not identically zero over A^n. Hence it is a weak sign representation of f. ∎

It follows from the proof that if any of the polynomials $P_i(X_{[n]})$ strongly represent f, then so does $Q(X_{[n]})$.

2.1. Descartes Rule of Signs

Theorem 2.3 [18, 4] Descartes Rule of Signs: *Let $P(X_1) = \sum_{i=0}^{n} c_i X_1^i$ be a real univariate polynomial. Let s denote the number of sign changes in the sequence $\{c_0, c_1, \cdots, c_n\}$. The number of positive roots of $P(X_1)$ counted with multiplicity is bounded by s.*

In fact, Descartes rule holds for the Newton basis, the Bernstein basis and many other bases of polynomials. We refer the reader to either [18, 4] for details. The next two statements are simple corollaries of Descartes rule.

Corollary 2.4 *Let $P(X_1) = \sum_{i=0}^{n} c_i X_1^i$ be a real univariate polynomial. Let s' denote the number of sign changes in the sequence $\{c_0, -c_1, \cdots, (-1)^n c_n\}$. The number of negative roots of $P(X_1)$ counted with multiplicity is bounded by s'.*

Corollary 2.5 *Let $P(X_1) = \sum_{i=0}^{n} c_i (X_1 - b)^i$ be a univariate polynomial written in the monomial basis shifted by b. Let s denote the number of sign changes in the sequence $\{c_0, c_1, \cdots, c_n\}$. The number of roots of $P(X_1)$ in $(b, +\infty)$ counted with multiplicity is bounded by s.*

Lemma 2.6 *[18] Let $d_1, \cdots, d_k$ be integers such that $0 \leq d_1 < d_2 \cdots < d_k$. Let $a_1, \cdots, a_k$ be positive real numbers such that $a_1 < a_2 \cdots < a_k$. Let*

$$V = \begin{pmatrix} a_1^{d_1} & a_1^{d_2} & \cdots & a_1^{d_k} \\ a_2^{d_1} & a_2^{d_2} & \cdots & a_2^{d_k} \\ \cdots & \cdots & \cdots & \cdots \\ a_k^{d_1} & a_k^{d_2} & \cdots & a_k^{d_k} \end{pmatrix}$$

Then $Det(V) > 0$.

We refer to this matrix as the generalized Vandermonde matrix. Note that the standard Vandermonde matrix corresponds to setting $d_1 = 0, d_2 = 1, \cdots d_k = k-1$.

Proof: We will instead work with the determinant

$$\begin{vmatrix} a_k^{d_k} & a_{k-1}^{d_{k-1}} & \cdots & a_1^{d_1} \\ a_{k-1}^{d_k} & a_{k-1}^{d_{k-1}} & \cdots & a_{k-1}^{d_1} \\ \cdots & \cdots & \cdots & \cdots \\ a_1^{d_k} & a_1^{d_{k-1}} & \cdots & a_1^{d_1} \end{vmatrix}$$

This determinant is obtained from the original one by an even number of row and column interchanges hence it has the same sign. The proof is by induction on k. The case $k = 1$ is trivial. Assume that the statement holds up to $k-1$. Now consider the polynomial

$$\begin{aligned} C(X) &= \begin{vmatrix} X^{d_k} & X^{d_{k-1}} & \cdots & X^{d_1} \\ a_{k-1}^{d_k} & a_{k-1}^{d_{k-1}} & \cdots & a_{k-1}^{d_1} \\ \cdots & \cdots & \cdots & \cdots \\ a_1^{d_k} & a_1^{d_{k-1}} & \cdots & a_1^{d_k} \end{vmatrix} \\ &= c_k X^{d_k} + c_{k-1} X^{d_{k-1}} \cdots c_1 X^{d_1} \end{aligned}$$

The sparsity of $C(X)$ is bounded by k, hence by Descartes rule, it has at most $k-1$ positive roots. But $a_1, \cdots, a_{k-1}$ are roots of $C(X)$. Hence there are no other roots. Hence the sign at a_k (or at any point to the right of a_{k-1}) is the same as the sign at $+\infty$. This in turn is the sing of the leading coefficient c_k of $C(X)$. But this coefficient is precisely

$$c_k = \begin{vmatrix} a_{k-1}^{d_{k-1}} & \cdots & a_{k-1}^{d_1} \\ \cdots & \cdots & \cdots \\ a_1^{d_{k-1}} & \cdots & a_1^{d_1} \end{vmatrix}$$

which is positive by the induction hypothesis. ∎

Lemma 2.7 *Let U be the inverse of the Generalized Vandermonde Matrix V. Then $sgn(u_{ij}) = (-1)^{i+j}$ where $0 \leq i, j < k$.*

Proof: From the formula for the inverse of a matrix,

$$u_{ij} = (-1)^{i+j} \frac{Det(V_{ji})}{Det(V)}$$

where V_{ji} is the minor of the entry v_{ji}. V_{ji} is also a generalized Vandermonde matrix. Hence by Lemma 2.6, both $Det(V_{ji})$ and $Det(V)$ are strictly positive. Hence $sgn(u_{ij}) = (-1)^{i+j}$. ∎

We will also need the case when $a_1 = 0$. If $d_1 > 0$, then clearly the first row is all 0s and the determinant vanishes. On the other hand, if $d_1 = 0$ we get the matrix

$$V_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & a_2^{d_2} & \cdots & a_2^{d_k} \\ 1 & \cdots & \cdots & \cdots \\ 1 & a_k^{d_2} & \cdots & a_k^{d_k} \end{pmatrix}$$

Lemma 2.8 *Let U be the inverse of the matrix V_0. Then $sgn(u_{ij}) = 0$ if $i = 0, j \geq 1$ and $(-1)^{i+j}$ otherwise.*

3. Lower Bounds

3.1. Strong Representations

We first consider the case when $A = \{0, 1\}$. Assume that $P(X_{[n]})$ sign represents parity over $\{0, 1\}^n$. Over 0-1 inputs, for $k \geq 1, X_i^k = X_i$. We can use this to reduce $P(X_{[n]})$ to a multilinear polynomial. This substitution can only reduce the degree and the sparsity of P. So we may assume that $P(X_{[n]})$ is multilinear.

Lemma 3.1 *If $P(X_{[n]})$ sign represents parity over $\{0, 1\}^n$,*

$$P(X_{[n]}) = X_n Q_1(X_{[n-1]}) + Q_0(X_{[n-1]}) \quad (1)$$

where $Q_0(X_{[n-1]})$ sign represents parity on $n-1$ variables and $Q_1(X_{[n-1]})$ sign represents the complement of parity.

