Rational Functions Are Optimal for Sign-Representing Intersections

ALEXANDER A. SHERSTOV*

October 16, 2008

Abstract

The *threshold degree* of a function $f : X \to \{-1,+1\}$, $X \subset \mathbb{R}^n$, is the least degree of a polynomial ϕ with $f(x) \equiv \operatorname{sgn}\phi(x)$. This notion has numerous applications in complexity theory and learning theory. Analyzing the threshold degree is a challenge, with few techniques currently available.

We develop a novel technique for estimating the threshold degree for a broad and natural class of problems. Specifically, fix nonconstant functions $f : X \to \{-1,+1\}$ and $g : Y \to \{-1,+1\}$ on any finite or compact infinite sets $X, Y \subset \mathbb{R}^n$. We prove that the conjunction $f(x) \wedge g(y)$ has threshold degree $\Theta(d)$ if and only if there exist degree-$\Theta(d)$ rational functions $F(x)$ and $G(y)$ with $\sup_X |f - F| + \sup_Y |g - G| < 1$. The "if" part is simple and well-known, and our contribution is to prove its converse. Our results extend to conjunctions $f_1 \wedge f_2 \wedge \cdots \wedge f_k$ of any Boolean functions $f_1, f_2, \ldots, f_k$ and further to compositions $h(f_1, f_2, \ldots, f_k)$ for various h such as halfspaces and read-once formulas. As an application, we prove the conjecture of O'Donnell and Servedio (2003) that the intersection of two majorities has threshold degree $\Theta(\log n)$. We discuss several other applications in communication complexity and learning.

At the heart of our proof is a novel method for analyzing polynomials $\phi(x, y)$ that sign-represent a given function $f(x) \wedge g(y)$, whereby we recover rational approximants of f and g from any such ϕ. This recovery crucially uses LP duality and succeeds regardless of how tightly x and y are coupled inside ϕ. Also employed here are classical results on compact sets of inequalities and a new, optimal lower bound for the rational approximation of MAJORITY of any given order.

*Department of Computer Sciences, The University of Texas at Austin, TX 78712 USA. Email: `sherstov@cs.utexas.edu`.

1 Introduction

Representations of Boolean functions by real polynomials play an important role in theoretical computer science, with applications ranging from complexity theory to quantum computing and learning theory. The surveys in [5, 34, 9, 36] offer a glimpse into the diversity of these results and techniques. In this paper, we focus on one such representation scheme known as *sign-representation*. Specifically, fix a Boolean function $f : X \to \{-1,+1\}$ for some set $X \subset \mathbb{R}^n$, such as the hypercube $X = \{0,1\}^n$ or the unit sphere $X = \mathbb{S}^{n-1}$. The *threshold degree* of f, denoted $\deg_\pm(f)$, is the least degree of a polynomial $\phi(x_1,\dots,x_n)$ such that

$$f(x) = \operatorname{sgn} \phi(x)$$

for each $x \in X$. In other words, the threshold degree of f is the least degree of a real polynomial that represents f in sign.

The formal study of this complexity measure and of sign-representations in general began in 1968 with the seminal monograph of Minsky and Papert [23]. Among other things, the authors of [23] settled the threshold degree of several common functions. Since then, the notion of threshold degree has found a variety of applications in circuit complexity. Paturi and Saks [28] and later Siu et al. [39] used Boolean functions with high threshold degree to obtain size-depth trade-offs for threshold circuits. The well-known result, due to Beigel et al. [7], that PP is closed under intersection is also naturally interpreted in terms of threshold degree. In another development, Aspnes et al. [2] used the notion of threshold degree and its relaxations to obtain oracle separations for PP and to give an insightful new proof of classical lower bounds for AC^0. Krause and Pudlák [21, 22] used random restrictions to show that the threshold degree gives lower bounds on the *weight* and *density* of perceptrons and their generalizations, which are well-studied computational models.

Learning theory is another area in which the threshold degree of Boolean functions is of considerable interest. In this context, low threshold degree translates into efficient learnability [17, 16]. Specifically, functions with low threshold degree can be efficiently PAC learned under arbitrary distributions via linear programming. The current fastest algorithm for PAC learning polynomial-size DNF formulas, due to Klivans and Servedio [17], is an illustrative example: it is based precisely on an upper bound on the threshold degree of this concept class. Klivans et al. [16] showed that intersections of low-weight halfspaces also have low threshold degree, thereby giving an efficient PAC algorithm for this class as well.

Recently, the notion of threshold degree has also become a versatile tool in communication complexity. The starting point in this line of research appears to be the author's *Degree/Discrepancy Theorem* [35, 37], which states that any Boolean function with high threshold degree induces a communication problem with low *discrepancy* and thus high communication complexity in almost all models. We used this result in [35] to show the optimality of Allender's simulation of AC^0 by majority circuits, thus solving an open problem of Krause and Pudlák [21, §6]. In follow-up work [38], we combined the notion of threshold degree with other analytic techniques to settle the communication complexity of symmetric functions in the *unbounded-error* model, which is considerably more powerful than the models above. In a more recent application of the threshold degree, Razborov and Sherstov [32] proved the separation $\Sigma_2^{cc} \not\subseteq \mathsf{UPP}^{cc}$ in communication complexity, thereby solving a long-standing open problem due to Babai et al. [3].

Note. Closely related to threshold degree are two other complexity measures of a sign-representation: *density* (the number of distinct monomials in the sign-representing polynomial) and *weight* (the maximum magnitude of an integer coefficient). These notions have been extensively studied and have found various applications [6, 40, 22, 16, 18, 20, 10, 30, 31]. We will not have occasion to discuss them in this paper, however, except for a brief note in Section 8.

1.1 Our Results

In summary, the threshold degree plays a significant role in circuit complexity, learning theory, and communication complexity. Despite several studies of this complexity measure [23, 2, 26, 27], it remains poorly understood, and few general methods are available for bounding it from above or below. The main contribution of this paper is a novel technique for estimating the threshold degree for a broad and natural class of problems. Specifically, let $f_1, f_2, \ldots, f_k$ be given Boolean functions on compact sets $X_1, X_2, \ldots, X_k \subset \mathbb{R}^n$, respectively. We study Boolean functions on $X_1 \times X_2 \times \cdots \times X_k$ of the form

$$h(f_1, f_2, \ldots, f_k),$$

where $h : \{-1, +1\}^k \to \{-1, +1\}$ is a given combining function. An important special case is the composition $f_1 \wedge f_2 \wedge \cdots \wedge f_k$, in other words, the conjunction of the original functions, each on an independent domain.

To set the stage for our results, consider the special but illustrative case $k = 2$. Here we are given functions $f : X \to \{-1, +1\}$ and $g : Y \to \{-1, +1\}$ for some compact sets $X, Y \subset \mathbb{R}^n$ and would like to determine the threshold degree of their conjunction, $(f \wedge g)(x, y) = f(x) \wedge g(y)$. A simple and well-known method [7, 16] for sign-representing $f \wedge g$ is to use rational approximation. Specifically, let $p_1(x)/q_1(x)$ and $p_2(y)/q_2(y)$ be rational functions of degree d that approximate f and g, respectively, in the following sense:

$$\sup_{x\in X}\left|f(x)-\frac{p_1(x)}{q_1(x)}\right| + \sup_{y\in Y}\left|g(y)-\frac{p_2(y)}{q_2(y)}\right| < 1. \tag{1.1}$$

Then clearly,

$$\begin{aligned} f(x)\wedge g(y) &\equiv \operatorname{sgn}\{1+f(x)+g(y)\} \\ &\equiv \operatorname{sgn}\left\{1+\frac{p_1(x)}{q_1(x)}+\frac{p_2(y)}{q_2(y)}\right\}. \end{aligned} \tag{1.2}$$

Multiplying the expression in braces by the positive quantity $q_1(x)^2q_2(y)^2$ gives

$$\begin{aligned} f(x)\wedge g(y) \equiv \operatorname{sgn}\Big\{&q_1(x)^2q_2(y)^2 \\ &+ p_1(x)q_1(x)q_2(y)^2 + p_2(y)q_1(x)^2q_2(y)\Big\}, \end{aligned}$$

whence $\deg_{\pm}(f\wedge g)\leqslant 4d$. In summary, if f and g can be approximated as in (1.1) by rational functions of degree at most d, then the conjunction $f\wedge g$ has threshold degree at most $4d$.

The question that we ask in this paper is: does there exist a better construction? After all, given a sign-representing polynomial $\phi(x,y)$ for $f(x)\wedge g(y)$, there is no reason to expect that ϕ arises from the sum of two independent rational functions as in (1.2). Indeed, x and y can be tightly coupled inside $\phi(x,y)$ and can interact in complicated ways. Our main result is that, surprisingly, no such interactions can beat the simple construction above. In other words, the sign-representation based on rational functions always achieves the optimal degree (up to a small constant factor):

Theorem 1.1 (Main Theorem). *Let $f : X\to\{-1,+1\}$ and $g : Y\to\{-1,+1\}$ be given functions, where $X,Y\subset\mathbb{R}^n$ are arbitrary finite or compact infinite sets. Assume that f and g are not identically false. Let $d=\deg_{\pm}(f\wedge g)<\infty$. Then there exist degree-$4d$ rational functions*

$$\frac{p_1(x)}{q_1(x)},\quad \frac{p_2(y)}{q_2(y)}$$

that satisfy (1.1).

Via repeated applications of Theorem 1.1, we are able to obtain analogous results for conjunctions $f_1\wedge f_2\wedge\cdots\wedge f_k$ for any Boolean functions $f_1,f_2,\dots,f_k$ and any k. Our results further extend to compositions $h(f_1,\dots,f_k)$ for various h other than $h=$ AND, such as halfspaces and read-once AND/OR/NOT formulas. We defer a more detailed description of these extensions to Section 6, limiting this overview to the following representative special case.

Theorem 1.2 (Extension to multiple functions). *Let $f_1, f_2, \ldots, f_k$ be nonconstant Boolean functions on compact sets $X_1, X_2, \ldots, X_k \subset \mathbb{R}^n$, respectively. Let $h : \{-1,+1\}^k \to \{-1,+1\}$ be a halfspace or a read-once* AND/OR/NOT *formula. Assume that h depends on all of its k inputs and that the composition $h(f_1, f_2, \ldots, f_k)$ has threshold degree $d < \infty$. Then there is a degree-D rational function p_i/q_i on X_i, $i = 1, 2, \ldots, k$, such that*

$$\sum_{i=1}^{k} \sup_{x_i \in X_i} \left| f_i(x_i) - \frac{p_i(x_i)}{q_i(x_i)} \right| < 1,$$

where $D = 8d \log 2k$.

Theorem 1.2 is close to optimal. For example, when $h =$ AND, the upper bound on D is tight up to a factor of $\Theta(k \log k)$; for all h in the statement of the theorem, it is tight up to a factor of $\mathrm{poly}(k)$. See Remark 6.4 for details.

Theorems 1.1 and 1.2 contribute a powerful technique for proving lower bounds on the threshold degree, via rational approximation. Prior to this paper, it was a substantial challenge to analyze the threshold degree even for compositions of the form $f \wedge g$. Indeed, we are only aware of the work in [23, 26], where the threshold degree of $f \wedge g$ was studied for the special case $f = g =$ MAJORITY. The main difficulty in those previous works was analyzing the unintuitive interactions between f and g. Our results remove this difficulty completely, even in the general setting of compositions $h(f_1, f_2, \ldots, f_k)$ for arbitrary $f_1, f_2, \ldots, f_k$ and various combining functions h. Specifically, Theorems 1.1 and 1.2 make it possible to study the base functions $f_1, f_2, \ldots, f_k$ individually, in isolation. Once their rational approximability is understood, one immediately obtains lower bounds on the threshold degree of $h(f_1, f_2, \ldots, f_k)$.

We further note that Theorems 1.1 and 1.2 can be applied with quite limited information about the rational approximability of the functions involved. To illustrate, let $\mathrm{rdeg}_\varepsilon(h)$ stand for the least degree of a rational function that approximates h within ε in the uniform (supremum) norm. Then it follows from Theorem 1.1 that

$$\deg_\pm(f \wedge g) \geqslant \frac{1}{4} \max_{0<\varepsilon<1} \min \Big\{ \mathrm{rdeg}_\varepsilon(f), \mathrm{rdeg}_{1-\varepsilon}(g) \Big\}. \tag{1.3}$$

Thus, it suffices to estimate $\mathrm{rdeg}_\varepsilon(f)$ and $\mathrm{rdeg}_{1-\varepsilon}(g)$ for *some* value of ε in order to get a lower bound on the threshold degree of $f \wedge g$. If f and g happen to be the *same* Boolean function, then Theorem 1.1 shows that

$$\deg_\pm(f \wedge g) = \Theta(\mathrm{rdeg}_{1/2}(f)). \tag{1.4}$$

Here, one just needs to estimate the quantity $\mathrm{rdeg}_{1/2}(f)$. Analogous remarks apply to the case of multiple functions, in Theorem 1.2.

