Complexity of Symmetric Functions in Perceptron-Like Models

Zhi-Li Zhang

Computer Science Department
University of Massachusetts
Amherst, Massachusetts MA 01003
USA

July, 1992

Abstract

We examine the size complexity of the symmetric boolean functions in two circuit models containing threshold gates: the d-perceptron model [BRS, ABFR] (a single threshold function of constant-depth AND/OR circuits) and the parity-threshold model studied by Bruck [Br] (a single threshold function of exclusive-ORs). These models are intermediate between the well-understood model of constant-depth AND/OR circuits and the still mysterious model of general constant-depth threshold circuits. In the d-perceptron model, we give an *if and only if* condition for a symmetric boolean function to be computable by a quasi-polynomial size d-perceptron: we show that a symmetric boolean function can be computed by a quasi-polynomial size d-perceptron *iff* it has only poly-log many sign changes, *i.e.* the number of times the function changes output value as the number of inputs on varies from zero through n (we call this parameter the *degree* of the symmetric function) is bounded above by $\log^c n$ for some c. This extends the work of Fagin *et al.* [FKPS] which gave a very nice characterization of symmetric functions computable by AC^0 circuits. An interesting consequence of our result is that a recent construction of Beigel [Be] is optimal. In the parity-threshold model, we find a similar parameter as a measure of size complexity, the *odd-even degree*, or number of output value changes as the number of inputs on varies through the odd numbers from 0 through n and then through the even numbers. We observe that poly-log odd-even degree implies quasi-polynomial size, conjecture the converse, and prove the converse in the presence of a certain technical condition on the function's Fourier coefficients. In particular, we prove that the modulo-q function for any constant $q > 2$ has more than quasi-polynomial size.

Acknowledement

This is a joint work with my advisor David A. Mix Barrington, I am grateful to him for many helpful discussions and patient guidance. I thank Neil Immerman for being the Second Reader. The proof of Theorem 10 in Section 3 was also independently discovered by Jun Tarui. I also thank him for many insightful comments. This work is supported by NSF grant CCR-8812567 and CCR-9008416.

1 Introduction

1.1 Perceptron-Like Models

The power of constant-depth circuits of unbounded fan-in AND and OR gates (*i.e.* the well-known AC^0 circuits) is by now fairly well understood [FSS, Aj, Hå, Ya]. One of the major open problems of complexity theory is to place any non-trivial bounds on the computing power of constant depth circuits of unbounded fan-in threshold or $MAJORITY$ gates. The class TC^0, of languages recognized by polynomial-size families of such circuits, might be equal to NP for all we can prove. A natural approach to bridging the gap between AND/OR circuits and threshold circuits is to consider models which combine the two kinds of gates.

One very old example of such a model is the *perceptron* of Minsky and Papert [MP], which can be viewed as a $MAJORITY$ gate whose inputs are ANDs of the input variables. These original perceptrons are rather limited and their computing power is well understood. But recently, perceptrons have been revived in a new form [BRS]. Along with a probabilistic version, there has emerged what we will call the d-perceptron, a constant-depth unbounded fan-in circuit which has AND and OR gates except for a single $MAJORITY$ gate at the output. It has been shown that such circuits require exponential size (exponentially many gates) to compute the MOD_2 function [Gr], to approximate the MOD_2 function [ABFR], or to compute or approximate the MOD_c function for any constant c [BS]. These d-perceptrons are closely linked to a model of computation which is interesting in its own right, where one evaluates a multilinear polynomial in the input variables, with coefficients in the integers or the reals, and outputs the sign of the result. (This can be extended to polynomials over the complex numbers, using an *ad hoc* notion of "sign" [BS].) Furthermore, the d-perceptron model is robust, in that other circuit models with a limited use of threshold gates can be mapped into it [BRS, ABFR, Be].

Another circuit model which can be placed in this framework is that used by Bruck [Br], which we will call the *parity-threshold model*. In this model a function with domain $\{-1,1\}^n$ and range $\{-1,1\}$ is computed by

evaluating a polynomial in the n input variables and taking its sign. A *polynomial threshold function* is one where this polynomial has only $n^{O(1)}$ nonzero coefficients. This model corresponds fairly closely to a circuit where the inputs are fed into MOD_2 gates (because multiplication in the $\{-1, 1\}$ domain corresponds to addition modulo 2) and the outputs of these gates are fed into a single $MAJORITY$ gate. The correspondence is exact if the coefficients are constrained to be integers bounded in absolute value by $n^{O(1)}$. Bruck studied this model using a form of harmonic analysis on boolean functions, which allowed him to show that certain natural functions are not polynomial threshold functions. This and other work place a geometrical structure on the boolean functions, so that a distance between functions can be defined. Linial *et al.* [LMN] used this geometry to show that any AC^0 function (any function computable by a poly-size constant-depth AND/OR circuit) can be closely approximated by a low-degree polynomial, a result which has consequences for the learnability of AC^0 functions.

1.2 Complexity of Symmetric Functions

The *symmetric* boolean functions are those which are invariant under any permutation of the inputs. We can describe a symmetric boolean function by giving its *spectrum*, which is the sequence $\langle f(0), \ldots, f(n) \rangle$, where each $f(i)$ is the value of the function when i of the n inputs are one. All symmetric functions are in TC^0, because they have linear-size depth-2 threshold circuits. In any model the complexity theory of the symmetric functions forms a subtheory of that of all boolean functions, and in some models this theory can be interesting and beautiful.

For example, consider the well-understood model of constant-depth unbounded fan-in AND/OR circuits. A theorem of Fagin *et al.* [FKPS], using the exponential lower bound for $PARITY$ due to Yao [Ya, Hå], gives an elegant characterization of the symmetric functions which have polynomial size in this model (are in the class AC^0). These functions are those whose spectra are constant except for a poly-log section at either end. That is, there is some function $g(n) = \log^{O(1)} n$ such that for each n, $f(i)$ is constant in the range $g(n) \leq i \leq n - g(n)$. Our principal question is whether similar

characterizations hold in the other models we consider.

1.3 Lower and Upper Bounds

In the d-perceptron model we add a single $MAJORITY$ gate to the AND/OR circuit, and thus we immediately allow new symmetric functions, such as $MAJORITY$ itself, to be computed. In previous work in this model [ABFR], two key parameters of a boolean function have proven to be its *strong degree* and *weak degree*. These are based on a space of polynomials over the real numbers, where the boolean domain is again taken to be $\{-1, 1\}$. The strong degree is the minimal degree of a polynomial whose sign always agrees with the target boolean function. The weak degree is the minimal degree of a polynomial, not identically zero, whose sign agrees with the target boolean function whenever the polynomial is nonzero. For symmetric functions, these two degrees are equal [ABFR], and furthermore are equal to the number of *sign changes* of the spectrum (the number of i for which $f(i) \neq f(i+1)$). This equivalence strongly suggests this parameter of symmetric functions (which we will call simply "degree") may be an important one.

Using a simple construction, we can show that any symmetric boolean function of poly-log degree can be computed by a quasi-polynomial size d-perceptron. We prove that the converse of this statement is also true — that if a symmetric boolean function has more than poly-log degree, it cannot be computed by a quasi-polynomial size d-perceptron. Before this work, only a few lower bounds on d-perceptron size were known, with very specific target functions such as $PARITY$ [Gr, ABFR] or MOD_c [BS], of linear degree.

In the parity-threshold model the linear-degree parity function is easy, so we need a different parameter if we are to get a similar result. The one we have in mind we call the *odd-even degree*, which we obtain by viewing the restrictions of the symmetric function to odd and even numbers of on-inputs respectively, and summing the two degrees (number of sign changes) of these restrictions. Equivalently, the odd-even degree is the number of i for which $f(i) \neq f(i+2)$.

In the parity-threshold model, we can show that a function with poly-log odd-even degree has quasi-polynomial size in the parity-threshold model

(in Bruck's language, it would be a "quasi-polynomial threshold function"). We conjecture the converse of this result is true — that if the odd-even degree is greater than poly-log, there is no parity-threshold circuit of quasi-polynomial size. Prior to this work Bruck had given a criterion for exponential (hence super-quasi-polynomial) parity-threshold size, based on the size of a function's Fourier coefficients. He gave a single example, the "complete quadratic" function, which has exponential size (this is very similar to the function proved not to have depth-2 threshold circuits by Hajnal *et al.* [HMPST]). We extend Bruck's work to allow some of the coefficients to be large, as long as the sum of the large ones is significantly bounded below one. This extension, Bruck's work, and our analysis of the Fourier coefficients of the MOD_p functions, let us prove that MOD_r requires exponential parity-threshold size for any constant $r > 2$. For general symmetric boolean functions of greater than poly-log odd-even degree, we can prove more that quasi-polynomial size in the presence of a certain technical condition — that the sum of the Fourier coefficients of sets of either poly-log size or more than n minus poly-log size is bounded significantly below one, and that the coefficients of all sets of other sizes are small.

2 Preliminaries

2.1 Symmetric Functions, Sign Change Spectra and Strong/Weak Representations

We will consider functions from $\{-1,1\}^n$ to the reals R (with boolean functions being the special case with range $\{-1,1\}$) as multilinear polynomials over R with input variables $\{x_1,\ldots,x_n\}$. The *size* of a polynomial is the number of nonzero coefficients, and the *degree* is the maximal number of variables appearing in a term with nonzero coefficient. We use $[n]$ to denote the set $\{1,2,\ldots,n\}$, and by $|\underline{x}|$ we mean the number of -1's in $\underline{x}$ (in general we think of -1 as "true" and 1 as "false").

