STABLE FAMILIES OF BASIS FUNCTIONS AND COMPLEXITY LOWER BOUNDS

Per Enflo
Department of Mathematics and Computer Sciences
Kent State University
Kent, OH 44240, USA
enflo@mcs.kent.edu

Meera Sitharam
Computer and Information Science and Engineering
University of Florida
Gainesville, FL 32611, USA
sitharam@cise.ufl.edu

Abstract.
It is now known so far whether polynomial size unbounded weighted threshold circuits of depth 2 differ from the class *NP*. In contrast several results are known in the case of polynomially bounded weights. This paper sheds light on this problem by defining and studying various notions of approximability by stable basis families.
– As a consequence, not only do we simplify and unify the proofs of several previous threshold and communication complexity lower bounds, we also obtain new and general complexity upper bounds by exploring approximation from Boolean bases and the transitivity of approximability relationships. For example, we show that close approximation of a Boolean function from a space spanned by *any* set of Boolean basis functions implies a close approximation from few of those basis functions. One consequence of this is a generalization to general Boolean bases, of a result by [5] that Boolean functions whose Fourier basis expansions have a small coefficient L_1 norm can be expressed as an unweighted threshold of parity functions.
– One of the examples of stable basis families indicates a direct method -using scalar product estimates - for proving lower bounds for an algebraic circuit model.
– We give a method for constructing unstable bases, and show that even

simple families of threshold functions, are unstable, thereby indicating a possible reason why lower bounds for weighted thresholds of thresholds are proving so difficult.

Key words. Circuit complexity; Communication complexity; Algebraic complexity; Stability; Complexity Lower bounds; Linear approximation; Nonlinear approximation.

Subject classifications. AMS classification 68Q15, 68Q99.

Table of Contents

1. Introduction

A distinct difference has been observed between the difficulties of proving circuit size complexity lower bounds for unweighted (polynomially bounded weight) versus weighted (unbounded weight) threshold circuits of depth 2. Several lower bounds are known, for instance, in [16], [5], [4], [27], [22] [19], [9], [11], for the size of circuits that compute an unweighted threshold of basis functions from various simple families, for example majorities, general symmetric functions, etc.

However, only a few lower bounds are known for weighted thresholds and only of weaker basis families, for example parities in [20], and $AC^0[d]$ functions in [8], (see also [2], [21] and [26]). Embarassingly, it is not yet known if the class of polynomial size weighted threshold circuits of depth 2 (or unweighted thresholds of depth 3) differs from *NP*!

One of the reasons suggested is that methods similar to the "correlation/ discriminator/ discrepancy lemma" of [16] and [11], which is used in most of the unweighted threshold lower bounds, are inapplicable in the weighted case. The correlation lemma states roughly that if a function is computable by a threshold of given basis functions, then it must have a high scalar product with at least one of the given basis functions over any distribution, and viceversa, *provided* the weights used by the threshold are small. This gives a clean way to construct a function that is hard to compute by an unweighted threshold of functions from the given basis family, roughly, by constructing a function that has small scalar products with all the functions in the family. Two methods similar to the "correlation method" have been named in the literature as the "communication complexity method," ([11]) and the "variation rank or geometric method" ([22]). Although these methods have been considered as distinct and different, these methods are closely related (as will be seen during the course of this paper), more or less reduce to the same general method that is based on scalar product estimates.

A direct application of such scalar product estimates for proving weighted threshold lower bounds does not, at first glance, seem viable for most bases. The following indirect application of this approach for weighted thresholds has been attempted, but has not met with success: first express a weighted threshold of a basis family as an unweighted threshold of a new basis family, and then then force small scalar products with the new family. The problem with this approach is that the new family is typically too powerful to allow the construction of a function having small scalar products with all the functions in the family. For example, [11] prove that two levels of weighted thresholds can be simulated by three levels of unweighted thresholds. Here, they have replaced the original basis family of weighted thresholds LT_1 by the significantly more powerful family consisting of unweighted thresholds of unweighted thresholds LT_2.

Scalar product estimates have been employed for proving weighted threshold lower bounds in the "spectral method" of [5] which is, however, applicable only to basis families that are orthonormal (at least over a large subdomain). We give a general result that subsumes the "spectral method."

1.1. Results. (We indicate general results by $*$, and specific complexity applications by $\bullet$). ($*$) The result around which this paper is constructed is the following: if basis families satisfy certain weak stability conditions, then a direct application of scalar product estimates *does yield* weighted threshold lower bounds.

We develop this result by formalizing several classical notions of approximability of functions from a space spanned by a given basis. Two of these notions correspond exactly to computability by weighted and unweighted thresholds of the basis functions; these two notions sandwich two intermediate notions of approximation, one of which we call "high-energy approximation," which isolates the use of scalar product estimates and provides the key to our result.

($*$) To illustrate the relationship between the various notions of approximation, we give general methods for proving nonapproximability. In some cases, the methods provide a *characterization* of (non)approximability as well. ($\bullet$) We point out how the methods that have been used for proving threshold and communication complexity lower bounds, reduce to these general methods.

($*$) Using these notions of approximation, we study the special properties of approximation from Boolean basis families, and the transitive nature of approximability. ($\bullet$) We illustrate the use of transitive approximability in several previous lower bound proofs, and in particular, we show how to bypass the communication complexity bounds which many previous results use as intermediaries (the "communication complexity method") in the process of proving threshold complexity lowerbounds. ($\bullet$) In addition, we obtain a general complexity upper bound: unweighted thresholds of functions that are linear combinations of arbitrarily many Boolean functions in a family B can in fact be expressed as an unweighted threshold of (polyomially many) functions in B, provided the linear combinations have small total weight. This is a generalization of a result of [5] that Boolean functions whose Fourier transforms have a small L_1 norm can be expressed as an unweighted threshold of parity functions.

($*$) Since many of the methods for showing nonapproximability and hence lower bounds involve finding good upper bounds on scalar products between functions, we give several general methods for doing so. ($\bullet$) In particular, we give an example of a read-once AC^0 function of depth 3 that has a small scalar product with every LT_1 function.

We formally define various notions of weak and strong stability of basis families and show that if a basis family is stable in the weakest sense, then lower bounds on the size of weighted thresholds of these basis functions can be proven using scalar product estimates. ($\bullet$) Stable basis families, in particular, include orthonormal basis families, and a special case of this result yields the "spectral method" of [5]. Using scalar product estimates, we give an alternative proof of an old lower bound of [20] on the size of weighted thresholds of parities (or *modr* functions) needed to compute an $AC^0[3]$ function. The proof in [20] does not take direct advantage of the stability of the basis. ($*$) In addition, we point out a general method of using the divide and conquer paradigm and adapting scalar product estimates to give lower bounds for stable and also for certain highly unstable bases. As one application, this yields the the "geometric method," or "variation rank" method of [22] which was used to prove, for example, that the function DIP_2 cannot be computed as an unweighted threshold of few symmetric functions. These examples motivate a conjecture that all of the basis families for which weighted threshold lower bounds are currently known, are in fact stable in a weak sense. This would permit the use of standard methods based on scalar product estimates even though these bounds have currently been proven using other methods.

