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

June 15, 2021

Abstract

A real polynomial $P(X_1, \ldots, X_n)$ sign represents $f : A^n \to \{0, 1\}$ if for every $(a_1, \ldots, a_n) \in A^n$, the sign of $P(a_1, \ldots, a_n)$ equals $(-1)^{f(a_1, \ldots, a_n)}$. Such sign representations are well-studied in computer science and have applications to computational complexity and computational learning theory. The work in this area aims to determine the minimum degree and sparsity possible for a polynomial that sign represents a function f. While the degree of such polynomials is relatively well-understood, far less is known about their sparsity. Known bounds apply only to the cases where $A = \{0, 1\}$ or $A = \{-1, +1\}$.

In this work, we present a systematic study of tradeoffs between degree and sparsity of sign representations through the lens of the parity function. We attempt to prove bounds that hold for any choice of set A. We show that sign representing parity over $\{0, \ldots, m-1\}^n$ with the degree in each variable at most $m - 1$ requires sparsity at least m^n. We show that a tradeoff exists between sparsity and degree, by exhibiting a sign representation that has higher degree but lower sparsity. We show a lower bound of $n(m-2) + 1$ on the sparsity of polynomials of any degree representing parity over $\{0, \ldots, m-1\}^n$. We prove exact bounds on the sparsity of such polynomials for any two element subset A. The main tool used is Descartes' Rule of Signs, a classical result in algebra, relating the sparsity of a polynomial to its number of real roots.

As an application, we use bounds on sparsity to derive circuit lower bounds for depth-two AND-OR-NOT circuits with a Threshold Gate at the top. We use this to give a simple proof that such circuits need size 1.5^n to compute parity, which improves the previous bound of $\frac{4}{3}^{n/2}$ [Gol97]. We show a tight lower bound of 2^n for the inner product function over $\{0,1\}^n \times \{0,1\}^n$.

1 Introduction

Let A be a subset of $\mathbb{Z}$ and let $f : A^n \to \{0,1\}$ be a function on A^n.

Definition 1.1 *A polynomial $P(X_1, \ldots, X_n) \in \mathbb{R}[X_1, \cdots, X_n]$ exactly represents f over A^n if for every $(a_1, \ldots, a_n) \in A^n$, $P(a_1, \ldots, a_n) = f(a_1, \ldots, a_n)$.*

Exact representations of functions by polynomials have been studied extensively in computer science, where they have numerous applications in circuit lower bounds[Raz87, Smo87], hardness of approximation [Hås01] and computational learning [MOS03]. In these applications, the set A is generally taken to be $\{-1, +1\}$ or $\{0, 1\}$.

In this paper, we study a less strict notion of representation of a function by a polynomial, which is called sign representation.

Definition 1.2 *[MP68] A polynomial $P(X_1, \ldots, X_n) \in \mathbb{R}[X_1, \ldots, X_n]$ sign represents f over A^n if for every $(a_1, \ldots, a_n) \in A^n$,*

$$f(a_1, \ldots, a_n) = 0 \quad \Rightarrow \quad P(a_1, \ldots, a_n) > 0,$$
$$f(a_1, \ldots, a_n) = 1 \quad \Rightarrow \quad P(a_1, \ldots, a_n) < 0.$$

Such polynomials are also referred to as Polynomial Threshold Functions for f or Perceptrons. Sign representations have been studied in computational complexity theory, where they were used by Beigel, Reingold and Spielman to show that the complexity class PP is closed under complement [BRS95]. Beigel *et al.* use such representations to show lower bounds on AC^0[BRS91]. We refer the reader to the survey by Beigel on applications of such polynomials in complexity theory[Bei93].

Further motivation for studying sign representations comes from Valiant's PAC model for computational learning [KV94]. If a class of functions on n variables can be sign represented by degree d polynomials, then that class can be learnt in time $n^{O(d)}$ in the PAC-learning model (see [KS01] for a precise statement of this result). Indeed, the best known algorithms for PAC-learning central concept classes like DNF formulas and intersections of halfspaces use this approach [KS01, KOS02]. For this application, there are two parameters of interest: the degree of the polynomial and the size of its coefficients. The former determines the running time of the algorithm, whereas the latter determines the number of samples required [KS04].

While much of the work on computational learning focuses on the Boolean case where the set A is taken to be $\{-1, +1\}$ or $\{0, 1\}$, it is quite natural to consider classes of functions such as decision trees when the variables take values from larger sets, especially $\{0, \ldots, m\}$. This is explicitly stated as an open problem in [MOS03] for a class of functions called *juntas* which are functions that depend only on some unknown subset of $X_1, \ldots, X_n$ of size k where $k = O(\log n)$. It is well-known and easy to show that juntas are a special class of decision trees [MOS03].

Finally, we will consider an even weaker notion of representing a function by a polynomial which is called weak representation, introduced by Aspnes, Beigel, Furst and Rudich.

Definition 1.3 *[ABFR94] A polynomial $P(X_1, \ldots, X_n) \in \mathbb{R}[X_1, \ldots, X_n]$ weakly sign represents f over A^n if for every $(a_1, \ldots, a_n) \in A^n$,*

$$f(a_1, \ldots, a_n) = 0 \quad \Rightarrow \quad P(a_1, \ldots, a_n) \geq 0,$$
$$f(a_1, \ldots, a_n) = 1 \quad \Rightarrow \quad P(a_1, \ldots, a_n) \leq 0.$$

and further $P(X_1, \ldots, X_n)$ does not vanish over the set A^n.

Weak sign representations (or weak representations for short) have been used in computational complexity to show circuit lower bounds. Aspnes *et al.* [ABFR94] and Klivans [Kli01] use weak representations to show that the parity function cannot even be approximated by AC^0 circuits.

Finally, polynomial representations have been studied 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. These measures, and tradeoffs between them have been investigated previously by many researchers [MP68, Bei93, KP95, OS03a, OS03b, KS01, KS04, KOS02]. Polynomial representations have also been studied over finite fields and rings of positive characteristic. This study has yielded useful insights into computational complexity [Raz87, Smo87, BBR94], computational learning [MOS03] and combinatorics [Gro00, Gro02, Gop06].

1.1 Our Results

Definition 1.4 *For $A \subset \mathbb{Z}$, the parity function* $\mathrm{PAR} : A^n \to \{0, 1\}$ *is defined as*

$$\mathrm{PAR}(a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i \pmod 2.$$

We will restrict our attention to the setting where A consists of non-negative integers, though our methods can be applied to arbitrary sets $A \subset \mathbb{Z}$. We define the sparsity $\mathrm{sp}(P)$ of a polynomial $P(X_1, \ldots, X_n)$ to be the number of monomials in its support when the polynomial is written in the standard monomial basis.

In this work, we present a systematic study of tradeoffs between degree and sparsity of sign representations through the lens of the parity function. Our methods also apply to related functions such as inner-product mod 2 (see Definition 6.7). While tradeoffs between degree and sparsity have been investigated by several researchers [Bei94, KP95, KS04], previous work focused on the case $A = \{0, 1\}$ or $A = \{-1, +1\}$. In contrast, we attempt to prove bounds that hold for any choice of set A. To motivate this, consider the problem of representing Boolean functions on the n-dimensional hypercube by polynomials. One could identify the hypercube with the set $\{a, b\}^n$ for any $a \neq b \in \mathbb{R}$. Indeed, this freedom to choose the set A is crucially used by the algorithm of Mossel *et al.* for learning juntas [MOS03]. Thus it is natural to study polynomial representations for arbitrary sets A. While it is known that the minimum degree of polynomials representing a function does not depend on the choice of a and b, it is unclear how this affects other parameters such as coefficient-size and sparsity.

Obtaining bounds in this general setting is challenging unlike degree, the minimum sparsity of polynomials representing a function is known to vary greatly with the choice of the set A. However, we show that one can completely classify the minimum sparsity required to represent parity for any set A of size 2. We obtain non-trivial lower bounds on the sparsity for arbitrary sets A of any size. We obtain tight sparsity lower bounds if we assume upper bounds on the degree of the polynomial. We show that there are tradeoffs between the degree and the sparsity of sign representations. Below we present exact statements of our main results.