A symmetric boolean function is a boolean function whose value only depends on $|\underline{x}|$. It can be proved that, over the reals, we can regard a

symmetric boolean function as a function of $x = |\underline{x}|$. Hence, in this way, we convert a n-variable symmetric boolean function $f(\underline{x})$, where $\underline{x} \in \{-1, 1\}^n$ into a univariate real function $f'(x)$, where $x = |\underline{x}|$ such that $deg(f) = deg(f')$. In the sequel, we will use f to denote $f(\underline{x})$ and $f'(x)$ interchangeably. Let $i \in [n]$, we say i is a *sign change* of a symmetric function f if $f(i) \neq f(i+1)$. The number of sign changes of f is equal to the cardinality of the set $\{i | f(i) \neq f(i+1)\}$. We will call $\langle f(0), f(1), \ldots, f(n) \rangle$ the *sign change spectrum* of f.

Following [ABFR], we define strong and weak representations of boolean functions as follows.

Definition 1 *We say a polynomial $F(\underline{x})$ over the reals R* strongly represents *a boolean function $f(\underline{x})$ if $sgn(F(\underline{x})) = f(\underline{x})$ for all $\underline{x} \in \{-1, 1\}^n$ where $sgn(F(\underline{x})) = 1$ if $F(\underline{x}) > 0$ and $sgn(F(\underline{x})) = -1$ if $F(\underline{x}) < 0$.*

And we say a polynomial $F(\underline{x})$ over the reals R weakly represents *a boolean function $f(\underline{x})$ if $F(\underline{x})$ is not identically zero and for all $\underline{x} \in \{-1, 1\}^n$ such that $F(\underline{x}) \neq 0$, $sgn(F(\underline{x})) = f(\underline{x})$.*

Definition 2 *Let $f(\underline{x})$ be a boolean function, the* strong degree *of $f(\underline{x})$ (denoted $d_s(f)$) is the minimum degree among all polynomials strongly representing $f(\underline{x})$, and the* weak degree *of $f(\underline{x})$ (denoted $w_s(f)$) is the minimum degree among all polynomials weakly representing $f(\underline{x})$.*

Notice that in general the weak degree of a boolean function may well be smaller than its strong degree. However, the following fact as first observed in [ABFR] says that for symmetric boolean functions, the two degrees are exactly the same, and equal to the number of sign changes of the symmetric boolean functions. We call this quantity the *degree* of a symmetric boolean function.

Lemma 1 ([ABFR]) *Let $f(\underline{x})$ be a symmetric boolean function with k sign changes, then $d_s(f) = d_w(f) = k$.*

The above lemma was proved in [ABFR] by exploring the duality relationship of certain function spaces. We note that it has a simpler proof

using the symmetrization technique [MP]: Let $F(\underline{x})$ be a strong or weak representation of a symmetric boolean function $f(\underline{x})$. Note that $F(\underline{x})$ is not necessarily symmetric itself; however, we can easily use $F(\underline{x})$ to construct a symmetric function $G(\underline{x})$. Formally, $G(\underline{x}) = \sum_{\sigma \in S_n} F^{\sigma}(\underline{x})$, where S_n is the nth symmetric group and $F^{\sigma}(\underline{x}) = F(\underline{x}^{\sigma}) = F(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. Obviously $G(\underline{x})$ strongly (weakly) represents $f(\underline{x})$ if $F(\underline{x})$ strongly (weakly) represents $f(\underline{x})$, and $G(\underline{x})$ has the same degree as $F(\underline{x})$. Since $G(\underline{x})$ is symmetric, $G(\underline{x})$ can be written as a univariate real function of $|\underline{x}|$. Therefore the degree of $G(\underline{x})$ is at least the number of sign changes of $f(\underline{x})$, and thus that of $F(\underline{x})$. But it is easy to construct a real polynomial strongly representing $f(\underline{x})$ such that its degree equals the number of sign changes of $f(\underline{x})$.

A variation of the degree concepts introduced above is the *odd-even* degree, which turns out to be a key parameter in our investigation of the parity-threshold model. The odd-even degree of a boolean symmetric function is the sum of number of sign chnages of the function when restricting to odd and even numbers of on-inputs respectively. Equivalently, the odd-even degree is the number of i for which $f(i) \neq f(i+2)$.

2.2 d-Perceptron Model and Parity-Threshold Model

We first define the d-perceptron model as introduced in [ABFR, BRS].

Definition 3 *A* d-perceptron *is a circuit with a MAJORITY gate at the top and depth-d AND/OR subcircuits feeding into the MAJORITY gate. The size of a d-perceptron is the number of gates in the circuit.*

We will be interested in d-perceptrons of polynomial and of *quasipolynomial* ($2^{\log^{O(1)} n}$) size. Note that by [FKPS], ordinary constant-depth AND/OR circuits of quasipolynomial size can compute no more symmetric functions than can circuits of polynomial size. For more on quasipolynomial size circuit classes, see [Ba].

Consider a polynomial strongly representing a boolean function on $\{0,1\}^n$ such that the coefficients are positive integers bounded by a quasipolynomial in n It is easy to see that such a polynomial corresponds to a quasipolynomial

size 1-perceptron whose gates on the first level are poly-log fan-in $ANDs$. In the following lem we show that any symmetric boolean function with only poly-log degree (*i.e.* sign changes) has such a low degree polynomial representation, and hence can be computed by such 1-perceptrons.

Lemma 2 *Any symmetric boolean function with only poly-log degree can be computed by a quasipolynomial size* 1*-perceptron with* $ANDs$ *of poly-log fan-in.*

Proof: Let $0 \leq c_1 < c_2 < \ldots < c_k < n$ be the positions of the sign changes of $f(\underline{x})$, where $k = \log^{O(1)} n$. Then the following function

$$F(\underline{x}) = (-1)^{\delta} \prod_{i=1}^{k} (c_i + \frac{1}{2} - \sum_{i=1}^{n} x_i), \text{ where } \delta = \begin{cases} 0 & \text{if } f(0) > 0 \\ 1 & \text{if } f(0) < 0 \end{cases}$$

agrees with $f(\underline{x})$ in sign (note here $\underline{x} \in \{0,1\}^n$).

F($\underline{x}$) is a poly-log degree polynomial with rational coefficients, but we can easily make $F(\underline{x})$ a polynomial with integer coefficients (without changing its sign) by multiplying it by an appropriate positive constant.

The problem left now is to convert the polynomial into one with only positive coefficients. Reversing the procedure used in [MP] to prove the *Positive Normal Form Theorem*, we can eliminate all the negative coefficients without blowing up either the degree or the size of the coefficients. ∎

Parity-threshold circuit is a depth-2 circuit with a $MAJORITY$ gate at the top level and $PARITY$ gates at the bottom level. The study of this model is inspired by the work of Bruck [Br] where he defined the polynomial threshold functions. Since we are considering functions defined on $\{-1,1\}^n$, each monomial $\prod_{i \in S} x_i$ corresponds to a $PARITY$ gate with input restricted to S. Clearly, the set of monomials $\prod_{i \in S} x_i$, $S \subseteq [n]$ forms a basis for functions defined on $\{-1,1\}^n$. We will denote the monomial $\prod_{i \in S} x_i$ by χ_S. A polynomial threshold function can be formally defined as follows.

Definition 4 *Let* $f(\underline{x})$ *be a boolean function, we say* $f(\underline{x})$ *is a* polynomial threshold function *if there is a real polynomial* $F(\underline{x})$ *such that*

$$f(\underline{x}) = sgn(F(\underline{x})) \text{ for all } \underline{x} \in \{-1,1\}^n \text{ and } F(\underline{x}) = \sum_{S \in [n]} w_S \chi_S$$

and the cardinality of the set $\mathbf{S} = \{S \subseteq [n] | w_S \neq 0\}$ *is bounded above by a polynomial in* n.

Similarly, we say a boolean function $f(\underline{x})$ *is a* quasi-polynomial threshold function *if the cardinality of* $\mathbf{S}$ *is bounded above by* $2^{\log^{O(1)} n}$.

Clearly, if the coefficients w_S in the above definition are integers bounded above by quasi-polynomial, the strong representation corresponds to a quasi-polynomial size parity-threshold circuit; Conversely, any quasi-polynomial size parity-threshold circuit yields a quasi-polynomial size representation of the boolean function it computes.

Analogous to the d-perceptron model, we have a similar upper bounds for symmetric boolean functions with poly-log odd-even degree in the parity-threshold models.

Theorem 3 *Let* $f(\underline{x})$ *be a symmetric boolean function with odd-even-degree* $\log^{O(1)} n$, *then* $f(\underline{x})$ *can be computed by a quasi-polynomial size parity-threshold circuit.*

Proof: Let $0 \leq c_2 < c_4 < \ldots < c_{2k} < n$ be the positions of sign changes at $|\underline{x}|$ even, where c_{2i}'s are even, $k = \log^{O(1)} n$, then the following function

$$F_{even}(\underline{x}) = (-1)^{\delta_{even}} \prod_{i=1}^{k} (c_{2i} + \frac{1}{2} - |\underline{x}|), \text{ where } \delta_{even} = \begin{cases} 0 & \text{if } f(0) > 0 \\ 1 & \text{if } f(0) < 0 \end{cases}$$

agrees with $f(\underline{x})$ in sign when $|\underline{x}|$ is even.