($\bullet$) We give examples of natural Boolean basis families that are stable. One of the examples of stable basis families involves threshold gates that take real values as input. This indicates a direct method -using scalar product estimates - for proving lower bounds for an algebraic circuit model, which is related to the more standard, arithmetic circuit, and the algebraic, linear decision tree model.

($*$) Finally, we give a method for constructing unstable bases, and
($\bullet$) show that even simple families of threshold functions, (all of which have the

same two symmetries), are unstable, thereby indicating a possible reason why lower bounds for weighted thresholds of thresholds are proving so difficult.

1.2. Organization. In Section 2, we give basic background and explain the notation that we use. In Section 3, we define four types of approximation, provide methods for establishing nonexistence of approximations, study the properties of approximations from a Boolean basis, explore the transitive nature of approximability, give methods for obtaining scalar product estimates, and sketch applications of all of the above, for proving threshold and communication complexity lower bounds. In Section 4, we give stability conditions on basis families that permit weighted threshold lower bounds to be proven using scalar product estimates; apply this result to give an alternative proof of the result that $AC^0[3] \not\subseteq QT_1$; and discuss how this result can be adapted to show uniform nonapproximability for certain highly unstable bases. In Section 5, we show several examples of Boolean basis families consisting of functions over discrete and continuous domains that are stable; and discuss how one of the continuous examples gives a method of proving lower bounds for an algebraic circuit model, using scalar product estimates. In Section 6, we give a method for establishing instability of a basis and show that there exist Boolean bases that are unstable; in fact, we show that a highly restricted family of threshold functions is unstable.

2. Preliminaries

Unless otherwise specified, all functions are from some subset $S \subseteq \mathbb{R}^n$ to $\mathbb{R}$. For example, S could be $\{-1,1\}^n$, or the continuous domain $(0,1)^n$, or $\{1,\ldots,n\}^n$. The n-tuples in $\{-1,1\}^n$ are viewed as subsets of both $\mathbb{R}^n$ *and* the finite vector space $\mathbb{F}_2^n$, with -1 mapping to $1_{\mathbb{F}_2}$ and 1 mapping to $0_{\mathbb{F}_2}$. The number of arguments of a function is often omitted and is assumed to be n. Similarly, the words "polynomially many" and "polynomially bounded" usually refers to a polynomial in n. For Boolean functions including characteristic functions of sets, the range is $\{1,-1\}$, viewed as a subset of $\mathbb{R}$. Thus, for example, the functions $\wedge$, $\vee$ etc. map from $\{1,-1\}^n$ to $\{1,-1\}$ in the obvious way, with -1 taking the place of the usual 1 and 1 taking the place of 0.

The Fourier transform of a function f from $\{-1,1\}^n$ or $\mathbb{F}_2^n$ to $\mathbb{R}$ is denoted $\hat{f}$ and is given by

$$\hat{f}(x) = 1/2^n \sum_{u \in \mathbb{F}_2^n} f(u)(-1)^{<x,u>};$$

thereby $f(x)$ can be written as $\sum\limits_{u\in\mathbb{F}_2^n} \hat{f}(u)(-1)^{<x,u>}$. The functions $\chi_u(x)$ are defined as $(-1)^{<x,u>}$, and are generally called parity functions, the function χ_{1^n} is called the *Parity* function, and χ_{0^n} is called the constant *One* function.

The Fourier coefficients of a function f from $(0,1)^n$ to $\mathbb{R}$ are denoted $\hat{f}(u)$ for $u \in \mathbb{N}^n$ defined as

$$\hat{f}(u) = \int\limits_S f(u)e^{2\pi i(ux)}dx$$

The space of functions from any subset S of $\mathbb{R}^n$ to $\mathbb{R}$ is denoted $\mathcal{F}_S$, and is equipped with the usual inner product: if S is a discrete and finite set i.e, the space is finite dimensional, then $< f,g >_S =_{def} 1/|S| \sum\limits_x f(x)g(x)$, and if S is $(0,1)^n$, i.e, the space is infinite dimensional, then $< f,g >_S =_{def} \int\limits_S f(x)g(x)dx$. Sometimes the inner product is defined with respect to a distribution $\mathcal{R}$ over S; i.e, $\mathcal{R}(x) \geq 0$ for $x \in S$ and $\sum\limits_x \mathcal{R}(x) = 1$, or $\int\limits_S \mathcal{R}(x)dx = 1$. Then $< f,g >_{S,\mathcal{R}} =_{def} \sum\limits_x \mathcal{R}(x)f(x)g(x)$ or $< f,g >_{S,\mathcal{R}} =_{def} \int\limits_S \mathcal{R}(x)f(x)g(x)dx$. The set of parity functions $\chi_u : u \in \mathbb{F}_2^n$ are mutually orthogonal in $\mathcal{F}_{\{-1,1\}^n}$, but not neccessarily in $\mathcal{F}_{\{-1,1\}^n,\mathcal{R}}$, for arbitrary distributions $\mathcal{R}$. However, these functions constitute a complete (possibly redundant) basis for $\mathcal{F}_{\{-1,1\}^n,\mathcal{R}}$, for any $\mathcal{R}$. The norms are defined as usual: $||f||_{1,S} =_{def} \sum\limits_{x\in S} |f(x)|$; or $\sum\limits_S |f(x)|dx$; and $||f||_{\infty,S} =_{def} sup_{x\in S} |f(x)|$. However, the 2-norm $||f||_{2,S} =_{def} \sqrt{< f,f >_S}$. The norms could also be defined with respect to a distribution $\mathcal{R}$ over S in the usual way. For example, $||f||_{2,S,\mathcal{R}} =_{def} \sqrt{< f,f >_{S,\mathcal{R}}}$. The projection $f|_X$ for X being a subspace of $\mathcal{F}_S$ is the component of f in X. Thus $f|_X + f|_{X^\perp} = f$ and if M forms an orthonormal basis for X, then $f|_X = \sum\limits_{g\in M} < f,g >_S g$.

The following are basic properties of the Fourier spectra of Boolean functions.

Fact 2.1. *For functions f and g over $\mathbb{F}_2^n$, the following hold.*

(i) Parseval's identity:

$$||f||_2^2 = (1/2^n) \sum_{x\in\mathbb{F}_2^n} f^2(x) \;=\; \sum_{x\in\mathbb{F}_2^n} \hat{f}^2(x) = ||\hat{f}||_2^2.$$

This identity holds also when the Fourier coefficients $\hat{f}(x)$ are replaced by the coefficients when f is expressed as a linear combination of any set of orthonormal basis functions.

(ii) The value of the transform at 0^n is the expected value of the function:

$$\hat{f}(0^n) = (1/2^n) \sum_u f(u).$$

The following theorem expresses nonapproximability of Boolean functions using equivalent statements of approximability. These are proven directly using the duality principle in [26] and are proven (and expressed) differently in [11] and [8] as generalizations of the "discriminator or correlation lemma" of [16].