We reprove the result of Minsky and Papert [MP68] that any polynomial that sign represents parity over $\{0, 1\}^n$ has degree n and sparsity 2^n. We generalize this to show that representing parity over $\{0, \ldots, m - 1\}^n$ with the degree in each variable at most $m - 1$ requires sparsity m^n. This result shows that low degree representations must have high sparsity. We show a tradeoff between degree and sparsity by exhibiting sign representations of lower sparsity but higher degree. We show a lower bound of $n(m - 2) + 1$ on the sparsity for polynomials of any degree representing parity over $\{0, \ldots, m - 1\}^n$. This allows us to prove tight upper and lower bounds for the case $|A| = 2$. For large sets A, we are unable to close the gap between our upper and lower bounds.

Our results indicate that studying sparsity gives useful insights into sign representations. For instance, consider the polynomials sign-representing $f : \{0,1\}^n \to \{0,1\}$. If we place the restriction that each variable X_i appears with degree at most 1, there is a unique polynomial that exactly represents every function f. However even with this restriction, the polynomials that sign represent a function are not unique: for any $0 < a < b$, the polynomial $\prod_{i=1}^n (a - bX_i)$ sign represents parity over $\{0,1\}^n$. We show that in any sign representation of parity, the sign of the coefficient corresponding to the monomial $\prod_{i \in S} X_i$ must be $(-1)^{|S|}$, as in the polynomial above. Thus all sign representations have some similar structure.

As an application of our methods, we show that lower bounds on the sparsity of sign representations can be used to prove circuit lower bounds for Thresholds of Ands circuits [Gol97] (see Definition 6.2). We give a simple proof that any such circuit for parity requires 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. While our methods are elementary, they give better lower bounds than those obtained by using the powerful random restriction method [Gol97]. Our sparsity-based approach also differs from most previous results which related the degree of sign-representations to the size of small depth circuits[ABFR94, BRS91]. Building on our work, Amano and Maruoka recently used LP-based methods to prove lower bounds on circuits comprising on Thresholds of symmetric gates that compute the Inner Product function [AM05].

1.2   Our Techniques

The degree of sign representations is better understood than the sparsity for a couple of reasons. Degree is less dependent than sparsity on the choice of A [MOS03]. A tool which helps in studying the degree of polynomials sign representing symmetric functions is symmetrization [MP68]: we can assume that the minimum degree polynomial sign representing a symmetric Boolean function is symmetric. Further, if $|A| = m$, we can assume that the minimum degree polynomial representing a function has degree at most $m - 1$ in each variable. However, such assumptions cannot be made in the context of sparsity.

Our main technical contribution is to show that non-trivial lower bounds on the sparsity of sign-representations can be obtained using some elementary techniques and a classical result from algebra called Descartes' Rule of Signs. Unlike over algebraically closed fields, Descrates' rule of signs shows that 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). 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 $\mathrm{sp}(P)$.

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 $\mathrm{sp}(P) - 1$.*

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 [BPR03]. 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 [Kho91, Bas99]. 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 [Stu98] and [LT97] for interesting conjectures and counter-examples and [LRW03] for results in special cases). A small first step in this direction might be to show tight sparsity bounds for multivariate sign representations of parity for arbitrary sets A.

1.3 Related Work

Minsky and Papert prove that representing parity over $\{0,1\}$ inputs needs degree n and sparsity 2^n [MP68]. Krause and Pudlak [KP95] 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 [OS03a] study various extremal properties of such representations. The sparsity of random Boolean functions on $\{-1,+1\}^n$ have been studied in [OS03a, Sak93].

2 Preliminaries

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

The degree of a polynomial $P(X_1, \ldots, X_n)$ denoted by $\deg(P)$ is the maximum of $\sum d_i$ over all monomials $\prod_i X_i^{d_i}$ that occur in the support of $P(X_1, \ldots, X_n)$. The degree in the variable X_i which is denoted $\deg_i(P)$ is the maximum of d_i over all monomials in the support of $P(X_1, \ldots, X_n)$. A multilinear polynomial is one where $\deg_i(P) \leq 1$ for all i. The sparsity of a polynomial P denoted $\mathrm{sp}(P)$ is the number of non-zero monomials in its support. We also define the sparsity in the variable X_i which we denote $\mathrm{sp}_i(P)$ to be the number of distinct powers of X_i that occur in $P(X_1, \ldots, X_n)$. Note that this is different from the number of monomials in which X_i appears. Given a function $f : A^n \to \{0,1\}$ define its complement $\overline{f} : A^n \to \{0,1\}$ by $\overline{f}(a_1, \ldots, a_n) = 1 - f(a_1, \ldots, a_n)$. If $P(X_1, \ldots, X_n)$ sign represents f, then $-P(X_1, \ldots, X_n)$ sign represents $\overline{f}$.

Lemma 2.1 *For $i \in [k]$, let $P_i(X_1, \ldots, X_n)$ be polynomials in $\mathbb{R}[X_1, \ldots, X_n]$ that sign represent f and let c_i be positive reals. Then*

$$Q(X_1, \ldots, X_n) = \sum_{i \in [k]} c_i P_i(X_1, \ldots, X_n)$$

sign represents f.

Proof: Let $(a_1, \ldots, a_n) \in A^n$. Suppose $f(a_1, \ldots, a_n) = 0$. Then since $P_i(a_1, \ldots, a_n) > 0$ for all i, $Q(a_1, \ldots, a_n) = \sum_{i=1}^{k} c_i P_i(a_1, \ldots, a_n) > 0$. Similarly if $f(a_1, \ldots, a_n) = 1, Q(a_1, \ldots, a_n) < 0$. $\blacksquare$

Similarly, one can show that if the polynomials $P_i(X_1, \ldots, X_n)$ weakly sign represent f, then $Q(X_1, \ldots, X_n)$ also weakly sign represents f.

Theorem 2.2 [PS76, BPR03] Descartes' Rule of Signs: *Let $P(X) = \sum_{i=0}^{n} c_i X^i$ be a real univariate polynomial. Let s denote the number of sign changes in the sequence $c_0, c_1, \ldots, c_n$. The number of positive roots of $P(X)$ counted with multiplicity is bounded by s.*

Let $d_0, \ldots, d_{k-1}$ be non-negative integers such that $d_0 < \cdots < d_{k-1}$. Let $a_0, \ldots, a_{k-1}$ be real numbers such that $a_0 < \cdots < a_{k-1}$. Define the corresponding generalized Vandermonde matrix as

$$V = \begin{pmatrix} a_0^{d_0} & a_0^{d_1} & \ldots & a_0^{d_{k-1}} \\ a_1^{d_0} & a_1^{d_1} & \ldots & a_1^{d_{k-1}} \\ \ldots & \ldots & \ldots & \ldots \\ a_{k-1}^{d_0} & a_{k-1}^{d_1} & \ldots & a_{k-1}^{d_{k-1}} \end{pmatrix}. \tag{1}$$

Our goal is to determine the signs of the entries in the inverse of such a matrix. For this we will use the following lemma:

Lemma 2.3 *[PS76] If $a_i > 0$ for all i, then $\det(V) > 0$.*

Proof: 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 univariate polynomial in $\mathbb{R}[X]$ defined as

$$
C(X) = \begin{vmatrix}
a_0^{d_0} & a_0^{d_1} & \dots & a_0^{d_{k-1}} \\
a_1^{d_0} & a_1^{d_1} & \dots & a_1^{d_{k-1}} \\
\dots & \dots & \dots & \dots \\
X^{d_0} & X^{d_1} & \dots & X^{d_{k-1}}
\end{vmatrix}
$$

Let $C(X) = c_{k-1}X^{d_{k-1}} + c_{k-2}X^{d_{k-2}} + \dots + c_0 X^{d_0}$. The sparsity of $C(X)$ is bounded by k, hence by Descartes' rule, it has at most $k - 1$ positive roots. But $a_0, \dots, a_{k-2}$ are roots of $C(X)$. Hence there are no other roots. Hence the sign at a_{k-1} (or at any point to the right of a_{k-2}) is the same as the sign at $+\infty$. This in turn is the sign of the leading coefficient c_{k-1} of $C(X)$, which is

$$
c_{k-1} = \begin{vmatrix}
a_0^{d_0} & \dots & a_0^{d_{k-2}} \\
\dots & \dots & \dots \\
a_{k-2}^{d_0} & \dots & a_{k-2}^{d_{k-2}}
\end{vmatrix}
$$

which is positive by the induction hypothesis. ∎

Let $V^{-1} = (v_{i,j}^{-1})$ denote inverse of V. Using Lemma 2.3 and the formula for inverse of a matrix, it is easy to see that for $0 \le i, j \le k - 1$, $\mathrm{sgn}(v_{ij}^{-1}) = (-1)^{i+j}$.