It is an interesting aspect of Theorems 1.1 and 1.2 that they apply to Boolean functions $f_1, f_2, \ldots, f_k$ on arbitrary finite or compact infinite sets $X_1, X_2, \ldots, X_k \subset \mathbb{R}^n$. This contrasts with several earlier results in the literature, such as the work of Nisan and Szegedy [25] and Beals et al. [4] that gives a polynomial relationship between the least degree required for representing a given function $f : \{0,1\}^n \to \{-1,+1\}$ exactly versus only approximating it pointwise within $1/3$. This equivalence critically depends on the structure of the full hypercube and need not hold for a Boolean function on a proper subset $X \subset \{0,1\}^n$. To illustrate, any exact representation of the majority function on the set $X = \{x \in \{0,1\}^n : \sum x_i \leqslant \frac{1}{3}n \text{ or } \sum x_i \geqslant \frac{2}{3}n\}$ requires degree $n/3$, whereas this function can be approximated on X to within $1/3$ by a polynomial of constant degree.

To illustrate Theorem 1.1, we study the well-known function on $\{0,1\}^n \times \{0,1\}^n$ given by $f(x,y) = \text{MAJORITY}(x_1, \ldots, x_n) \wedge \text{MAJORITY}(y_1, \ldots, y_n)$. Beigel et al. [7] showed that f has threshold degree at most $O(\log n)$. In fact, the same upper bound holds for the intersection of any constant number of majority functions. On the lower bound front, Minsky and Papert [23] proved that f has threshold degree $\omega(1)$. O'Donnell and Servedio [26] considerably improved this lower bound to $\Omega(\log n / \log\log n)$. They conjectured that the true answer is $\Theta(\log n)$, meaning that the construction of Beigel et al. is optimal. We settle this conjecture using Theorem 1.1:

Theorem 1.3 (Intersection of majorities). *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ be given by $f(x,y) = \text{MAJORITY}(x_1, \ldots, x_n) \wedge \text{MAJORITY}(y_1, \ldots, y_n)$. Then*

$$\deg_{\pm}(f) = \Omega(\log n).$$

Our techniques are completely different from those used in earlier work [23, 26].

To obtain Theorem 1.3 from Theorem 1.1, we determine, for any given ε, the least degree required for approximating MAJORITY by rational functions within ε. This result appears to be of independent interest:

Theorem 1.4 (Approximation of MAJORITY). *Let $\mathrm{MAJ}_n : \{0,1\}^n \to \{-1,+1\}$ denote the majority function. For $2^{-n} < \varepsilon < 1$, let $d(n,\varepsilon)$ be the least degree of a* *rational function that approximates* MAJ_n *pointwise within* ε. *Then*

$$d(n,\varepsilon) = \begin{cases} \Theta\left(\log\left\{\dfrac{2n}{\log(1/\varepsilon)}\right\}\cdot\log\dfrac{1}{\varepsilon}\right) & \text{if } 2^{-n} < \varepsilon < 1/2, \\[2ex] \Theta\left(\dfrac{\log n}{\log\{1/(1-\varepsilon)\}}\right) + 1 & \text{if } 1/2 \leqslant \varepsilon < 1. \end{cases}$$

We actually derive a more precise statement; see Theorem 7.3. The upper bound in Theorem 1.4 generalizes well-known earlier constructions [28, 7], all based on the seminal work of Newman [24]. The lower bound, on the other hand, has not been previously considered and requires technical novelty. Our solution is based on casting the rational approximation of MAJORITY as a linear program and constructing an explicit solution to its dual. We discuss our techniques in detail in Section 1.3.

1.2 Additional Applications and Discussion

Functions of the form $f_1 \wedge f_2 \wedge \cdots \wedge f_k$ have important applications in computational learning theory. It is a central open problem in the area to PAC learn the intersection of k halfspaces, i.e., Boolean functions of the form $f(x_1, x_2, \ldots, x_n) = \mathrm{sgn}(\sum a_i x_i - \theta)$ for fixed reals $a_1, a_2, \ldots, a_n, \theta$. This problem is wide open even for the case $k = 2$. One promising approach [16], as we discussed earlier, would be to show that intersections of k halfspaces (k small) have small threshold degree. Our work shows that this project is exactly equivalent to understanding the approximability of a *single* halfspace by rational functions. This has several consequences. On the upper bound side, it is reassuring to know that the rational approximation route is optimal and one need not look for more exotic ways to derive a sign-representation. Another reassuring consequence of our work is that the case $k = 2$ really is the hardest case to solve, in that a degree-d sign-representing polynomial for the intersection of two halfspaces immediately gives a degree-$(8dk \log k)$ sign-representing polynomial for the intersection of k halfspaces, for any k. (See Theorem 6.5 for a more general statement.) Thus, it suffices to bound the threshold degree for the case $k = 2$, by whatever means, in order to solve the general problem. No such implication was known prior to this paper.

On the lower bound side, much effort has been invested to show the hardness of learning intersections of halfspaces. Several conditional results have been obtained, both with restrictions on the output hypothesis [8, 1, 15] and without [19]. Unconditional lower bounds are available [20] in less powerful models, such as the SQ model. However, no unconditional, structural lower bounds are known for

PAC learning the intersection of $k = 2$ halfspaces. Our results pave the way to proving just such a lower bound: it suffices to prove that a halfspace cannot be approximated pointwise to within $1/3$ by a rational function of nontrivial degree.

Intersections of halfspaces also arise naturally in complexity theory. To illustrate, the result of Beigel et al. [7] that PP is closed under intersection can be viewed precisely as an upper bound on the threshold degree of the intersection of two halfspaces. Using similar techniques, one can prove that the analogue of this class in communication complexity, known as PP^{cc}, is also closed under intersection. These important advances contrast with our poor understanding of UPP^{cc}, a linear-algebraic counterpart of PP^{cc} with numerous applications to learning, circuits, communication, and matrix analysis [13, 14, 38, 32]. In particular, it is an important open problem to prove or disprove the closure of UPP^{cc} under intersection. In view of the recent techniques in [38, 32], a strong lower bound on the threshold degree of the intersection of two halfspaces would constitute substantial progress toward disproving the closure of UPP^{cc} under intersection. Theorems 1.1 and 1.2 in this paper greatly facilitate the former task.

1.3 Our Techniques

Minsky and Paper's proof [23] of an $\omega(1)$ lower bound on the threshold degree of $\text{MAJORITY}(x_1, \ldots, x_n) \wedge \text{MAJORITY}(y_1, \ldots, y_n)$ uses a compactness argument and basic properties of real univariate polynomials. O'Donnell and Servedio [26] strengthen this lower bound to a near-optimal $\Omega(\log n / \log\log n)$ with a substantially more sophisticated proof. Specifically, the authors of [26] construct an explicit solution to the dual problem, i.e., an explicit probability distribution on the Boolean hypercube (an *orthogonalizing* distribution, as it is sometimes called) with respect to which the given function has zero correlation with all low-degree monomials.

Our proof of Theorem 1.1 introduces a novel technique that is completely unrelated to the above two approaches. In some sense, our strategy is the exact opposite of the approach of O'Donnell and Servedio. Namely, we work directly with the original, primal problem to show its infeasibility. To carry out this program, we also end up appealing to linear-programming duality but for a smaller and much different intermediate problem. The problem in question is that of finding, in the positive spans of two given matrices, two vectors whose corresponding entries have comparable magnitude. By an analytic argument, we are able to prove that this intermediate problem is precisely the missing link between sign-representation and rational approximation.

Intuitively, our proof decomposes any sign-representation $\phi(x, y)$ of the function $f(x) \wedge g(y)$ into individual rational approximants for f and g. The intermediate problem about positive spans is the tool that enables this decomposition of $\phi(x, y)$, regardless of how tightly the x and y parts are coupled inside ϕ. In this way, we complete the proof without ever constructing an orthogonalizing distribution, explicitly or implicitly. It is this feature of our proof that allows us to handle conjunctions of *arbitrary* functions. Indeed, constructing an orthogonalizing distribution—already a considerable challenge in the seemingly well-structured case [26] of two majority functions—would probably be altogether unrealistic for subtler examples.

The above solution refers to functions on *finite* sets in $\mathbb{R}^n$. To generalize it to arbitrary compact sets, we use a classical result [33] from approximation theory and the theory of linear inequalities. This result states that any infeasible compact set of strict inequalities has a finite infeasible subset. By suitably formulating sign-representation and rational approximation as compact sets of such inequalities, we find ourselves in the familiar finite case which we have already solved.

We conclude by outlining our techniques in Theorem 1.4 on the rational approximation of MAJORITY. While the upper bound in this theorem extends well-known previous constructions [28, 7], the lower bound is novel to the best of our knowledge. Here, the closest previous line of research concerns *continuous* approximation of the sign function on $[-1, -\varepsilon] \cup [\varepsilon, 1]$, which unfortunately gives no insight into the discrete case. For example, the lower bound derived by Newman [24] in the continuous setting is based on the integration of relevant rational functions with respect to a suitable measure, which has no meaningful discrete analogue. We obtain our discrete lower bounds in a quite different way, by reformulating the discrete case as a linear program and solving its dual from scratch.

1.4 Organization

After a review of necessary preliminaries in Section 2, we start our proof in Section 3 with a number of observations and intermediate results. A complete solution for functions of the form $f_1 \wedge f_2 \wedge \cdots \wedge f_k$ on finite sets is given in Section 4. We generalize this result in Section 5 to conjunctions $f_1 \wedge f_2 \wedge \cdots \wedge f_k$ on arbitrary compact sets. Section 6 further generalizes these results to compositions $h(f_1, f_2, \ldots, f_k)$ for various h as well as states a number of other observations.

In the concluding part of the paper, we illustrate our technique by settling the threshold degree of the intersection of two majorities, stated above as Theorem 1.3. Specifically, in Section 7, we prove Theorem 1.4 on the least error achievable by rational functions of any given degree in approximating MAJORITY. The sought Theorem 1.3 is then settled in Section 8, along with an optimal lower bound on the threshold *density* of the function in question.

2 Preliminaries

Throughout this work, the symbol t refers to a real variable, whereas u, v, w, x, y, z refer to vectors in $\mathbb{R}^n$ and in particular $\{0,1\}^n$. We adopt the following standard definition of the sign function:

$$\operatorname{sgn} t = \begin{cases} -1 & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ 1 & \text{if } t > 0. \end{cases}$$

Equations and inequalities involving vectors in $\mathbb{R}^n$, such as $x < y$ or $x \geqslant 0$, must be interpreted component-wise, as usual. We let "log" stand for the logarithm to base 2.

We view Boolean functions as mappings $X \to \{-1,+1\}$, where X is an arbitrary set and -1 corresponds to the logical "true." Given a function $f : X \to \{-1,+1\}$ and a subset $A \subseteq X$, we let $f|_A$ denote the restriction of f to A. In other words, the function $f|_A : A \to \{-1,+1\}$ is given by $f_A(x) = f(x)$.

We say that a set $X \subseteq \mathbb{R}^n$ is *closed under negation* if $x \in X \Leftrightarrow -x \in X$. Given a function $f : X \to \{-1,+1\}$, where $X \subseteq \mathbb{R}^n$, we say that f is *odd* if X is closed under negation and $f(-x) = -f(x)$ for all $x \in X$.

By the *degree* of a multivariate polynomial p on $\mathbb{R}^n$, we shall always mean the total degree of p, i.e., the greatest total degree of any monomial of p. Given a function $f : X \to \{-1,+1\}$, where $X \subseteq \mathbb{R}^n$ is an arbitrary set, the *threshold degree* $\deg_\pm(f)$ of f is defined as the least degree of a multivariate polynomial p such that $f(x)p(x) > 0$ for all $x \in X$. In words, the threshold degree of f is the least degree of a polynomial that represents f in sign. If no such polynomial p can be found, we put $\deg_\pm(f) = \infty$. Observe that if X is finite, then $\deg_\pm(f)$ is finite as well. Threshold degree is also known in the literature as "strong degree" [2], "voting polynomial degree" [21], "PTF degree" [27], and "sign degree" [10].

Given functions $f : X \to \{-1,+1\}$ and $g : Y \to \{-1,+1\}$, recall that the function $f \wedge g : X \times Y \to \{-1,+1\}$ is given by $(f \wedge g)(x, y) = f(x) \wedge g(y)$. The function $f \vee g$ is defined analogously. Observe that in this notation, $f \wedge f$ and f are completely different functions, the former having domain $X \times X$ and the latter X. These conventions extend in the obvious way to any number of functions. For example, $f_1 \wedge f_2 \wedge \cdots \wedge f_k$ is a Boolean function with domain $X_1 \times X_2 \times \cdots \times X_k$, where X_i is the domain of f_i. Generalizing further, we let the symbol $h(f_1, \ldots, f_k)$ denote the Boolean function on $X_1 \times X_2 \times \cdots \times X_k$ obtained by composing a given function $h : \{-1,+1\}^k \to \{-1,+1\}$ with the functions $f_1, f_2, \ldots, f_k$. Finally, recall that the negated function $\overline{f} : X \to \{-1,+1\}$ is given by $\overline{f}(x) = -f(x)$.