Proof: Since $P(X_{[n]})$ is multilinear,

$$\begin{aligned} P(X_{[n]}) &= X_n Q_1(X_{[n-1]}) + Q_0(X_{[n-1]}) \\ P(X_{[n-1]}, 0) &= Q_0(X_{[n-1]}) \\ P(X_{[n-1]}, 1) &= Q_1(X_{[n-1]}) + Q_0(X_{[n-1]}) \\ \Rightarrow Q_1(X_{[n-1]}) &= P(X_{[n-1]}, 1) - P(X_{[n-1]}, 0) \end{aligned}$$

The polynomial $P(X_{[n-1]}, 0)$ represents parity on $n-1$ variables. The polynomial $P(X_{[n-1]}, 1)$ represents its complement. Hence $Q_1(X_{[n-1]}) = P(X_{[n-1]}, 1) - P(X_{[n-1]}, 0)$ is the sum of two polynomials that sign represent the complement of parity. Hence Q_0 and Q_1 are as claimed. ∎

The polynomial $\prod_{i=1}^{n}(1 - 2X_i)$ sign represents parity over $\{0, 1\}^n$. We will show that one cannot do better with regard to degree or sparsity.

Theorem 3.2 *[15] If $P(X_{[n]})$ sign represents parity over $\{0, 1\}^n$, then it must have degree n and sparsity 2^n.*

Proof: Observe that the sparsity bound of 2^n implies that every monomial including $\prod_{i=1}^{n} X_i$ has a non-zero coefficient, hence the degree is n. So it is sufficient to prove the sparsity bound. The proof is by induction on n.

When $n = 1$, we want $P(X_1) = aX_1 + b$ so that $P(0) = b > 0$ and $P(1) = a + b < 0$. This implies $b > 0$ and $a < -b < 0$. Hence the claim is true. Assume inductively that the claim holds for $n-1$ variables. Write $P(X_{[n]})$ as in Lemma 3.1. Observe that $sp(P) = sp(Q_0) + sp(Q_1)$, since there cannot be cancelations between the monomials in $X_n Q_1(X_{[n-1]})$ and $Q_0(X_{[n-1]})$. By the induction hypothesis $sp(Q_0) = sp(Q_1) = 2^{n-1}$, hence $sp(P) = 2^n$. ∎

We can strengthen the claim to show that not only must all monomials have non-zero coefficients but even their signs must the same as in $\prod_{i=1}^{n}(1 - 2X_i)$. For $S \subset [n]$, we denote the coefficient corresponding to the monomial $\prod_{i \in S} X_i$ by c_S.

Theorem 3.3 *If $P(X_{[n]})$ sign represents parity on $\{0, 1\}^n$, the sign of the coefficient c_S corresponding to the monomial $\prod_{i \in S} X_i$ is $(-1)^{|S|}$.*

Proof: We have already verified the case $n = 1$ in Theorem 3.2. Assume inductively that the claim holds for $n-1$ variables. Write $P(X_{[n]})$ as in Lemma 3.1. The monomials involving X_n come from $X_n Q_1(X_{[n-1]})$ while those not involving X_n come from $Q_0(X_{[n-1]})$. Now consider $S \subset [n]$ such that $n \notin S$. The coefficient c_S in $P(X_{[n]})$ is the same as the coefficient in $Q_0(X_{[n-1]})$. Since $Q_0(X_{[n-1]})$ represents parity on $n-1$ variables, hence $sgn(c_S) = (-1)^{|S|}$ by the induction hypothesis. For $S \subset [n]$ such that $n \in S$, the coefficient c_S in $P(X_{[n]})$ is equal to the coefficient $c_{S \setminus \{n\}}$ in $Q_1(X_{[n-1]})$. Since $Q_1(X_{[n-1]})$ represents the complement of parity, $sgn(c_S) = -(-1)^{|S|-1} = (-1)^{|S|}$ by induction ∎

We will now generalize the above theorem to the case when $A = \{0, \cdots, m-1\}$ and the degree in each variable is at most $m-1$. To construct a polynomial sign representing parity satisfying these conditions, for $0 \leq j \leq m-2$, pick $\alpha_j \in \{j, j+1\}$. Set

$$P(X_{[n]}) = \prod_{i=1}^{n} \prod_{j=0}^{m-2} (-1)^m (X_i - \alpha_j)$$

It can be verified that $P(X_{[n]})$ indeed sign represents parity on A^n. For $i \in [n]$, let $M(X_i) = \prod_{j=0}^{m-1}(X_i - j)$. Clearly these polynomials are zero at every point in A^n. Given any polynomial $P(X_{[n]})$ which sign represents parity over A^n, we can divide $P(X_{[n]})$ by the polynomials $M(X_i)$ so that the remainder has degree at most $m-1$ in each X_i. However we will show that such polynomials require sparsity m^n.

Lemma 3.4 *If $P(X_{[n]})$ sign represents parity over $\{0, 1, \cdots, m-1\}^n$, and if the degree in X_n is bounded by $m-1$, then*

$$P(X_{[n]}) = \sum_{i=0}^{m-1} X_n^i Q_i(X_{[n-1]}) \quad (2)$$

where $(-1)^i Q_i(X_{[n-1]})$ represents parity on $n-1$ variables.

Proof: Since $deg(X_n) \leq m-1$, by collecting monomials by the power of X_n which occurs in them,

$$P(X_{[n]}) = \sum_{i=0}^{m-1} X_n^i Q_i(X_{[n-1]})$$

By substituting values 0 through $m-1$ for X_n, we get

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1^{m-1} \\ 1 & \cdots & \cdots & \cdots \\ 1 & m-1 & \cdots & (m-1)^{m-1} \end{pmatrix} \begin{pmatrix} Q_0(X_{[n-1]}) \\ Q_1(X_{[n-1]}) \\ \cdots \\ Q_{m-1}(X_{[n-1]}) \end{pmatrix} = \begin{pmatrix} P(X_{[n-1]},0) \\ P(X_{[n-1]},1) \\ \cdots \\ P(X_{[n-1]},m-1) \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} Q_0(X_{[n-1]}) \\ Q_1(X_{[n-1]}) \\ \cdots \\ Q_{m-1}(X_{[n-1]}) \end{pmatrix} = U \begin{pmatrix} P(X_{[n-1]},0) \\ P(X_{[n-1]},1) \\ \cdots \\ P(X_{[n-1]},m-1) \end{pmatrix}$$

where $U = (u_{ij})$ is the inverse of the Vandermonde matrix. By Lemma 2.8, $sgn(u_{ij}) = 0$ if $i = 0, j \geq 1$ and $(-1)^{i+j}$ otherwise. This implies

$$Q_0(X_{[n-1]}) = cP(X_{[n-1]}, 0)$$

where $c > 0$. Hence Q_0 sign represents parity on $n-1$ variables. For $i \geq 1$, $u_{ij} = (-1)^{i+j} c_{ij}$ where $c_{ij} > 0$.

$$\begin{aligned} Q_i(X_{[n-1]}) &= \sum_{j=0}^{m-1} (-1)^{i+j} c_{ij} P(X_{[n-1]}, j) \\ (-1)^i Q_i(X_{[n-1]}) &= \sum_{j=0}^{m-1} (-1)^{j} c_{ij} P(X_{[n-1]}, j) \end{aligned}$$

For each j, the polynomial $(-1)^j c_{ij} P(X_{[n-1]}, j)$ represents parity on $n-1$ variables. Hence by Lemma 2.2, each $(-1)^i Q_i$ also represents parity on $n-1$ variables. ∎

By using an induction argument similar to Theorem 3.2, we can prove the following claim.