THEOREM 2.2. *Let f be a Boolean function and B a set of Boolean functions, and let $M \subseteq B$ consist of independent Boolean functions.*

(i) The following are equivalent.
• There does not exist an approximation $g \in span(M)$ with $sign(f) = sign(g)$.
• There exists an approximation $l \in span(M)^{\perp}$ with $||l||_1 > 0$ and $sign(f(x)) = sign(l(x))$, whenever $l(x) \neq 0$.

(ii) The following are equivalent.
• There does not exist an approximation $g = \sum_{h \in M} a_h h$ with $\sum_{h \in M} |a_h| \leq 1$, and $\epsilon \leq |g(x)| \leq 1$ everywhere, and $sign(f) = sign(g)$.
• There exists an approximation l close to $span(M)^{\perp}$ with $||l||_1 = 1$ and $sign(f) = sign(l)$, where ever $l \neq 0$. By "close to $span(M)^{\perp}$" we mean that $|\sum_x l(x)h(x)| \leq \epsilon$ for all $h \in M$.

Complexity (resource) bounds on a function are always understood as being in terms of the number of its variables. In the case of threshold complexity classes, the complexity of functions is given by the dimension of a good approximating space spanned by specific kinds of basis functions. Some common basis functions besides *Parity*, *One* and the parity functions χ_s are the following: the functions $\wedge_{u,v}$ for disjoint $u, v \in \mathbb{F}_2^n$ are the *And* functions, and are defined as

$$\wedge_{u,v}(x) := \bigwedge_{i \in u} x_i \bigwedge_{i \in v} \bar{x}_i;$$

When viewed as mapping from $\{1, -1\}^n$ to $\{1, -1\}$, $\wedge_{u,v}(x)$ takes the value -1 exactly when all the x_i's with $i \in u$ are -1's, and all the x_i's with $i \in v$ are 1's.

Some complexity classes of functions over $\{-1, 1\}^n$ that the paper deals with are the following. The class PT_1 (QT_1) consists of Boolean functions f that are approximable by a function g in the span of (quasi)polynomially many basis

parity functions χ_s, with $||f-g||_\infty < 1$.
The class LT_1 consists of Boolean functions f that are approximable by a function g in the span of basis parity functions χ_s, with $|s| \leq 1$ and $||f-g||_\infty < 1$.
In general, LT_d is the class of Boolean functions f that are approximable by a function g in the span of polynomially many basis functions in LT_{d-1}, with with $||f-g||_\infty < 1$. The class $\hat{LT}_d$ is the class of Boolean functions f that are approximable by a function g which is in the span of polynomially many basis functions g_i in $\hat{LT}_{d-1}$, with the additional condition that when $g = \sum_i a_i g_i$, the coefficients a_i are normalized to $\sum_i |a_i| \leq 1$, and each a_i is a rational whose denominator is polynomially bounded. It should be noted that often in the literature, the normalization of $\sum_i |a_i|$ is removed, the condition $||f-g||_\infty < 1$ is simply written as $sign(f) = sign(g)$, and the a_i's are taken to be polynomially bounded integers. Finally, $AC^0[d]$ is the class of functions computable by (constant) depth d $\{\wedge, \vee, \neg\}$-circuits of polynomially bounded size.

Some algebraic complexity classes of functions over $(0,1)^n$ that the paper deals with are described by threshold circuits, where the gates at the bottom level take inputs from $(0,1)^n$. These algebraic models of computation are related to arithmetic circuits, (see [7] and see [23]), and algebraic, fixed degree (linear) decision trees and algebraic computation trees studied in, for example, [29] and [6].

3. Types of approximation

In this section, we define four types of approximation, provide methods for establishing nonexistence of approximations, study the properties of approximations from a Boolean basis, explore the transitive nature of approximability, and give methods for obtaining scalar product estimates. We also provide applications of all of the above, for proving threshold and communication complexity lower bounds. The applications are usually sandwiched by ◯'s

Definition 3.1. *Let f be a Boolean function and B a set of Boolean functions, over a domain $S \subseteq \mathbb{R}^n$, and let $M \subseteq B$ consist of independent Boolean functions.*

(1) A function $g \in span(M)$ is a **uniform** *approximation to f from $span(M)$ if $sign(f) = sign(g)$.*

(2) *A function $g \in span(M)$ is an ϵ-**close 2-norm** approximation to f from $span(M)$ if $||g||_2 \leq 1$ and $|\sum\limits_x f(x)g(x)|$ (or $|\int\limits_S f(x)g(x)dx|$) is at least ϵ.*

(3) *A function $g \in span(M)$ is an ϵ-**high-energy** approximation to f from $span(M)$ if $g = \sum\limits_{h \in M} a_h h$, with $\sum\limits_{h \in M} |a_h| \leq 1$, and $|\sum\limits_x f(x)g(x)|$ (or $|\int\limits_S f(x)g(x)dx|$) is at least ϵ.*

(4) *A function $g \in span(M)$ is a ϵ-**close** approximation to f from $span(M)$ if $g = \sum\limits_{h \in M} a_h h$, with $\sum\limits_{h \in M} |a_h| \leq 1$, and $sign(f) = sign(g)$, with $|g(x)| \geq \epsilon$ everywhere.*

Remark 3.2.

- *An ϵ-close approximation is an ϵ-high energy approximation which is in turn an ϵ-close 2-norm approximation.*
- *An ϵ-high energy approximation is an ϵ-close approximation over some subdomain whose measure is an ϵ fraction of the measure of S.*
- *An ϵ-close approximation is a uniform approximation, but high energy or close, 2-norm approximations need not be uniform approximations, nor viceversa.*
- *The first two kinds of approximation depend only on the space $span(M)$ and are independent (even of the Boolean-ness) of the basis M, but the latter two depend on the actual basis M.*

We now sketch some instances where these approximations arise in complexity applications. (See [26] for a unified approximation theoretic treatment of these and other complexity applications).

○ Here, we only consider the case where the domain S of the function f being approximated is $\{-1, 1\}^n$, so the space of functions is 2^n-dimensional. The complexity applications for continuous domains S is discussed in the Section 5. The function f has a uniform approximation from $span(M)$ if f can be expressed as a weighted threshold of functions in M. Similarly, f has an ϵ-close approximation from $span(M)$, with ϵ being at least $1/poly(n)$ for some polynomial $poly$ if f is an unweighted threshold of functions in M and $|M|$ is bounded by a polynomial. Many natural Boolean functions are unweighted thresholds of linear functions, or simply linear combinations of polynomially many Boolean functions, for example, any function with a low communication complexity, is a simple linear combination of cross product functions by the following fact.

FACT 3.3. *If the deterministic communication complexity of a Boolean function f is at most $\log m$, then f is exactly interpolated by $\sum\limits_{i \leq m} r_i + (m-1)$, where the r_i are cross-product functions (characteristic functions of cross-product sets, or "rectangles,") not including the constant function.*

In addition, close approximations naturally arise from probabilistic communication complexity, using the following fact.