We will need to consider the case when $a_0 = 0$. If $d_0 > 0$, then clearly the first row is all 0s and the determinant vanishes. On the other hand, if $d_0 = 0$ we get the matrix

$$
W = \begin{pmatrix}
1 & 0 & 0 & 0 \\
1 & a_1^{d_1} & \dots & a_1^{d_{k-1}} \\
1 & \dots & \dots & \dots \\
1 & a_{k-1}^{d_1} & \dots & a_{k-1}^{d_{k-1}}
\end{pmatrix}. \tag{2}
$$

Lemma 2.4 *Let $W^{-1} = (w_{i,j}^{-1})$ denote the inverse of the matrix W. For $0 \le i, j \le k - 1$,*

$$
\mathrm{sgn}(w_{ij}^{-1}) = \begin{cases} 0 & \text{if } i = 0, j \ge 1, \\ (-1)^{i+j} & \text{otherwise.} \end{cases}
$$

Proof: The minors W_{ij} for $j = 0$ and $i \ge 1$ are 0 since their top row consists entirely of 0s. Hence the entries in W^{-1} for $i = 0$ and $j \ge 1$ are 0. For the other minors, we can apply Lemma 2.3 to show that they are positive. Also $\det(W) > 0$, hence by the formula for matrix inverses, $\mathrm{sgn}(w_{ij}^{-1}) = (-1)^{i+j}$. ∎

3 Lower Bounds

We first consider the case when $A = \{0, 1\}$. Assume that $P(X_1, \dots, X_n)$ sign represents parity over $\{0, 1\}^n$. If the variable X takes values in $\{0, 1\}$, then $X^k = X$ for $k \ge 2$. So we can use the relation $X_i^k = X_i$ for $k \ge 2$ to reduce the polynomial $P(X_1, \dots, X_n)$ to a multilinear polynomial. These substitutions can only decrease $\mathrm{sp}(P)$ and $\deg(P)$.

Lemma 3.1 *If $P(X_1, \ldots, X_n)$ sign represents parity over $\{0,1\}^n$,*

$$P(X_1, \ldots, X_n) = X_n Q_1(X_1, \ldots, X_{n-1}) + Q_0(X_1, \ldots, X_{n-1}) \tag{3}$$

where $Q_0(X_1, \ldots, X_{n-1})$ and $-Q_1(X_1, \ldots, X_{n-1})$ sign represent parity on $n-1$ variables.

Proof: Since $P(X_1, \ldots, X_n)$ is multilinear, by grouping together monomials which involve X_i, we can write

$$P(X_1, \ldots, X_n) = X_n Q_1(X_1, \ldots, X_{n-1}) + Q_0(X_1, \ldots, X_{n-1})$$

By substituting values for X_n, we get

$$P(X_1, \ldots, X_{n-1}, 0) = Q_0(X_1, \ldots, X_{n-1}),$$
$$P(X_1, \ldots, X_{n-1}, 1) = Q_1(X_1, \ldots, X_{n-1}) + Q_0(X_1, \ldots, X_{n-1}).$$

We now use the so-called self-reducibility of the parity function:

$$\mathrm{PAR}(a_1, \ldots, a_{n-1}, 0) = \mathrm{PAR}(a_1, \ldots, a_{n-1}),$$
$$\mathrm{PAR}(a_1, \ldots, a_{n-1}, 1) = \overline{\mathrm{PAR}}(a_1, \ldots, a_{n-1}).$$

From this it follows that $P(X_1, \ldots, X_{n-1}, 0) = Q_0(X_1, \ldots, X_{n-1})$ and $-P(X_1, \ldots, X_{n-1}, 1)$ sign represent parity on $n-1$ variables. Also, we have

$$-Q_1(X_1, \ldots, X_{n-1}) = P(X_1, \ldots, X_{n-1}, 0) - P(X_1, \ldots, X_{n-1}, 1).$$

Hence $Q_1(X_1, \ldots, X_{n-1})$ sign represents parity by Lemma 2.1. ∎

The polynomial $\prod_{i=1}^{n}(1 - 2X_i)$ sign represents parity over $\{0,1\}^n$. We will show that the degree and sparsity cannot be lower for any sign representation.

Theorem 3.2 *[MP68] If $P(X_1, \ldots, X_n)$ 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. For $n = 1$, let $P(X_1) = aX_1 + b$. $P(X_1)$ must satisfy the conditions

$$P(0) = b > 0, \quad P(1) = a + b < 0.$$

This implies $b > 0$ and $a < -b < 0$, hence $\mathrm{sp}(P) = 2$.

Assume inductively that the claim holds for $n - 1$ variables. Write $P(X_1, \ldots, X_n)$ as in Lemma 3.1. Observe that $\mathrm{sp}(P) = \mathrm{sp}(Q_0) + \mathrm{sp}(Q_1)$, since there cannot be cancellations between the monomials in $X_n Q_1(X_1, \ldots, X_{n-1})$ and $Q_0(X_1, \ldots, X_{n-1})$. By the induction hypothesis $\mathrm{sp}(Q_0) = \mathrm{sp}(Q_1) = 2^{n-1}$, hence $\mathrm{sp}(P) = 2^n$. ∎

We can strengthen the claim to show that the sign of the coefficient of every monomial is fixed. For $S \subset [n]$, we denote the coefficient corresponding to the monomial $\prod_{i \in S} X_i$ by c_S. Thus

$$P(X_1, \ldots, X_n) = \sum_{S \subseteq [n]} c_S \prod_{i \in S} X_i.$$

7

Theorem 3.3 *If $P(X_1, \ldots, X_n)$ sign represents parity on $\{0, 1\}^n$, then $\mathrm{sgn}(c_S) = (-1)^{|S|}$.*

Proof: The proof is by induction on n. The case $n = 1$ follows from the Proof of Theorem 3.2. Assume inductively that the claim holds for $n-1$ variables. Write $P(X_1, \ldots, X_n)$ as in Lemma 3.1. The monomials involving X_n come from $X_n Q_1(X_1, \ldots, X_{n-1})$ while those not involving X_n come from $Q_0(X_1, \ldots, X_{n-1})$. Now consider $S \subset [n]$ such that $n \notin S$. The coefficient c_S in $P(X_1, \ldots, X_n)$ is the same as the coefficient in $Q_0(X_1, \ldots, X_{n-1})$. Since $Q_0(X_1, \ldots, X_{n-1})$ represents parity on $n-1$ variables, hence $\mathrm{sgn}(c_S) = (-1)^{|S|}$ by the induction hypothesis. For $S \subset [n]$ such that $n \in S$, the coefficient c_S in $P(X_1, \ldots, X_n)$ is equal to the coefficient $c_{S \setminus \{n\}}$ in $Q_1(X_1, \ldots, X_{n-1})$. Since $Q_1(X_1, \ldots, X_{n-1})$ represents the complement of parity, $\mathrm{sgn}(c_S) = -(-1)^{|S|-1} = (-1)^{|S|}$ by induction. $\blacksquare$

One can similarly show a bound on the sum of the coefficient sizes for polynomials with integer coefficients. We omit the proof.

Next we generalize Theorem 3.3 to the case when $A = \{0, \ldots, m-1\}$ and the degree in each variable is at most $m-1$. To construct a polynomial sign representing parity satisfying these conditions, for $0 \leq j \leq m-2$, let $\alpha_j = j + \frac{1}{2}$. Let

$$P(X_1, \ldots, X_n) \quad = \quad \prod_{i=1}^{n} \prod_{j=0}^{m-2} (-1)^m (X_i - \alpha_j)$$

It can be verified that $P(X_1, \ldots, X_n)$ indeed sign represents parity on A^n and $\mathrm{sp}(P) = m^n$.

Define the univariate polynomial $M(X) \in \mathbb{R}[x]$ by $M(X) = \prod_{j=0}^{m-1}(X - j)$. Note that $M(X)$ is a monic polynomial of degree m which vanishes on the set $A = \{0, \ldots, m-1\}$. By Euclidean division, for any $d \geq m$, we can write

$$X^d = Q_d(X) M(X) + R_d(X)$$

where $\deg(R_d) \leq m - 1$.