The symbol P_k stands for the family of all univariate real polynomials of degree up to k. The following combinatorial identity is well-known.

Fact 2.1. *For every integer $n \geqslant 1$ and every polynomial $p \in P_{n-1}$,*

$$\sum_{i=0}^{n} \binom{n}{i} (-1)^i p(i) = 0.$$

This fact can be verified by repeated differentiation of the real function

$$(t-1)^n = \sum_{i=0}^{n} \binom{n}{i} (-1)^{n-i} t^i$$

at $t = 1$, as explained in [26].

2.1 Symmetric Functions

Let S_n denote the symmetric group on n elements. For $\sigma \in S_n$ and $x \in \{0,1\}^n$, we let σx denote the string $(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \in \{0,1\}^n$. A function $\phi : \{0,1\}^n \to \mathbb{R}$ is called *symmetric* if $\phi(x) = \phi(\sigma x)$ for every $x \in \{0,1\}^n$ and every $\sigma \in S_n$. Equivalently, ϕ is symmetric if $\phi(x)$ is uniquely determined by $\sum x_i$. Observe that for every $\phi : \{0,1\}^n \to \mathbb{R}$ (symmetric or not), the derived function

$$\phi_{\text{sym}}(x) = \mathop{\mathbf{E}}_{\sigma \in S_n} [\phi(\sigma x)]$$

is symmetric. Symmetric functions on $\{0,1\}^n$ are intimately related to univariate polynomials, as demonstrated by Minsky and Papert's well-known *symmetrization argument* [23]:

Proposition 2.2 (Minsky & Papert). *Let $\phi : \{0,1\}^n \to \mathbb{R}$ be a multivariate polynomial of degree d. Then there is $p \in P_d$ such that for all $x \in \{0,1\}^n$,*

$$\mathop{\mathbf{E}}_{\sigma \in S_n} [\phi(\sigma x)] = p\left(\sum x_i\right).$$

An important symmetric function is the *majority* function on $\{0,1\}^n$, given by

$$\text{MAJ}_n(x) = \begin{cases} 1 & \text{if } \sum x_i > n/2, \\ -1 & \text{otherwise.} \end{cases}$$

2.2 Rational Approximation

Consider a function $f : X \to \{-1, +1\}$, where $X \subseteq \mathbb{R}^n$ is an arbitrary set. For $d \geqslant 0$, we define

$$R(d, f) = \inf_{p,q} \sup_{x \in X} \left| f(x) - \frac{p(x)}{q(x)} \right|,$$

where the infimum is over multivariate polynomials p and q of degree up to d such that q does not vanish on X. In words, $R(d, f)$ is the least error in an approximation of f by a multivariate rational function of degree up to d. We will also take an interest in the related quantity

$$R^+(d, f) = \inf_{p,q} \sup_{x \in X} \left| f(x) - \frac{p(x)}{q(x)} \right|,$$

where the infimum is over multivariate polynomials p and q of degree up to d such that q is positive on X. These two quantities are related in a straightforward way:

$$R^+(2d, f) \leqslant R(d, f) \leqslant R^+(d, f). \tag{2.1}$$

The second inequality here is trivial. The first follows from the fact that every rational approximant $p(x)/q(x)$ of degree d gives rise to a degree-$2d$ rational approximant with the same error and a positive denominator, namely, $\{p(x)q(x)\}/q(x)^2$.

The infimum in the definitions of $R(d, f)$ and $R^+(d, f)$ cannot in general be replaced by a minimum [33], even when X is finite subset of $\mathbb{R}$. This is in contrast with the more familiar setting of polynomials, where least-error approximants are guaranteed to exist in any normed linear space.

3 Auxiliary Results

In this section, we prove a number of auxiliary facts about rational approximation and sign-representation. This preparatory work will set the stage for the proofs of our main results in later sections.

We start by spelling out the exact relationship between the rational approximation and sign-representation of a Boolean function.

Theorem 3.1. *Let $f : X \to \{-1, +1\}$ be a given function, where $X \subset \mathbb{R}^n$ is compact. Then for every integer d,*

$$\deg_{\pm}(f) \leqslant d \quad \Leftrightarrow \quad R^+(d, f) < 1.$$

Proof. For the forward implication, let p be a polynomial of degree at most d such that $f(x)p(x) > 0$ for every $x \in X$. Letting $M = \max_{x\in X} |p(x)|$ and $m = \min_{x\in X} |p(x)|$, we have

$$R^+(d, f) \leqslant \sup_{x\in X} \left| f(x) - \frac{p(x)}{M} \right| \leqslant 1 - \frac{m}{M} < 1.$$

For the converse, fix a degree-d rational function $p(x)/q(x)$ such that $q(x) > 0$ for $x \in X$ and

$$\sup_{x\in X} \left| f(x) - \frac{p(x)}{q(x)} \right| < 1.$$

Then clearly $f(x)p(x) > 0$ for all $x \in X$. $\square$

Our next observation amounts to reformulating the rational approximation of Boolean functions in a way that is more analytically pleasing.

Theorem 3.2. *Let $f : X \to \{-1, +1\}$ be a given function, where $X \subset \mathbb{R}^n$ is compact and $\deg_\pm(f) < \infty$. Then for every integer $d \geqslant \deg_\pm(f)$, one has*

$$R^+(d, f) = \inf_{c\geqslant 1} \frac{c^2 - 1}{c^2 + 1},$$

where the infimum is over all $c \geqslant 1$ for which there exist polynomials p, q of degree up to d such that

$$0 < \frac{1}{c} q(x) \leqslant f(x)p(x) \leqslant cq(x), \qquad x \in X. \tag{3.1}$$

Observe that the infimum in the statement of the lemma is over a nonempty set. For example, one could let p be any degree-d polynomial with $f(x)p(x) > 0$ for $x \in X$, and then put $q = 1$ and $c = \max_{x\in X}\{|p(x)| + |1/p(x)|\}$.

Proof of Theorem 3.2. Let p, q be arbitrary polynomials of degree up to d that obey (3.1) for some $c \geqslant 1$. Such polynomials do exist, by the remark preceding this proof. Then

$$R^+(d, f) \leqslant \sup_{x\in X} \left| f(x) - \frac{2c}{c^2 + 1} \cdot \frac{p(x)}{q(x)} \right| \leqslant \frac{c^2 - 1}{c^2 + 1}.$$

For the converse, we first note that $R^+(d, f) < 1$ by the construction just given. Fix $\delta > 0$ small enough that $R^+(d, f) + \delta < 1$. Take polynomials $\tilde{p}, \tilde{q}$ of degree up to d such that $\tilde{q}$ is positive on X and

$$\sup_{x \in X} \left| f(x) - \frac{\tilde{p}(x)}{\tilde{q}(x)} \right| < R^+(d, f) + \delta.$$

Letting ε stand for the right member of this inequality, we find that (3.1) holds with

$$p = \frac{1}{\sqrt{1 - \varepsilon^2}}\, \tilde{p}, \quad q = \tilde{q}, \quad c = \sqrt{\frac{1 + \varepsilon}{1 - \varepsilon}}.$$

Hence,

$$R^+(d, f) + \delta = \varepsilon = \frac{c^2 - 1}{c^2 + 1} \geqslant \inf_{c \geqslant 1} \frac{c^2 - 1}{c^2 + 1},$$

where the infimum is as in the statement of the lemma. Letting $\delta \to 0$ completes the proof. $\square$

Our next result shows that if a degree-d rational approximant achieves error ε in approximating a given Boolean function, then a degree-$2d$ approximant can achieve error as small as ε^2. The exact constants in this result are important for us. *Asymptotic* dependencies of this flavor can be found, for example, in earlier work by Beigel et al. [7, §3].

Theorem 3.3. *Let $f : X \to \{-1, +1\}$ be a given function, where $X \subseteq \mathbb{R}^n$. Let d be a given integer. Then*

$$R^+(2d, f) \leqslant \left(\frac{\varepsilon}{1 + \sqrt{1 - \varepsilon^2}} \right)^2,$$

where $\varepsilon = R(d, f)$.

Proof. The theorem is clearly true for $\varepsilon = 1$. For $0 \leqslant \varepsilon < 1$, consider the univariate rational function

$$S(t) = \frac{4\sqrt{1 - \varepsilon^2}}{1 + \sqrt{1 - \varepsilon^2}} \cdot \frac{t}{t^2 + (1 - \varepsilon^2)}.$$

Calculus shows that

$$\max_{1 - \varepsilon \leqslant |t| \leqslant 1 + \varepsilon} |\operatorname{sgn} t - S(t)| = \left(\frac{\varepsilon}{1 + \sqrt{1 - \varepsilon^2}} \right)^2.$$

Fix a sequence $A_1, A_2, \ldots, A_m, \ldots$ of rational functions of degree at most d such that $\sup_{x \in X} |f(x) - A_m(x)| \to \varepsilon$ as $m \to \infty$. Then

$$\lim_{m \to \infty} \sup_{x \in X} |f(x) - S(A_m(x))| \leqslant \left(\frac{\varepsilon}{1 + \sqrt{1 - \varepsilon^2}} \right)^2.$$

It remains to note that each $S(A_m(x))$ is a rational function of degree at most $2d$ with a positive denominator. □

Corollary 3.4. *Let $f : X \to \{-1, +1\}$ be a given function, where $X \subseteq \mathbb{R}^n$. Then for all integers $d \geqslant 1$ and reals $t \geqslant 1$,*

$$R^+(td, f) \leqslant R(d, f)^{t/2}.$$

Proof. If $t = 2^k$ for some integer k, then repeated applications of Theorem 3.3 yield

$$R^+(2^k d, f) \leqslant R(2^{k-1} d, f)^2 \leqslant \cdots \leqslant R(d, f)^{2^k}.$$

The general case follows because the interval $[t/2, t]$ contains a power of 2. □

Recall that a key challenge in this work will be, given a sign-representation $\phi(x, y)$ of a composite function $f(x) \wedge g(y)$, to suitably break down ϕ and recover individual rational approximants of f and g. We now present an ingredient of our solution, namely, a certain fact about pairs of matrices based on Farkas' Lemma. For the time being, we will formulate this fact in a clean and abstract way.

Theorem 3.5. *Fix matrices $A, B \in \mathbb{R}^{m \times n}$ and a real $c \geqslant 1$. Consider the following system of linear inequalities in $u, v \in \mathbb{R}^n$:*

$$\left.\begin{aligned} \frac{1}{c} Au \leqslant &Bv \leqslant cAu, \\ &u \geqslant 0, \\ &v \geqslant 0. \end{aligned}\right\} \qquad (3.2)$$

If $u = v = 0$ is the only solution to (3.2), *then there exist vectors $w \geqslant 0$ and $z \geqslant 0$ such that*

$$w^\mathsf{T} A + z^\mathsf{T} B > c(z^\mathsf{T} A + w^\mathsf{T} B).$$

Proof. If $u = v = 0$ is the only solution to (3.2), then linear-programming duality implies the existence of vectors $w \geqslant 0$ and $z \geqslant 0$ such that $w^\mathsf{T} A > cz^\mathsf{T} A$ and $z^\mathsf{T} B > cw^\mathsf{T} B$. Adding the last two inequalities completes the proof. □

We close this section with a natural topological property of Boolean functions on compact sets. This property will be key to extending our main results from finite sets to arbitrary compact sets.

Theorem 3.6. *Let $f : X \to \{-1, +1\}$ be a given function, where $X \subset \mathbb{R}^n$ is compact. If* $\deg_\pm(f) < \infty$, *then the sets $f^{-1}(-1)$ and $f^{-1}(1)$ are compact.*

Proof. By symmetry, it suffices to prove the claim for $f^{-1}(-1)$. Since $f^{-1}(-1) \subseteq X$ is bounded, it remains to verify its closure. Let $x_1, x_2, \ldots, x_m, \ldots$ be a sequence in $f^{-1}(-1)$ such that $x_m \to x^*$ as $m \to \infty$. By hypothesis, there exists a polynomial p such that $f(x)p(x) > 0$ for all $x \in X$. By continuity, we have $p(x^*) \leqslant 0$, which means that $x^* \notin f^{-1}(1)$. Since X is closed, we have $x^* \in X$ and therefore $x^* \in X \setminus f^{-1}(1) = f^{-1}(-1)$. $\square$

4 Solution for the Finite Case

In this section, we prove our main results on conjunctions of Boolean functions on finite sets. For clarity of exposition, we first settle the case of *two* Boolean functions, at least one of which is *odd.* While this case seems restricted, we will see that it captures the full complexity of the problem.