FACT 3.4. *If the $(1-\epsilon)$-error probabilistic communication complexity of a Boolean function f is at most $\log m$, then there is a very close approximation g with the same sign as f, of the form $g = \sum\limits_{i \leq m^2} a_i r_i$, where, as usual, $\sum\limits_{i \leq m} |a_i| \leq 1$, the the r_i are cross-product functions, and $|g(x)| \geq \epsilon/m$ everywhere.*

○

The next proposition shows that establishing non-existence of high-energy and close approximations is fairly straightforward, while establishing non-existence of uniform approximations is not, which is not surprising since f has a uniform approximation from $span(M)$ if and only if f is computable by a weighted threshold of functions in M.

PROPOSITION 3.5. *Let f be a Boolean function and B a set of Boolean functions, over a domain $S \subseteq \mathbb{R}^n$, and let $M \subseteq B$ consist of independent Boolean functions.*
(1) *Establishing non-existence of a uniform approximation to f, is equivalent to showing that there is a function l such that $< l, g >= 0$ for every $g \in M$ (or in $span(M)$) and $sign(f) = sign(l)$ wherever l is non-zero.*
(2) *To establish the non-existence of ϵ-close 2-norm approximation to f from $span(M)$ is equivalent to showing that the projection $f|_{span(M)}$ has 2-norm strictly less than ϵ. This shows that there is no high-energy approximation to f from any normal* basis *for $span(M)$.*
(3) *To establish non-existence of ϵ- high energy approximations from $span(M)$ to f, and therefore of ϵ-close approximations, it is sufficient to show that for all functions h in M, the scalar product $< f, h >< \epsilon$.*
(4) *Establishing non-existence of ϵ-close approximations is equivalent to showing that there is some distribution $\mathcal{R}$ on the domain such that for all functions h in M, the scalar product $< f, h >_{\mathcal{R}} \leq \epsilon$.*

The validity of all of these methods is either straightforward or follows from Theorem 2.2. We discuss implications of these methods for proving complexity lower bounds.

○ Applying Method (1) is by no means easy, and hence uniform nonapproximability results and the resulting lower bounds for weighted thresholds are notoriously difficult and rare (recall that LT_2 is not known to differ from NP). Exceptions include a few results in [2] (relying mainly on univariate techniques) and [20] for weighted thresholds of parities, [21] for weighted thresholds of ands, [8] for weighted thresholds of $AC^0[d]$ functions, and [26] for stronger lower bounds for weighted thresholds of these and other classes of functions. All of these papers, in effect, use Method 1, although all but the last two papers do not state so explicitly. The last paper explicitly constructs a function l as in the proposition. It should, however, be noted that in 4.5 the result of [20] for weighted thresholds of parities is given an alternate proof, this time explicitly using the stability of the parity functions as a basis family and a method close to Method (2) above.

The last method is the most commonly used, follows directly from Theorem 2.2(ii), and is referred to as a generalization of the "discriminator/correlation/ discrepancy method" in [11] and [16]. Methods 2 and 3 are fairly easy to use for proving lower bounds, if the hard function f can be simply constructed to have small scalar products with all the functions in the basis family B, perhaps with respect to some distribution $\mathcal{R}$. These methods have been used in [16], [19], [22], [9], [11], to show non-existence of close approximations, and therefore, to show unweighted threshold lower bounds.

Often, the stronger second and third methods of showing non-existence of close 2-norm and high energy approximations are used for showing non-existence of close approximations, but high energy approximations have not been considered explicitly. The second method is used, in effect, in [22], but is referred to as the "variation rank or geometric method," (see explanation following Observation 4.9).

The last three methods are also viable for showing the communication complexity lower bounds of [16], [11], [15], [18], [24], [10], [12], [3], and [9], although the original proofs of these bounds use more ad hoc methods.○

3.1. Boolean bases and transitive approximability. The next result shows that close approximations from Boolean functions have a surprising property which results in a general complexity upper bound as corollary. The proof of the theorem follows closely along the lines of the proof of [5]. They showed a special case of the corollary, namely that $PL_1 \subseteq PT_1$.

Theorem 3.6. *If f is closely approximable by $g \in span(B)$, with $|g(x)| \geq 1/poly(n)$ for all x, where B is a Boolean basis, then f is closely approximable by $g' \in span(M)$ with $|g'(x)| \geq 1/poly^3(n)$ everywhere, where $M \subseteq B$, and $|M| \leq poly^3(n)$.*

Proof. The idea of the proof is to obtain a random function g' which is constructed as a linear combination of at most $poly^3(n)$ functions in B, and show that with non-zero probability, g' has the same sign as g everywhere. Let $g = \sum\limits_{h \in B} a_h h$, where $\sum\limits_{h \in B} |a_h| = 1$. Define a probability distribution over B, as $p(h) =_{def} |a_h|$. Construct the random function g' as

$$g' =_{def} \sum_{i=1}^{N} Z_i,$$

where the Z_i are N i.i.d. random variable, taking the values h or $-h$ for $h \in B$, and defined as:

$$Z_i = sign(a_h)h \text{ with probability } p(h).$$

I.e, the terms that go into g' are picked at random.

Notice that for each fixed x, the *value* $g'(x)$ is a random variable, given as the sum of N i.i.d. random variables corresponding to the *values* $Z_i(x)$, $1 \leq i \leq N$, each distributed as follows:

$$Z_i(x) = \begin{cases} 1 & \text{with prob } \sum\limits_{h: a_h h(x) > 0} p(h), \\ -1 & \text{with prob } \sum\limits_{h: a_h h(x) < 0} p(h). \end{cases}$$

It is not hard to see that at any given x, the expected value of $Z_i(x)$ is

$$\sum_{h: a_h h(x) > 0} p(h) - \sum_{h: a_h h(x) < 0} p(h),$$

which is nothing but $g(x)$, and the variance of $Z_i(x)$ is

$$E(Z_i(x)^2) - g^2(x) = 1 - g^2(x).$$

Hence for any x, $E(g'(x)) = Ng(x)$, and $Var(g'(x)) = N(1 - g^2(x))$.

Choosing $N \geq 2n/min_x g^2(x) \geq poly^3(n)$, using the central limit theorem, we obtain that for each x, the probability that $|g(x) - g'(x)| \geq |g(x)|$, (i.e, the probability that $g'(x)$ does not have the same sign as $g(x)$) is at most $e^{-n} < 2^{-n}$, and by the union bound, the probability that for some x the signs of g and

g' differ is strictly less than 1. Thus with non-zero probability, a function g' exists that is a linear combination of at most $poly^3(n)$ functions in B, and has the same sign as g everywhere. Furthermore, by the nature of the construction of g' as a linear combination with polynomially bounded *integer* weights, it follows that f has a $1/poly^3(n)$- close approximation (obtained by normalizing g'), from the span of polynomially many functions in B, and in fact, g' directly provides an expression for f as an unweighted threshold of functions in B. □

Next, we give results of the following form: "f is approximable from the span of a set of m_1 functions g_i in some basis B, and g_i are all approximable from the span of a set of m_2 "simple" functions h, then f is approximable from the span of $m_1 m_2$ "simple" functions." *These results will follow directly from 3.5 and 3.1.*

Results of this nature are useful for building on previous nonapproximability results. For example, a result of the above form, together with the nonapproximability of f from the span of $m_1 m_2$ simple functions would imply that one of the approximability hypotheses is false.