Theorem 3.5 *Let $P(X_{[n]})$ be a polynomial that sign represents parity over $\{0,1,\cdots,m-1\}^n$, with $deg(X_i) \leq m-1$ for all i. Then $P(X_{[n]})$ has sparsity m^n and the sign of the coefficient of the monomial $\prod_j X_j^{i_j}$ is $(-1)^{\sum_j i_j}$.*

Proof: The base case $n = 1$ is an application of Descartes rule. Let

$$P(X_1) = \sum_{i=0}^{m-1} c_i X_1^i$$

Let $0 \leq k \leq m-2$. Since $P(k)$ and $P(k+1)$ have opposite signs, $P(X_1)$ has a root α_k in the interval $(k, k+1)$. Since the degree of $P(X_1)$ is bounded by $m-1$,

$$P(X_1) = c_{m-1} \prod_{k=0}^{m-2} (X_1 - \alpha_k)$$

To determine the sign of c_{m-1}, substitute $X_1 = 0$.

$$P(0) = (-1)^{m-1} c_{m-1} \prod_{k=0}^{m-2} \alpha_k$$

Since $P(X_1)$ represents parity, $P(0) > 0$. Since all the α_k are positive, we must have $sgn(c_{m-1}) = (-1)^{m-1}$. Now applying Descartes rule, since P has $m-1$ positive roots, there must be m sign changes in the sequence $\{c_0, \cdots, c_{m-1}\}$. Hence, $sgn(c_i) = (-1)^i$. This implies that $sp(P) = m$.

The inductive case proceeds using Lemma 3.4 exactly as in Theorem 3.3. We skip the proof. ∎

Corollary 3.6 *If $P(X_{[n]})$ sign represents parity over $\{0,1,\cdots,m-1\}^n$, it has degree at least $n(m-1)$.*

Proof: We can divide $P(X_{[n]})$ by the polynomials $M(X_i)$ so that the degree in each X_i is bounded by $m-1$. This can only reduce the overall degree. By Theorem 3.5, the sparsity is at least m^n. This implies that every monomial $\prod_i X_i^{d_i}, d_i \leq m-1$ occurs with non-zero coefficients. This proves that the overall degree is at least $n(m-1)$. ∎

The same proof extends to sets of the form $A = \{a, a+1, \cdots, a+m-1\}$ for $a > 0$. This implies the following corollary, by taking $a = 1$ and $m = 2$.

Corollary 3.7 *If $P(X_{[n]})$ is a multilinear polynomial that sign represents parity over $\{1,2\}^n$, it has sparsity 2^n.*

A natural question is what happens if we allow polynomials of higher degree. It might be that there are polynomials of high degree and low sparsity and applying the degree reduction causes the sparsity to blow up to m^n. We will address this question in Section 4.

However, we can relax the requirement in Theorem 3.5 that the degree in each variable is bounded by $m-1$ to the sparsity in each variable is bounded by m and show that the overall sparsity still needs to be m^n. Hence if locally in each variable the sparsity is small, overall the sparsity needs to be large. We will prove this by showing that an analogue of Lemma 3.4 holds.

Theorem 3.8 *Let $P(X_{[n]})$ be a polynomial that sign represents parity over $\{0,1,\cdots,m-1\}^n$, with the sparsity in each variable X_i at most m. Then $P(X_{[n]})$ has sparsity m^n.*

Proof: Since $P(X_{[n]})$ has at most m distinct power $d_0, \cdots, d_{m-1}$ of X_n, by grouping monomials by the power of X_n, we get

$$P(X_{[n]}) = \sum_{i=0}^{m-1} X_n^{d_i} Q_i(X_{[n-1]})$$

We will show that $(-1)^i Q_i(X_{[n-1]})$ represents parity on $n-1$ variables. Substituting values $0, \cdots, m-1$ for

X_n gives us a system of equations similar to what we had before, except that we have a generalized Vandermonde matrix V_g. By Lemma 2.8, the sign of the ij^{th} entry in V_g^{-1} is the same as in U. Hence the same proof holds and $(-1)^i Q_i$ represents parity on $n-1$ variables. The sparsity bound follows by induction. ∎

Finally we consider sparsity in the *shifted* basis $\{1, (X-b), (X-b)^2, \cdots, \}$. For instance parity over $\{0,1\}$ can be represented with sparsity 1 by taking $b = 0.5$ but we have seen that it needs sparsity 2^n in the usual monomial basis. However for large m, the sparsity does not change too much by shifting.

Corollary 3.9 *Let $P(X)$ be a polynomial that sign represents parity over $\{0,1,\cdots,m-1\}^n$. If the sparsity of each X_i in P is bounded by $\lceil \frac{m}{2} \rceil$ in the monomial basis shifted by b, then P has sparsity at least $\lceil \frac{m}{2} \rceil^n$.*

This follows trivially from the observation that for any b, the polynomial sign represents parity on $\lceil \frac{m}{2} \rceil$ points either to the left or to the right of b. This shows that the exponential decrease in the $\{0,1\}$ case produced by shifting to $b = 0.5$ was because m was 2. For $m > 2$, the sparsity in any shifted basis is exponential in n.

3.2. Weak Representations

We now study weak representations for parity with low degree. Over $A = \{0,1\}$, the polynomial $(-1)^n \prod_i X_i$ gives a weak representation with sparsity 1, and in fact this is optimal with regard to degree too.

Lemma 3.10 *[15, 1] Any polynomial that weakly sign represents parity over $\{0,1\}^n$ has degree n.*

We show that over $A = \{0,1,\cdots,m-1\}^n$, a lower bound of $(m-1)^n$ still applies for weak representations when the degree in each variable is at most $m-1$.

Lemma 3.11 *If $P(X_{[n]})$ weakly sign represents parity over $A = \{0,1,\cdots,m-1\}^n$, and if the degree in X_n is bounded by $m-1$, then*

$$P(X_{[n]}) = \sum_{i=0}^{m-1} X_n^i Q_i(X_{[n-1]}) \qquad (3)$$

where for $i \geq 1$, $(-1)^{i-1} Q_i(X_{[n-1]})$ weakly represents parity on $n-1$ variables.