Similarly, let $1 \leq c_1 < c_3 < \ldots < c_{2l+1} < n$ be the positions of sign changes at $|\underline{x}|$ odd, where c_{2i+1}'s are odd, $l = \log^{O(1)} n$, then the following function

$$F_{odd}(\underline{x}) = (-1)^{\delta_{odd}} \prod_{i=0}^{l} (c_{2i+1} + \frac{1}{2} - |\underline{x}|), \text{ where } \delta_{odd} = \begin{cases} 0 & \text{if } f(0) > 0 \\ 1 & \text{if } f(0) < 0 \end{cases}$$

agrees with $f(\underline{x})$ in sign when $|\underline{x}|$ is odd.

Therefore $F(\underline{x}) = (1 + \chi_{[n]}(\underline{x}))F_{even}(\underline{x}) + (1 - \chi_{[n]}(\underline{x}))F_{odd}(\underline{x})$ strongly represents $f(\underline{x})$. Moreover, since F_{even} and F_{odd} are of degree $\log^{O(1)} n$, we can rewrite $F(\underline{x})$ as $F(\underline{x}) = \sum_{S \in \mathbf{S}} w_S \chi_S$, where $w_S \neq 0$ for $S \in \mathbf{S}$, such that $|\mathbf{S}| = 2^{\log^{O(1)} n}$. Therefore, $f(\underline{x})$ can be computed by a quasi-polynomial size parity-threshold circuit. ∎

2.3 Fourier Transform For Symmetric Functions

Harmonic Analysis has been applied to the study of boolean functions under various complexity measures and models and many interesting results have been obtained (see, *e.g.* [Br, KKL, LMN]). The essence of Harmonic Analysis is to find an orthonormal basis for the function space in question. We will follow the notations[1] used in [LMN]. For the space of functions defined on $\{-1,1\}^n$, it is not hard to see that the set of 2^n functions χ_S, $S \subseteq [n]$ forms an orthonormal basis.

The inner product defined on this function space is:

$$< f, g >= 2^{-n} \sum_{\underline{x} \in \{-1,1\}^n} f(\underline{x}) g(\underline{x})$$

As usual, the L_2 norm of a function is defined as $\| f \|_2 = \sqrt{< f, f >}$. For any $S \subseteq [n]$, the Fourier coefficient $\hat{f}(S)$ (or $\hat{f}_S$) of a function $f(\underline{x})$ is defined as

$$\hat{f}(S) =< f, \chi_S >$$

Hence, $f(\underline{x}) = \sum_S \hat{f}(S) \chi_S$.

The L_2 norm of a function and its Fourier coefficients are related by the following famous identity.

Fact 1 *Parseval's identity*

$$\| f \|_2^2 = \sum_{S \subseteq [n]} \hat{f}(S)^2$$

The Fourier transform on the boolean functions can also be succinctly expressed in a matrix form. Let $\mathbf{f^n}$ denote the vector of the 2^n values of a boolean function $f(\underline{x})$ and $\hat{\mathbf{f}}^\mathbf{n}$ denote the vector of the 2^n Fourier coefficients of $f(\underline{x})$. Then $\hat{\mathbf{f}}^\mathbf{n}$ and $\mathbf{f^n}$ are related by the so-called *Sylvester type Hadamard* matrix as illustrated below:

[1] In [Br], Bruck used a different set of notations, which are more suitable in talking about polynomial threshold functions. However, for consistency's sake, we will use LMN's notations throughout the paper.

Fact 2

$$\hat{\mathbf{f}}^{\mathbf{n}} = \frac{1}{2^n} H_{2^n} \mathbf{f}^{\mathbf{n}}$$

where the Sylvester type Hadamard matrix is defined recursively for all 2^k, $k \geq 0$ *as follows*

$$H_1 = [1] \quad H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad H_{2^{k+1}} = \begin{bmatrix} H_{2^k} & H_{2^k} \\ H_{2^k} & -H_{2^k} \end{bmatrix}$$

As we are mostly interested in symmetric functions, we will develop a Fourier Transform for symmetric boolean functions and show that the vector of Fourier coefficients and the sign change spectrum of a symmetric boolean function are related by a matrix which processes many interesting properties.

For $S \subseteq [n]$ and $t \in [n]$, we define $\psi_S(t) = \sum_{|\underline{x}|=t} \chi_S(\underline{x})$. The following properties of $\psi_S(t)$ can be easily checked.

Fact 3 $\psi_S(t)$ *has the following properties*

1) For $S, S' \subseteq [n]$ *such that* $|S| = |S'| = s$, *and for* $t \in [n]$, *we have*

$$\psi_S(t) = \psi_{S'}(t) = \sum_{i=\max(s+t-n,0)}^{min(s,t)} (-1)^i \binom{s}{i} \binom{n-s}{t-i}$$

Hence, we will use the lower case letter s *in* $\psi_S(t)$ *for all* S *with* $|S| = s$.

2) $\psi_s(t) = (-1)^s \psi_s(n-t)$

3) $\psi_s(t) = (-1)^t \psi_{n-s}(t)$

For $0 \leq s \leq n$, let $\sigma_s(\underline{x})$ denote the sth elementary symmetric function on $\{-1,1\}^n$, *i.e.* $\sigma_s(\underline{x}) = \sum_{\substack{S \subseteq [n] \\ |S|=s}} \chi_S(\underline{x})$. σ_s has very similar properties as ψ_s, which reveals some kind of duality between σ_s and ψ_s.

Fact 4 $\sigma_s(\underline{x})$ *has the following properties*

1) Let $\underline{x} \in \{-1,1\}^n$, $t = |\underline{x}| \in [n]$, *then*

$$\sigma_s(\underline{x}) = \sigma_s(t) = \sum_{i=\max(s+t-n,0)}^{min(s,t)} (-1)^i \binom{t}{i} \binom{n-t}{s-i}$$

2) $\sigma_s(t) = (-1)^s \sigma_s(n-t)$ *for* $0 \leq t \leq n$
3) $\sigma_s(t) = (-1)^t \sigma_{n-s}(t)$

Using σ_s and ψ_s, we can write any symmetric boolean function $f(\underline{x})$ as $f(\underline{x}) = \sum_{s=0}^{n} \hat{f}_s \sigma_s(\underline{x})$ where $\hat{f}_s = \hat{f}_S = 2^{-n} \sum_{t=0}^{n} f(t)\psi_s(t)$ for any $S \subseteq [n]$ with $|S| = s$, since $f(\underline{x})$ is symmetric. Obviously, $\{\sigma_s(\underline{x}) | 0 \leq s \leq n\}$ forms a basis for the space of all symmetric functions on $\{-1,1\}^n$. Viewing these symmetric functions as univariate functions of $x = |\underline{x}| \in [n]$, the induced inner product on this function space can then be defined via the original inner product.

Let $f(\underline{x})$ and $g(\underline{x})$ be two symmetric functions on $\{-1,1\}^n$

$$< f, g >= \frac{1}{2^n} \sum_{t=0}^{n} \binom{n}{t} f(t)g(t)$$

The orthonormality of $\{\chi_S(\underline{x}) | S \subseteq [n]\}$ under the original inner product leads to the orthogonality of $\{\sigma_s | 0 \leq s \leq n\}$ under this induced inner product.

Fact 5

$$< \sigma_s, \sigma_{s'} >= \begin{cases} 0 & when \quad s \neq s' \\ \binom{n}{s} & when \quad s = s' \end{cases}$$

$\sigma_s(t)$'s are the so-called Krawtchouk polynomials, please refer to [MS] for other interesting properties.

For any symmetric boolean function $f(\underline{x})$, the relationship between its sign change spectrum and its Fourier coefficients is stated in the following theorem

Theorem 4 *For any symmetric boolean function* $f(\underline{x})$, *let* $\hat{\mathbf{f}}^{\mathbf{n}} = [\hat{f}_0, \hat{f}_1, \ldots, \hat{f}_n]^{\mathbf{t}}$, *the vector of its Fourier coefficients and* $\mathbf{f}^{\mathbf{n}} = [f(0), f(1), \ldots, f(n)]^{\mathbf{t}}$, *the sign change spectrum of* $f(\underline{x})$ *on* $\underline{x} \in \{-1,1\}^n$. *Then*

$$\hat{\mathbf{f}}^{\mathbf{n}} = \frac{1}{2^n} T^n \cdot \mathbf{f}^{\mathbf{n}}$$

where T^n is an $(n+1)$ by $(n+1)$ matrix defined as

$$T^n = [\psi_s(t)]$$

i.e. the (s,t) entry $T^n_{s,t}$ of T^n is $\psi_s(t)$, $0 \le s,t \le n$. We call T^n the nth order Fourier Transform Matrix for symmetric functions.

Proof: The relationship follows directly from the relationship between $\hat{f}_s$ and $\psi_s(t)$. ∎

T^n enjoys many interesting properties. For example, by the properties of $\psi_s(t)$, we see T^n has many symmetries: for $0 \le s,t \le n$, $T^n_{s,t} = (-1)^s T^n_{s,n-t} = (-1)^t T^n_{n-s,t} = (-1)^{s+t} T^n_{n-s,n-t}$. And $|T^n_{s,t}| \le |T^n_{0,t}| = |T^n_{n,t}| = \binom{n}{t}$. Also observe that, by the duality between ψ_s and σ_t, the rows of T^n correspond to ψ_s and columns to σ_t. The orthogonality of σ_t yields that T^n is almost self-inverse, that is, we have $(T^n)^{-1} = \frac{1}{2^n} T^n$.