Theorem 4.1. *Let $f : X \to \{-1, +1\}$ and $g : Y \to \{-1, +1\}$ be given functions, where $X, Y \subset \mathbb{R}^n$ are arbitrary finite sets. Assume that $f \not\equiv 1$ and $g \not\equiv 1$. Let $d = \deg_\pm(f \wedge g)$. If f is odd, then*

$$R^+(2d, f) + R^+(d, g) < 1.$$

Proof. We first collect some basic observations. Since $f \not\equiv 1$ and $g \not\equiv 1$, we have $\deg_\pm(f) \leqslant d$ and $\deg_\pm(g) \leqslant d$. Therefore, Theorem 3.1 implies that

$$R^+(d, f) < 1, \qquad R^+(d, g) < 1. \tag{4.1}$$

In particular, the theorem holds if $R^+(d, g) = 0$. In the remainder of the proof, we assume that $R^+(d, g) = \varepsilon$, where $0 < \varepsilon < 1$.

By hypothesis, there exists a degree-d polynomial ϕ such that $f(x) \wedge g(y) = \operatorname{sgn} \phi(x, y)$ for all $x \in X$, $y \in Y$. Define

$$X^- = \{x \in X : f(x) = -1\}.$$

Since X is closed under negation and f is odd, we have $f(x) = 1 \Leftrightarrow -x \in X^-$. We will make several uses of this fact in what follows, without further mention.

Put

$$c = \sqrt{\frac{1 + (1 - \delta)\varepsilon}{1 - (1 - \delta)\varepsilon}},$$

where $\delta \in (0, 1)$ is sufficiently small. Since $R^+(d, g) > (c^2 - 1)/(c^2 + 1)$, we know by Theorem 3.2 that there cannot exist polynomials p, q of degree up to d such that

$$0 < \frac{1}{c} q(y) \leqslant g(y)p(y) \leqslant cq(y), \qquad y \in Y. \tag{4.2}$$

We claim, then, that there cannot exist reals $a_x \geqslant 0$, $x \in X$, not all zero, such that

$$\frac{1}{c} \sum_{x \in X^-} a_{-x}\phi(-x, y) \leqslant g(y) \sum_{x \in X^-} a_x \phi(x, y) \leqslant c \sum_{x \in X^-} a_{-x}\phi(-x, y), \qquad y \in Y.$$

Indeed, if such reals a_x were to exist, then (4.2) would hold for the polynomials $p(y) = \sum_{x \in X^-} a_x \phi(x, y)$ and $q(y) = \sum_{x \in X^-} a_{-x}\phi(-x, y)$. In view of the nonexistence of the a_x, Theorem 3.5 applies to the matrices

$$\Big[\phi(-x, y)\Big]_{y \in Y,\, x \in X^-}, \qquad \Big[g(y)\phi(x, y)\Big]_{y \in Y,\, x \in X^-}$$

and guarantees the existence of nonnegative reals λ_y, μ_y for $y \in Y$ such that

$$\begin{aligned} &\sum_{y \in Y} \lambda_y \phi(-x, y) + \sum_{y \in Y} \mu_y g(y)\phi(x, y) \\ &\qquad > c \left(\sum_{y \in Y} \mu_y \phi(-x, y) + \sum_{y \in Y} \lambda_y g(y)\phi(x, y) \right), \qquad x \in X^-. \end{aligned} \tag{4.3}$$

Define polynomials α, β on X by

$$\begin{aligned} \alpha(x) &= \sum_{y \in g^{-1}(-1)} \{\lambda_y \phi(-x, y) - \mu_y \phi(x, y)\}, \\ \beta(x) &= \sum_{y \in g^{-1}(1)} \{\lambda_y \phi(-x, y) + \mu_y \phi(x, y)\}. \end{aligned}$$

Then (4.3) can be restated as

$$\alpha(x) + \beta(x) > c\{-\alpha(-x) + \beta(-x)\}, \qquad x \in X^-.$$

Both members of this inequality are nonnegative, and thus $\{\alpha(x) + \beta(x)\}^2 > c^2\{-\alpha(-x) + \beta(-x)\}^2$ for $x \in X^-$. Since in addition $\alpha(-x) \leqslant 0$ and $\beta(-x) \geqslant 0$ for $x \in X^-$, we have

$$\{\alpha(x) + \beta(x)\}^2 > c^2\{\alpha(-x) + \beta(-x)\}^2, \qquad x \in X^-.$$

Letting $\gamma(x) = \{\alpha(x) + \beta(x)\}^2$, we see that

$$R^+(2d, f) \leqslant \sup_{x \in X} \left| f(x) - \frac{c^2+1}{c^2} \cdot \frac{\gamma(-x) - \gamma(x)}{\gamma(-x) + \gamma(x)} \right| \leqslant \frac{1}{c^2} < 1 - \varepsilon,$$

where the final inequality holds for all $\delta \in (0,1)$ small enough. $\square$

Remark 4.2. In Theorem 4.1 and elsewhere in this paper, the degree of a multivariate polynomial $p(x_1, x_2, \dots, x_n)$ is defined as the greatest total degree of any monomial of p. A related notion is the *partial degree* of p, which is the maximum degree of p in any one of the variables $x_1, x_2, \dots, x_n$. One readily sees that the proof of Theorem 4.1 applies unchanged to this alternate notion. Specifically, if the conjunction $f(x) \wedge g(y)$ can be sign-represented by a polynomial of partial degree d, then there exist rational functions $F(x)$ and $G(y)$ of partial degree $2d$ such that $\|f - G\|_\infty + \|g - F\|_\infty < 1$. In the same way, the entire program of this paper carries over, with cosmetic changes, to the notion of partial degree. Analogously, our proofs apply to hybrid definitions of degree, such as partial degree over blocks of variables. Other, more abstract notions of degree can also be handled. In the remainder of the paper, we will maintain our focus on total degree and will not further elaborate on its generalizations.

As promised, we will now remove the assumption, made in Theorem 4.1, about one of the functions being odd.

Theorem 4.3. *Let $f : X \to \{-1, +1\}$ and $g : Y \to \{-1, +1\}$ be given functions, where $X, Y \subset \mathbb{R}^n$ are arbitrary finite sets. Assume that $f \not\equiv 1$ and $g \not\equiv 1$. Let $d = \deg_\pm(f \wedge g)$. Then*

$$R^+(4d, f) + R^+(2d, g) < 1 \tag{4.4}$$

and, by symmetry,

$$R^+(2d, f) + R^+(4d, g) < 1.$$

Proof. It suffices to prove (4.4). Define $X' \subset \mathbb{R}^{n+1}$ by

$$X' = \{(x, 1), (-x, -1) : x \in X\}.$$

It is clear that X' is closed under negation. Let $f' : X' \to \{-1, +1\}$ be the odd Boolean function given by

$$f'(x, b) = \begin{cases} f(x) & \text{if } b = 1, \\ -f(-x) & \text{if } b = -1. \end{cases}$$

Let ϕ be a degree-d polynomial such that $f(x) \wedge g(y) \equiv \operatorname{sgn} \phi(x, y)$. Fix an input $\tilde{x} \in X$ such that $f(\tilde{x}) = -1$. Then

$$f'(x, b) \wedge g(y) \equiv \operatorname{sgn} \{K(1 + b)\phi(x, y) + \phi(-x, y)\phi(\tilde{x}, y)\}$$

for a sufficiently large constant $K \gg 1$, whence

$$\deg_{\pm}(f' \wedge g) \leqslant 2d.$$

Applying Theorem 4.1 to $f' \wedge g$ yields

$$R^+(4d, f') + R^+(2d, g) < 1.$$

Since $R^+(4d, f) \leqslant R^+(4d, f')$ by definition, the proof is complete. □

Finally, we obtain an analogue of Theorem 4.3 for $k \geqslant 3$ functions.

Theorem 4.4. *Let $f_1, f_2, \ldots, f_k$ be given Boolean functions on finite sets $X_1, X_2, \ldots, X_k \subset \mathbb{R}^n$, respectively. Assume that $f_i \not\equiv 1$ for $i = 1, 2, \ldots, k$. Let $d = \deg_{\pm}(f_1 \wedge f_2 \wedge \cdots \wedge f_k)$. Then*

$$\sum_{i=1}^{k} R^+(D, f_i) < 1$$

for $D = 8d \log 2k$.

Proof. Since $f_1, f_2, \ldots, f_k \not\equiv 1$, it follows that for each pair of indices $i < j$, the function $f_i \wedge f_j$ is a subfunction of $f_1 \wedge f_2 \wedge \cdots \wedge f_k$. Theorem 4.3 now shows that for each $i < j$,

$$R^+(4d, f_i) + R^+(4d, f_j) < 1. \tag{4.5}$$

Without loss of generality, $R^+(4d, f_1) = \max_{i=1,\ldots,k} R^+(4d, f_i)$. Abbreviate $\varepsilon = R^+(4d, f_1)$. By (4.5),

$$R^+(4d, f_i) < \min\left\{1 - \varepsilon, \frac{1}{2}\right\}, \qquad i = 2, 3, \ldots, k.$$

Now Corollary 3.4 implies that

$$\sum_{i=1}^{k} R^+(D, f_i) \leqslant \varepsilon + \sum_{i=2}^{k} R^+(4d, f_i)^{1+\log k} < 1.$$ □

5 Extension to Functions on Compact Sets

Typically, one studies Boolean functions on finite sets. There are cases, however, when Boolean functions naturally arise on infinite sets, such as the unit sphere $\mathbb{S}^{n-1}$ in learning theory. The purpose of this section is to generalize the results of Section 4 from finite sets to arbitrary compact sets. We do so by applying compactness and duality arguments to our work in the finite case. We will treat the results of the previous section in a black-box manner and will not need to reinspect their proofs.

5.1 Preliminaries

We will be concerned with systems of linear inequalities of the form

$$\langle u, x\rangle > 0, \qquad u \in S, \tag{5.1}$$

for some $S \subseteq \mathbb{R}^n$. We call S the *characteristic set* of (5.1). The *convex hull* of a set $S \subseteq \mathbb{R}^n$, denoted $\operatorname{conv} S$, is the smallest convex set in $\mathbb{R}^n$ that contains S. Constructively, the convex hull of S is the set of all finite convex combinations of vectors in S, where by a convex combination of vectors is meant a linear combination with positive coefficients that sum to 1.

An important ingredient in this section is the following classical result from the theory of linear inequalities; see Cheney [11, §1.5] for an elegant proof.

Fact 5.1. *Let $S \subset \mathbb{R}^n$ be compact. Then $0 \in \operatorname{conv} S$ if and only if there does not exist $x \in \mathbb{R}^n$ such that $\langle u, x\rangle > 0$ for all $u \in S$.*

Corollary 5.2. *Let $S \subset \mathbb{R}^n$ be compact. Suppose that the system $\langle u, x\rangle > 0$, $u \in S$, has no solution in $x \in \mathbb{R}^n$. Then there exists a finite subset $A \subseteq S$ such that the system $\langle u, x\rangle > 0$, $u \in A$, has no solution in $x \in \mathbb{R}^n$.*

Proof. By Fact 5.1, there is a finite subset $A \subseteq S$ such that $0 \in \operatorname{conv} A$. Again by Fact 5.1, the system $\langle u, x\rangle > 0$, $u \in A$, has no solution. $\square$

5.2 Approximation on Compact vs. Finite Sets

Our strategy will be to reduce the problem of rational approximation on a compact set X to the analogous problem on a finite subset of X. More precisely, we will show that if a Boolean function f is difficult to approximate on a given compact set X, then f must be just as difficult to approximate on some finite subset $A \subseteq X$. This general phenomenon arises frequently in approximation theory [11, 33].

We start by proving an analogous result for sign-representation, which arises in this development as a degenerate case of rational approximation. The result on sign-representation is of independent interest, however, and we state it on its own.

Theorem 5.3 (Sign-representation on compact vs. finite sets). *Let a function $f : X \to \{-1, +1\}$ be given, where $X \subset \mathbb{R}^n$ is compact and $\deg_\pm(f) < \infty$. Then there exists a finite subset $A \subseteq X$ such that*

$$\deg_\pm(f|_A) = \deg_\pm(f).$$

Proof. Put $d = \deg_\pm(f) - 1$. Let D denote the set of all n-tuples $(d_1, \ldots, d_n)$ of nonnegative integers $d_1, \ldots, d_n$ with $\sum d_i \leqslant d$. Consider the following system of inequalities in the real variables $a_{d_1,\ldots,d_n}$ for $(d_1, \ldots, d_n) \in D$:

$$f(x) \sum_D a_{d_1,\ldots,d_n} x_1^{d_1} x_2^{d_2} \cdots x_n^{d_n} > 0, \qquad x \in X. \tag{5.2}$$

Since $d < \deg_\pm(f)$, this system has no solution. Theorem 3.6 implies that the sets $f^{-1}(-1)$ and $f^{-1}(1)$ are compact, which means that the characteristic set of (5.2) is a compact subset of $\mathbb{R}^{|D|}$. In view of the last two observations, Corollary 5.2 implies that some finite subset of the inequalities (5.2) has no solution. In other words, there is a finite subset $A \subseteq X$ with $\deg_\pm(f|_A) \geqslant d + 1$. Since $\deg_\pm(f|_A) \leqslant \deg_\pm(f) = d + 1$, the proof is complete. $\square$

Remark 5.4. The hypothesis that $\deg_\pm(f) < \infty$ in Theorem 5.3 is necessary since the restriction of f to any finite subset $A \subseteq X$ has finite threshold degree.