The polynomials $M(X_i)$ for $i \in [n]$ vanish on the set A^n. Given any polynomial $P'(X_1, \ldots, X_n)$ which sign represents parity over A^n, we can reduce $P'(X_1, \ldots, X_n)$ modulo the polynomials $M(X_i)$ using

$$X_i^d \equiv R_d(X_i) \pmod{M(X_i)}$$

to obtain a polynomial $P(X_1, \ldots, X_n)$ such that

$$P(X_1, \ldots, X_n) \equiv P'(X_1, \ldots, X_n) \pmod{M(X_1), \ldots, M(X_n)}$$

The polynomial $P(X_1, \ldots, X_n)$ agrees with $P'(X_1, \ldots, X_n)$ over the set A^n, and $\deg_i(P) \leq m-1$. However we will show that such polynomials where $deg_i(P) \leq m - 1$ require sparsity m^n.

Lemma 3.4 *Assume that $P(X_1, \ldots, X_n)$ sign represents parity over $\{0, 1, \ldots, m-1\}^n$. If $\deg_n(P) \leq m-1$, then*

$$P(X_1, \ldots, X_n) = \sum_{i=0}^{m-1} X_n^i Q_i(X_1, \ldots, X_{n-1}) \tag{4}$$

where $(-1)^i Q_i(X_1, \ldots, X_{n-1})$ represents parity on $n-1$ variables.

Proof: Since $\deg_n(P) \leq m - 1$, grouping monomials by powers of X_n,

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

8

Let

$$
W = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1^{m-1} \\ 1 & \dots & \dots & \dots \\ 1 & m-1 & \dots & (m-1)^{m-1} \end{pmatrix}.
$$

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

$$
W \cdot \begin{pmatrix} Q_0(X_1,\dots,X_{n-1}) \\ Q_1(X_1,\dots,X_{n-1}) \\ \dots \\ Q_{m-1}(X_1,\dots,X_{n-1}) \end{pmatrix} = \begin{pmatrix} P(X_1,\dots,X_{n-1},0) \\ P(X_1,\dots,X_{n-1},1) \\ \dots \\ P(X_1,\dots,X_{n-1},m-1) \end{pmatrix}
$$

$$
\Rightarrow \begin{pmatrix} Q_0(X_1,\dots,X_{n-1}) \\ Q_1(X_1,\dots,X_{n-1}) \\ \dots \\ Q_{m-1}(X_1,\dots,X_{n-1}) \end{pmatrix} = W^{-1} \cdot \begin{pmatrix} P(X_1,\dots,X_{n-1},0) \\ P(X_1,\dots,X_{n-1},1) \\ \dots \\ P(X_1,\dots,X_{n-1},m-1) \end{pmatrix}
$$

We now expand the LHS. Consider the top row of W^{-1}, which is indexed by $i = 0$. By Lemma 2.4, the first entry is some number $w_{00}^{-1} > 0$, and the other entries are 0. This implies

$$
Q_0(X_1,\dots,X_{n-1}) = w_{00}^{-1} P(X_1,\dots,X_{n-1},0),
$$

so $Q_0(X_1,\dots,X_{n-1})$ sign represents parity on $n-1$ variables.

For $i \geq 1$, $w_{ij}^{-1} = (-1)^{i+j}|w_{ij}^{-1}|$. Hence

$$
Q_i(X_1,\dots,X_{n-1}) = \sum_{j=0}^{m-1} (-1)^{i+j}|w_{ij}^{-1}|P(X_1,\dots,X_{n-1},j)
$$

$$
\Rightarrow (-1)^i Q_i(X_1,\dots,X_{n-1}) = \sum_{j=0}^{m-1} (-1)^j |w_{ij}^{-1}|P(X_1,\dots,X_{n-1},j). \tag{5}
$$

We now use the self-reducibility of the parity function:

$$
\mathrm{PAR}(a_1,\dots,a_{n-1},j) = \begin{cases} \mathrm{PAR}(a_1,\dots,a_{n-1}) & \text{if } j \equiv 0 \pmod 2 \\ \overline{\mathrm{PAR}}(a_1,\dots,a_{n-1}) & \text{if } j \equiv 1 \pmod 2 \end{cases}
$$

Hence the polynomial $(-1)^j|w_{ij}|P(X_1,\dots,X_{n-1},j)$ sign represents parity on $n-1$ variables for all j. Hence by Lemma 2.1, $(-1)^i Q_i(X_1,\dots,X_n)$ also represents parity on $n-1$ variables for every i. ∎

Theorem 3.5 *Let $P(X_1,\dots,X_n)$ be a polynomial that sign represents parity over $\{0,1,\dots,m-1\}^n$, with $\deg(X_i) \leq m-1$ for all i. Then $P(X_1,\dots,X_n)$ 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 proof is by induction. 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 $m - 1$,

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

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

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

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

The inductive case proceeds using Lemma 3.4 exactly as in Theorem 3.2. We skip the proof. ∎

Corollary 3.6 *If $P'(X_1, \ldots, X_n)$ sign represents parity over $\{0, 1, \ldots, m-1\}^n$, then $\deg(P') \geq n(m-1)$.*

Proof: We quotient out the $P'(X_1, \ldots, X_n)$ by the polynomials $M(X_i), \ldots, M(X_n)$, to get $P(X_1, \ldots, X_n)$ where $\deg_i(P) \leq m-1$. Note that this only reduces the total degree, hence $\deg(P') \geq \deg(P)$. By Theorem 3.5, $\mathrm{sp}(P) \geq m^n$. This implies that for every tuple $(d_1, \ldots, d_n)$ where $d_i \leq m-1$, the monomial $\prod_i X_i^{d_i}$ occurs with non-zero coefficients. Thus the monomial $\prod_{i \in [n]} X_i^{m-1}$ is in the support, which implies that $\deg(P) \geq n(m-1)$. ∎

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

Corollary 3.7 *If $P(X_1, \ldots, X_n)$ is a multilinear polynomial that sign represents parity over $\{1, 2\}^n$, it has sparsity 2^n.*

A natural question is what happens to the sparsity if we allow polynomials of higher degree. It might be that there are polynomials of high degree and low sparsity and quotienting by the $M(X_i)$s causes the sparsity to increase. We will address this question in Section 4. We next turn our attention to weak representations.

3.1 Weak Representations

We first consider weak representations for parity with low degree. Over $\{0, 1\}^n$, the polynomial $P(X_1, \ldots, X_n) = (-1)^n \prod_i X_i$ gives a weak representation with sparsity 1, and in fact this is optimal with regard to degree too.

Lemma 3.8 *[ABFR94] Any polynomial that weakly sign represents parity over $\{0, 1\}^n$ has degree n.*

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

Lemma 3.9 *If $P(X_1, \ldots, X_n)$ weakly sign represents parity over $A = \{0, 1, \ldots, m-1\}^n$, and if $\deg_n(P) \leq m-1$, then*

$$P(X_1, \ldots, X_n) = \sum_{i=0}^{m-1} X_n^i Q_i(X_1, \ldots, X_{n-1}) \tag{6}$$

where for $i \geq 1$, the polynomial $(-1)^i Q_i(X_1, \ldots, X_{n-1})$ weakly represents parity on $n-1$ variables.

Proof: The proof is similar to that of Lemma 3.4, the difference being that we need to show that the polynomials $Q_i(X_1, \ldots, X_{n-1})$ do not vanish over the set A^{n-1}. By substituting values 0 through $m-1$ for X_n and inverting the Vandermonde matrix, we get

$$
\begin{pmatrix}
Q_0(X_1, \ldots, X_{n-1}) \\
Q_1(X_1, \ldots, X_{n-1}) \\
\ldots \\
Q_{m-1}(X_1, \ldots, X_{n-1})
\end{pmatrix}
= W^{-1} \cdot
\begin{pmatrix}
P(X_1, \ldots, X_{n-1}, 0) \\
P(X_1, \ldots, X_{n-1}, 1) \\
\ldots \\
P(X_1, \ldots, X_{n-1}, m-1)
\end{pmatrix}.
$$

For $i \geq 1$, by Equation (5), we have $w_{ij}^{-1} \neq 0$ and

$$
(-1)^i Q_i(X_1, \ldots, X_{n-1}) = \sum_{j=0}^{m-1} (-1)^j |w_{ij}^{-1}| P(X_1, \ldots, X_{n-1}, j).
$$

Since $P(X_1, \ldots, X_n)$ weakly represents parity on A^n, the polynomial $(-1)^j P(X_1, \ldots, X_{n-1}, j)$ either weakly represents parity, or it vanishes over A^{n-1}. Since $P(X_1, \ldots, X_n)$ is a weak representation of parity, it does not vanish on A^n. Hence there is a point $(a_1, \ldots, a_n) \in A^n$ so that $P(a_1, \ldots, a_n) \neq 0$. Hence the polynomial $P(X_1, \ldots, X_{n-1}, a_n)$ does not vanish over A^{n-1}. Hence by Lemma 2.1, the polynomials $(-1)^i Q_i(X_1, \ldots, X_{n-1})$ weakly sign represent parity on A^{n-1} for $i \geq 1$. ∎

The condition $i \geq 1$ in the statement of Lemma 3.9 is in fact necessary: take the polynomial $P(X_1, \ldots, X_n) = (-1)^n \prod_i X_i$ that weakly sign represents parity on $\{0,1\}^n$. In this case, $Q_0(X_1, \ldots, X_{n-1}) = 0$, so it does not represent parity even weakly.