Proof: The main difference from Lemma 3.4 is that we need to show that the polynomials $Q_i(X_{[n-1]})$ are not zero at some point in A^{n-1}. By substituting values 0 through $m-1$ for X_n and inverting the Vandermonde matrix, we get

$$\begin{pmatrix} Q_0(X_{[n-1]}) \\ Q_1(X_{[n-1]}) \\ \cdots \\ Q_{m-1}(X_{[n-1]}) \end{pmatrix} = U \begin{pmatrix} P(X_{[n-1]}, 0) \\ P(X_{[n-1]}, 1) \\ \cdots \\ P(X_{[n-1]}, m-1) \end{pmatrix}$$

For $i \geq 1$ and suitable constants $c_{ij} > 0$, we get

$$\begin{aligned} Q_i(X_{[n-1]}) &= \sum_{j=0}^{m-1} (-1)^{i+j} c_{ij} P(X_{[n-1]}, j) \\ (-1)^i Q_i(X_{[n-1]}) &= \sum_{j=0}^{m-1} (-1)^j c_{ij} P(X_{[n-1]}, j) \end{aligned}$$

Since $P(X_{[n]})$ weakly represents parity on A^n, each polynomial $(-1)^i P(X_{[n-1]}, i)$ for $i \in A$ either weakly represents parity, or it is identically 0 over A^{n-1}. Since $P(X_{[n]})$ being a weak representation is not identically zero on A^n, at least one of the polynomials $P(X_{[n-1]}, i)$ is not identically zero over A^{n-1}. Hence, for $i \geq 1$ the polynomials $(-1)^i Q_i(X_{[n-1]})$ weakly sign represent parity on A^{n-1} by Lemma 2.2.

Note that $Q_0(X_{[n-1]}) = cP(X_{[n-1]}, 0)$ could be identically zero on A^{n-1}. ∎

We will use this to show a lower bound of $(m-1)^n$ on the sparsity of weak representations over A^n. The base case $n = 1$ is proved using Lemma 3.10.

Lemma 3.12 *Any univariate polynomial $P(X_1)$ that weakly sign represents parity over $\{0,\cdots,m-1\}$ has degree at least $m-1$.*

Proof: Assume that $deg(P(X_1)) < m-1$. Let $Y_1, \cdots, Y_{m-1} \in \{0,1\}$. Set

$$Q(Y_{[m-1]}) = P(\sum_{i=1}^{m-1} Y_i)$$

Clearly $Q(Y_{[m-1]})$ weakly sign represents parity on $\{0,1\}^{m-1}$ and has degree less than $m-1$, which contradicts Lemma 3.10. ∎

Lemma 3.13 *Any univariate polynomial $P(X_1)$ that weakly sign represents parity over $\{0,\cdots,m-1\}$ must have $m-2$ roots in the interval $(0, m-1]$.*

Proof: Consider the factorization of $P(X_1)$ over the reals. It may have irreducible factors of degree 2. Irreducible factors do not have roots in $(-\infty, +\infty)$ and hence they do not change sign. So we will have a sign representation even after factoring out such irreducible factors. Similarly, consider a linear factor of the form $(X_1 - \alpha)$ where $\alpha \notin [0, m-1]$. $(X_1 - \alpha)$ is non-zero over $[0, m-1]$ and its sign remains the same hence it can be factored out. This leaves us with a product of the form $\prod(X_1 - \alpha_i)$, where $\alpha_i \in [0, m-1]\ \forall i$. Further we may assume that there is a root of multiplicity at most 1 at 0. If not, we can write $P(X_1) = X_1^k Q(X_1)$ for $k \geq 2$. The polynomial $X_1 Q(X_1)$ has the same sign over each point in $[0, m-1]$ and only smaller degree.

We are now left with a polynomial $P(X_1) = \prod_i (X_1 - \alpha_i)$, $\alpha_i \in [0, m-1]$ which weakly sign represents parity on A. By Lemma 3.12, this polynomial has degree at least $m-1$. Since 0 is a root of multiplicity at most 1, at least $m-2$ of the roots α_i lie in the interval $(0, m-1]$. ∎

Theorem 3.14 *Let $P(X_{[n]})$ be a polynomial that weakly represents parity over $\{0, 1, \cdots, m-1\}^n$. If the degree of each X_i is bounded by $m-1$, then P has sparsity at least $(m-1)^n$.*

Proof: The proof is by induction. When $n = 1$, by Lemma 3.13 $P(X_1)$ has $m-2$ roots in $(0, m-1]$. Hence by Descartes rule, it has sparsity $m-1$. We now use Lemma 3.11 and go by induction on n. From Equation 3 it follows that

$$sp(P) \geq \sum_{i=1}^{m-1} sp(Q_i)$$

By 3.11, for $i \geq 1$, each Q_i weakly represents parity or its complement, hence $sp(Q_i) \geq (m-1)^{n-1}$ by induction. Hence $sp(P) \geq (m-1)^n$. ∎

In the case when $A = \{1, \cdots, m\}$, if we restrict the degree in each variable to $m-1$, we have shown that a lower bound of m^n holds for strong representations. We will show that exactly the same bound holds even for weak representations.

Corollary 3.15 *Let $P(X_{[n]})$ be a polynomial that weakly represents parity over $\{1, \cdots, m\}^n$. If the degree of each X_i is bounded by $m-1$, then P has sparsity at least m^n.*

Proof: Let $P(X_{[n]})$ be as above. We claim that the polynomial

$$Q(X_{[n]}) = P(X_{[n]}) \cdot \prod_{i=1}^{n} X_i$$

weakly represents parity over $\{0, 1, \cdots, m\}^n$. This is because, if any of the inputs $X_i = 0$, then $Q(X_{[n]}) = 0$. Otherwise, since $X_i > 0$ for each i, $Q(X_{[n]})$ has the same sign as $P(X_{[n]})$. The degree in each variable in $Q(X_{[n]})$ is at most m. Hence, by applying Theorem 3.14, we get $sp(Q) \geq m^n$. But $sp(P) = sp(Q)$. Hence we get a lower bound on the sparsity of $P(X_{[n]})$. ∎

4. Upper Bounds

Does the lower bound of m^n in Theorem 3.5 hold for all polynomials? Or can we tradeoff degree and sparsity and find high degree polynomials that beat the m^n bound? At least for some choices of A, such a tradeoff is indeed possible.

Theorem 4.1 *There exists a polynomial $P(X_{[n]})$ that sign represents parity over $\{1, 2\}^n$ with degree n^2 and sparsity $n+1$.*

Proof: Set $W = \prod_{i=1}^n X_i$. W maps $\{1, 2\}^n$ to $n+1$ distinct points $\{1, 2, \cdots, 2^n\}$ in $\mathbb{R}$. It maps all permutations of a given vector to the same point, but if two vectors are not permutations of each other, they are mapped to different points. Choose points $\alpha_k \in (2^{k-1}, 2^k)$ for $1 \leq k \leq n$ and let

$$Q(W) = \prod_{k=1}^{n} (W - \alpha_k)$$

The polynomial $P(X_{[n]}) = Q(\prod_{i=1}^n X_i)$ sign represents parity. Its sparsity is $n+1$ and its degree is n^2. ∎

In contrast by Corollary 3.7 we know that for multilinear polynomials, the overall sparsity should be 2^n. A similar construction is possible for any set A, take $W = \prod_i X_i$ and then construct a univariate polynomial in W as above. This may not give a strong representation for all A. For instance if $n = 2$ and $A = \{1, 2, 3, 4\}$, then taking $W = X_1 X_2$ maps both $(2, 2)$ and $(1, 4)$ to the same point, though their parities are different. However this happens because $4 = 2^2$. In general if the above construction does not work, then there must be integers $c_1, \cdots, c_m$ not all zero such that $\sum_i c_i \log(a_i) = 0$. We can use this observation to get a sign representation over the monomial basis shifted by ε.