Though T^n has many nice properties, the construction of T^n appears to be formidable. This actually is not the case: we can construct T^n from T^{n-1} in a Pascal-Triangle-like way (more precisely, we are building a *Pascal Pyramid*), as demonstrated in the following theorem.

Theorem 5 *For $n > 1$, we have*

1. *For $0 \le s \le n$, $1 \le t \le n-1$,*
   $T^n_{s,t} = T^{n-1}_{s,t-1} + T^{n-1}_{s,t}, T^n_{s,0} = 1, T^n_{s,n} = T^{n-1}_{s,n-1} = (-1)^s$
2. *For $1 \le t \le n-1$,*
   $T^n_{n,t} = -T^{n-1}_{n,t-1} + T^{n-1}_{n,t}, T^n_{n,0} = 1, T^n_{n,n} = -T^{n-1}_{n,n-1} = (-1)^n$
3. *For $1 \le s \le n$, $1 \le t \le n-1$,*
   $T^n_{s,t} = -T^{n-1}_{s-1,t-1} + T^{n-1}_{s-1,t}, T^n_{s,0} = 1, T^n_{s,n} = -T^{n-1}_{s-1,n-1} = (-1)^s$
4. *For $1 \le t \le n-1$,*
   $T^n_{0,t} = T^{n-1}_{0,t-1} + T^{n-1}_{s,t}, T^n_{0,0} = 1, T^n_{0,n} = T^{n-1}_{0,n-1} = 1$

Proof: The above identities are very easy to prove by using the identity $\binom{m}{r} = \binom{m-1}{r-1} + \binom{m-1}{r}$ and the properties of $\psi_s(t)$. ∎

Examples

$$T^1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad T^2 = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 0 & -1 \\ 1 & -2 & 1 \end{bmatrix}, \quad T^3 = \begin{bmatrix} 1 & 3 & 3 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -3 & 3 & -1 \end{bmatrix},$$

$$T^4 = \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 1 & 2 & 0 & -2 & -1 \\ 1 & 0 & -2 & 0 & 1 \\ 1 & -2 & 0 & 2 & -1 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}, \quad T^5 = \begin{bmatrix} 1 & 5 & 10 & 10 & 5 & 1 \\ 1 & 3 & 2 & -2 & -3 & -1 \\ 1 & 1 & -2 & -2 & 1 & 1 \\ 1 & -1 & -2 & 2 & 1 & -1 \\ 1 & -3 & 2 & 2 & -3 & 1 \\ 1 & -5 & 10 & -10 & -5 & -1 \end{bmatrix},$$

$$T^6 = \begin{bmatrix} 1 & 6 & 15 & 20 & 15 & 6 & 1 \\ 1 & 4 & 5 & 0 & -5 & -4 & -1 \\ 1 & 2 & -1 & -4 & -1 & 2 & 1 \\ 1 & 0 & -3 & 0 & 3 & 0 & -1 \\ 1 & -2 & -1 & 4 & -1 & -2 & 1 \\ 1 & -4 & 5 & 0 & -5 & 4 & -1 \\ 1 & -6 & 15 & -20 & 15 & -6 & 1 \end{bmatrix}$$

One application of the matrix T^n is the estimation of Fourier coefficients for certain symmetric boolean functions. For example, by using T^n, we can obtain the following estimation for the modular functions, we leave the detailed proof in the appendix.

Theorem 6 *For any p prime bigger than 2, the Fourier coefficients of* $MOD_p(\underline{x})$*, when repersented as a function from* $\{-1,1\}^n$ *to* $\{-1,1\}$*, have the following properties:*

There exists $\epsilon > 0$ *such that*

$$\begin{aligned} \hat{f}_0 &= \frac{p-2}{p} \pm O(\frac{1}{2^{\epsilon n}}) < \frac{p-1}{p} \\ |\hat{f}_s| &= O(\frac{1}{2^{\epsilon n}}) \quad \textit{when } s \neq 0 \end{aligned}$$

3 Lower Bounds in the d-Perceptron Model

In this section, we will prove that any symmetric boolean function with more than poly-log sign changes cannot be computed by any quasi-polynomial size d-perceptron. Together with Theorem 2 in section 3, we have an *if and only if* condition for a symmetric boolean function to be computed by a quasi-polynomial size d-perceptron. The proof makes use of some key observations by Linial, et al [LMN], and uses the very technique of *random restriction*, first introduced in [FSS] and refined in [Ya, Hå], that gave the first exponential lower bound for AND/OR circuits.

A random restriction is a random mapping of the input variables to 0, 1 and $*$ according to some probability distribution. The function obtained from $f(x_1, \ldots, x_n)$ by applying a random restriction ρ is denoted by f^ρ, and its variables are those x_i for which $\rho(x_i) = *$. For our purpose, we will assume that ρ assigns values to each input variable independently and $Pr[0] = Pr[1] = \frac{1-Pr[*]}{2}$.

A simple observation is that any random restriction of a symmetric boolean function is still symmetric; furthermore, its sign change spectrum is a subinterval of that of the original function.

Recall that a *minterm* of a boolean function is a set of variables such that a partial assignment to the variables in the set makes the function identically 1, but no partial assignment to any subset of the set makes the function identically 1. Similarly, a *maxterm* is a set of variables such that a partial assignment to the variables in the set makes the function identically 0, but no partial assignment to a subset of the set makes the function identically 0. Linial, *et al* [LMN] observed that if a boolean function has both small size minterms and maxterms, then all of its high degree Fourier coefficients vanish, and hence it can be represented by a low degree polynomial over the reals. More formally, we have

Lemma 7 ([LMN]) *If all the minterms and the maxterms of a boolean function f have size at most t, then for any subset S with $|S| > t^2$, the Fourier coefficient of f on S, $\hat{f}(S)$ is equal to 0.*

Therefore,

$$f = \sum_{S \subset [n], |S| \leq t^2} \hat{f}(S)\chi_S.$$

The above lemma is proved by using decision trees. The following fact was independently discovered in [BI,HH, Ta], and explicitly stated in [LMN]. A relevant fact was observed in [Hå], and our proof is a simple adaptation of the proof there.

Lemma 8 *If all the minterms and maxterms of a boolean function f have size at most s and t respectively, then f can be evaluated by a decision tree of depth at most st.*

Proof: Observe first that any minterm has a nonempty intersection with any maxterm, and vice versa (this follows easily from the definitions of minterms and maxterms). Now we prove the lemma by induction on t.

When $t = 1$, then there exists an $S \subseteq [n]$ such that $|S| \leq s$ and $f = \prod_{i \in S} x_i$. Obviously f has a decision tree of depth at most s.

Suppose the lemma is true for functions with maxterm size at most $t-1$. Let δ be a minterm of f, consider all possible assignments to δ. Let β be such an assignment, denote by f_β the resulting function. Since δ intersects any maxterm of f nonemptily, any maxterm of f_β is of size at most $t-1$. By the inductive hypothesis, f_β has a decision tree of depth at most $s(t-1)$ for all β. But assignments to δ correspond to decision trees of depth $|\delta| \leq s$. Adjoining these decision trees together, we have a decision tree for f of depth at most $s + s(t-1) = st$. ∎

Proof of Lemma 7: By Lemma 8, we see that f has a decision tree of depth t^2. Let T be a decision tree for f of depth at most t^2 and let S be a set of size larger than t^2. Since the path to any leaf depends on at most t^2 variables, there is a variable in S not queried along this path, hence exactly half of the inputs that end in the leaf agree with χ_S. Since the leaves induce a partition of the inputs, the function computed by T agrees with χ_S on exactly half of the inputs. Therefore, the correlation between T and χ_S is zero, and the lemma follows. ∎

It is well-known that with high probability, a random restriction of an AC^0 function will have small minterm size and maxterm size. This can be proved by a repeated applications of Håstad's switching lemma [BoS, LMN]. We state this fact formally as follows.

Lemma 9 ([LMN]) *Let f be a boolean function computed by an AND/OR circuit of size M and depth d. Then*

$$Pr[f^\rho \text{ has a minterm or a maxterm of size } > t] \leq M2^{-t}$$

where ρ is random restriction such that $Pr[*] = 1/(10t)^d$.

Theorem 10 *Let f be a symmetric boolean function on n variables. Suppose f can be computed by a d-perceptron such that the fan-in of the $MAJORITY$ gate is N, the size of each AC^0 subcircuit is at most M, and the depth is at most d, where $N, M \leq 2^{t-1}$. Then for a positive fraction of the random restrictions ρ in a distribution with $Pr[*]\ldots$, f^ρ is a function of at least $np = \frac{n}{(10t)^d}$ variables, the number of sign changes of f^ρ is at most $O(t^2)$, and the sign change spectrum of f^ρ is a subinterval at most $O(\sqrt{n})$ off the center of the sign change spectrum of f.*

Proof: Let ρ be a random restriction with $Pr[*] = p = \frac{1}{(10t)^d}$. Denote by f_i, $1 \leq i \leq N$, the subfunctions computed by the AND/OR subcircuits, and by f_i^ρ, $1 \leq i \leq N$, the functions obtained by the random restriction. Applying Lemma 9 to each f_i, we have

$$Pr[f_i^\rho \text{ has a minterm or maxterm of size } > 2t] \leq M2^{-2t}.$$

Hence,

$$\begin{aligned} & Pr[\bigwedge_{i=1}^{N} f_i^\rho \text{ has only minterms and maxterms of size } \leq 2t] \\ \geq\ & 1 - \sum_{i=1}^{N} Pr[f_i^\rho \text{ has a minterm or maxterm of size } > 2t] \\ \geq\ & 1 - NM2^{-2t} \geq 1 - 2^{2t-2}2^{-2t} \geq \frac{3}{4} \end{aligned}$$

On the other hand, the expected number of variables assigned $*$ is $np = \frac{n}{(10t)^d}$. By the normal approximation to binomial distribution, for n large, we see that with probability at least $\frac{1}{4}$, ρ will assign $*$'s to at least np variables and an almost equal number of 0 and 1's to the rest of the input variables.