We use Lemma 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 3.8.

Lemma 3.10 *Any univariate polynomial $P(X) \in \mathbb{R}[X]$ that weakly sign represents parity over $\{0, \ldots, m-1\}$ must have $m-2$ roots in the interval $(0, m-1]$.*

Proof: We first show that $\deg(P) \geq m-1$. Assume that this is not so. Let $X_1, \ldots, X_{m-1}$ be variables that take values in $\{0,1\}$. Then the polynomial

$$
Q(X_1, \ldots, X_{m-1}) = P\left(\sum_{i=1}^{m-1} X_i\right)
$$

weakly sign represents parity on $\{0,1\}^{m-1}$ and $\deg(Q) < m-1$, contradicting Lemma 3.8.

Now consider the factorization of $P(X)$ over the reals. Assume that this contains an irreducible polynomial $D(X) \in R[X]$ with $\deg(D) = 2$. The polynomial $D(X)$ does not have real roots its sign stays unchanged in the interval $[0, \ldots, m-1]$. Hence we can replace $D(X)$ by the constant $\operatorname{sgn}(D(0))$, and get a sign representation of lower degree. Similarly, consider a linear factor of the form $(X - \alpha)$ where $\alpha \notin [0, m-1]$. Such linear factors can also be replaced by their signs at 0. Further we may assume that there is a root of multiplicity at most 1 at 0. If not, we can write $P(X) = X^k Q(X)$ for $k \geq 2$. The polynomial $X Q(X)$ has the same sign at each point in $[0, m-1]$ and only smaller degree. We are left with a polynomial of the form

$$
P'(X) = \prod (X - \alpha_i) \qquad \alpha_i \in [0, m-1]
$$

which weakly represents parity over $\{0, \ldots, m-1\}$, hence $\deg(P') \geq m-1$. 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 3.11 *Let $P(X_1, \ldots, X_n)$ be a polynomial that weakly represents parity over $\{0, \ldots, m-1\}^n$. If $\deg_i(P) \leq m-1$ for every $i \in [n]$, then $\mathrm{sp}(P) \geq m^n$.*

Proof: The proof is by induction on n. When $n = 1$, by Lemma 3.10 $P(X_1)$ has $m-2$ roots in $(0, m-1]$, hence by Descartes' rule, $\mathrm{sp}(P) \geq m-1$.

For the inductive case, we use Lemma 3.9. From Equation (6) it follows that

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

By Lemma 3.9, the polynomial $(-1)^i Q_i(X_1, \ldots, X_{n-1})$ weakly represents parity on A^{n-1}. Hence by induction, $\mathrm{sp}(Q_i) \geq (m-1)^{n-1}$, and so $\mathrm{sp}(P) \geq (m-1)^n$. ∎

This bound is in fact tight.

Lemma 3.12 *There is a polynomial $P(X_1, \ldots, X_n)$ that weakly represents parity over $\{0, \ldots, m-1\}^n$ where $\deg_i(P) \leq m-1$ for every $i \in [n]$, and $\mathrm{sp}(P) = (m-1)^n$.*

Proof: Take $Q(X_1, \ldots, X_n)$ to be a polynomial that sign represents parity on $\{1, \ldots, m-1\}^n$ satisfying $\mathrm{sp}(Q) = (m-1)^n$, and $\deg_i(Q) \leq m-2$. We claim that the polynomial

$$P(X_1, \ldots, X_n) = Q(X_1, \ldots, X_n) \cdot \prod_{i=1}^{n} X_i$$

weakly represents parity over $\{0, \ldots, m-1\}^n$. This is because, for $a_1, \ldots, a_n \in \{0, \ldots, m-1\}^n$,

$$\mathrm{sgn}(P(a_1, \ldots, a_n)) = \begin{cases} 0 & \text{if } a_i = 0 \text{ for some } i, \\ \mathrm{sgn}(Q(a_1, \ldots, a_n)) & \text{otherwise.} \end{cases}$$

Further, since $\deg_i(Q) \leq m-2$, $\deg_i(P) \leq m-1$ for all $i \in [n]$. Also, $\mathrm{sp}(P) = \mathrm{sp}(Q) = (m-1)^n$. ∎

The proof of Lemma 3.12 crucially uses the fact that $0 \in A$. Indeed we will show that if $A = \{1, \ldots, m\}$, then weak representations of parity require sparsity m^n.

Corollary 3.13 *Let $P(X_1, \ldots, X_n)$ be a polynomial that weakly represents parity over $\{1, \ldots, m\}^n$. If $\deg_i(P) \leq m-1$ for every $i \in [n]$, then $\mathrm{sp}(P) \geq m^n$.*

Proof: Let $P(X_1, \ldots, X_n)$ be as above. The polynomial

$$P'(X_1, \ldots, X_n) = P(X_1, \ldots, X_n) \cdot \prod_{i=1}^{n} X_i$$

weakly represents parity over $\{0, \ldots, m\}^n$. Further, since $\deg_i(P) \leq m-1$, $\deg_i(P') \leq m$ for all $i \in [n]$. Hence we can apply Theorem 3.11, which implies

$$\mathrm{sp}(P) = \mathrm{sp}(P') \geq m^n.$$

∎

4 Upper Bounds

Does the lower bound of m^n in Theorem 3.5 hold for all polynomials? Or are there polynomials with higher degree but lower sparsity? We show that such a tradeoff is indeed possible.

Theorem 4.1 *There exists a polynomial $P(X_1, \ldots, X_n)$ that sign represents parity over $\{1, 2\}^n$ with $\deg(P) = n^2$ and $\mathrm{sp}(P) = n + 1$.*

Proof: Define $w : \{1, 2\}^n \rightarrow \{2^k\}_{k=0}^n$ by $w(a_1, \ldots, a_n) = \prod_i a_i$. If $w(a_1, \ldots, a_n) = 2^k$, then $a_i = 2$ for exactly k co-ordinates $i \in [n]$ hence $\mathrm{PAR}(a_1, \ldots, a_n) \equiv n - k \pmod 2$.
  Choose points $\alpha_j \in (2^{j-1}, 2^j)$ for $1 \le j \le n$ and let

$$P(X_1, \ldots, X_n) = \prod_{j=1}^n \left(\prod_{i=1}^n X_i - \alpha_j \right),$$

so that $\deg(P) = n^2$ and $\mathrm{sp}(P) = n + 1$. We claim that $P(X_1, \ldots, X_n)$ sign represents parity on $\{1, 2\}^n$. Note that

$$P(a_1, \ldots, a_n) \;=\; \prod_{j=1}^n (w(a_1, \ldots, a_n) - \alpha_j).$$

If $w(a_1, \ldots, a_n) = 2^k$, then

$$P(a_1, \ldots, a_n) = \prod_{j=1}^n (2^k - \alpha_j)$$

$$\Rightarrow \; \mathrm{sgn}(P(a_1, \ldots, a_n)) = (-1)^{n-k}.$$

Thus the polynomial $P(X_1, \ldots, X_n)$ sign represents parity. Its sparsity is $n + 1$ and its degree is n^2. ∎

In contrast, Corollary 3.7 shows a lower bound of 2^n on the sparsity for sign representations by multilinear polynomials.
  We can extend Theorem 4.1 to show that for any set A of non-negative integers of size m, there are polynomials that weakly sign represent parity whose sparsity is less m^n, but which have high degree.