We now turn to rational approximation. We will need the following technical lemma.

Lemma 5.5. *Let $f : X \to \{-1, +1\}$ be given, where $X \subset \mathbb{R}^n$ is compact and $\deg_\pm(f) < \infty$. Let d be an integer with $d \geqslant \deg_\pm(f)$. Then there do not exist polynomials p, q of degree up to d such that*

$$0 < \frac{1}{c} q(x) < f(x)p(x) < cq(x), \qquad x \in X,$$

where

$$c = \sqrt{\frac{1 + R^+(d, f)}{1 - R^+(d, f)}}.$$

Proof. For the sake of contradiction, suppose that such p and q do exist. By the compactness of X, there exist $\varepsilon > 0$ and $M > 0$ such that

$$\varepsilon \leqslant q(x) \leqslant M, \qquad x \in X,$$

and

$$\frac{1}{c}q(x) + \varepsilon \leqslant f(x)p(x) \leqslant cq(x) - \varepsilon, \qquad x \in X.$$

In particular, there exists a real b such that $1 \leqslant b < c$ and

$$0 < \frac{1}{b}q(x) \leqslant f(x)p(x) \leqslant bq(x), \qquad x \in X.$$

Applying Theorem 3.2, we arrive at the following contradiction:

$$R^+(d,f) \leqslant \frac{b^2-1}{b^2+1} < \frac{c^2-1}{c^2+1} = R^+(d,f). \qquad \square$$

We are now in a position to give the promised result for rational approximation.

Theorem 5.6 (Rational approximation on compact vs. finite sets). *Let* $f : X \to \{-1,+1\}$ *be given, where* $X \subset \mathbb{R}^n$ *is compact and* $\deg_\pm(f) < \infty$. *Then for every integer* d, *there exists a finite subset* $A \subseteq X$ *such that*

$$R^+(d,f) = R^+(d,f|_A). \tag{5.3}$$

Proof. Abbreviate $\varepsilon = R^+(d,f)$. If $\varepsilon = 1$, then this theorem follows from the previously established Theorem 5.3, in view of Theorem 3.1.

We now examine the complementary case $\varepsilon < 1$. Let D denote the set of all n-tuples $(d_1,\dots,d_n)$ of nonnegative integers $d_1,\dots,d_n$ with $\sum d_i \leqslant d$. Put $c = \sqrt{(1+\varepsilon)/(1-\varepsilon)}$ and consider the following system of inequalities in the $2|D|$ real variables $a_{d_1,\dots,d_n}$ and $b_{d_1,\dots,d_n}$ for $(d_1,\dots,d_n) \in D$:

$$\left.\begin{aligned}
&f(x)\sum_D a_{d_1,\dots,d_n} x_1^{d_1}\cdots x_n^{d_n} > \frac{1}{c}\sum_D b_{d_1,\dots,d_n} x_1^{d_1}\cdots x_n^{d_n}, && x \in X,\\
&f(x)\sum_D a_{d_1,\dots,d_n} x_1^{d_1}\cdots x_n^{d_n} < c\sum_D b_{d_1,\dots,d_n} x_1^{d_1}\cdots x_n^{d_n}, && x \in X,\\
&\sum_D b_{d_1,\dots,d_n} x_1^{d_1}\cdots x_n^{d_n} > 0, && x \in X.
\end{aligned}\right\} \tag{5.4}$$

By Lemma 5.5, this system has no solution. Theorem 3.6 implies that the sets $f^{-1}(-1)$ and $f^{-1}(1)$ are compact, which means that the characteristic set of (5.4) is a compact subset of $\mathbb{R}^{2|D|}$. In view of the last two observations, Corollary 5.2 implies that some finite subset of the inequalities (5.4) has no solution. By Theorem 3.2, this gives a finite subset $A \subseteq X$ such that

$$R^+(d, f|_A) \geqslant \frac{c^2-1}{c^2+1} = \varepsilon.$$

Since $R^+(d, f|_A) \leqslant \varepsilon$ by definition, the proof is complete. $\square$

5.3 Final Results

Using Theorem 5.6, we will now show that the results of Section 4 are valid for arbitrary compact sets.

Theorem 5.7 (Main Theorem, two functions). *Let $f : X \to \{-1, +1\}$ and $g : Y \to \{-1, +1\}$ be given functions, where $X, Y \subset \mathbb{R}^n$ are compact sets. Assume that $f \not\equiv 1$ and $g \not\equiv 1$. Assume further that $d = \deg_\pm(f \wedge g) < \infty$. Then*

$$R^+(2d, f) + R^+(4d, g) < 1, \tag{5.5}$$

$$R^+(4d, f) + R^+(2d, g) < 1. \tag{5.6}$$

If in addition f is odd, then

$$R^+(2d, f) + R^+(d, g) < 1. \tag{5.7}$$

Proof. We prove (5.7) first. By Theorem 5.6, there exist finite subsets $A \subseteq X$ and $B \subseteq Y$ such that $R^+(2d, f) = R^+(2d, f|_A)$ and $R^+(d, g) = R^+(d, g|_B)$. We may assume that $f|_A \not\equiv 1$, since otherwise we can extend A to contain a point from $f^{-1}(-1)$. Similarly, we may assume that A is closed under negation and that $g|_B \not\equiv 1$. Since the restriction of $f \wedge g$ to $A \times B$ has threshold degree is d or less, (5.7) now follows from Theorem 4.1.

The proofs of (5.5) and (5.6) are analogous, except one invokes Theorem 4.3 instead of Theorem 4.1. $\square$

In the same way, one obtains the following compact analogue of Theorem 4.4.

Theorem 5.8. *Let $f_1, f_2, \ldots, f_k$ be given Boolean functions on compact sets $X_1, X_2, \ldots, X_k \subset \mathbb{R}^n$, respectively. Assume that $f_i \not\equiv 1$ for $i = 1, 2, \ldots, k$. Let $d = \deg_\pm(f_1 \wedge f_2 \wedge \cdots \wedge f_k)$. Then*

$$\sum_{i=1}^{k} R^+(D, f_i) < 1$$

for $D = 8d \log 2k$.

Proof. Analogous to the proof of Theorem 5.7. $\square$

6 Disjunctions and Other Combining Functions

In Sections 4 and 5, we were concerned exclusively with *conjunctions* of Boolean functions. As we will now see, the results of the previous sections apply just as well to *disjunctions* and many other combining functions.

Disjunctions are an illustrative starting point. Consider two Boolean functions $f : X \to \{-1,+1\}$ and $g : Y \to \{-1,+1\}$, where $X, Y \subset \mathbb{R}^n$ are compact sets and $f, g \not\equiv -1$. Assume further that the function $f \vee g$ has finite threshold degree, d. Then, we claim that

$$R^+(4d, f) + R^+(4d, g) < 1. \tag{6.1}$$

To see this, note first that the function $f \vee g$ has the same threshold degree as its negation, $\overline{f} \wedge \overline{g}$. Applying Theorem 5.7 to the latter function shows that

$$R^+(4d, \overline{f}) + R^+(4d, \overline{g}) < 1.$$

This is equivalent to (6.1) since approximating a function is the same as approximating its negation: $R^+(4d, \overline{f}) = R^+(4d, f)$ and $R^+(4d, \overline{g}) = R^+(4d, g)$. As in the case of conjunctions, (6.1) can be strengthened to

$$R^+(2d, f) + R^+(2d, g) < 1$$

if at least one of f, g is known to be odd. These observations carry over to disjunctions of multiple functions, $f_1 \vee f_2 \vee \cdots \vee f_k$.

The above discussion is still too specialized. In what follows, we consider composite functions $h(f_1, f_2, \ldots, f_k)$, where $h : \{-1,+1\}^k \to \{-1,+1\}$ is any given Boolean function. We will shortly see that the results of the previous sections hold for various h other than $h = \text{AND}$ and $h = \text{OR}$.

We start with some notation and definitions. Let $f, h : \{-1,+1\}^k \to \{-1,+1\}$ be given Boolean functions. Recall that f is called a *subfunction* of h if for some fixed strings $y, z \in \{-1,+1\}^k$, one has

$$f(x) = h(\ldots, (x_i \wedge y_i) \vee z_i, \ldots)$$

for each $x \in \{-1,+1\}^k$. In words, f can be obtained from h by replacing some of the variables $x_1, x_2, \ldots, x_k$ with fixed values (-1 or $+1$).

Definition 6.1. A function $h : \{-1,+1\}^k \to \{-1,+1\}$ is AND-*equivalent* if for each pair of indices i, j, where $1 \leqslant i \leqslant j \leqslant k$, at least one of the eight functions

$$\begin{array}{ll} x_i \wedge x_j, & x_i \vee x_j, \\ x_i \wedge \overline{x_j}, & x_i \vee \overline{x_j}, \\ \overline{x_i} \wedge x_j, & \overline{x_i} \vee x_j, \\ \overline{x_i} \wedge \overline{x_j}, & \overline{x_i} \vee \overline{x_j} \end{array}$$

is a subfunction of $h(x)$.

Theorem 6.2. *Let $f_1, f_2, \dots, f_k$ be nonconstant Boolean functions on compact sets $X_1, X_2, \dots, X_k \subset \mathbb{R}^n$, respectively. Let $h : \{-1,+1\}^k \to \{-1,+1\}$ be an* AND-*equivalent function. Assume that $d = \deg_\pm(h(f_1, f_2, \dots, f_k)) < \infty$. Then*

$$\sum_{i=1}^{k} R^+(D, f_i) < 1$$

for $D = 8d \log 2k$.

Proof. Since h is AND-equivalent, it follows that for each pair of indices $i < j$, one of the following eight functions is a subfunction of $h(f_1, \dots, f_k)$:

$$\begin{array}{ll} f_i \wedge f_j, & f_i \vee f_j, \\ f_i \wedge \overline{f_j}, & f_i \vee \overline{f_j}, \\ \overline{f_i} \wedge f_j, & \overline{f_i} \vee f_j, \\ \overline{f_i} \wedge \overline{f_j}, & \overline{f_i} \vee \overline{f_j}. \end{array}$$

By Theorem 5.7 (and the opening remarks of this section),

$$R^+(4d, f_i) + R^+(4d, f_j) < 1.$$

The remainder of the proof is identical to the proof of Theorem 4.4, starting at equation (4.5). □

In summary, the results of the previous sections apply to compositions of the form $h(f_1, f_2, \dots, f_k)$ for various h. For a function $h : \{-1,+1\}^k \to \{-1,+1\}$ to be AND-equivalent, h must clearly depend on all of its inputs. This obvious necessary condition is often sufficient, for example when h is a read-once AND/OR/NOT formula or a halfspace. Hence, Theorem 1.2 from the Introduction is a corollary of Theorem 6.2.

If more information is available about the combining function h, Theorem 6.2 can be generalized to let some of $f_1, \dots, f_k$ be constant functions. For example, some or all of the functions $f_1, \dots, f_k$ in Theorem 5.8 can be identically true. Another direction for generalization is as follows. In Definition 6.1, one considers all the $\binom{k}{2}$ distinct pairs of indices (i, j). If one happens to know that f_1 is harder to approximate than $f_2, \dots, f_k$, then one can relax Definition 6.1 to examine only the $k-1$ pairs $(1,2), (1,3), \dots, (1,k)$. We do not formulate these extensions as theorems, the fundamental technique being already clear.

Our results so far can be viewed as a technique for proving lower bounds on the threshold degree of composite functions $h(f_1, f_2, \dots, f_k)$. We make this view explicit in the following statement, which is the contrapositive of Theorem 6.2.