Therefore, there must be a ρ such that $f^\rho = MAJORITY(f_1^\rho, \ldots, f_N^\rho)$ is a function on at least np variables, and each f_i^ρ has both minterms and maxterms of size $\leq 2t$. By Lemma 7, it follows that f_i^ρ can be represented by a $(-1,1)$-valued real function of degree at most $4t^2$, hence $g^\rho = \sum_{i=1}^N f_i^\rho - \frac{1}{2}$ is a strong representation of f^ρ. Since g^ρ has degree at most $4t^2$, f^ρ can have at most $4t^2$ sign changes. ∎

Remark: If we choose $t = \log^{O(1)} n$, then f^ρ can have only poly-log many sign changes. Therefore for the original function f, there must exist a subinterval of length at least $\frac{n}{(10)^d} = \frac{n}{\log^{O(1)} n}$, near the center of the sign change spectrum of f, such that f has at most poly-log many sign changes in that interval.

To prove the result that any symmetric function of more than poly-log sign changes cannot be computed by any quasipolynomial size d-perceptron, we need to use a shifting technique to locate an interval in the sign change spectrum of the function such that we can apply the above lemma to obtain a contradiction.

Theorem 11 *If f is a symmetric boolean function of more than poly-log sign changes, then f cannot be computed by a quasipolynomial size d-perceptron for any constant d.*

Proof: Suppose the opposite is true: there exists a quasipolynomial size d-perceptron for some constant d. Let c be such that $N, M \leq 2^{\log^c n - 1}$ where N, M are as in Theorem 10. Let $s(n)$ be the sign change function of f, by the hypothesis $s(n) = \log^{\omega(1)} n$.

Consider the interval $[s^{\frac{1}{2}}(n), n - s^{\frac{1}{2}}(n)]$ of the sign change spectrum of f, the number of sign changes in this interval is $\Omega(s(n))$. Without loss of generality, we assume that there are $\Omega(s(n))$ sign changes in $[s(n)^{\frac{1}{2}}, \frac{n}{2}]$. Partition this interval into k intervals of the form $[2^i s^{\frac{1}{2}}(n), 2^{i+1} s^{\frac{1}{2}}(n)]$, where $0 \leq i \leq k-1$ and $k = \log n - \frac{1}{2}\log s(n) - 1 = O(\log n)$.

We further partition each of the intervals $[2^i s^{\frac{1}{2}}(n), 2^{i+1} s^{\frac{1}{2}}(n)]$ into δ subintervals of length $\frac{2^i s^{\frac{1}{2}}(n)}{\delta}$ where $\delta = (10t)^d$ and $t = \log^c n$, *i.e.* $\delta = (10 \log^c n)^d = O(\log^{dc} n)$. We contend that one of the subintervals must have $\omega(t^2)$ sign changes, since otherwise, the total number of sign changes in $[s^{\frac{1}{2}}(n), \frac{n}{2}]$ is at most $\sum_{i=0}^{k-1} O(t^2)\delta = O(t^2 \delta k) = \log^{O(1)} n$, a contradiction. Therefore for some i, $0 \leq i \leq k-1$, a subinterval of length $\frac{2^i s^{\frac{1}{2}}(n)}{\delta} = \log^{\omega(1)} n$ has $\omega(t^2)$ sign changes.

By an appropriate partial assignment to the input variables, we obtain from f a function f', of $2^i s^{\frac{1}{2}}(n)$ variables, whose sign change spectrum is identical to an interval of the sign change spectrum of f which contains the aforementioned subinterval at center. Note that the circuit for f induces a circuit for f' of size at most that for f and thus its N', M' are bounded by $2^{t-1} = 2^{\log^c n - 1}$. Therefore, applying Theorem 10 to f', we have that there exists a random restriction ρ such that f'^{ρ} contains the subinterval as its sign change spectrum. However, f'^{ρ} can have only $O(t^2) = \log^{O(1)} n$ many sign changes, hence we arrive at a contradiction. ∎

We conclude this section with a consequence of Theorem 11 — the optimality of a recent construction of Beigel [Be].

Beigel shows that d-perceptrons serve as a normal form for AND/OR circuit s augmented by a small number of $MAJORITY$ gates. In particular, a circuit of unbounded fan-in, quasipolynomially many AND, OR, and NOT gates, and poly-log many $MAJORITY$ gates can be converted into a d-perceptron of quasipolynomial size. We can now show there is no general way to eliminate more than poly-log many $MAJORITY$s in this way.

Corollary 12 *Let $m = \log^{\omega(1)} n$. There exists a symmetric boolean function computable by circuits of $\Theta(m)$ $MAJORITY$ gates (with no other gates) but not computable by any d-perceptron family of quasipolynomial size.*

Proof: By Theorem 11, any symmetric function with exactly m sign changes will do. ∎

4 Lower Bounds in the Parity-Threshold Model

Our conjecture is that any symmetric boolean function with greater than poly-log odd-even degree needs more than quasi-polynomial size in the parity-thresholdmodel. We have some small steps toward such a result, extending the earlier work of Bruck [Br].

Bruck's main theorem gives a criterion to test when a boolean function is a polynomial threshold functions based on the size of the Fourier coefficients of a boolean function — the size is at least the inverse of the largest Fourier coefficient. This result can be easily translated into a size lower bound in parity-threshold model.

Theorem 13 (Bruck's Theorem) *Fix any $\epsilon > 0$. Let $f(\underline{x})$ be a boolean function of n variables. If $|\hat{f}_S| \leq 2^{-\epsilon n}$ for all $S \subseteq [n]$, then $f(\underline{x})$ is not a polynomial threshold function.*

In particular, Bruck defined the *Complete Quadratic* function $CQ(\underline{x})$ as follows

$$CQ(\underline{x}) = \begin{cases} 1 & |\underline{x}| = 0 \text{ or } 1 \bmod 4 \\ -1 & \text{otherwise} \end{cases}$$

and showed that when n is even, all Fourier coefficients $\hat{f}_S$ of $CQ(\underline{x})$ is $\pm 2^{-n/2}$; when n is odd, all Fourier coefficients are either zero or $2^{-(n-1)/2}$. Hence, $CQ(\underline{x})$ is not a polynomial threshold function, *i.e.* any parity-threshold circuit computing $CQ(\underline{x})$ has exponential size.

We extend Bruck's theorem to allow ourselves to ignore some large Fourier coefficients, as long as they do not come close to summing to one. Recall, however, that the sum of all the Fourier coefficients might be as high as $2^{n/2}$. Limited though it is, this extension and our analysis of the Fourier coefficients of the MOD functions give us some significant results. Before we give the proof of the extension, let us first cite two useful lemmas from [Br].

Lemma 14 *Let $F(\underline{x}) = \sum_{S \subseteq [n]} w_S \chi_S$. strongly represents a boolean function $f(\underline{x})$. Define $\mathbf{S} = \{S \subseteq [n] | w_S \neq 0\}$, then*

$$\sum_{\underline{x} \in \{-1,1\}^n} |F(\underline{x})| = 2^n \sum_{S \in \mathbf{S}} w_S \chi_S$$

Lemma 15 *Let $F(\underline{x})$ and $f(x)$ be as in Lemma 13. For all $S \in \mathbf{S}$*

$$2^n|w_S| \leq \sum_{\underline{x}\in\{-1,1\}^n} |F(\underline{x})|$$

Hence,

$$|w_S| \leq \sum_{S\in\mathbf{S}} w_S \hat{f}_S$$

The following theorem is an extension of Bruck's main theorem.

Theorem 16 *Suppose $F(\underline{x}) = \sum_{S\subseteq[n]} w_S\chi_S$ strongly represents a boolean function $f(\underline{x})$. As before, let $\mathbf{S} = \{S \subseteq [n] | w_S \neq 0\}$. For a fixed constant ϵ, $0 < \epsilon \leq 1$, define $\mathbf{T}, \mathbf{R} \subseteq \mathbf{S}$ as follows:*

$$\mathbf{T} = \{S \in \mathbf{S} || \hat{f}_S| \leq 2^{-n^\epsilon}\} \text{ and } \mathbf{R} = \mathbf{S}\backslash\mathbf{T}$$

If $T \neq \emptyset$ and $\sum_{S\in\mathbf{R}} |\hat{f}_S| \leq \delta$ for some constant δ, $0 \leq \delta < 1$, then $|\mathbf{S}| = 2^{n^{O(1)}}$.

Before we prove Theorem 16, we need a lemma.