Theorem 4.2 *For any set A of cardinality m, $\exists\ \varepsilon > 0$ such that parity can be sign represented over A^n in the monomial basis shifted by ε by a polynomial $P_\varepsilon(X_{[n]})$ of sparsity at most $\binom{n+m-1}{n}$ and degree $n \cdot \binom{n+m-1}{n}$.*

Proof: Let $a_{[n]} = (a_1, \cdots, a_n)$ and $b_{[n]} = (b_1, \cdots, b_n)$ be inputs in A^n which are not permutations of each other. Consider the set of ε such that

$$\prod_i (a_i - \varepsilon) \quad = \quad \prod_i (b_i - \varepsilon) \tag{4}$$

The polynomial in ε, $\prod_i (a_i - \varepsilon)$ has roots $\{a_1, \cdots, a_n\}$ whereas the polynomial $\prod_i (b_i - \varepsilon)$ has roots $\{b_1, \cdots, b_n\}$. Since they don't have the same multi-set of roots, they must be distinct polynomials. Their difference is a non-zero polynomial of degree at most n in ε. Hence there are at most n solutions. Hence there is only a finite set S of values for ε such that equality holds in Equation(4) for some pair a, b.

Pick an $\varepsilon \in \overline{S}$. Set $W = \prod_i (X_i - \varepsilon)$. If $a_{[n]}$ and $b_{[n]}$ are permutations of each other, they are mapped to the same point in $\mathbb{R}$. If not, by our choice of ε, they are mapped to distinct points. Hence, the points of A^n are mapped to $\binom{n+m-1}{n}$ distinct points in $\mathbb{R}$. We now construct a polynomial $Q_\varepsilon(W)$ that has the right sign at each point W. We get the desired polynomial $P_\varepsilon(X_{[n]})$ by setting $W = \prod_i (X_i - \varepsilon)$ in $Q_\varepsilon(W)$. Its sparsity is bounded by $\binom{n+m-1}{n}$ in the monomial basis shifted by ε. The degree is at most $n \cdot \binom{n+m-1}{n}$. ∎

In the standard monomial basis, we can still get a *weak representation* of parity.

Theorem 4.3 *For any set A of non-negative integers of cardinality m, parity can be weakly sign represented over A^n by a polynomial $P_\omega(X_{[n]})$ that has sparsity at most $\binom{n+m-1}{n}$ and degree $n \cdot \binom{n+m-1}{n}$.*

Proof: Set $W = \prod_i X_i$. This maps points in A^n to at most $\binom{n+m-1}{n}$ distinct points in $\mathbb{Z}$. We now construct a polynomial $P_\omega(W)$ as follows. It may happen that points in A^n that differ in parity are mapped to the same point. P_ω will output 0 on such points. At the other points, P_ω outputs a positive or negative value depending on the parity of the point in A^n. We need to show that P_ω is not always 0. Let $c > 0$ be the maximum value in A. Only the point $(c, c, \cdots, c)$ is mapped to c^n, and here P_ω is non-zero. Hence P_ω is a weak representation with the desired parameters. ∎

We have bounded the sparsities of the polynomials P_ε and P_ω by $\binom{n+m-1}{n}$. We will study these constructions in more detail when $A = \{1, \cdots, m\}$. In fact, for n fixed the sparsity of the polynomial $P_\omega(X_{[n]})$ is $o(m^n)$. This is a consequence of Erdos' multiplication table theorem which states that the number of distinct integers less than m^n which can be expressed as the product of n numbers each less than m is $o(m^n)$ [2]. We will show a nearly tight lower bound.

Call a point w in $\mathbb{Z}$ *ambiguous* if there exist $a_{[n]}, b_{[n]}$ in A^n which differ in parity, but however $\prod_i a_i = \prod_i b_i = w$. We first show that there are many ambiguous points.

Lemma 4.4 *For m sufficiently large, the number of ambiguous points is at least $\Omega\left(\frac{m}{\log m}\right)^n$.*

Proof: We prove the Lemma for $n = 2$, the case of larger n is similar. Consider all points of the form $4pq$ where p, q are distinct odd primes less than $\frac{m}{4}$. Such points are indeed ambiguous since $(2p, 2q)$ and $(4p, q)$ are both mapped to $4pq$. Further $Par(2p, 2q) \neq Par(4p, q)$. Since $p, q \leq \frac{m}{4}$, this ensures $(4p, q)$ and $(2p, 2q)$ are in $\{1, \cdots, m\}^2$. Each choice of a pair of distinct primes gives a different ambiguous point. The number of such pairs is $\Omega\left(\frac{m}{\log m}\right)^2$ by the prime number theorem [9]. ∎

Corollary 4.5 *For $A = \{0, 1, \cdots, m\}$ and m sufficiently large, the sparsity of the polynomial P_ω constructed in Theorem 4.3 is $\Omega\left(\frac{m}{\log m}\right)^n$.*

Proof: The polynomial P_ω must output 0 on every ambiguous point w. Hence it has $\Omega\left(\frac{m}{\log m}\right)^n$ positive roots by Lemma 4.4. Now apply Descartes rule. ∎

Corollary 4.6 *For $A = \{0, 1, \cdots, m\}$ and m sufficiently large, for sufficiently small $\varepsilon > 0$, the sparsity of the polynomial P_ε constructed in Theorem 4.2 is $\Omega\left(\frac{m}{\log m}\right)^n$.*

Proof: Consider the sequence S_ε consisting of the numbers $\prod_i (a_i - \varepsilon)$ for $a_{[n]} \in A^n$ in sorted order. To each point in S_ε, we associate a sign according to $Par(a_{[n]})$. Note that by the choice of ε, we can associate these signs unambiguously. The number of sign changes in S_ε is a lower bound on the sparsity of P_ε by Descartes rule.