Theorem 6.3. *Let $f_1, f_2, \dots, f_k$ be nonconstant Boolean functions on compact sets $X_1, X_2, \dots, X_k \subset \mathbb{R}^n$, respectively. Let $h : \{-1,+1\}^k \to \{-1,+1\}$ be an* AND-*equivalent function. Suppose that*

$$\sum_{i=1}^{k} R^+(D, f_i) \geqslant 1$$

for some integer D. Then

$$\deg_\pm(h(f_1, f_2, \dots, f_k)) > \frac{D}{8 \log 2k}. \tag{6.2}$$

Remark 6.4 (*On the tightness of Theorem* 6.3). Theorem 6.3 is close to optimal. For example, when $h = \text{AND}$, the lower bound in (6.2) is tight up to a factor of $\Theta(k \log k)$. This can be seen by the well-known argument [7, 16] described in the Introduction. Specifically, fix an integer D such that $\sum R^+(D, f_i) < 1$. Then there exists a rational function $p_i(x_i)/q_i(x_i)$ on X_i, for $i = 1, 2, \dots, k$, such that q_i is positive on X_i and

$$\sum_{i=1}^{k} \sup_{x_i \in X_i} \left| f_i(x_i) - \frac{p_i(x_i)}{q_i(x_i)} \right| < 1.$$

As a result,

$$\begin{aligned} \bigwedge_{i=1}^{k} f_i(x_i) &\equiv \operatorname{sgn}\left(k - 1 + \sum_{i=1}^{k} f_i(x_i)\right) \\ &\equiv \operatorname{sgn}\left(k - 1 + \sum_{i=1}^{k} \frac{p_i(x_i)}{q_i(x_i)}\right). \end{aligned}$$

Multiplying the expression in parentheses by the positive quantity $\prod q_i(x_i)$ yields

$$\bigwedge_{i=1}^{k} f_i(x_i) \equiv \operatorname{sgn}\left((k-1) \prod_{i=1}^{k} q_i(x_i) + \sum_{i=1}^{k} p_i(x_i) \prod_{j \in \{1,\dots,k\} \setminus \{i\}} q_j(x_j) \right),$$

whence

$$\deg_{\pm}(f_1 \wedge f_2 \wedge \cdots \wedge f_k) \leqslant kD.$$

This settles our claim regarding $h = \text{AND}$. For arbitrary AND-equivalent functions $h : \{-1,+1\}^k \to \{-1,+1\}$, a similar argument (cf. Theorem 31 of Klivans et al. [16]) shows that the lower bound in (6.2) is tight up to a factor of $\text{poly}(k)$.

We close this section by revisiting the discussion in Section 1.2 on learning intersections of halfspaces. We have:

Theorem 6.5. *Let $f : X \to \{-1,+1\}$ be a given function, where $X \subset \mathbb{R}^n$ is compact. Suppose that $f \wedge f$ has finite threshold degree. Then for every integer $k \geqslant 2$,*

$$\deg_{\pm}(\underbrace{f \wedge f \wedge \cdots \wedge f}_{k}) \leqslant (8k \log k) \cdot \deg_{\pm}(f \wedge f). \tag{6.3}$$

Proof. Put $d = \deg_{\pm}(f \wedge f)$. Theorem 5.7 implies that $R^+(4d, f) < 1/2$, whence $R^+(8d \log k, f) < 1/k$ by Corollary 3.4. By the argument in Remark 6.4, this proves the theorem. ☐

To illustrate, let $\mathscr{C}$ be the concept class of halfspaces on $\{0,1\}^n$. Theorem 6.5 shows that the task of constructing a sign-representation for the intersection of k halfspaces reduces to the case $k = 2$. In other words, solving the problem for $k = 2$ solves it at once for all k. The dependence on k in (6.3) is tight up to a factor of $16 \log k$, even in the simple case when f is the OR function on n bits [23].

7 Rational Approximation of the Majority Function

The study of rational approximation dates back to the remarkable 1877 article by E. I. Zolotarev [12], a student of P. L. Chebyshev. Interest in the subject was revived a century later when D. J. Newman [24] obtained surprisingly accurate rational approximations for several common functions in $\mathscr{C}[-1,1]$, such as $|t|$ and t^α for rational $\alpha > 0$. Newman's discovery inspired a considerable body of work in the area; see the monograph of Petrushev and Popov [29] for an excellent introduction.

Newman's work has also found important applications in theoretical computer science. These applications are based on the rational approximation of $\operatorname{sgn} t$, obtained from Newman's approximation of $|t|$ by dividing by t, see [28]. Paturi and Saks [28] used this fact to develop a novel technique for studying the size of threshold circuits. Siu et al. [39] continued this line of work with new lower bounds on circuit size. Beigel et al. [7] used the rational approximation to the sign function to establish the closure of PP under intersection and various other reductions. Klivans et al. [16] used rational approximation to obtain much improved algorithms for PAC learning intersections of low-weight halfspaces under arbitrary distributions.

The goal of this section is to determine $R^+(d, \mathrm{MAJ}_n)$ for each integer d, i.e., to determine the least error to which a degree-d multivariate rational function can approximate the majority function. We will then combine this result with Theorem 4.3 to prove a conjecture of O'Donnell and Servedio [26]. As is frequently the case with symmetric Boolean functions such as majority, we first reduce the multivariate problem of analyzing $R^+(d, \mathrm{MAJ}_n)$ to a univariate question. Specifically, given an integer d and a set $S \subset \mathbb{R}$, we define

$$R^+(d, S) \;=\; \inf_{p,q} \sup_{t \in S} \left| \operatorname{sgn} t - \frac{p(t)}{q(t)} \right|,$$

where the infimum ranges over $p, q \in P_d$ such that q is positive on S. We have:

Theorem 7.1. *For every integer d,*

$$R^+(d, \mathrm{MAJ}_n) \leqslant R^+(d-2, \{\pm 1, \pm 2, \dots, \pm\lceil n/2 \rceil\}), \tag{7.1}$$
$$R^+(d, \mathrm{MAJ}_n) \geqslant R^+(d, \{\pm 1, \pm 2, \dots, \pm\lfloor n/2 \rfloor\}). \tag{7.2}$$

Proof. We prove (7.1) first. Fix a degree-$(d-2)$ approximant $p(t)/q(t)$ to $\operatorname{sgn} t$ on $S = \{\pm 1, \dots, \pm\lceil n/2 \rceil\}$, where q is positive on S. For small $\delta > 0$, define

$$A_\delta(t) = \frac{t^2 p(t) - \delta}{t^2 q(t) + \delta}.$$

Then A_δ is a rational function of degree at most d whose denominator is positive on $S \cup \{0\}$. Finally, we have $A_\delta(0) = -1$ and

$$\lim_{\delta \to 0} \max_{t \in S} |\operatorname{sgn} t - A_\delta(t)| = \max_{t \in S} \left| \operatorname{sgn} t - \frac{p(t)}{q(t)} \right|,$$

which yields the corresponding approximant for MAJ_n, namely, $A_\delta(\sum x_i - \lfloor n/2 \rfloor)$.

We now turn to the proof of the lower bound, (7.2). Fix a degree-d approximant $P(x)/Q(x)$ for MAJ_n, where $Q(x) > 0$ for $x \in \{0,1\}^n$. Let ε denote the error of this approximant, $\varepsilon \leqslant 1$. Then

$$(1-\varepsilon) Q(x) \leqslant -P(x) \leqslant (1+\varepsilon) Q(x)$$

whenever $\sum x_i \in \{0, 1, \dots, \lfloor n/2 \rfloor\}$, and

$$(1 - \varepsilon)Q(x) \leqslant P(x) \leqslant (1 + \varepsilon)Q(x)$$

whenever $\sum x_i \in \{\lfloor n/2 \rfloor + 1, \dots, n\}$. Now, Proposition 2.2 guarantees the existence of univariate polynomials $p, q \in P_d$ such that for all $x \in \{0,1\}^n$, one has $p(\sum x_i) = \mathbf{E}_{\sigma \in S_n}[P(\sigma x)]$ and $q(\sum x_i) = \mathbf{E}_{\sigma \in S_n}[Q(\sigma x)]$. In view of the previous two inequalities for P and Q, we obtain:

$$\begin{aligned} (1 - \varepsilon)q(t) \leqslant -p(t) \leqslant (1 + \varepsilon)q(t), \qquad & t = 0, 1, \dots, \lfloor n/2 \rfloor; \\ (1 - \varepsilon)q(t) \leqslant p(t) \leqslant (1 + \varepsilon)q(t), \qquad & t = \lfloor n/2 \rfloor + 1, \dots, n. \end{aligned}$$

Thus,

$$\max_{t = \pm 1, \pm 2, \dots, \pm \lfloor n/2 \rfloor} \left| \operatorname{sgn} t - \frac{p(t + \lfloor n/2 \rfloor)}{q(t + \lfloor n/2 \rfloor)} \right| \leqslant \varepsilon.$$

Since q is positive on $\{0, 1, \dots, n\}$, this completes the proof of (7.2). □

Remark 7.2. The proof that we gave for the upper bound, (7.1), illustrates a useful property of univariate rational approximants $A(t) = p(t)/q(t)$ on a finite set S. Specifically, given such an approximant and a point $t^* \notin S$, there exists an approximant A' with $A'(t^*) = a$ for any prescribed value a and $A' \approx A$ everywhere on S. One such construction is

$$A'(t) = \frac{(t - t^*)p(t) + a\delta}{(t - t^*)q(t) + \delta}$$

for an arbitrarily small constant $\delta > 0$. Note that A' has degree only 1 higher than the degree of the original approximant, A. This phenomenon is in sharp contrast to approximation by polynomials, which do not possess this corrective ability.

Theorem 7.1 states that rational approximation of the majority function is essentially equivalent to rational approximation of the sign function over the corresponding finite support. We give a detailed solution to the latter problem:

Theorem 7.3 (Rational approximation of MAJORITY). *Let n, d be positive integers. Abbreviate $R = R^+(d, \{\pm 1, \pm 2, \dots, \pm n\})$. For $1 \leqslant d \leqslant \log n$,*

$$\exp\left\{-\Theta\left(\frac{1}{n^{1/(2d)}}\right)\right\} \leqslant R < \exp\left\{-\frac{1}{n^{1/d}}\right\}.$$

For $\log n < d < n$,

$$R = \exp\left\{-\Theta\left(\frac{d}{\log(2n/d)}\right)\right\}.$$

For $d \geqslant n$,

$$R = 0.$$

Moreover, the rational approximant is constructed explicitly in each case.

Theorem 7.3 is the main result of this section. We establish it in the next two subsections, giving separate treatment to the cases $d \leqslant \log n$ and $d > \log n$ (see Theorems 7.5 and 7.10, respectively). This gives a rather complete answer to how well rational functions can approximate majority.

Before we delve into the proof of Theorem 7.3, we take a moment to survey related results. Long before our work, an upper bound of $\exp\{-\Omega(d/\log n)\}$ on the approximation error for $d \geqslant \log n$ was known and used in the complexity literature, e.g., [28, 39, 7, 16]. We improve on that construction and extend it to all possible d. Our primary contribution in Theorem 7.3, however, is a matching *lower* bound on the error for each d. Indeed, it is only the lower bounds of Theorem 7.3 that we need in this paper. Here, we are not aware of any previous work. The closest previous line of research concerns *continuous* approximation of the sign function on $[-1, -\varepsilon] \cup [\varepsilon, 1]$, which unfortunately gives no insight into the discrete case. For example, the lower bound derived by Newman [24] in the continuous setting is based on the integration of relevant rational functions with respect to a suitable measure, which has no meaningful discrete analogue. We obtain our discrete lower bounds in a quite different way, by reformulating the discrete case as a linear program and providing an explicit solution to its dual.

7.1 Low-Degree Approximation

We start by deriving a method for lower bounds on the approximation error.

Theorem 7.4. *Let d be an integer, $0 \leqslant d \leqslant 2n-1$. Fix a nonempty subset $S \subseteq \{\pm 1, \pm 2, \dots, \pm n\}$ closed under negation. Suppose that there exists a real $\alpha \in [0,1]$ and a polynomial $r \in P_{2n-d-1}$ such that*

$$r(t) = 0, \qquad t \in \{-n, \dots, n\} \setminus S, \tag{7.3}$$

and

$$(-1)^t r(t) > \alpha |r(-t)|, \qquad t \in S \cap \{1, 2, \dots, n\}. \tag{7.4}$$

Then

$$R^+(d, S) \geqslant \frac{2\alpha}{1+\alpha}.$$

Proof. Fix $p, q \in P_d$ such that q is positive on S. Put

$$\varepsilon = \max_{t \in S} \left| \operatorname{sgn} t - \frac{p(t)}{q(t)} \right|.$$

We assume that $\varepsilon < 1$ since otherwise there is nothing to show. For $t \in S \cap \{1, 2, \ldots, n\}$, we have

$$(1 - \varepsilon)q(t) \leqslant p(t) \leqslant (1 + \varepsilon)q(t) \tag{7.5}$$

and

$$(1 - \varepsilon)q(-t) \leqslant -p(-t) \leqslant (1 + \varepsilon)q(-t). \tag{7.6}$$

Consider the polynomial $u(t) = q(t) + q(-t) + p(t) - p(-t)$. Equations (7.5) and (7.6) show that for $t \in S \cap \{1, 2, \ldots, n\}$, one has $u(t) \geqslant (2 - \varepsilon)\{q(t) + q(-t)\}$ and $|u(-t)| \leqslant \varepsilon\{q(t) + q(-t)\}$, whence

$$u(t) \geqslant \left(\frac{2}{\varepsilon} - 1\right) |u(-t)|, \qquad t \in S \cap \{1, 2, \ldots, n\}. \tag{7.7}$$

We also note that

$$u(t) > 0, \qquad t \in S \cap \{1, 2, \ldots, n\}. \tag{7.8}$$

Since $r(t)u(t)$ has degree at most $2n - 1$,

$$\begin{aligned} 0 &= \sum_{t=-n}^{n} \binom{2n}{n+t} (-1)^t r(t)u(t) && \text{by Fact 2.1} \\ &= \sum_{t \in S \cap \{1,\ldots,n\}} \binom{2n}{n+t} \Big\{ (-1)^t r(t)u(t) + (-1)^t r(-t)u(-t) \Big\} && \text{by (7.3).} \end{aligned}$$

At the same time, inequalities (7.4), (7.7), and (7.8) show that

$$(-1)^t r(t)u(t) > \alpha \left(\frac{2}{\varepsilon} - 1\right) |r(-t)u(-t)|, \qquad t \in S \cap \{1, 2, \ldots, n\}.$$

We immediately obtain $\alpha\left(\frac{2}{\varepsilon} - 1\right) < 1$, as was to be shown. □

The method of Theorem 7.4 amounts to reformulating (7.7) and (7.8) as a linear program and exhibiting a solution to its dual. The presentation above does not explicitly use the language of linear programs or appeal to duality, however, because our goal is solely to prove the *correctness* of our method and not its *completeness*.