Lemma 17 *Let $\mathbf{S}$, $\mathbf{T}$ and $\mathbf{R}$ be defined as in Theorem 17, and let $|w_{S_0}| = \max_{S\in\mathbf{R}}\{|w_S|\}$, then*

$$\frac{|w_{S_0}|}{\sum_{S\in\mathbf{T}} |w_S|} \leq 2^{-n^\epsilon}(1 - \sum_{S\in\mathbf{R}} |\hat{f}_S|)^{-1} \leq 2^{-n^\epsilon}(1-\delta)^{(-1)}$$

Proof: By Lemma 15, we have $\sum_{S\in\mathbf{S}} |w_S||\hat{f}_S| \geq |w_{S_0}|$, hence,

$$\sum_{S\in\mathbf{T}} |w_S||\hat{f}_S| + \sum_{S\in\mathbf{R}} |w_S||\hat{f}_S| \geq |w_{S_0}|$$

Note that for $S \in \mathbf{T}$, $|\hat{f}_S| \leq 2^{-n^\epsilon}$, therefore

$$\begin{aligned}(\sum_{S\in\mathbf{T}} |w_S|)2^{-n^\epsilon} &\geq \sum_{S\in\mathbf{T}} |w_S||\hat{f}_S| \geq |w_{S_0}| - \sum_{S\in\mathbf{R}} |w_S||\hat{f}_S| \\ &\geq |w_{S_0}| - (\sum_{S\in\mathbf{R}} |\hat{f}_S|)|w_{S_0}| \geq |w_{S_0}|(1 - \sum_{S\in\mathbf{R}} |\hat{f}_S|)\end{aligned}$$

Thus the lemma follows. ∎

Proof of Theorem 16: Again by Lemma 15, we have

$$|w_{S_0}| \leq \sum_{S \in \mathbf{S}} w_S \hat{f}_S \leq \sum_{S \in \mathbf{S}} |w_S||\hat{f}_S|$$

Then

$$\begin{aligned} \sum_{S \in \mathbf{T}} |w_S| &\leq \sum_{S \in \mathbf{S}} |w_S| \leq |\mathbf{S}|(\sum_{S \in \mathbf{S}} |w_S||\hat{f}_S|) \\ &\leq |\mathbf{S}|(\sum_{S \in \mathbf{T}} |w_S||\hat{f}_S| + \sum_{S \in \mathbf{R}} |w_S||\hat{f}_S|) \end{aligned}$$

Hence

$$|\mathbf{S}| \geq \left(\frac{\sum_{S \in \mathbf{T}} |w_S||\hat{f}_S| + \sum_{S \in \mathbf{R}} |w_S||\hat{f}_S|}{\sum_{S \in \mathbf{T}} |w_S|} \right)^{-1}$$

But since

$$\begin{aligned} &\frac{\sum_{S \in \mathbf{T}} |w_S||\hat{f}_S| + \sum_{S \in \mathbf{R}} |w_S||\hat{f}_S|}{\sum_{S \in \mathbf{T}} |w_S|} \\ \leq &\frac{(\sum_{S \in \mathbf{T}} |w_S|)2^{-n^\epsilon} + |w_{S_0}|(\sum_{S \in \mathbf{R}} |\hat{f}_S|)}{\sum_{S \in \mathbf{T}} |w_S|} \\ &\leq 2^{-n^\epsilon} + \frac{|w_{S_0}|}{\sum_{S \in \mathbf{T}} |w_S|} \cdot (\sum_{S \in \mathbf{R}} |w_S|) \\ &\leq 2^{-n^\epsilon} + 2^{-n^\epsilon} \frac{\delta}{1-\delta} \\ &\leq 2^{-n^\epsilon}(1-\delta)^{-1} \end{aligned}$$

Therefore

$$|\mathbf{S}| \geq 2^{n^\epsilon}(1-\delta)$$

■

Remark Instead of requiring δ be a constant < 1, it is enough to have $\delta \leq 1 - o(2^{-n^{O(1)}})$, say $\delta \leq 1 - n^{-c}$ or $\delta \leq 1 - 2^{-\log^c n}$ for some constant c. In other words, we only need to bound $\sum_{S \in \mathbf{R}} |\hat{f}_S|$ significantly away from 1 (*i.e.* more than an exponential fraction away from 1).

Corollary 18 *Let $f(\underline{x})$ be a boolean function. If any strong representation $F(\underline{x})$ of $f(\underline{x})$, where $F(\underline{x}) = \sum_{S\subseteq[n]} w_S\chi_S$, satisfies the condition in Theorem 16, then $f(\underline{x})$ cannot be computed by a quasi-polynomial size parity-threshold circuit.*

On the other hand, all quasi-polynomial threshold functions do not satisfy the condition in Theorem 16, that is, the large Fourier coefficients must sum up either very close to 1 or bigger than 1.

Theorem 19 *Let $f(\underline{x}) = \sum_{S\subseteq[n]} \hat{f}_S\chi_S$ be a boolean function, define $\mathbf{R} = \{S||\hat{f}_S^{-1}| = O(2^{\log^{O(1)}})\}$. If $f(\underline{x})$ can be computed by a quasi-polynomial size parity-threshold circuit, then $\sum_{S\in\mathbf{R}} |\hat{f}_S| \geq 1 - 2^{-n^\epsilon}$ for some $\epsilon > 0$.*

Proof: Clearly, if for any S not in $\mathbf{R}$, $\hat{f}_S = 0$, then the theorem is trivially true, since $1 = \sum_{S\subseteq[n]} |\hat{f}_S^2 \leq \sum_{S\subseteq[n]} |\hat{f}_S$.

Suppose there is an S' not in $\mathbf{R}$ such that $\hat{f}_S \neq 0$. Let $F(\underline{x}) = \sum_{S\in\mathbf{S}} w_S\chi_S$, where $w_S \neq 0$ for all $S \in \mathbf{S}$, be a strong representation of $f(\underline{x})$. It maybe be the case that $S' \notin \mathbf{R}$, in order to make use of Theorem 16, we construct another polynomial that contains S'. Without loss of generality, we assume w_S's are integers, hence $|F(\underline{x})| \geq 1$. Define $F'(\underline{x}) = \sum_{S\in\mathbf{S}} w_S\chi_S + w_{S'}\chi_{S'}$ where $|w_{S'}| < \frac{1}{2}$, it is easy to see that $F'(\underline{x})$ is also strong representation of $f(\underline{x})$. Since $f(\underline{x})$ can be computed by a quasi-polynomial size parity-threshold circuit, applying Corollary 18, we prove the theorem.

■

Corollary 20 *For any prime $p > 2$, MOD_p cannot be computed by a quasi-polynomial size parity-threshold circuit.*

Proof: In the appendix, we will show that for any prime $p > 2$, $MOD_p(\underline{x})$ has only one "large" coefficient, the Fourier coefficient of the constant term, which is smaller than $\frac{p-1}{p}$; all other Fourier coefficients are exponentially smaller than 1, hence by Theorem 19, MOD_p cannot be computed by a quasi-polynomial size parity-threshold circuit. ■

Note that for any p and r such that $p|r$, if MOD_r can be computed by a quais-polynomial parity-threshold circuit, then so can $MOD_p(\underline{x})$. Therefore we have shown that for any $r \neq 2^k$ for some $k > 0$, $MOD_r(\underline{x})$ cannot be computed by a quasi-polynomial size parity-threshold circuit. By a reduction from $CQ(\underline{x})$ to $MOD_4(\underline{x})$, where $CQ(\underline{x})$ is the *Complete Quadratic* function defined by Bruck, we can show that MOD_4 cannot be computed by a quasi-polynomial size parity-threshold circuit, either.

Theorem 21 *$MOD_4(\underline{x})$ cannot be computed by a quasi-polynomial size parity-threshold circuit.*

Proof: By Kummar's Lemma ([Kn], Exercise 1.2.6-11), $x \equiv 0 \bmod 4$ *iff* $x \equiv 0 \bmod 2$ and $\binom{x}{2} \equiv 0 \bmod 2$. Let $F_n(\underline{x})$ and $\chi_{[n]}(\underline{x})$ be the representation of $MOD_4(\underline{x})$ and $MOD_2(\underline{x})$ functions on the domain $\{-1,1\}^n$, we have

$$CQ(\underline{x}) = -\frac{1}{2}(1+\chi_{[n]}(\underline{x}))F_n(\underline{x}) + \frac{1}{2}(1-\chi_{[n]}(\underline{x}))F_{n+1}(\underline{x},-1)$$

Note that $F_n(\underline{x}) = -1$ *iff* $|\underline{x}| \equiv 0 \bmod 4$ for $\underline{x} \in \{-1,1\}^n$, hence $F_{n+1}(\underline{x},-1) = -1$ *iff* $|\underline{x}| \equiv 3 \bmod 4$ for $\underline{x} \in \{-1,1\}^n$. Since $CQ(\underline{x})$ cannot be computed by a quasi-polynomial size parity-threshold circuit, neither is $MOD_4(\underline{x})$. ∎

Combining Corollary 20 and Theorem 21, we have

Theorem 22 *For any constant $r > 2$, $MOD_r(\underline{x})$ cannot be computed by a quasi-polynomial size parity-threshold circuit.*

Now we give a more general criterion for when a symmetric boolean function cannot be computed by a quasi-polynomial size parity-threshold circuit. Recall that the odd-even-degree of a symmetric boolean function $f(\underline{x})$ is sum of the number of sign changes of $f(\underline{x})$ when restricting $|\underline{x}|$ to even and odd respcetively; equivalently, the odd-even degree of $f(\underline{x})$ is the number of i for which $f(i) \neq f(i+2)$.