Consider the sequence S consisting of the integers $\prod_i a_i$ in sorted order. We will show that the number of ambiguous points in S is a lower bound on the number of sign changes in S_ε. Corresponding to each ambiguous point w in S, we have points $a_w = \prod_i (a_1 - \varepsilon)$ and $b_w = \prod_i (b_1 - \varepsilon)$ in S_ε such that $\prod_i a_i = \prod_i b_i = w$, and $Par(a_w) = 1, Par(b_w) = 0$. For each ambiguous point w, by alternately picking a_w or b_w from S_ε, we can get a subsequence where the sign alternates. We have to show that this is an increasing subsequence of S_ε. We will show that this is true for sufficiently small $\varepsilon > 0$. Consider $c_{[n]}$ and $d_{[n]}$ such that

$$\begin{aligned} & \prod_i c_i > \prod_i d_i \\ \Rightarrow\ & \prod_i c_i - \prod_i d_i \geq 1 \\ \Rightarrow\ & \lim_{\varepsilon \to 0} \prod_i (c_i - \varepsilon) - \prod_i (d_i - \varepsilon) \geq 1 \end{aligned}$$

Hence for ε sufficiently small, for every such pair $c_{[n]}$ and $d_{[n]}$,

$$\prod_i (c_i - \varepsilon) \quad > \quad \prod_i (d_i - \varepsilon)$$

By Lemma 4.4, this implies that S_ε has $\Omega\left(\frac{m}{\log m}\right)^n$ sign changes. ∎

On the other hand, for some sets A, the sparsity can be significantly smaller. For instance let $A = \{1, 2, \cdots 2^{m-1}\}$. Setting $W = \prod_{i=1}^n X_i$ maps A^n to the set $\{1, 2, \cdots, 2^{(m-1)n}\}$, which contains only $n(m-1)+1$ distinct points. So in this case, the sparsity of the polynomial P_ω is upper bounded by $n(m-1)+1$.

5. Lower Bounds without Degree Restrictions

We will now show a lower bound which holds for all polynomials strongly representing parity on $\{1, \cdots, m\}^n$ without any restrictions on the degree or sparsity of each variable. The proof is a generalization of the proof idea of Theorem 3.5.

Theorem 5.1 *Let $P(X_{[n]})$ be a polynomial which sign represents parity over $\{1, \cdots, m\}^n$. Then P has sparsity at least $n(m-1)+1$.*

Proof: The proof is by induction on n. When $n = 1$, the claim follows by Descartes rule. Assume it is true for $n-1$. Let the sparsity in X_n be k. If we set all the other variables to 0, the resulting univariate polynomial sign represents parity on $\{1, \cdots, m\}$, hence it must have sparsity $\geq m$. Hence $k \geq m$. If $k > n(m-1)+1$, there is nothing to prove. Hence we may assume $m \leq k \leq n(m-1)$.

Let $P(X_{[n]})$ be a polynomial in n variables that sign represents parity. Grouping monomials by the power of X_n they contain,

$$P(X_{[n]}) \quad = \quad \sum_{i=1}^{k} X_n^{d_i} Q_i(X_{[n-1]})$$

By substituting values 1 through m for X_n, we get

$$\begin{pmatrix} 1 & \cdot\cdot & 1 & 1 & \cdot\cdot & 1 \\ 2^{d_1} & \cdot\cdot & 2^{d_m} & 2^{d_{m+1}} & \cdot\cdot & 2^{d_k} \\ \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ m^{d_1} & \cdot\cdot & m^{d_m} & m^{d_{m+1}} & \cdot\cdot & m^{d_k} \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \\ \cdots \\ Q_k \end{pmatrix} = \begin{pmatrix} P(X_{[n-1]}, 1) \\ P(X_{[n-1]}, 2) \\ \cdots \\ P(X_{[n-1]}, m) \end{pmatrix}$$

We denote the $m \times k$ matrix by A. We have an under determined system of equations. Hence we will not be able to show that each $Q_i(X_{[n-1]})$ represents parity or its complement. However we will show that appropriate linear combinations of the Q_i sign represent parity. To begin with we pre-multiply by U which is the inverse of the $m \times m$ generalized Vandermonde matrix consisting of the first m columns of A.

$$\begin{pmatrix} 1 & \cdot\cdot & 0 & b_{1,m+1} & \cdot\cdot & b_{1,k} \\ 0 & \cdot\cdot & 0 & b_{2,m+1} & \cdot\cdot & b_{2,k} \\ \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot & \cdot\cdot \\ 0 & \cdot\cdot & 1 & b_{m,m+1} & \cdot\cdot & b_{m,k} \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \\ \cdots \\ Q_k \end{pmatrix} = U \begin{pmatrix} P(X_{[n-1]}, 1) \\ P(X_{[n-1]}, 2) \\ \cdots \\ P(X_{[n-1]}, m) \end{pmatrix}$$

Using the sign alternations of the entries of U, we conclude that for $1 \leq i \leq m$ the polynomials

$$R_i = Q_i + \sum_{j=m+1}^{k} b_{ij} Q_j$$

sign represent parity or its complement on $n-1$ variables. Hence by applying the induction hypothesis, $sp(R_i) \geq (n-1)(m-1)+1$. Clearly the sparsity of R_i is no more than $sp(Q_i) + \sum_{j=m+1}^{k} sp(Q_j)$. Hence we get

$$sp(Q_i) + \sum_{j=m+1}^{k} sp(Q_j) \geq (n-1)(m-1)+1$$

By choosing the matrix U to be the inverse of an appropriate sub-matrix, we can obtain a similar equation for any subset of the Q_is of cardinality $k-m+1$. There are $\binom{k}{m-1}$ such subsets. Each Q_i occurs in exactly $\binom{k-1}{m-1}$ of them. Hence we get

$$\binom{k-1}{m-1} \sum_{i=1}^{k} sp(Q_i) \geq \binom{k}{m-1}((n-1)(m-1)+1)$$
$$\Rightarrow \sum_{i=1}^{k} sp(Q_i) \geq \frac{k}{k-m+1}((n-1)(m-1)+1)$$

The quantity $\frac{k}{k-m+1}$ monotonically decreases as k increases. In the range $m \leq k \leq n(m-1)$, it is always greater than $\frac{n(m-1)+1}{(n-1)(m-1)+1}$ which is the value it takes for $k = n(m-1)+1$. Hence

$$\sum_{i=1}^{k} sp(Q_i) \geq n(m-1)+1$$

But $sp(P) = \sum_{i=1}^{k} sp(Q_i)$, hence the claim is proved. ∎

Corollary 5.2 *Any polynomial that sign represents parity over $\{1, 2\}^n$ must have sparsity at least $n+1$.*

This follows by substituting $m = 2$ in Theorem 5.1. This tells us that the construction of Theorem 4.1 is optimal. With this in hand, for any set $\{a, b\}$, we can prove optimal lower bounds on polynomials sign representing parity.

- If $\{sgn(a), sgn(b)\} = \{-1, +1\}$, $\exists\ P(X)$ with sparsity 1 that represents parity.
- If $\{sgn(a), sgn(b)\} = \{+1, +1\}$ or $\{-1, -1\}$, then any such polynomial has sparsity at least $n+1$.
- Over $\{0, a\}$ any such polynomial has sparsity at least 2^n.

While the lower bound of $n(m-1)+1$ in Theorem 5.1 is tight for $m = 2$, this is far from the upper bound of Theorem 4.1 for large m since this upper bound is at least $\left(\frac{m}{\log m}\right)^n$.

Corollary 5.3 *Let $P(X_{[n]})$ be a polynomial which sign represents parity over $\{0, 1, \cdots, m-1\}^n$. Then P has sparsity at least* $\max(2^n, n(m-2)+1)$.

Proof: A strong representation over $\{0, 1, \cdots, m-1\}^n$ gives a strong representation over $\{1, \cdots, m-1\}^n$. Hence Theorem 5.1 implies a lower bound of $n(m-2)+1$. It is also a strong representation over $\{0, 1\}^n$, hence Theorem 3.2 implies a lower bound of 2^n.