Theorem 4.2 *For any set A of non-negative integers of cardinality m, parity can be weakly sign represented over A^n by a polynomial $P(X_1, \ldots, X_n)$ that has sparsity at most $\binom{n+m-1}{n}$ and degree $n \cdot \binom{n+m-1}{n}$.*

Proof: Define the function $w : A^n \rightarrow \mathbb{Z}$ by $w(a_1, \ldots, a_n) = \prod_i a_i$. This maps A^n to a set S of size at most $\binom{n+m-1}{n}$ in $\mathbb{Z}$. Let a denote the largest integer in the set A. Note that $w(a, \ldots, a) = a^n$ is the largest integer in S. Further $(a, \ldots, a)$ is the unique point in A^n that is mapped to a^n by w. We claim that the polynomial

$$P(X_1, \ldots, X_n) = (-1)^{na} \cdot \prod_{\alpha \in S \setminus \{a^n\}} \left(\prod_{i \in [n]} X_i - \alpha \right)$$

weakly represents parity on A^n. To prove this, note that

$$P(a_1, \ldots, a_n) = (-1)^{na} \cdot \prod_{\alpha \in S \setminus \{a^n\}} (w(a_1, \ldots, a_n) - \alpha).$$

13

Hence the polynomial vanishes for every point in $A^n \setminus \{(a, \ldots, a)\}$. Further

$$P(a, \ldots, a) = (-1)^{na} \cdot \prod_{\alpha \in S \setminus \{a^n\}} (a^n - \alpha)$$

$$\Rightarrow \operatorname{sgn}(P(a, \ldots, a)) = (-1)^{na}.$$

Thus $P(X_1, \ldots, X_n)$ weakly represents parity on A^n. ∎

In general $\operatorname{sp}(P) = |S|$ and $|S|$ depends on the set A. For some sets A, $\operatorname{sp}(P)$ can be significantly smaller that the bound stated $\binom{n+m-1}{n}$ stated above. In the case when $A = \{0, \ldots, m-1\}$ and n is a fixed constant, one can in fact show that $\operatorname{sp}(P) = o(m)^n$. This is a consequence of Erdös' 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 $o(m)^n$ [BPV98].

5 Lower Bounds without Degree Restrictions

We will now show a lower bound which holds for all polynomials strongly representing parity on $\{1, \ldots, m\}^n$ 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 $P(X_1, \ldots, X_n)$ be a polynomial which sign represents parity over $\{1, \ldots, m\}^n$. Then $\operatorname{sp}(P) \geq n(m-1) + 1$.*

Proof: The proof is by induction on n. When $n = 1$, the claim follows by Descartes' rule. Assume it is true for $n-1$. Recall that $\operatorname{sp}_n(P)$ is the number of distinct powers of X_n that occur in monomials in the support of P. If we set all the other variables to 1, the univariate polynomial $Q(X_n) = (-1)^{n-1} P(1, \ldots, 1, X_n)$ sign represents parity on $\{1, \ldots, m\}$, hence it must have sparsity at least m. Hence $\operatorname{sp}_n(P) \geq \operatorname{sp}(Q) \geq m$. If $k > n(m-1) + 1$, there is nothing to prove. Hence we may assume $m \leq k \leq n(m-1)$.

Grouping monomials in $P(X_1, \ldots, X_n)$ by the power of X_n they contain, we can write

$$P(X_1, \ldots, X_n) = \sum_{i=1}^{k} X_n^{d_i} Q_i(X_1, \ldots, X_{n-1}).$$

By substituting values 1 through m for X_n, we get

$$\begin{pmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 \\ 2^{d_1} & \cdots & 2^{d_m} & 2^{d_{m+1}} & \cdots & 2^{d_k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ m^{d_1} & \cdots & m^{d_m} & m^{d_{m+1}} & \cdots & m^{d_k} \end{pmatrix} \begin{pmatrix} Q_1(X_1, \ldots, X_{n-1}) \\ Q_2(X_1, \ldots, X_{n-1}) \\ \ldots \\ Q_k(X_1, \ldots, X_{n-1}) \end{pmatrix} = \begin{pmatrix} P(X_1, \ldots, X_{n-1}, 1) \\ P(X_1, \ldots, X_{n-1}, 2) \\ \ldots \\ P(X_1, \ldots, X_{n-1}, m) \end{pmatrix} \quad (7)$$

We denote the $m \times k$ matrix by A. While we cannot prove that each $Q_i(X_1, \ldots, X_{n-1})$ represents parity (or its complement), we will show that appropriate linear combinations of the $Q_i(X_1, \ldots, X_{n-1})$ sign represent parity. We pre-multiply each side of Equation (7) by U, which is the inverse of the $m \times m$ generalized Vandermonde matrix consisting of the first m columns of A.

$$\begin{pmatrix} 1 & \cdots & 0 & b_{1,m+1} & \cdots & b_{1,k} \\ 0 & \cdots & 0 & b_{2,m+1} & \cdots & b_{2,k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 1 & b_{m,m+1} & \cdots & b_{m,k} \end{pmatrix} \begin{pmatrix} Q_1(X_1, \ldots, X_{n-1}) \\ Q_2(X_1, \ldots, X_{n-1}) \\ \ldots \\ Q_k(X_1, \ldots, X_{n-1}) \end{pmatrix} = U \cdot \begin{pmatrix} P(X_1, \ldots, X_{n-1}, 1) \\ P(X_1, \ldots, X_{n-1}, 2) \\ \ldots \\ P(X_1, \ldots, X_{n-1}, m) \end{pmatrix}$$

14

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

$$R_i(X_1, \ldots, X_{n-1}) \;=\; Q_i(X_1, \ldots, X_{n-1}) + \sum_{j=m+1}^{k} b_{ij} Q_j(X_1, \ldots, X_{n-1})$$

sign represent parity or its complement on $n-1$ variables. Hence by applying the induction hypothesis,

$$\mathrm{sp}(R_i) \geq (n-1)(m-1) + 1.$$

But we also have

$$\mathrm{sp}(R_i) \leq \mathrm{sp}(Q_i) + \sum_{j=m+1}^{k} \mathrm{sp}(Q_j).$$

Hence we get

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

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

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

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

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

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

But $\mathrm{sp}(P) = \sum_{i=1}^{k} \mathrm{sp}(Q_i)$, 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 shows that the construction of Theorem 4.1 is optimal with regard to sparsity. We can now prove tight lower bounds on polynomials sign representing parity on A^n for any set A of size 2. Let $A = \{a, b\}$ where $0 \leq a < b$.

- If $a = 0$, then any polynomial which sign represents parity has sparsity at least 2^n.

- If $a > 0$, then any polynomial which sign represents parity has sparsity at least $n + 1$.

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. It would be interesting to close this gap.

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.

Definition 6.1 *A Threshold function $f : \{0,1\}^n \to \{0,1\}$ is defined as*

$$f(a_1, \ldots, a_n) = \begin{cases} 1 & \text{if } w_0 + \sum_{i=1}^n w_i a_i < 0 \\ 0 & \text{if } w_0 + \sum_{i=1}^n w_i a_i > 0 \end{cases}$$

where $w_0, \ldots, w_n \in \mathbb{R}$. The coefficients w_i are called the weights of the Threshold function. A gate computing a threshold function is called a Threshold gate and is denoted by THR.

In our definition, we assume that $w_0, \ldots, w_n$ are such that $w_0 + \sum_{i=1}^n w_i a_i \neq 0$ for $a_1, \ldots, a_n \in \{0,1\}^n$.

Definition 6.2 *A Threshold of Ands circuit (denoted* THR $\circ$ AND*) is a depth-two circuit with Boolean inputs $X_1, \ldots, X_n$. The top level of the circuit consists of a single* THR *gate, while the bottom level consists of* AND *gates. The inputs to the* AND *gates are the inputs $X_1, \ldots, X_n$ and their complements $\neg X_1, \ldots, \neg X_n$. The size of a* THR $\circ$ AND *circuit is defined to be the number of* AND *gates in the bottom level of the circuit. The minimum size of a* THR $\circ$ AND *circuit needed to compute a Boolean function f is denoted by $S(f)$.*

These circuits are well-studied (see [Gol97] and the references therein). By De Morgan's law, such circuits can simulate OR gates at the bottom level. We will show that $S(f)$ corresponds to minimum sparsity required to sign represent f over a certain *basis*. Thus proving circuit lower bounds is equivalent to proving bounds on the sparsity of sign representations.

To begin with, assume that the inputs to the AND gates were only the variables X_i, not their complements. 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. Such a circuit computing parity corresponds to a sign representation of parity in the standard monomial basis. The number of AND gates is exactly the number of non-constant monomials required. By Theorem 3.2, this is $2^n - 1$.