Using Theorem 7.4, we will now determine the optimal error in the approximation of the majority function by rational functions of degree up to $\log n$. The case of higher degrees will be settled in the next subsection.

Theorem 7.5 (Low-degree rational approximation of MAJORITY). *Let d be an integer, $1 \leqslant d \leqslant \log n$. Then*

$$\exp\left\{-\Theta\left(\frac{1}{n^{1/(2d)}}\right)\right\} \leqslant R^+(d, \{\pm1, \pm2, \ldots, \pm n\}) < \exp\left\{-\frac{1}{n^{1/d}}\right\}.$$

Proof. Put $\Delta = \lfloor n^{1/d} \rfloor \geqslant 2$ and $S = \{\pm 1, \pm\Delta, \pm\Delta^2, \dots, \pm\Delta^d\}$. Define $r \in P_{2n-d-1}$ by

$$r(t) = (-1)^n \prod_{i=0}^{d-1} (t - \Delta^i \sqrt{\Delta}) \prod_{i \in \{-n,\dots,n\} \setminus S} (t - i).$$

For $j = 0, 1, 2, \dots, d$,

$$\begin{aligned} \frac{|r(\Delta^j)|}{|r(-\Delta^j)|} &= \prod_{i=0}^{j-1} \frac{\Delta^j - \Delta^i\sqrt{\Delta}}{\Delta^j + \Delta^i\sqrt{\Delta}} \prod_{i=j}^{d-1} \frac{\Delta^i\sqrt{\Delta} - \Delta^j}{\Delta^i\sqrt{\Delta} + \Delta^j} > \left(\prod_{i=1}^{\infty} \frac{\Delta^{i/2} - 1}{\Delta^{i/2} + 1} \right)^2 \\ &> \exp\left\{ -5 \sum_{i=1}^{\infty} \frac{1}{\Delta^{i/2}} \right\} > \exp\left\{ -\frac{18}{\sqrt{\Delta}} \right\}, \end{aligned}$$

where we used the bound $(a-1)/(a+1) > \exp(-2.5/a)$, valid for $a \geqslant \sqrt{2}$. Since $\operatorname{sgn} r(t) = (-1)^t$ for positive $t \in S$, we conclude that

$$(-1)^t r(t) > \exp\left\{ -\frac{18}{\sqrt{\Delta}} \right\} |r(-t)|, \qquad t \in S \cap \{1, 2, \dots, n\}.$$

Since in addition r vanishes on $\{-n, \dots, n\} \setminus S$, we infer from Theorem 7.4 that $R^+(d, S) \geqslant \exp\{-18/\sqrt{\Delta}\}$.

We now turn to the proof of the upper bound. Following [24, 28], we let

$$p(t) = \prod_{i=1}^{d} \left(t + n^{(2i-1)/(2d)} \right).$$

By considering every interval $[n^{i/(2d)}, n^{(i+1)/(2d)}]$, where $i = 0, 1, \dots, 2d-1$, one sees that

$$p(t) \geqslant \frac{n^{1/(2d)} + 1}{n^{1/(2d)} - 1} |p(-t)|, \qquad t = 1, 2, \dots, n.$$

We conclude that $R^+(d, \{\pm 1, \pm 2, \dots, \pm n\}) \leqslant 1 - n^{-1/d}$, the approximant in question being $n^{-1/(2d)}\{p(t) - p(-t)\}/\{p(t) + p(-t)\}$. $\square$

7.2 High-Degree Approximation

In the previous subsection, we determined the least error in approximating the majority function by rational functions of degree up to $\log n$. Our goal here is to solve the case of higher degrees.

We start with some preparatory work. First, we need to accurately estimate products of the form $\prod_i (\Delta^i + 1)/(\Delta^i - 1)$ for all $\Delta > 1$. A suitable *lower* bound was already given by Newman [24, Lem. 1]:

Lemma 7.6 (Newman). *For all* $\Delta > 1$,

$$\prod_{i=1}^{n} \frac{\Delta^i + 1}{\Delta^i - 1} > \exp\left\{ \frac{2(\Delta^n - 1)}{\Delta^n(\Delta - 1)} \right\}.$$

Proof. Immediate from the bound $(a+1)/(a-1) > \exp(2/a)$, valid for $a > 1$. $\square$

We will need a matching upper bound:

Lemma 7.7. *For all* $\Delta > 1$,

$$\prod_{i=1}^{\infty} \frac{\Delta^i + 1}{\Delta^i - 1} < \exp\left\{ \frac{4}{\Delta - 1} \right\}.$$

Proof. Let $k \geqslant 0$ be an integer. By the binomial theorem, $\Delta^i \geqslant (\Delta - 1)i + 1$ for integers $i \geqslant 0$. As a result,

$$\prod_{i=1}^{k} \frac{\Delta^i + 1}{\Delta^i - 1} \leqslant \prod_{i=1}^{k} \frac{1}{i} \left(i + \frac{2}{\Delta - 1} \right) \leqslant \binom{k + \lceil \frac{2}{\Delta-1} \rceil}{k}.$$

Also,

$$\prod_{i=k+1}^{\infty} \frac{\Delta^i + 1}{\Delta^i - 1} < \prod_{i=0}^{\infty} \left(1 + \frac{2}{(\Delta^{k+1} - 1)\Delta^i} \right) < \exp\left\{ \frac{2\Delta}{(\Delta^{k+1} - 1)(\Delta - 1)} \right\}.$$

Setting $k = k(\Delta) = \lfloor \frac{2}{\Delta-1} \rfloor$, we conclude that

$$\prod_{i=1}^{\infty} \frac{\Delta^i + 1}{\Delta^i - 1} < \exp\left\{ \frac{C}{\Delta - 1} \right\},$$

where

$$C = \sup_{\Delta > 1} \left\{ (\Delta - 1) \ln \binom{k(\Delta) + \lceil \frac{2}{\Delta-1} \rceil}{k(\Delta)} + \frac{2\Delta}{\Delta^{k(\Delta)+1} - 1} \right\} < 4. \qquad \square$$

We will also need the following binomial estimate.

Lemma 7.8. *Put* $p(t) = \prod_{i=1}^{n} \left(t - i - \frac{1}{2} \right)$. *Then*

$$\max_{t=1,2,\dots,n+1} \left| \frac{p(-t)}{p(t)} \right| \leqslant \Theta(16^n).$$

Proof. For $t = 1, 2, \ldots, n+1$, we have

$$|p(t)| = \frac{(2t-2)!(2n-2t+2)!}{4^n (t-1)!(n-t+1)!}, \quad |p(-t)| = \frac{t!(2n+2t+1)!}{4^n (2t+1)!(n+t)!}.$$

As a result,

$$\left|\frac{p(-t)}{p(t)}\right| = \frac{t}{2t+1} \cdot \frac{\binom{2n+2t+1}{2t}\binom{2n+1}{n+t}}{\binom{2t-2}{t-1}\binom{2n-2t+2}{n-t+1}} \leqslant \frac{\Theta\left(\frac{2^{4n}}{\sqrt{n}}\right)\Theta\left(\frac{2^{2n}}{\sqrt{n}}\right)}{\Theta\left(\frac{2^{2n}}{n}\right)},$$

which gives the sought bound. $\square$

Our construction requires the following claim, which we settle in advance.

Lemma 7.9. *Let n, d be integers, $1 \leqslant d \leqslant n/55$. Put $p(t) = \prod_{i=1}^{d-1}(t - d\Delta^i\sqrt{\Delta})$, where $\Delta = (n/d)^{1/d}$. Then*

$$\min_{j=1,\ldots,d}\left|\frac{p(\lfloor d\Delta^j \rfloor)}{p(-\lfloor d\Delta^j \rfloor)}\right| > \exp\left\{-\frac{4\ln 3d}{\ln(n/d)} - \frac{8}{\sqrt{\Delta}-1}\right\}.$$

Proof. Fix $j = 1, 2, \ldots, d$. Then for each $i = 1, 2, \ldots, j-1$,

$$d\Delta^j - d\Delta^i\sqrt{\Delta} \geqslant d\left(\Delta^{j-i-\frac{1}{2}} - 1\right) \geqslant \frac{1}{2}(j-i)\ln\frac{n}{d},$$

and thus

$$\begin{aligned}\prod_{i=1}^{j-1}\left(1 - \frac{1}{d\Delta^j - d\Delta^i\sqrt{\Delta}}\right) &\geqslant \exp\left\{-\frac{4}{\ln(n/d)}\sum_{i=1}^{j-1}\frac{1}{j-i}\right\} \\ &\geqslant \exp\left\{-\frac{4\ln 3d}{\ln(n/d)}\right\}. \end{aligned} \tag{7.9}$$

For brevity, let ξ stand for the final expression in (7.9). Since $d \leqslant n/55$, we have $\lfloor d\Delta^j \rfloor - d\Delta^{j-1}\sqrt{\Delta} > 1$. As a result,

$$\begin{aligned}\left|\frac{p(\lfloor d\Delta^j \rfloor)}{p(-\lfloor d\Delta^j \rfloor)}\right| &\geqslant \prod_{i=1}^{j-1}\frac{d\Delta^j - 1 - d\Delta^i\sqrt{\Delta}}{d\Delta^j + d\Delta^i\sqrt{\Delta}} \prod_{i=j}^{d-1}\frac{d\Delta^i\sqrt{\Delta} - d\Delta^j}{d\Delta^i\sqrt{\Delta} + d\Delta^j} \\ &\geqslant \xi\prod_{i=1}^{j-1}\frac{d\Delta^j - d\Delta^i\sqrt{\Delta}}{d\Delta^j + d\Delta^i\sqrt{\Delta}} \prod_{i=j}^{d-1}\frac{d\Delta^i\sqrt{\Delta} - d\Delta^j}{d\Delta^i\sqrt{\Delta} + d\Delta^j} \qquad \text{by (7.9)} \\ &> \xi\left(\prod_{i=1}^{\infty}\frac{\Delta^{i/2}-1}{\Delta^{i/2}+1}\right)^2 \\ &\geqslant \xi\exp\left\{-\frac{8}{\sqrt{\Delta}-1}\right\},\end{aligned}$$

where the last inequality holds by Lemma 7.7. $\square$

We have reached the main result of this subsection.

Theorem 7.10 (High-degree rational approximation of MAJORITY). *Let d be an integer, $\log n < d \leqslant n - 1$. Then*

$$R^+(d, \{\pm 1, \pm 2, \ldots, \pm n\}) = \exp\left\{-\Theta\left(\frac{d}{\log(2n/d)}\right)\right\}.$$

Also,

$$R^+(n, \{\pm 1, \pm 2, \ldots, \pm n\}) = 0.$$

Proof. The final statement in the theorem follows at once by considering the rational function $\{p(t) - p(-t)\}/\{p(t) + p(-t)\}$, where $p(t) = \prod_{i=1}^{n}(t + i)$.

Now assume that $\log n < d < n/55$. Let

$$k = \left\lceil \frac{d}{\log(n/d)} \right\rceil, \qquad \Delta = \left(\frac{n}{d}\right)^{1/d}.$$

Define sets

$$\begin{aligned} S_1 &= \{\pm 1, \pm 2, \ldots, \pm k\}, \\ S_2 &= \{\pm \lfloor d\Delta^i \rfloor \,:\, i = 1, 2, \ldots, d\}, \\ S \ &= S_1 \cup S_2. \end{aligned}$$

Consider the polynomial

$$r(t) = (-1)^n r_1(t) r_2(t) \prod_{i \in \{-n, \ldots, n\} \setminus S} (t - i),$$

where

$$r_1(t) = \prod_{i=1}^{k}\left(t - i - \frac{1}{2}\right), \qquad r_2(t) = \prod_{i=1}^{d-1}(t - d\Delta^i \sqrt{\Delta}).$$

We have:

$$\begin{aligned} \min_{t \in S \cap \{1,2,\ldots,n\}} \left|\frac{r(t)}{r(-t)}\right| &\geqslant \min_{i=1,\ldots,k+1} \left|\frac{r_1(i)}{r_1(-i)}\right| \cdot \min_{i=1,\ldots,d} \left|\frac{r_2(\lfloor d\Delta^i \rfloor)}{r_2(-\lfloor d\Delta^i \rfloor)}\right| \\ &> \exp\left\{-\frac{Cd}{\log(n/d)}\right\} \end{aligned}$$

by Lemmas 7.8 and 7.9, where $C > 0$ is an absolute constant. Since $\operatorname{sgn} p(t) = (-1)^t$ for positive $t \in S$, we can restate this result as follows:

$$(-1)^t r(t) > \exp\left\{-\frac{Cd}{\log(n/d)}\right\} |r(-t)|, \qquad t \in S \cap \{1, 2, \ldots, n\}.$$

Since r vanishes on $\{-n, \ldots, n\} \setminus S$ and has degree $\leqslant 2n - 1 - d$, we infer from Theorem 7.4 that $R^+(d, S) \geqslant \exp\{-Cd/\log(n/d)\}$. This proves the lower bound for the case $\log n < d < n/55$.