6. Circuit Lower Bounds

We shall use bounds on the sparsity of parity to derive lower bounds on the size of certain restricted circuits. The circuits we consider are rather weak, however the proof of the lower bound is simple and yields better parameters than were previously known. From now onwards, we will only consider Boolean inputs coming from $\{0, 1\}^n$.

Definition 6.1 *A threshold gate with inputs $X_{[n]}$ computes a function $f(X_{[n]}) = \sum_i w_i X_i + w_0$ for some real values $w_0, \cdots, w_n$ and outputs 1 if f is positive and 0 otherwise.*

We shall study depth two circuits which comprise of the inputs and their complements, followed by a single level of AND and OR gates, followed by a Threshold gate at the top. It is easy to see that without loss of generality, one may assume that the circuit has only AND gates and no OR gates. We will refer to such circuits as a threshold of ANDs following [8]. We will show that circuit size in this model exactly corresponds to sparsity over a certain *basis*. Thus proving circuit lower bounds is equivalent to proving bounds on the sparsity of sign representations.

To begin with, we take a toy example. Assume the inputs are available only in uncomplemented form. Each AND gate computes a function of the form $\prod_{i \in A} X_i$ where A is the set of inputs into the gate. A threshold of ANDs circuit for computing parity corresponds to a sign representation of parity in the monomial basis. The number of AND gates is exactly the number of non-constant monomials required. By Theorem 3.2, this is $2^n - 1$.

We now allow inputs and their complements at level 0. The expression computed by each AND gate is of the form

$$\prod_{i \in I} X_i \cdot \prod_{j \in J} (1 - X_j) \tag{5}$$

where I and J are disjoint subsets of $\{1, \cdots, n\}$. This looks similar to monomials in the Bernstein basis for multilinear polynomials in n variables [4] where the monomials are of the form $\prod_{i \in A} X_i \prod_{j \in A^c} (1 - X_j)$. It is easy to show that any polynomial that sign represents parity in the multilinear Bernstein basis has sparsity 2^n.

We define a set of polynomials B_n^* that span the $\mathbb{R}$-vector space of polynomials of degree at most n. In 1 variable $B_1^* = \{X_1, 1 - X_1, 1\}$. We define B_n^* to be the product of the bases in each variable. Hence the monomials in B_n^* are all expressions of the form $\prod_{i \in I} X_i \cdot \prod_{j \in J} (1 - X_j)$ where $I \cap J = \phi$. There are 3^n monomials in B_n^* and they include the monomials of the Bernstein basis. Since there is a linear dependence among the generators, we call B_n^* a generating set for multilinear polynomials. This linear dependence implies that there are many ways to express any polynomial as a combination of the generators. So for a polynomial P, $sp(P)$ denotes the minimum sparsity over all different ways of expressing P as a linear combination of elements in B_n^*.

Lemma 6.2 *The minimum size of a threshold of ANDs circuit for a Boolean function f is exactly the minimum sparsity of a sign representation of f over the generating set B_n^*.*

This follows from the observation made earlier that the expressions computed by the AND gates correspond to the generators in B_n^*.

Theorem 6.3 *Any circuit which is a threshold of AND gates and computes parity on $\{0, 1\}^n$ has size at least $\left(\frac{3}{2}\right)^n$.*

Proof: We will show that any polynomial sign representing parity has sparsity greater than $(3/2)^n$ over the generating set B_n^* . The proof is by induction on n. For $n = 1$, since none of the generators $\{X_1, 1 - X_1, 1\}$ alone or their multiples gives a sign representation, any representation has sparsity at least 2. Assume the statement holds for $n - 1$. Let $P(X_{[n]})$ sign represent parity. Consider the sparsest representation of P over B_n^*.

$$P([X_n]) = \sum_{I \cap J = \phi} c_{I,J} \prod_{i \in I} X_i \prod_{j \in J} (1 - X_j)$$

Grouping together monomials where X_n appears, monomials where $(1 - X_n)$ appears, and those where neither appears, we get

$$\begin{aligned} P(X_{[n]}) = X_n A(X_{[n-1]}) + (1 - X_n) B(X_{[n-1]}) \\ + C(X_{[n-1]}) \end{aligned}$$

Writing A, B and C for short, we have $sp(P) = sp(A) + sp(B) + sp(C)$. Substituting for X_n,

$$\begin{aligned} P(X_{[n-1]}, 0) &= B + C \\ P(X_{[n-1]}, 1) &= A + C \\ P(X_{[n-1]}, 0) - P(X_{[n-1]}, 1) &= B - A \end{aligned}$$

All the polynomials on the LHS represent either parity or its complement on $n - 1$ variables. By applying the induction hypothesis,

$$\begin{aligned} sp(B) + sp(A) \ &\geq \ sp(B - A) \ > \ (3/2)^{n-1} \\ sp(B) + sp(C) \ &\geq \ sp(B + C) \ > \ (3/2)^{n-1} \\ sp(A) + sp(C) \ &\geq \ sp(A + C) \ > \ (3/2)^{n-1} \\ \Rightarrow sp(P) = sp(A) &+ sp(B) + sp(C) > (3/2)^n \end{aligned}$$

∎

Proposition 6.4 *There is a threshold of ANDs circuit computing parity on* $\{0,1\}^n$ *of size* $O(5^{\frac{n}{3}})$.

Proof: Consider the polynomial

$$Q(X_{[3]}) = X_1X_2X_3 + X_1(1-X_2)(1-X_3) \\ +X_2(1-X_3)(1-X_1) + X_3(1-X_1)(1-X_2)$$

Note that Q is 1 iff $Par(X_{[3]})$ is odd. The polynomial $P(X_{[3]}) = 1 - 2Q(X_{[3]})$ sign represents parity on 3 variables and has sparsity 5. To sign represent parity on n variables, divide them into $\frac{n}{3}$ groups of 3 variables each. Apply P to each group and multiply. This construction has sparsity $O(5^{\frac{n}{3}})$. ▮

Define the inner product function on $\{0,1\}^n \times \{0,1\}^n$ by

$$IP(X,Y) = \sum_i X_iY_i \bmod 2$$

Theorem 6.5 *Any circuit which is a threshold of AND gates and computes the inner product on* $\{0,1\}^n \times \{0,1\}^n$ *has size at least* 2^n.

The proof is similar to the proof for parity and can be found in the Appendix. In fact this bound is tight.

Proposition 6.6 *There is a threshold of ANDs circuit computing the inner product on* $\{0,1\}^n \times \{0,1\}^n$ *of size* 2^n.