Theorem 23 *Let $f(\underline{x})$ be a symmetric boolean function with odd-even-degree $\log^{\omega(1)} n$ satisfying the following condition:*

There exists $t = O(\log^{O(1)} n)$, $\epsilon > 0$ *and* $c > 0$ *such that*

$$|\hat{f}_S|^{-1} = 2^{n^\epsilon} \text{ for all } S \subseteq [n],\ t < |S| < n - t$$

and

$$\sum_{\substack{S \subseteq [n] \\ |S| \leq t \text{ or } |S| \geq n-t}} |\hat{f}_S| < 1 - \frac{1}{2^{\log^c n}}$$

Then $f(\underline{x})$ *cannot be computed by a quasi-polynomial size parity-threshold circuit.*

Proof: Let $F(\underline{x}) = \sum_{S \in \mathbf{S}} w_S \chi_S$, where $w_S \neq 0$ for $S \in \mathbf{S}$, be any polynomial strongly representing $f(\underline{x})$, we will show that $|\mathbf{S}|$ is exponentially large.

First notice that for any $S \subseteq [n]$, $\chi_S(\underline{x}) = (-1)^{|\underline{x}|}\chi_{\overline{S}}(\underline{x})$ where $\overline{S}$ is the complement of S.

Suppose for any $S \in \mathbf{S}$, $|S| \leq t$ or $|S| \geq n - t$, then

$$\begin{aligned} F(\underline{x}) &= \sum_{|S| \leq t} w_S \chi_S + \sum_{|S| \geq n-t} w_S \chi_S \\ &= \sum_{|S| \leq t} (w_S + (-1)^{|\underline{x}|} w_{\overline{S}}) \chi_S \end{aligned}$$

Without loss of genrality, we assume n is even, and $f(\underline{x})$ has more than poly-log sign changes when restricted to $|\underline{x}|$ even. Define

$$\begin{aligned} F'(x_1, x_2, \ldots, x_{\frac{n}{2}}) &= F(x_1, x_1, x_2, x_2, \ldots, x_{\frac{n}{2}}, x_{\frac{n}{2}}) \\ &= \sum_{|S| \leq t} (w_S + w_{\overline{S}}) \chi_S(x_1, x_1, x_2, x_2, \ldots, x_{\frac{n}{2}}, x_{\frac{n}{2}}) \end{aligned}$$

It is not too hard to see that $F'(x_1, x_2, \ldots, x_{\frac{n}{2}})$ strongly represents the $n/2$ variable boolean function $f'(x_1, x_2, \ldots, x_{\frac{n}{2}}) = f(x_1, x_1, x_2, x_2, \ldots, x_{\frac{n}{2}}, x_{\frac{n}{2}})$, which is obviously symmetric. Since $f(\underline{x})$ has more than poly-log sign changes when restricted to $|\underline{x}|$ even, $f'(x_1, x_2, \ldots, x_{\frac{n}{2}})$ has more than poly-log sign changes on $\{x_1, x_2, \ldots, x_{\frac{n}{2}}\} \in \{-1, 1\}^{\frac{n}{2}}$. However, since $deg(F') = t = \log^{O(1)} n$, we arrive at a contradiction.

Hence, for any polynomial $F(\underline{x}) = \sum_{S \in \mathbf{S}} w_S \chi_S$ strongly representing $f(\underline{x})$, $\mathbf{S}$ must contain an S such $t < |S| < n - t$, by Corollary 18, the condition in the theorem implies that $f(\underline{x})$ cannot be computed by a quasi-polynomial size parity-threshold circuit. ∎

In contrast to d-perceptron model, in this model currently we are not able to match the upper and lower bounds, due to the technical condition in Theorem 23. One might suspect that all symmetric boolean functions with odd-even-degree $\log^{\omega(1)} n$ satisfy the condition, since their Fourier coefficients is very likely to spread out in the middle. However, this is not true, there are symmetric boolean functions with odd-even-degree $\log^{\omega(1)} n$ violating the condition in Theorem 23. $MOD_4(\underline{x})$ is such an example, as it can be proved that $|\hat{f}_0| + |\hat{f}_n| = 1$, by using the Fourier Transform Matrix for symmetric boolean function defined in section 2.3. We hope that more careful analysis of Fourier coefficients of symmetric functions may allow us to prove the following conjecture:

Conjecture 1 *A symmetric boolean function $f(\underline{x})$ is computable by a quasi-polynomial size parity-threshold circuit iff its odd-even-degree is $\log^{O(1)} n$.*

5 Conclusion and Open Problems

In this thesis, we studied two circuit models containing a single *Majority* gate, the d-perceptron model and the parity-threshold model, using polynomials over reals as methods of boolean function representation.

In the d-perceptron case, we proved an *if and only if* condition for a symmetric boolean function to be computable by a quasi-polynomial size d-perceptron circuit. This work extends the line of an earlier work by Fagin, et al [FKPS] where they gave an *if and only if* condition for a symmetric function to be computable by a polynomial size AC^0.

In the parity-threshold case, we conjectured an analogous *if and only if* condition exists, but we were only able to partially resolve the problem under a certain technical condition. One particular case of interest is that for any constant $p > 2$, MOD_p is not computable by any quasi-polynomial size parity-threshold circuit.

The analysis of threshold computation by algebra over fields of characteristic zero has proved somewhat fruitful. The computation of circuits of AND, OR, and MOD_p gates has been very well explained using algebra over fields

of characteristic p [Ra, Sm]. Is it possible to combine the two methods, or otherwise place limits on the power of the following perceptron-like model: a $MAJORITY$ gate, whose inputs are constant-depth $AND/OR/MOD_p$ circuits?

References

[Aj] M. Ajtai. $\sum_1^1$-formulae on Finite Structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[ABFR] J. Aspnes, R. Beigel, M. Furst and S. Rudich. On the Expressive Power of Voting Polynomials. *Proceedings of the 23rd Annual Symposium on Theory of Computing*, 1991

[Ba] D. A. M. Barrington. Quasipolynomial size circuit classes. *Proceedings: Structure in Complexity Theory, Seventh Annual Conference*, pages 86-93, June 1992.

[Be] R. Beigel. Do Extra Threshold Gates Help. *Proceedings of the 24th Annual Symposium on Theory of Computing*, 1992

[BI] M. Blum and R. Impagliazzo. Generic Oracles and Oracle Classes. *Proceedings of the 28th Annual Symposium on Foundations of Computer Science*, pages 118-126, October 1987

[BoS] R. Boppana and M. Sipser. The Complexity of Finite Functions. *Handbook of Theoretical Computer Science*, Vol. 1, ed. by van Leeuwen, 1990.

[Br] J. Bruck. Harmonic Analysis of Polynomial Threshold Functions. *SIAM J. DISC. MATH.* Vol. 3 No.2, May 1990.

[BRS] R. Beigel, N. Reingold and D. Spielman. The Perceptron Strikes Back. *Proceedings of the 6th Annual Conference on Structrure in Complexity Theory* pages 286-291, 1991.

[BS] D. A. Mix Barrington and H. Straubing. Complex Polynomials and Circuit Lower Bounds for Modular Counting. *LATIN '92*, also Technical Report BCCS-91-5, Boston College, June, 1991.

[FKPS] R. Fagin, M. M. Klawe, N. J. Pippenger, and L. Stockmeyer. Bounded Depth, Polynomial Size Circuits for Symmetric functions. *Theoretical Computer Science* **36** pages 239-250, 1985.

[FSS] M. Furst, J. Saxe, and M. Sipser. Parity, Circuits, and the Polynomial Time Hierarchy. *Math. System Theory*, 17:13–27, 1984.

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

[Hå] J. Håstad. *Computational Limitations of Small-Depth Circuits.* MIT PRESS, 1986.

[HH] J. Hartmanis and L. A. Hemachandra. One-way Functions, Robustness and Non-isomorphism of NP-complete sets. Technical Report DCS TR86-796, Cornell University, 1987

[HMPST] A. Hajnal, W. Maass, P. Pudlak, M. Szegedy, and G. Turan. Threshold Circuits of Bounded Depth. *Proceedings 28th Annual IEEE Symposium on Foundation of computer science*, pages 99–110, 1987

[KKL] J. Kahn, G. Kalai, and N. Linial. The Influence of Variables on Boolean Function. *Proceedings of 29th Annual ACM Symposium on Theory of Computing*, pages 68–80, 1988.

[Kn] D. E. Knuth. *The Art of Computer Programming vol. Fundamental Algorithms.* Addison-Wesley, Reading, MA, 1973

[LMN] N. Linial, Y. Mansour and N. Nisan. Constant Depth Circuit, Fourier Transform and Learnability. *Proceedings of the 30th Annual IEEE Symposium on Foundations of Computer Science* pages 574–579, 1989

[MP] M. L. Minsky and S. Papert. *Perceptrons* Cambridge, MA, MIT Press, 1988. Original edition 1968.

[MS] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes.* North-Holland, 1977.

[Ra] A. A. Razborov. Lower Bounds for the the Size of Circuits of Bounded Depth with Basis $\wedge$, $\oplus$. *Math. Zametki* **41:4**,1987, 598-607 (in Russian). English translation *Math. Notes Acad. Sci. USSR* **41:4**, pages 333-338, 1987.

[Sm] R. Smolensky. Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity. *Proceedings of 19th Annual ACM Symposium on Theory of Computing*, pages 77–82, 1987.

[Ta] G. Tardos. Query Complexity, or Why Is It Difficult to Separate $NP^A \cap$ co-NP^A from P^A by a Random Oracle A? Manuscript, 1988.