In a general THR $\circ$ AND circuit, and AND gate computes the function $\wedge_{i \in I} X_i \wedge_{j \in J} \neg X_j$. We can assume that $I \cap J$ is empty, else the AND gate computes the function 0. Thus the AND gate computes the polynomial

$$B(X_1, \ldots, X_n) = \prod_{i \in I} X_i \prod_{j \in J} (1 - X_j), \quad I \cap J = \phi \tag{8}$$

Let $\mathcal{B}_n$ denote the set of all such polynomials taken over all choices of the sets I and J. It is easy to show that $|\mathcal{B}_n| = 3^n$. Since $\mathcal{B}_n$ contains the standard monomial basis, so it spans the $\mathbb{R}$-vector space of multilinear polynomials in $\mathbb{R}[X_1, \ldots, X_n]$. Since this vector space has dimension 2^n, there are many ways to write a multilinear polynomial $P(X_1, \ldots, X_n)$ as a linear combination of polynomials in $\mathcal{B}_n$. We will define $\text{sp}_{\mathcal{B}}(P)$ as the minimum possible sparsity over all such linear combinations. Formally:

Definition 6.3 *Let $P(X_1, \ldots, X_n)$ be a multilinear polynomial in $\mathbb{R}[X_1, \ldots, X_n]$. We define the sparsity of $P(X_1, \ldots, X_n)$ over $\mathcal{B}_n$ as*

$$\text{sp}_{\mathcal{B}}(P) = \{\min k \mid P(X_1, \ldots, X_n) = \sum_{i=1}^k c_i B_i(X_1, \ldots, X_n), \quad B_i(X_1, \ldots, X_n) \in \mathcal{B}_n\}.$$

The following lemma relating circuit-size for THR ∘ AND circuits computing f and sparsity over $\mathcal{B}$ of polynomials that sign represent f follows from the preceding discussion.

Lemma 6.4 *For any Boolean function* $f : \{0,1\}^n \to \{0,1\}$,

$$S(f) = \min \operatorname{sp}_{\mathcal{B}}(P)$$

over all polynomials $P(X_1, \ldots, X_n)$ *that sign represent* f.

Theorem 6.5 *Every* THR ∘ AND *circuit computing the parity function on* $\{0,1\}^n$ *has size at least* $\left(\frac{3}{2}\right)^n$.

Proof: We will show that if $P(X_1, \ldots, X_n)$ sign represents parity, then $\operatorname{sp}_{\mathcal{B}}(P) \geq (\frac{3}{2})^n$.

The proof is by induction on n. For $n = 1$, $\mathcal{B}_1 = \{X_1, 1 - X_1, 1\}$. Since none of these polynomials or their multiples sign represents parity on 1 variable, $\operatorname{sp}_{\mathcal{B}}(P) \geq 2$.

Now assume the claim holds for $n - 1$. Let $P(X_1, \ldots, X_n)$ sign represent parity. Consider the sparsest representation of P over $\mathcal{B}$.

$$P(X_1, \ldots, X_n) = \sum_{I \cap J = \phi} c_{I,J} \prod_{i \in I} X_i \prod_{j \in J} (1 - X_j)$$

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

$$P(X_1, \ldots, X_n) = X_n A(X_1, \ldots, X_{n-1}) + (1 - X_n) B(X_1, \ldots, X_{n-1}) + C(X_1, \ldots, X_{n-1}). \qquad (9)$$

The best (sparsest) way to write $P(X_1, \ldots, X_n)$ as a linear combination of polynomials in $\mathcal{B}_n$ is to use the best (sparsest) expression for each of $A(X_1, \ldots, X_{n-1}), B(X_1, \ldots, X_{n-1})$ and $C(X_1, \ldots, X_{n-1})$ as linear combinations of polynomials in $\mathcal{B}_{n-1}$. Hence

$$\operatorname{sp}_{\mathcal{B}}(P) = \operatorname{sp}_{\mathcal{B}}(A) + \operatorname{sp}_{\mathcal{B}}(B) + \operatorname{sp}_{\mathcal{B}}(C).$$

Substituting for X_n in Equation (9),

$$
\begin{aligned}
P(X_1, \ldots, X_{n-1}, 0) &= B(X_1, \ldots, X_{n-1}) + C(X_1, \ldots, X_{n-1}), \\
P(X_1, \ldots, X_{n-1}, 1) &= A(X_1, \ldots, X_{n-1}) + C(X_1, \ldots, X_{n-1}), \\
P(X_1, \ldots, X_{n-1}, 0) - P(X_1, \ldots, X_{n-1}, 1) &= B(X_1, \ldots, X_{n-1}) - A(X_1, \ldots, X_{n-1})
\end{aligned}
$$

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

$$
\begin{aligned}
\operatorname{sp}_{\mathcal{B}}(B) + \operatorname{sp}_{\mathcal{B}}(A) &\geq \operatorname{sp}_{\mathcal{B}}(B - A) > (3/2)^{n-1}, \\
\operatorname{sp}_{\mathcal{B}}(B) + \operatorname{sp}_{\mathcal{B}}(C) &\geq \operatorname{sp}_{\mathcal{B}}(B + C) > (3/2)^{n-1}, \\
\operatorname{sp}_{\mathcal{B}}(A) + \operatorname{sp}_{\mathcal{B}}(C) &\geq \operatorname{sp}_{\mathcal{B}}(A + C) > (3/2)^{n-1}.
\end{aligned}
$$

Adding these equations, we get

$$\operatorname{sp}_{\mathcal{B}}(P) = \operatorname{sp}_{\mathcal{B}}(A) + \operatorname{sp}_{\mathcal{B}}(B) + \operatorname{sp}_{\mathcal{B}}(C) > (3/2)^n$$

which completes the proof of the Theorem. ∎

Proposition 6.6 *There is a* THR ∘ AND *circuit computing the parity function on* $\{0,1\}^n$ *of size* $O(5^{\frac{n}{3}})$.

Proof: The polynomial

$$Q(X_1, X_2, X_3) = X_1 X_2 X_3 + X_1(1 - X_2)(1 - X_3) + X_2(1 - X_3)(1 - X_1) + X_3(1 - X_1)(1 - X_2)$$

exactly represents parity on $\{0,1\}^3$. Hence, the polynomial $P(X_1, X_2, X_3) = 1 - 2Q(X_1, X_2, X_3)$ sign represents parity on 3 variables and has sparsity 5. Hence the polynomial

$$R(X) = \prod_{i=1}^{\frac{n}{3}} P(X_{3i-2}, X_{3i-1}, X_{3i})$$

sign represents parity on n variables, and $\mathrm{sp}_{\mathcal{B}}(R) = 5^{\frac{n}{3}}$. ∎

Definition 6.7 *The inner product function* IP : $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *is defined as*

$$\mathrm{IP}(a_1, \ldots, a_n, b_1, \ldots, b_n) = \sum_i a_i b_i \pmod 2.$$

Theorem 6.8 *Every* THR ∘ AND *circuit computing the inner product function on* $\{0,1\}^n \times \{0,1\}^n$ *has size at least* 2^n.

Proof: The proof is by induction on n. The base case is trivial. Assume the claim holds for $n - 1$.

Let $P(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ sign represent IP on $\{0,1\}^n \times \{0,1\}^n$. Consider the sparsest way to write $P(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ over $\mathcal{B}_{2n}$. Grouping the monomials according to X_n, Y_n, where each A_i is a polynomial in $X_1, Y_1, \ldots, X_{n-1}, Y_{n-1}$,

$$P(X_1, \ldots, X_n, Y_1, \ldots, Y_n) = X_n Y_n A_1 + X_n(1 - Y_n)A_2 + (1 - X_n)Y_n A_3 + (1 - X_n)(1 - Y_n)A_4$$
$$+ X_n A_5 + Y_n A_6 + (1 - X_n)A_7 + (1 - Y_n)A_8 + A_9$$

Now substituting for X_n, Y_n and writing $P(0,1)$ for $P(X_1, \ldots, X_{n-1}, 0, Y_1, \ldots, Y_{n-1}, 1)$ and so on,

$$P(0,1) = A_3 + A_6 + A_7 + A_9, \tag{10}$$
$$P(1,0) = A_2 + A_5 + A_8 + A_9, \tag{11}$$
$$P(1,1) = A_1 + A_5 + A_6 + A_9. \tag{12}$$