Moreover, statements of this type remain meaningful when the word "not approximable" is replaced by "small scalar product." This will be used in Subsection 3.2.

THEOREM 3.7. *Let f be a Boolean function, B a set of Boolean functions, and Q a set of "simple" Boolean functions.*

(1) If
• f has a high energy approximation g, with $|\sum_x f(x)g(x)| \geq \delta$, from the span of functions in B
and
• every function in B is the linear combination of Boolean functions in Q, with the sum of the absolute values of the coefficients bounded by m,
then
$| < f, h > | \geq \delta/m$, for some function $h \in Q$.

(2) If
• f has a high energy approximation g from the span of functions in B with $|\sum_x g(x)f(x)| \geq \epsilon$,
and
• for every $g \in B$ there is a close approximation h, with the same sign as g, and in the span of m Boolean functions in Q, satisfying $|h(x)| \geq \delta$ for all x
then

(i) there is a set of m of functions $h^* \in Q$, and a subset S consisting of at least $(1+\epsilon)/2$ fraction of the domain, such that for every distribution $\mathcal{R}$ over S, $| < f, h^* > |_{\mathcal{R}} \geq \delta$, for at least one of the m functions h^*; and

(ii) $| < f, h^* > | \geq \epsilon + \delta - 1$, for some $h^* \in Q$.

(3) If
• f has a close approximation g from the span of a set of m_1 functions g_i, (i.e, g has the same sign as f with $|g(x)| \geq \delta$ for all x,)
and
• each g_i has a close approximation h_i from the span of a set of m_2 functions h_{ij}, with $|h_i(x)| \geq \epsilon$ for all x, and $1 - \epsilon < \delta/m_1$,
then
there is a close approximation h^* to f from the span of the m_1m_2 resulting functions $h_{ij} \in Q$ with $|h^*(x)| \geq \epsilon\delta$ for all x.

(4) If
• f has a close approximation g from the span of m_1 functions $g_i \in B$ with with $|g(x)| \geq \delta$ for all x,
and
• each g_i can be expressed as a linear combination of functions in a set Q with the sum of the absolute values of the coefficients bounded by m_2,
then
f has a close approximation h^*, with $|h^*(x)| \geq \delta/m_2$ everywhere. It follows from 3.5(3) that for every distribution $\mathcal{R}$, there is a function $h \in Q$ such that $< f, h >_{\mathcal{R}} \geq \delta/m_2$. Moreover, by 3.6 f has a close approximation h^* from the span of at most $4m_1^2m_2^2$ functions in Q.

(5) If
• f has a high energy approximation g, with the same sign as f, from the span of m_1 functions in B with $|\sum_x g(x)f(x)| \geq \epsilon$,
and
• each function in B has an approximation with the same sign from the span of m_2 simple functions $h \in Q$,
then
f is approximable on some subset S consisting of at least $(1+\epsilon)/2$ fraction of the domain from the span of m_1m_2 functions $h \in Q$.

PROOF. For (1) (2) and (5), since f has a high energy approximation $g \in$ *span*(B), by 3.5(2), there is a function $g^* \in B$ such that $|\sum_x f(x)g^*(x)| \geq \delta$. (1) now follows immediately.

For (2) and (5), notice that since both f and g^* are Boolean, f and g^* must coincide in sign on at least a $(1+\epsilon)/2$ fraction of the domain. (5) now follows immediately.

For (2), since g^* has a close approximation h from the span of m functions $h^* \in Q$, (i) follows by 3.5(3). The consequence (ii) follows from the fact that h is a high energy approximation to g^* with $|\sum_x g^*(x)h(x)| \geq \delta$. Thus we obtain the existence of 3 Boolean functions f, g^* and h^* such that

$$< f, g^* > \geq \epsilon, \text{ and } < g^*, h^* > \geq \delta.$$

Therefore $< f, h^* > \geq \epsilon + \delta - 1$.

For (3), we use the fact that the functions g_i have a very close approximation h_i from the span of m_2 functions h_{ij}, since $1 - \epsilon \leq \delta/m_1$. Now, to form the required close approximation to f from the span of the $m_1 m_2$ functions h_{ij} in Q, modify g as follows: simply replace each of the g_i's that form g, with the corresponding approximation h_i.

For (4), choose the close approximation h^* to f as g/m_2. □

Next, we give some complexity applications of the properties of Boolean bases in 3.6 and the transitive nature of approximability in 3.7, in conjunction with the nonapproximability characterizations in 3.5. Each application is sandwiched by $\bigcirc$'s.

$\bigcirc$ 3.7 (1) and (2) form the backbone of the lower bounds (nonapproximability results) of [16] and [11] that $\hat{LT}_3 \not\subseteq \hat{LT}_2$, $LT_1 \not\subseteq \hat{PT}_1$, and $PT_1 \not\subseteq \hat{LT}_2$. [11] uses communication complexity upper bounds to prove approximability of the functions in the relevant class B by the span of a few cross-product functions (in Q), and then, in effect, uses of 2.2 (2) to show that f is not appropriately approximable from the span of few cross-product functions. The desired nonapproximability of f from the span of few functions in B then follows from 3.7. The papers, especially [11] employ the communication complexity paradigm throughout instead of treating the issue as approximability from cross-product functions.

It should be noted that while it is often easier to show that *specific* functions are appropriately approximable from the span of cross-product functions by giving a *direct* upper bound on the communication complexity, the word "communication complexity" can otherwise be removed from all (lower bound) proofs involving threshold functions, or linear (non) approximability results, without making the proofs any more difficult, or any less intuitive. In fact, translating "low communication complexity" as "appropriate approximability by few cross-product functions" allows one to take advantage of transitive approximability. This, in turn, allows a natural extension of approximability notions, such as 3.7 that are already being employed and and often makes the proofs that employ the "communication complexity method" shorter and more transparent.

The hypothesis of 3.7 (2) also holds when the functions g are functions in $\hat{LT}_1$, with $h = \sum\limits_i a_i x_i + a_0$ and, as usual, $\sum\limits_i |a_i| \leq 1$ and the a_i are rationals with denominators bounded above by $1/\delta$. Here the simple functions in Q are the linear monomials and the constant function. In this case, showing the the negation of 3.7 (2) (ii) for $1/\delta$ being polynomially bounded would simply mean that $g \notin \hat{LT}_1$. ○

○ An example application of 3.7(3) is the following result proved in [11] using communication complexity. This can be obtained directly from 3.7(3) using the useful fact in [11] that every LT_1 function has a very close approximation from the span of polynomially many $\hat{LT}_1$ functions.
"A circuit with an unweighted linear threshold gate on top, arbitrary linear threshold gates at the middle level, and gates from a class C in the lowest level can be simulated by a circuit with exactly the same gate on top, unweighted linear threshold gates in the middle level and exactly the same gates from C at the bottom."