[Ya] A. C.-C. Yao. Separating the Polynomial-time Hierarchy by Oracles. *Proceedings 26th Annual IEEE Symposium on Foundations of Computer Science*, pages 1–10, 1985.

A An Estimate of Fourier Coefficients of MOD_p Functions

In this appendix, we are going to prove Theorem 6, the proof utilizes the property of the Fourier Transform Matrix for symmetric functions defined in section 2.3.

Proof of Theorem 6: Let us first remark that the (s,t) entry $T^n_{s,t}$ of T^n, the nth order Fourier Transform Matrix for symmetric functions, is the coefficient of x^t in the generating function $(1-x)^s(1+x)^{n-s}$.

Another useful fact is about the relationship between Fourier coefficients of boolean functions on the ranges $\{1,-1\}$ and $\{0,1\}$. Let $f(\underline{x}) = \sum_{S\subseteq[n]} \hat{f}\chi_S$ be a boolean function from $\{-1,1\}^n$ to $\{-1,1\}$, and let $f'(\underline{x}) = \sum_{S\subseteq[n]} \hat{f}'_S\chi_S$ be the corresponding boolean function from $\{-1,1\}^n$ to $\{0,1\}$, where $\hat{f}_S$ and $\hat{f}'_S$ are the Fourier coefficients of χ_S respectively. Note that here 0 corresponds to 1, and 1 to -1. Then we have

$$\begin{aligned} \hat{f}_0 &= 1 - 2\hat{f}'_0 \\ \hat{f}_s &= -2\hat{f}'_s \qquad s \neq 0 \end{aligned}$$

Now we proceed with computation of the Fourier coefficients of $MOD_p(\underline{x})$ function, where p is a prime number bigger than 2. For notational simplicity, let $f(\underline{x})$ and $f'(\underline{x})$ be representations of $MOD_p(\underline{x})$ from $\{-1,1\}^n$ to $\{-1,1\}$ and from $\{-1,1\}^n$ to $\{0,1\}$ respectively.

Let ω be the pth primitive root of 1, *i.e.* $\omega = e^{\frac{2\pi i}{p}}$, by the Fourier Transformation for $f(\underline{x})$ and $f'(\underline{x})$, we have

$$\begin{aligned} \hat{f}_s &= \frac{1}{2^n}\sum_{k=0}^{\lfloor\frac{n}{p}\rfloor}(-T^n_{s,pk} + T^n_{s,pk+1} + \cdots + T^n_{s,pk+p-1}) \\ \hat{f}'_s &= \frac{1}{2^n}\sum_{k=0}^{\lfloor\frac{n}{p}\rfloor} T^n_{s,pk} \end{aligned}$$

Note also that the relation

$$\left.\begin{array}{l} \hat{f}_0 = 1 - 2\hat{f}'_0 \\ \hat{f}_s = -2\hat{f}'_s \quad s \neq 0 \end{array}\right. \quad \text{is equivalent to} \quad \left.\begin{array}{l} \sum_{t=0}^n T^n_{0,t} = 2^n \\ \sum_{t=0}^n T^n_{s,t} = 0 \quad s \neq 0 \end{array}\right\} \quad (**)$$

For $1 \leq l < p$, we have that $T^n_{s,t}$ is the coefficient of $(\omega^l)^t$ in $(1-\omega^l)^s(1+\omega^l)^{n-s}$. *i.e.*

$$(1-\omega^l)^s(1+\omega^l)^{n-s} = \sum_{t=0}^{n} T^m_{0,t}\omega^t$$

Let us use a^s_l denote the sum $\sum_{k=0}^{\lfloor \frac{n}{p} \rfloor} T^n_{s,pk+l}$, and use b^s_l denote the value $(1-\omega^l)^s(1+\omega^l)^{n-s}$, for $0 \leq s \leq n$ and $0 \leq l < p$. These relations together with $(*)$ above yield

$$\vec{b}^s = F(\omega)\vec{a}^s \quad \text{for } 0 \leq s \leq n$$

where $\vec{a}_s = [a^s_0, a^s_1, \ldots, a^s_{p-1}]^{\mathbf{t}}$, $\vec{b}_s = [b^s_0, b^s_1, \ldots, b^s_{p-1}]^{\mathbf{t}}$ and $F(\omega)$ is the $p \times p$ *Discrete Fourier Transform* matrix, defined as $F(\omega)_{ij} = \omega^{(i-1)(j-1)}$.

Since $F^{-1}(\omega) = \frac{1}{p} \cdot F(\omega^{-1})$, we can solve $\vec{a}^s$ in terms of $\vec{b}^s$. In particular, we have

$$\begin{aligned} a^0_0 &= \frac{1}{p}(b^0_0 + b^0_1 + \cdots + b^0_{p-1}) = \frac{1}{p}[2^n + \sum_{l=1}^{p-1}(1+\omega^l)^n] \\ a^s_0 &= \frac{1}{p}(b^s_0 + b^0_1 + \cdots + b^s_{p-1}) = \frac{1}{p}\sum_{l=1}^{p-1}(1-\omega^l)^s(1+\omega^l)^{n-s} \quad s \neq 0 \end{aligned}$$

Hence

$$\begin{aligned} \hat{f}_0 &= 1 - \frac{1}{2^{n-1}}a^0_0 = \frac{p-2}{p} - \frac{2}{p} \cdot \frac{1}{2^n}\sum_{l=1}^{p-1}(1+\omega^l)^n \\ \hat{f}_s &= -\frac{1}{2^{n-1}}a^s_0 = -\frac{2}{p} \cdot \frac{1}{2^n}\sum_{l=1}^{p-1}(1-\omega^l)^s(1+\omega^l)^{n-s} \quad s \neq 0 \end{aligned}$$

The final task we need to complete is to find an estimate of $|(1-\omega^l)^s(1+\omega^l)^n|$, for $1 \leq l \leq p-1$ and $0 \leq s \leq n$. First notice that, for $1 \leq l \leq p-1$,

$$(1-\omega^l)^p = \sum_{j=0}^{p}(-1)^j\binom{p}{j}\omega^{lj} = -\sum_{j=0}^{\frac{p-1}{2}}\binom{p}{j}(\omega^{lj} - \omega^{l(p-j)}) = -2(\sum_{j=1}^{\frac{p-1}{2}}\binom{p}{j}\sin\frac{2\pi jl}{p})i$$

Similarly,

$$(1+\omega^l)^p = \sum_{j=0}^{p}\binom{p}{j}\omega^{lj} = \sum_{j=0}^{\frac{p-1}{2}}\binom{p}{j}(\omega^{lj} + \omega^{l(p-j)}) = 2(\sum_{j=0}^{\frac{p-1}{2}}\binom{p}{j}\cos\frac{2\pi jl}{p})$$

Observe that, for $1 \leq l \leq p-1$,

$$|\sum_{j=1}^{\frac{p-1}{2}} \binom{p}{j} \sin \frac{2\pi jl}{p}| < \sum_{j=1}^{\frac{p-1}{2}} \binom{p}{j} < 2^{p-1}$$
$$|\sum_{j=0}^{\frac{p-1}{2}} \binom{p}{j} \cos \frac{2\pi jl}{p}| < \sum_{j=0}^{\frac{p-1}{2}} \binom{p}{j} = 2^{p-1}$$

Let α be the largest of these $2(p-1)$ numbers, then $\alpha < 2^{p-1}$.

Suppose $n = 2pn' + k$, $s = 2ps' + t$, where $0 \leq k < 2p$ and $0 \leq t < 2p$, then

$$\begin{aligned}
&|(1-\omega^l)^s(1+\omega^l)^{n-s}| \\
&\quad = |(1-\omega^l)^{2ps'}(1+\omega^l)^{2p(n'-s')}||(1-\omega^l)^t(1+\omega^l)^{k-t}| \\
&\quad = \left[-2(\sum_{j=1}^{\frac{p-1}{2}} \binom{p}{j} \sin \frac{2\pi jl}{p})i\right]^{2s'} \left[2(\sum_{j=0}^{\frac{p-1}{2}} \binom{p}{j} \cos \frac{2\pi jl}{p})\right]^{2(n'-s')} \beta \\
&\quad \leq 2^{2n'}\alpha^{2n'}\beta
\end{aligned}$$

where

$$\beta = \max_{\substack{1 \leq l \leq p-1 \\ 0 \leq t,k < 2p}} |(1-\omega^l)^t(1+\omega^l)^{k-t}|$$

Hence, for $0 \leq s \leq n$, we have

$$\begin{aligned}
\frac{1}{2^n}|\sum_{l=1}^{p-1}(1-\omega^l)^s(1+\omega^l)^{n-s}| &\leq (p-1)\beta 2^{-n+2n'(1+\log\alpha)} \\
&\leq (p-1)\beta 2^{-\epsilon n}
\end{aligned}$$

where $\epsilon = 1 - \frac{1}{p}(1+\log\alpha) > 0$, since $n' \leq \frac{n}{2p}$ and $\log\alpha < p-1$.

Therefore, we conclude that, for some $\epsilon > 0$,

$$\begin{aligned}
\hat{f}_0 &= \frac{p-2}{p} \pm O(\frac{1}{2^{\epsilon n}}) < \frac{p-1}{p} \\
|\hat{f}_s| &= O(\frac{1}{2^{\epsilon n}}) \quad \text{when } s \neq 0
\end{aligned}$$

■