Subtracting Equation (12) from (10) and (11) respectively,

$$P(0,1) - P(1,1) = -A_1 + A_3 - A_5 + A_7, \tag{13}$$
$$P(1,0) - P(1,1) = -A_1 + A_2 - A_6 + A_8. \tag{14}$$

In the above equations the polynomials on the LHS represents IP or its complement on $\{0,1\}^{n-1} \times \{0,1\}^{n-1}$, so each has sparsity at least 2^{n-1} by the induction hypothesis. Applying this observation to equations 10, 11, 13 and 14,

$$\mathrm{sp}_{\mathcal{B}}(A_3) + \mathrm{sp}_{\mathcal{B}}(A_6) + \mathrm{sp}_{\mathcal{B}}(A_7) + \mathrm{sp}_{\mathcal{B}}(A_9) \geq 2^{n-1},$$
$$\mathrm{sp}_{\mathcal{B}}(A_2) + \mathrm{sp}_{\mathcal{B}}(A_5) + \mathrm{sp}_{\mathcal{B}}(A_8) + \mathrm{sp}_{\mathcal{B}}(A_9) \geq 2^{n-1},$$
$$\mathrm{sp}_{\mathcal{B}}(A_1) + \mathrm{sp}_{\mathcal{B}}(A_3) + \mathrm{sp}_{\mathcal{B}}(A_5) + \mathrm{sp}_{\mathcal{B}}(A_7) \geq 2^{n-1},$$
$$\mathrm{sp}_{\mathcal{B}}(A_1) + \mathrm{sp}_{\mathcal{B}}(A_2) + \mathrm{sp}_{\mathcal{B}}(A_6) + \mathrm{sp}_{\mathcal{B}}(A_8) \geq 2^{n-1}.$$

Adding these equations, we get

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

Hence

$$\mathrm{sp}_{\mathcal{B}}(P) = \sum_{i=1}^{9} \mathrm{sp}_{\mathcal{B}}(A_i) \geq 2^n$$

which completes the proof. ∎

Proposition 6.9 *There is a* THR ∘ AND *circuit computing the inner product function on* $\{0,1\}^n \times \{0,1\}^n$ *of size* 2^n.

Proof: The polynomial

$$Q(X_1, X_2, Y_1, Y_2) = X_1 Y_1 + X_2 Y_2 - 2X_1 Y_1 X_2 Y_2$$

exactly represents IP on $\{0,1\}^2 \times \{0,1\}^2$. The polynomial

$$P(X_1, X_2, Y_1, Y_2) = 1 - 2Q(X_1, X_2 Y_1, Y_2)$$

sign represents IP and on $\{0,1\}^2 \times \{0,1\}^2$ and $\mathrm{sp}_{\mathcal{B}}(P) = 4$. For arbitrary n, the polynomial

$$R(X_1, \ldots, X_n, Y_1, \ldots, Y_n) = \prod_{i=1}^{\frac{n}{2}} P(X_{2i-1}, X_{2i}, Y_{2i-1}, Y_{2i})$$

sign represents IP on $\{0,1\}^n \times \{0,1\}^n$ and $\mathrm{sp}_{\mathcal{B}}(R) = 4^{\frac{n}{2}} = 2^n$. ∎

Acknowledgments

References

[ABFR94] James Aspnes, Richard Beigel, Merrick Furst, and Steven Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994.

[AM05] Kazuyuki Amano and Akira Maruoka. On the complexity of depth-2 circuits with threshold gates. In *Mathematical Foundation of Computer Science (MFCS'05)*, pages 107–118, 2005.

[Bas99] Saugata Basu. On bounding the Betti numbers and computing the Euler characteristics of semi-algebraic sets. *Discrete and Computational Geometry*, 22:1–18, 1999.

[BBR94]   David A. Barrington, Richard Beigel, and Steven Rudich. Representing Boolean functions as polynomials modulo composite numbers. *Computational Complexity*, 4:367–382, 1994.

[Bei93]   Richard Beigel. The polynomial method in circuit complexity. In *Structures in Complexity Theory: 8th Annual Conference*, pages 82–95, 1993.

[Bei94]   Richard Beigel. Perceptrons, PP, and the polynomial hierarchy. *Computational Complexity*, 4:339–349, 1994.

[BPR03]   Saugata Basu, Richard Pollack, and Marie-Françoise Roy. *Algorithms in Real Algebraic Geometry*. Springer, Algorithms and Computation in Mathematics , Vol. 10, 2003.

[BPV98]   László Babai, Carl Pomerance, and Peter Vertesi. The mathematics of Paul Erdös. *Notices of the AMS*, January 1998.

[BRS91]   Richard Beigel, Nick Reingold, and Daniel A. Spielman. The Perceptron strikes back. In *Structure in Complexity Theory Conference*, pages 286–291, 1991.

[BRS95]   Richard Beigel, Nick Reingold, and Daniel A. Spielman. PP is closed under intersection. *J. Comput. Syst. Sci.*, 50(2):191–202, 1995.

[Gol97]   Mikael Goldmann. On the power of a threshold gate at the top. *Information Processing Letters*, pages 63(6): 287–293, (1997).

[Gop06]   Parikshit Gopalan. Constructing Ramsey graphs from Boolean function representations. In *Proceedings of the 21st IEEE Conference on Computational Complexity (CCC'06)*, 2006.

[Gro00]   Vince Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs. *Combinatorica*, 20(1):71–86, 2000.

[Gro02]   Vince Grolmusz. Constructing set systems with prescribed intersection sizes. *Journal of Algorithms*, 44(2):321–337, 2002.

[Hås01]   Johann Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[Kho91]   A.G. Khovanskii. *Fewnomials*. Translations of Mathematical Monographs, Vol.88, AMS, 1991.

[Kli01]   Adam Klivans. On the derandomization of constant depth circuits. In *International Workshop on Randomization and Computation (RANDOM'01)*, pages 249–260, 2001.

[KOS02]   Adam Klivans, Ryan O'Donnell, and Rocco Servedio. Learning intersections and thresholds of halfspaces. In *Proceedings of the 43rd Annual Symposium on Foundations of Computer Science*, pages 177–186, 2002.

[KP95]   Matthias Krause and Pavel Pudlak. On computing Boolean functions by sparse real polynomials. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, pages 682–691, 1995.

[KS01]   Adam Klivans and Rocco Servedio. Learning DNF in time $2^{O(n^{\frac{1}{3}})}$. In *Proceedings of the 33rd Annual Symposium on Theory of Computing*, pages 258–265, 2001.

[KS04]   Adam Klivans and Rocco Servedio. Toward attribute efficient learning of decision lists and parities. In *Proceedings of Conference on Computational Learning Theory*, pages 234–248, 2004.

[KV94]   Michael Kearns and Umesh Vazirani. *An Introduction to Computational Learning Theory*. MIT Press, 1994.

[LRW03]   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.

[LT97]   J.C Lagaraias and Richardson T.J. Multivariate Descartes' rule of signs and Sturmfels's challenge problem. *The Mathematical Intelligencer*, 19(3):9–15, 1997.

[MOS03]   Elchannan Mossel, Ryan O'Donnell, and Rocco Servedio. Learning juntas. In *Proceedings of the 35th Annual Symposium on Theory of Computing*, pages 206–212, 2003.

[MP68]   Marvin Minsky and Seymour Papert. *Perceptrons: an Introduction to Computational Geometry*. MIT Press, 1968.

[OS03a]   Ryan O'Donnell and Rocco Servedio. Extremal properties of polynomial threshold functions. In *Proceedings of the 18th Annual Conference on Computational Complexity*, pages 3–12, 2003.

[OS03b]   Ryan O'Donnell and Rocco Servedio. New degree bounds for polynomial threshold functions. In *Proceedings of the 35th Annual Symposium on Theory of Computation*, pages 325–334, 2003.

[PS76]   George Pólya and Gabor Szegö. *Problems and Theorems in Analysis, Vol. II*. Springer-Verlag, New York, 1976.

[Raz87]   Alexander Razborov. Lower bounds for the size of circuits of bounded depth with basis $\{\wedge, \oplus\}$. *Mathematical Notes of the Academy of Science of the USSR*, 41:333–338, 1987.

[Sak93]   Michael Saks. Slicing the hypercube. *Surveys in Combinatorics*, pages 211–255, 1993.

[Smo87]   Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the 19th Annual ACM Symposium on Theoretical Computer Science*, pages 77–82, 1987.

[Stu98]   Bernd Sturmfels. Polynomial equations and convex polytopes. *American Mathematical Monthly*, 105(10):907–922, 1998.