Another example application of 3.7(3) is the following: LT_1 functions have a very close approximation from the span of few $\hat{LT}_1$ functions by a result of [11]. The approximation is sufficiently close that the weighted gates at the middle level can be removed by replacing them by their approximation which is a linear combination of unweighted threshold gates. Moreover, $\hat{LT}_1$ functions have a very close approximation from the span of few cross-product functions by a simple probabilistic communication complexity upper bound, (in fact they are even interpolable from the span of a few more cross-product functions by the straightforward deterministic communication complexity upper bound). Therefore, LT_1 functions have a very close approximation from the span of few cross-product functions, and this approximation does yield a probabilistic communication complexity upperbound for LT_1 functions, although this is not directly a consequence of Fact 3.4 alone. In general, it seems that approximability results are a viable method for proving communication complexity upper bounds as well, in addition to the usual method of finding an appropriate communication protocol. ○

○ Non-approximability results from cross-product functions, on the other hand, are theoretically stronger than lower bounds on communication complexity, but nevertheless provide a viable method of proving such lower bounds, since in practice, several of the known lower bounds on communication complexity such as the results of [16], [11], [15], [18], [24], [9], [12], [3], [9] actually yield the stronger nonapproximability results from cross product functions, which can be obtained independently using the methods of this Section. For example, using 3.7(1) and (2), a lower bound of $\log m$ on the communication complexity of a function g can be obtained by showing g is not a linear combination -with small coefficients - of m cross-product functions or by showing that there is a function f such that $| < f, s > | \leq \epsilon$ for cross-product functions s, and yet $| < f, g > | \geq m\epsilon$. Similarly, a lower bound of $\log m$ can be obtained on the $(1-\delta)/2$-error probabilistic communication complexity of g as follows: show the negation of 3.7 (2) (i) that for each set of m cross-product functions, for some ϵ and each subset S with $|S| \geq (1+\epsilon)/2$ fraction of the domain, there is a distribution $\mathcal{R}$ over S with $< f, s >_{\mathcal{R}} \leq \delta$, but $< f, g > \geq \epsilon$. ○

○ A special version of 3.7(4) is used in the proofs of [20] and [17] i.e, the *approximability* result that $AC^0[2] \subseteq \hat{PT}_1$. [20] gives a separate probabilistic argument. [17] gives an algorithm to find the approximating function. Our proof of 3.7(4) subsumes this result, is straightforward from 3.5(3), and 3.6, and clarifies exactly where the orthonormality of the functions in Q is needed, and moreover gives the new result complexity result below.

Corollary 3.8. *Let f be expressible as an unweighted threshold of (polynomially many) functions g_i, and let each g_i be a linear combination $\sum_{h\in B} a_h h$ for any Boolean family B, with $\sum_{h\in B} |a_h|$ being polynomially bounded. Then f can be expressed as an unweighted threshold of (polynomially many) functions in B. For example, taking B to be the family of parity functions, it follows that $\hat{LT}_1$- $PL_1 \subseteq \hat{PT}_1$. I.e, an unweighted threshold of polynomially many functions in PL_1 can be simulated by an unweighted threshold of polynomially many Parity functions. Here the orthonormality of the Parity functions is purely incidental and irrelevant. Only their Boolean-ness is relevant.*

○

3.2. Small correlation and nonapproximability. The next two theorems provide general methods for showing that the scalar product $< f, g >$ is small, for some fixed f, and for all functions g in some class B which is modelled after LT_1. Such scalar product estimates are needed for employing Methods (2) and (3) of Proposition 3.5 to prove nonexistence of close and high energy approximations to f. In the next section, we will show that when the family B is a "stable basis family", then in fact, scalar product estimates and Method (2) and (3) can be used for proving uniform nonapproximability of f as well, (which is much more valuable in the quest of proving weighted threshold lower bounds).

Viceversa, many of the methods given below - for obtaining scalar product estimates - will, in their turn, rely on nonapproximability results.

It is generally assumed that every $g \in B$ is either in the span of simple functions or is approximable (to some degree of closeness) in the $||.||_\infty$ norm from the span of simple functions.

Showing that the scalar product $< f, g >$ is small is a combination of two tasks. First, a transitive approximability relationship is shown roughly of the form: if $< f, g >$ is large, and g is closely approximable from the span of simple functions, then f must be also be closely approximable in some sense from the span of few simple functions. Second, a strong nonapproximability result is proven that f cannot be thus approximated. The second part is stronger than what is required: we only require that f should not be approximable by linear combinations of those simple functions that are used to approximate the functions $g \in B$. But such sets of simple functions are hard to isolate, so one is usually forced to consider all sets. The second part could be based on any of the methods of the previous four subsections. The next theorem presents natural combinations of these two parts. The proofs are straightforward and use Theorem 3.7 and Theorem 2.2.

THEOREM 3.9. *Let f be a Boolean function, B a set of Boolean functions, and Q as set of "simple" Boolean functions.*

(1) If every $g \in B$ is the linear combination of functions h in Q with the coefficients summing, in absolute value, to at most m, and $< f, h >\leq \epsilon$ for each $h \in Q$. Then $< f, g >\leq m\epsilon$. (Application of 3.7(1)).

(2) If for every $g \in B$ there is a function h, with the same sign as g, and in the span of m functions in Q, satisfying $|h(x)| \geq \delta$ for all x; and furthermore, if f is a function such that $< f, q >\leq \epsilon$, for every $q \in Q$, then $< f, g >\leq \epsilon + 1 - \delta$. (Application of 3.7(2)).

(3) If g is in the span of m functions $h \in Q$, and if for every subset S of the domain that contains more than an $(1+\epsilon)/2$ fraction of the points and

any set M of m functions in Q, f is not approximable from $span(M)$ on S, then $< f, g > \leq \epsilon$. Any appropriate method of this paper (including those that follow in the next section) could be used to show that f is not approximable on S from $span(M)$.

○ ○ From 3.9(3) we get the following for the special case where the functions in B are LT_1 functions, and Q is the set of the linear monomials.

We use the following straightforward observation which reduces nonapproximability into a purely geometric problem that is amenable to decomposition.

Observation 3.10. *Any set M of Boolean functions provides a map T_M from $\{-1,1\}^n$ to $\{-1,1\}^{|M|}$ by mapping $x \rightarrow (h_1(x), \ldots, h_{|M|}(x))$. Thus any Boolean function f over $\{-1,1\}^n$ is transformed into a function f_M over the image of T_M. I.e, $f_M(x) =_{def} f_M(T_M(x))$. Now, the convex hulls of $f_M^{-1}(1)$ and $f_M^{-1}(-1)$ intersect if and only if f does not have an approximation with the same sign from $span(M)$. Two convex sets intersect if any of their subsets intersect, and thus this observation allows a nonapproximability question to be decomposed.*

Theorem 3.11. *Given a Boolean function f, and a function $g \in LT_1$. if for each subset $S \subseteq \{-1,1\}^n$ with $|S|$ containing more than an $(1+\epsilon)/2$ fraction of points, one of the following holds, then $< f, g > \leq \epsilon$.*

(i) $ConvexHull(f^-1(1) \cap S) \cap ConvexHull(f^-1(-1) \cap S)$ is non-empty (follows from 3.9(3) and 3.10).