To handle the case $n/55 \leqslant d \leqslant n - 1$, a different argument is needed. Let

$$r(t) = (-1)^n t \prod_{i=1}^{d} \left(t - i - \frac{1}{2}\right) \prod_{i=d+2}^{n} (t^2 - i^2).$$

By Lemma 7.8, there is an absolute constant $C > 1$ such that

$$\left|\frac{r(t)}{r(-t)}\right| > C^{-d}, \qquad t = 1, 2, \ldots, d + 1.$$

Since $\operatorname{sgn} r(t) = (-1)^t$ for $t = 1, 2, \ldots, d + 1$, we conclude that

$$(-1)^t r(t) > C^{-d} |r(-t)|, \qquad t = 1, 2, \ldots, d + 1.$$

Since the polynomial r vanishes on $\{-n, \ldots, n\} \setminus \{\pm 1, \pm 2, \ldots, \pm(d+1)\}$ and has degree $2n - 1 - d$, we infer from Theorem 7.4 that

$$R^+(d, \{\pm 1, \pm 2, \ldots, \pm(d + 1)\}) \geqslant C^{-d}.$$

This settles the lower bound for the case $n/55 \leqslant d \leqslant n - 1$.

It remains to prove the upper bound for the case $\log n < d \leqslant n - 1$. Here we always have $d \geqslant 2$. Letting $k = \lfloor d/2 \rfloor$ and $\Delta = (n/k)^{1/k}$, define $p \in P_{2k}$ by

$$p(t) = \prod_{i=1}^{k} (t + i) \prod_{i=1}^{k} (t + k\Delta^i).$$

Fix any point $t \in \{1, 2, \ldots, n\}$ with $p(-t) \neq 0$. Letting i^* be the integer with $k\Delta^{i^*} < t < k\Delta^{i^*+1}$, we have:

$$\begin{aligned} \frac{p(t)}{|p(-t)|} &> \prod_{i=0}^{i^*} \frac{k\Delta^{i^*+1} + k\Delta^i}{k\Delta^{i^*+1} - k\Delta^i} \prod_{i=i^*+1}^{k} \frac{k\Delta^i + k\Delta^{i^*}}{k\Delta^i - k\Delta^{i^*}} \geqslant \prod_{i=1}^{k} \frac{\Delta^i + 1}{\Delta^i - 1} \\ &> \exp\left\{\frac{2(\Delta^k - 1)}{\Delta^k(\Delta - 1)}\right\}, \end{aligned}$$

where the last inequality holds by Lemma 7.6. Substituting $\Delta = (n/k)^{1/k}$, we obtain $p(t) > A|p(-t)|$ for $t = 1, 2, \ldots, n$, where

$$A = \exp\left\{\frac{k}{4\ln(n/k)}\right\}.$$

As a result, $R^+(2k, \{\pm 1, \pm 2, \ldots, \pm n\}) \leqslant 2A/(A^2+1)$, the approximant in question being

$$\frac{A^2-1}{A^2+1} \cdot \frac{p(t)-p(-t)}{p(t)+p(-t)}. \qquad \square$$

8 Threshold Degree of the Intersection of Two Majorities

Consider the function $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ given by

$$f(x,y) = \mathrm{MAJ}_n(x_1,\ldots,x_n) \wedge \mathrm{MAJ}_n(y_1,\ldots,y_n).$$

Recall from the Introduction that f has threshold degree at most $O(\log n)$, by a result of Beigel et al. [7]. We now prove that this upper bound is tight, confirming a conjecture of O'Donnell and Servedio [26].

Theorem 1.3 (Restated from p. 5). *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ be given by $f(x,y) = \mathrm{MAJ}_n(x_1,\ldots,x_n) \wedge \mathrm{MAJ}_n(y_1,\ldots,y_n)$. Then*

$$\deg_\pm(f) = \Omega(\log n).$$

Proof. Theorems 7.1 and 7.3 show that $R^+(\alpha \log n, \mathrm{MAJ}_n) \geqslant 1/2$ for a constant $\alpha > 0$. In view of Theorem 4.3, this completes the proof. $\square$

In addition to threshold degree, several other complexity measures are of interest when sign-representing Boolean functions by real polynomials. One such complexity measure is *density,* i.e., the number of distinct monomials in any polynomial that sign-represents a given function. Constructions in [7, 16] show that the function $f(x,y) = \mathrm{MAJ}_n(x) \wedge \mathrm{MAJ}_n(y)$ can be sign-represented by a linear combination of $n^{O(\log n)}$ monomials, namely, the monomials of degree up to $O(\log n)$. Klivans and Sherstov [20, Thm. 1.2] complement this with a lower bound of $n^{\Omega(\log n/\log\log n)}$ on the number of distinct monomials needed. Our next result improves this lower bound to a tight $n^{\Theta(\log n)}$.

Theorem 8.1 (Threshold density of MAJ $\wedge$ MAJ). *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ be given by $f(x,y) = \mathrm{MAJ}_n(x_1,\dots,x_n) \wedge \mathrm{MAJ}_n(y_1,\dots,y_n)$. Assume that*

$$f(x,y) \equiv \mathrm{sgn}\left(\sum_{i=1}^{k} \lambda_i \phi_i(x,y)\right)$$

for some reals $\lambda_1,\dots,\lambda_k$, where each ϕ_i computes the parity or conjunction of a subset of the literals $x_1,\overline{x_1},\dots,x_n,\overline{x_n},y_1,\overline{y_1},\dots,y_n,\overline{y_n}$. Then

$$k = n^{\Omega(\log n)}.$$

Proof. Identical to the proof of Klivans and Sherstov [20, §3.3, Thm. 1.2], with the only difference that Theorem 1.3 should be used in place of O'Donnell and Servedio's earlier result [26] that $\deg_{\pm}(f) = \Omega(\log n/\log\log n)$. □

References

[1] M. Alekhnovich, M. Braverman, V. Feldman, A. R. Klivans, and T. Pitassi. Learnability and automatizability. In *Proc. of the 45th Symposium on Foundations of Computer Science (FOCS)*, pages 621–630, 2004.

[2] J. Aspnes, R. Beigel, M. L. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994.

[3] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proc. of the 27th Symposium on Foundations of Computer Science (FOCS)*, pages 337–347, 1986.

[4] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[5] R. Beigel. The polynomial method in circuit complexity. In *Proc. of the Eigth Annual Conference on Structure in Complexity Theory*, pages 82–95, 1993.

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

[7] R. Beigel, N. Reingold, and D. A. Spielman. PP is closed under intersection. *J. Comput. Syst. Sci.*, 50(2):191–202, 1995.

[8] A. L. Blum and R. L. Rivest. Training a 3-node neural network is NP-complete. *Neural Networks*, 5:117–127, 1992.

[9] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: A survey. *Theor. Comput. Sci.*, 288(1):21–43, 2002.

[10] H. Buhrman, N. K. Vereshchagin, and R. de Wolf. On computation and communication with small bias. In *Proc. of the 22nd Conf. on Computational Complexity (CCC)*, pages 24–32, 2007.

[11] E. W. Cheney. *Introduction to Approximation Theory*. Chelsea Publishing, New York, 2nd edition, 1982.

[12] Е. И. Золотарёв. Приложение эллиптических функций к вопросам о функциях, наименее и наиболее отклоняющихся от нуля. Известия Имп. Академии Наук, 1877, т. XXX, вып. 5. // E. I. Zolotarev. Application of elliptic functions to questions of functions deviating least and most from zero. *Izvestiya Imp. Akad. Nauk*, 30(5), 1877. In Russian.

[13] J. Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.*, 65(4):612–625, 2002.

[14] J. Forster, M. Krause, S. V. Lokam, R. Mubarakzjanov, N. Schmitt, and H.-U. Simon. Relations between communication complexity, linear arrangements, and computational complexity. In *Proc. of the 21st Conf. on Foundations of Software Technology and Theoretical Computer Science (FST TCS)*, pages 171–182, 2001.

[15] S. Khot and R. Saket. On hardness of learning intersection of two halfspaces. In *Proc. of the 40th Symposium on Theory of Computing (STOC)*, pages 345–354, 2008.

[16] A. R. Klivans, R. O'Donnell, and R. A. Servedio. Learning intersections and thresholds of halfspaces. *J. Comput. Syst. Sci.*, 68(4):808–840, 2004.

[17] A. R. Klivans and R. A. Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. Syst. Sci.*, 68(2):303–318, 2004.

[18] A. R. Klivans and R. A. Servedio. Toward attribute efficient learning of decision lists and parities. *J. Machine Learning Research*, 7:587–602, 2006.

[19] A. R. Klivans and A. A. Sherstov. Cryptographic hardness for learning intersections of halfspaces. In *Proc. of the 47th Symposium on Foundations of Computer Science (FOCS)*, pages 553–562, 2006.

[20] A. R. Klivans and A. A. Sherstov. Unconditional lower bounds for learning intersections of halfspaces. *Machine Learning*, 69(2–3):97–114, 2007.

[21] M. Krause and P. Pudlák. On the computational power of depth-2 circuits with threshold and modulo gates. *Theor. Comput. Sci.*, 174(1-2):137–156, 1997.

[22] M. Krause and P. Pudlák. Computing Boolean functions by polynomials and threshold circuits. *Comput. Complex.*, 7(4):346–370, 1998.

[23] M. L. Minsky and S. A. Papert. *Perceptrons: Expanded edition*. MIT Press, Cambridge, Mass., 1988.

[24] D. J. Newman. Rational approximation to $|x|$. *Michigan Math. J.*, 11(1):11–14, 1964.

[25] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[26] R. O'Donnell and R. A. Servedio. New degree bounds for polynomial threshold functions. In *Proc. of the 35th Symposium on Theory of Computing (STOC)*, pages 325–334, 2003.

[27] R. O'Donnell and R. A. Servedio. Extremal properties of polynomial threshold functions. *J. Comput. Syst. Sci.*, 74(3):298–312, 2008.

[28] R. Paturi and M. E. Saks. Approximating threshold circuits by rational functions. *Inf. Comput.*, 112(2):257–272, 1994.

[29] P. P. Petrushev and V. A. Popov. *Rational Approximation of Real Functions*. Cambridge University Press, Cambridge, 1987.

[30] V. V. Podolskii. Perceptrons of large weight. In *Proc. of the Second International Computer Science Symposium in Russia (CSR)*, pages 328–336, 2007.

[31] V. V. Podolskii. A uniform lower bound on weights of perceptrons. In *Proc. of the Third International Computer Science Symposium in Russia (CSR)*, pages 261–272, 2008.

[32] A. A. Razborov and A. A. Sherstov. The sign-rank of AC^0. In *Proc. of the 49th Symposium on Foundations of Computer Science (FOCS)*, 2008. To appear.

[33] T. J. Rivlin. *An Introduction to the Approximation of Functions*. Dover Publications, New York, 1981.

[34] M. E. Saks. Slicing the hypercube. *Surveys in Combinatorics*, pages 211–255, 1993.

[35] A. A. Sherstov. Separating AC^0 from depth-2 majority circuits. In *Proc. of the 39th Symposium on Theory of Computing (STOC)*, pages 294–301, 2007.

[36] A. A. Sherstov. Communication lower bounds using dual polynomials. *Bulletin of the EATCS*, 95:59–93, 2008.

[37] A. A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Proc. of the 40th Symposium on Theory of Computing (STOC)*, pages 85–94, 2008.

[38] A. A. Sherstov. The unbounded-error communication complexity of symmetric functions. In *Proc. of the 49th Symposium on Foundations of Computer Science (FOCS)*, 2008. To appear.

[39] K.-Y. Siu, V. P. Roychowdhury, and T. Kailath. Rational approximation techniques for analysis of neural networks. *IEEE Transactions on Information Theory*, 40(2):455–466, 1994.

[40] N. K. Vereshchagin. Lower bounds for perceptrons solving some separation problems and oracle separation of AM from PP. In *Proc. of the Third Israel Symposium on Theory of Computing and Systems (ISTCS)*, pages 46–51, 1995.