Proof: We construct a polynomial of sparsity 4 that sign represents IP on $\{0,1\}^2 \times \{0,1\}^2$. The general upper bound follows by dividing the inputs into blocks of 4. Let

$$Q(X_{[2]}, Y_{[2]}) = X_1Y_1 + X_2Y_2 - 2X_1Y_1X_2Y_2$$

Note that Q exactly computes IP. The polynomial $P(X_{[2]}, Y_{[2]}) = 1 - 2Q(X_{[2]}Y_{[2]})$ sign represents IP and has sparsity 4. ▮

7. Conclusions

Our results indicate that studying the sparsity of sign representations raises many interesting questions. For large values of m, there is a considerable gap between the upper and lower bounds for polynomials representing parity over A^n with no restrictions on the degree. We would like to be able to extend our techniques to prove lower bounds on the size of circuits of larger depth. Finally, can one say anything interesting about the sparsity of polynomials that represent the mod-k function, when k is not a power of 2?

Acknowledgments

We would like to thank Richard Beigel for pointing out the upper bound of Proposition 6.4, as well as many pointers to literature. We would like to thank Ernie Croot for the reference to Erdos' multiplication table theorem [2].

References

[1] James Aspnes, Richard Beigel, Merrick Furst, and Steven Rudich. The expressive power of voting polynomials. *Combinatorica*, pages 14(2):1–14, 1994.

[2] Laszlo Babai, Carl Pomerance, and Peter Vertesi. The mathematics of paul erdos. *Notices of the AMS*, January 1998.

[3] Saugata Basu. On bounding the betti numbers and computing the euler characteristics of semi-algebraic sets. *Discrete and Computational Geometry*, 22:1–18, 1999.

[4] Saugata Basu, Richard Pollack, and Marie-Francoise Roy. *Algorithms in Real Algebraic Geometry*. Springer, Algorithms and Computation in Mathematics , Vol. 10, 2003.

[5] Richard Beigel. The polynomial method in circuit complexity. *Structures in Complexity Theory: 8th Annual Conference*, pages 82–95, 1993.

[6] Richard Beigel. Perceptrons, pp, and the polynomial hierarchy. *Computational Complexity*, 4:339–349, 1994.

[7] Richard Beigel, Nick Reingold, and Daniel Spielman. PP is closed under intersection. *Journal of Computer and System Sciences*, pages 50(2):191–202, 1995.

[8] Mikael Goldmann. On the power of a threshold gate at the top. *Information Processing Letters*, pages 63(6): 287–293, (1997).

[9] G.H. Hardy and E.M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, 1979.

[10] A.G. Khovanskii. *Fewnomials*. Translations of Mathematical Monographs, Vol.88, AMS, 1991.

[11] Adam Klivans and Rocco Servedio. Learning dnf in time $2^{O(n^{\frac{1}{3}})}$. *Proceedings of the Thirty-Third Annual Symposium on Theory of Computing*, pages 258–265, 2001.

[12] Matthias Krause and Pavel Pudlak. On computing boolean functions by sparse real polynomials. *Proceedings of the Thirty-sixth Annual Symposium on Foundations of Computer Science*, pages 682–691, 1995.

[13] J.C Lagaraias and Richardson T.J. Multivariate descartes rules of signs and sturmfels's challenge problem. *The Mathematical Intelligencer*, 19(3):9–15, 1997.

[14] T.Y Li, M. Rojas, and X. Wang. Counting real connected components of trinomials curve intersections and monomial hypersurfaces. *Discrete and Computational Geometry*, 30:379–414, 2003.

[15] Marvin Minsky and Seymour Papert. *Perceptrons: an Introduction to Computational Geometry*. MIT Press, 1968.

[16] Elchannan Mossel, Ryan O'Donnell, and Rocco Servedio. Learning juntas. *Proceedings of the Thirty-Fifth Annual Symposium on Theory of Computing*, pages 206–212, 2003.

[17] Ryan O'Donnell and Rocco Servedio. Extremal properties of polynomial threshold functions. *Proceedings of the Eighteenth Annual Conference on Computational Complexity*, pages 3–12, 2003.

[18] George Polya and Gabor Szego. *Problems and Theorems in Analysis, Vol. II.* Springer-Verlag, New York, 1976.

[19] Michael Saks. Slicing the hypercube. *Surveys in Combinatorics*, pages 211–255, 1993.

[20] Bernd Sturmfels. Polynomial equations and convex polytopes. *American Mathematical Monthly*, 105(10):907–922, 1998.

Appendix

Theorem 6.5: Any circuit which is a threshold of AND gates and computes the inner product on $\{0,1\}^n \times \{0,1\}^n$ has size at least 2^n.

Proof: The proof is by induction on n. Note that $IP(X,Y)$ is a function defined on $2n$ variables. When $n = 1$, $IP(X,Y) = X_1 \cdot Y_1$. It is easy to verify that no single monomial in B_2^* sign represents IP. Hence the sparsity is at least 2. Assume the statement holds for $n-1$. Consider the sparsest way to express P over B_n^*. Grouping the monomials according to X_n, Y_n, where each A_i is a polynomial in $X_1, Y_1, \cdots, X_{n-1}, Y_{n-1}$,

$$\begin{aligned} P(X_{[n]}, Y_{[n]}) \quad = \quad & X_n Y_n A_1 + X_n(1-Y_n)A_2 + \\ & (1-X_n)Y_n A_3 + (1-X_n)(1-Y_n)A_4 + X_n A_5 \\ & + Y_n A_6 + (1-X_n)A_7 + (1-Y_n)A_8 + A_9 \end{aligned}$$

Now substituting for X_n, Y_n and writing $P(0,1)$ for $P(X_{[n-1]}, 0, Y_{[n-1]}, 1)$ and so on,

$$\begin{aligned} P(0,1) &= A_3 + A_6 + A_7 + A_9 && (6) \\ P(1,0) &= A_2 + A_5 + A_8 + A_9 && (7) \\ P(1,1) &= A_1 + A_5 + A_6 + A_9 && (8) \end{aligned}$$

Subtracting equation (8) from (6) and (7),

$$\begin{aligned} P(0,1) - P(1,1) &= -A_1 + A_3 - A_5 + A_7 && (9) \\ P(1,0) - P(1,1) &= -A_1 + A_2 - A_6 + A_8 && (10) \end{aligned}$$

In the above equations the LHS represents IP or its complement on $2n-2$ variables, so their sparsity is at least 2^{n-1} by the induction hypothesis. Applying this observation to equations (6), (7), (9) and (10)

$$\begin{aligned} sp(A_3) + sp(A_6) + sp(A_7) + sp(A_9) \geq 2^{n-1} \\ sp(A_2) + sp(A_5) + sp(A_8) + sp(A_9) \geq 2^{n-1} \\ sp(A_1) + sp(A_3) + sp(A_5) + sp(A_7) \geq 2^{n-1} \\ sp(A_1) + sp(A_2) + sp(A_6) + sp(A_8) \geq 2^{n-1} \end{aligned}$$

Adding these equations, we get

$$2\left(\sum_{i=1}^{9} sp(A_i)\right) - 2sp(A_4) \geq 4 \cdot 2^{n-1}$$

$$\Rightarrow sp(P) = \sum_{i=1}^{9} sp(A_i) \geq 2^n$$

∎