(ii) The linear monomials and the constant function One continue to remain a polystable basis on S, but $\mid < f, x_i >_S \mid$ and $\mid < f, One >_S \mid$ are $\leq 1/(n+1)poly(n+1)$ (follows from 3.9(3) and 4.2 in the next section).

(iii) There is a set Π of permutations of the variables such that f is invariant under the permutations in Π, and for all linear functions g, $\sum_{\pi\in\Pi} g(\pi)(x)$ has the symmetric form $a\sum_i x_i + a_0$, for some a and a_0. Finally, f is not approximable by any symmetric function $a\sum_i x_i + a_0$, over $\bigcap_{\pi\in\Pi} S(\pi)$ (follows from 3.9(3), and 4.9 in the next section).

(iv) There is a set Π of permutations of the variables such that f is invariant under the permutations in Π, and for some point a in $f^-1(1) \cap S$ and b in $f^-1(-1) \cap S$, $\sum_{\pi\in\Pi,\pi(a)\in S} \pi(a)_i = 0$, and $\sum_{\pi\in\Pi,\pi(b)\in S} \pi(b)_i = 0$, for all i; (or $\sum_{\pi\in\Pi,\pi(a)\in S} \pi(a)_i = \sum_{\pi\in\Pi,\pi(b)\in S} \pi(b)_i$ for all i, and $|\{\pi \in \Pi : \pi(a) \in S\}| = |\{\pi \in \Pi : \pi(b) \in S\}|$) (also follows from 3.9(3), and 4.9 in the next section).

As a direct application of any one of the methods of 3.11, we obtain the following.

FACT 3.12. *The read-once $AC^0[3]$ functions of depth 3 satisfy: $< RO[3], g > \leq < RO[3], One >= 2\hat{RO}[3](0^n)$, for all $g \in LT_1$.*

PROOF. Any of the methods of 3.11 can be used to prove this result. We illustrate the convex-hull intersection argument (neccessary for employing 3.11(i)) We prove the straightforward result that $RO^0[2] \notin LT_1$. The proof of the fact is carried out along the same lines, by showing $RO^0[3]|_S \notin LT_{1,S}$ for any large subdomain S. We will consider the canonical $AC^0[2]$ function $RO[2](x) := \bigvee\limits_{j=1}^{k} \bigwedge\limits_{i=1}^{k} x_{ij}$. If an LT_1 function g equals $RO[2]$, then, in particular, $g(x) = -1$ when $x_{i1} = -1$, for $1 \leq i \leq k$ and $x_{ij} = 1$, for all i and all $j \neq 1$; and $g(y) = -1$ when $y_{i2} = -1$, for $1 \leq i \leq k$ and $y_{ij} = 1$, for all i and all $j \neq 2$.
Moreover, for every z, w such that $x + y = z + w$ (where the $+$ stands for addition in $\mathbb{R}^n$), either $g(z) = -1$ or $g(w) = -1$, since g, being in LT_1, is the characteristic function of a halfspace of $\mathbb{R}^n$. However, for our chosen function $RO[2]$, we can find z and w with $z + w = x + y$, with both $f(z) = f(w) = 1$. For example, choose z with $z_{i1} = -1$ and $z_{i2} = -1$ for $1 \leq i \leq k/2$ and $z_{ij} = 1$, for all i and all $j \neq 1, 2$, and choose w such that $x+y = z+w$. This contradicts the assumption that $g = f$. □

○

In the next section we discuss conditions on the basis family B under which proving non-existence of high-energy approximations (Method 3 above) is sufficient to show non-existence of *uniform* approximations, and therefore sufficient to show weighted threshold lower bounds (which, in general, requires the more difficult Method 1).

4. Stable basis families and uniform approximation

In this section, we define four notions of stability and show that if a basis is stable even in the weakest sense, then showing uniform nonapproximability reduces to showing "high energy" nonapproximability. We apply this idea to give an alternative proof of the result that $AC^0[3] \not\subseteq QT_1$. In addition, we show how this idea can be adapted to show uniform nonapproximability for certain specific unstable bases. Finally, we show how stable basis families are useful for proving approximability results and use this to give a new complexity result that unweighted thresholds of PL_1 functions can be obtained as unweighted thresholds of parities.

Definition 4.1. *(1) A set B of functions forms a poly* **stable** *basis family, if for all $M \subseteq B$ of independent functions h, and values a_h, the functions satisfy $|| \sum_{h \in M} a_h h||_2^2 \geq 1/poly(|M|) \sum_{h \in M} a_h^2$. When poly is just 1, then B is an orthonormal basis.*

(2) A set B of functions forms a poly **quasistable** *basis if for all $M \subseteq B$ of independent functions, and values a_h, there exists $M' \subset B$ of independent functions, and values a'_h, with $|M'| \leq poly(|M|)$, $sign(\sum_{h \in M'} a'_h h) = sign(\sum_{h \in M} a_h h)$, and furthermore, for all values a'_h, $|| \sum_{h \in M'} a'_h h||_2^2 \geq 1/poly(|M'|) \sum_{h' \in M'} a'^2_h$.*

(3) A set B of functions forms a poly **strong stable** *basis* **for approximating** *a function f, if for all $M \subset B$ of independent functions and for all values a_h that satisfy $sign(\sum_{h \in M} a_h h) = sign(f)$, it holds that $|| \sum_{h \in M} a_h h||_2^2 \geq 1/poly(|M|) \sum_{h \in M} a_h^2$. This notion can also be defined for quasistability.*

(4) A set B of functions forms a poly **weak stable** *basis* **for approximating** *a function f, if for all $M \subset B$ of independent functions, for at least one value of a_h that satisfies $sign(\sum_{h \in M} a_h h) = sign(f)$, it holds that $|| \sum_{h \in M} a_h h||_2^2 \geq 1/poly(|M|) \sum_{h \in M} a_h^2$. This notion can also be defined for quasistability.*

Next we show that if a Boolean basis B is stable even in the weakest sense, then showing uniform nonapproximability reduces to showing high energy nonapproximability, which, as pointed out earlier, is much easier to show, using the Method 3 in Proposition 3.5.

Theorem 4.2. *Let f be a Boolean function, B a family of poly-stable Boolean functions, and let $M \subseteq B$. Then any approximation $g \in span(M)$ with $sign(f) = sign(g)$, $g = \sum_{h \in M} a_h h$, and $\sum_{h \in M} |a_h| \leq 1$, is a high energy approximation satisfying $||g||_1 = |\sum_x f(x) g(x)| \geq 1/(|M| poly(|M|))$.*

This result extends to the case where B forms a quasistable basis in the weak sense with respect to approximating f.

Proof. We show that any function g, which is a linear combination as in the theorem, satisfies $||g||_1 \geq 1/(|M|poly(|M|))$, using the fact that B is *poly* stable. Since, in addition, $sign(f) = sign(g)$, it follows that $|\sum_x f(x)g(x)| \geq 1/(|M|poly(|M|))$, since f is Boolean.

Since B is stable,

$$\sum_{h \in M} a_h^2 \leq poly(|M|) \sum_x g^2(x) = 2^n poly(|M|)||g||_2^2.$$

Assuming without loss that $\sum_{h \in M} |a_h| = 1$, it follows that

$$\sum_{h \in M} a_h^2 \geq 1/|M|,$$

and hence $2^n||g||_2^2 \geq 1/(|M|poly(|M|))$. Moreover, the functions h are Boolean, and therefore $||g||_\infty \leq 1$, since $\sum_{h \in M} |a_h| = 1$; thus

$$||g||_1 \geq 2^n||g||_2^2 \geq 1/(|M|poly(|M|)).$$

□

○ Notice that Theorem 4.2, together with the Method 3 in Proposition 3.5 directly imply, as a special case, the "spectral method" of [5] that the number of *Parities* or monomials, or any other orthonormal set needed to approximate a function exceeds the inverse of its maximum Fourier coefficient, or respectively, the maximum scalar product of the function with an element of the orthonormal set, which, in turn, yields that $PT_1 \subseteq PL_\infty^{-1}$ as a corollary. ○

Below, we obtain a generalization of this result that gives a similar lower bound on the number of elements of a polynomially stable basis that are needed to approximate a function, in terms of the maximum scalar product of the function with elements of the basis.

In general, we will show that stable basis families reduce the difficulty of Methods 1 and 2 in in Proposition 3.5 to the level of Method 3. The proof follows directly from Propositions 3.5 and 4.2.

Proposition 4.3. *Let f be a Boolean function, B a set of poly stable (not neccessarily Boolean) functions, and let $M \subseteq B$.*

- *To establish non-existence of a polynomially-close 2-norm approximation to f from $span(M)$, say a $1/p(n)$-close approximation, it is sufficient to show that* *for all functions $h \in M$, $| < f, h > |$ is at most an inverse polynomial in $|M|$, which depends on poly and p.*

• *If, in addition, B consists of Boolean functions, then to establish nonexistence of any uniform approximation to f from $span(M)$, it is sufficient to show that for all functions $h \in M$, $| < f, h > | \leq 1/|M| poly(|M|)$. This also holds when the basis is quasistable in the weak sense with respect to approximating f.*

Therefore, once a Boolean basis B is shown to be *poly*stable, then clearly f cannot be computed by a weighted threshold of any set of at most m functions in B, if $| < f, h > | \leq 1/(m \; poly(m))$ holds for *all* $h \in B$. This however, may not be easy to show, or may not even be true. To help with this difficulty, at least in the case where *poly* is linear or even 1 (which happens if B is actually orthonormal), the next theorem shows that with stable basis families, an even weaker estimate on the scalar products is sufficient for showing nonexistence of uniform approximations.

THEOREM 4.4. *Let f be a Boolean function, B a family of poly stable Boolean functions, and let $M \subseteq B$. To establish nonexistence of any uniform approximation to f from $span(M)$, it is sufficient to show that $||f|_{span(M)}||_\infty = || \sum_{h \in M} < f, h > h||_\infty < 1$; or that $\sum_{h \in M} | < f, h > | < 1/poly(|M|)$.*

PROOF. The proof uses the following claim which follows by Proposition 3.5(1).

Claim. Let f be Boolean and let X be an arbitrary space of functions. If $f \notin X$ and the projection $f|_X$ satisfies $||f|_X||_\infty \leq 1$ then f is not approximable in the ∞ norm from X, i.e, there is no $g \in X$, with the same sign as f.

The proof of the theorem follows immediately from the fact that the basis family B is both Boolean and *poly* stable. □

○ Next, we turn to a complexity application: we use Theorem 4.4 to give a simpler proof of a result of [20] that $AC^0[3] \not\subseteq QT_1$.

THEOREM 4.5. *If the $AC^0[3]$ function $f(x) := \bigvee_{i=1}^{l_1} \bigwedge_{j=1}^{l_2} \bigvee_{k=1}^{l_3} x_{ijk}$, where $n = l_1 l_2 l_3$ and $l_1 = l_2 = l_3$ has an approximant $g \in span(M)$ where M consists of parity functions χ_s, with $||f - g||_\infty < 1$, then $|M| \geq \Omega(2^{n^c})$, for any constant $c < 1$.*

PROOF. Since the parity functions form a stable, even orthonormal basis, it is sufficient to show, using Theorem 4.4 that for all sets M consisting of parity functions, $\sum_{\chi_s \in M} |\hat{f}(s)| \leq 1$, unless $|M| \geq \Omega(2^{n^c})$.

Set the quantities $q_1 := 1/2^{l_1}$, $q_2 := (1-q_1)^{l_2}$ and $q_3 := (1-q_2)^{l_3}$. It is not hard to see (see [25]) that $|\hat{f}(x)|$ is largest for $|\hat{f}(1^n)| = |1-2q_3|$ and for x for which $x_{ijk} = 1$ except for a single value of i and j. In the latter case, $|\hat{f}(x)| \leq 2q_1q_2q_3/((1-q_1)(1-q_2))$. Since

$$\sum_{\chi_s \in M} |\hat{f}(s)| \leq 2q_3 - 1 + |M| * 2q_1q_2/(1-q_1)q_3/(1-q_2)$$

the conditions on $|M|$ corresponding to $\sum_{\chi_s \in M} |\hat{f}(s)| \leq 1$ depend on the choice of q_3:
(i) $q_3 \geq 1/2$ or
(ii) $q_3 \leq 1/2$.

In case (i), $\sum_{\chi_s \in M} |\hat{f}(s)|$ is bounded by 1 as long as

$$|M| \leq (1/q_1 - 1)(1/q_2 - 1)(1/q_3 - 1);$$

and in case (ii), $\sum_{\chi_s \in M} |\hat{f}(s)|$ is bounded by 1 as long as

$$|M| \leq (1/q_1 - 1)(1/q_2 - 1);$$

We choose case (i), and ignore the latter. Now for any $c < 1$, l_1, l_2, l_3 can be chosen such that both $q_3 \geq 1/2$, and $(1/q_1 - 1)(1/q_2 - 1)(1/q_3 - 1) \geq \Omega(2^{n^c})$. which completes the proof. □

○

Is the converse of Theorem 4.4 true? The answer is no. In other words, it could be that f is not approximable from X and yet $f|_{X^\perp}$ does not behave like f. However, a version of the converse does hold, thereby giving another equivalent condition to nonapproximability, via Theorem 2.2 (1).

FACT 4.6. *We use $f|_{S,X}$ to denote the projection of f_S on X_S. Notice that this is different from taking the projection $f|_X$ and then restricting it to S. In other words, as mentioned in the background section, the space $X_S^\perp$ is not the same as taking $X^\perp$ and restricting to S. The following are equivalent.*

- *There is a nonempty subdomain S with $||f|_{S,X}||_{\infty,S} \leq 1$*
- *There is no $g \in X$, with the same sign as f.*

Proof. For the forward direction, define $l \in X^\perp$ to be 0 outside S, and $f|_{X^\perp_S} = f_S - f|_{S,X}$ on S. Since $f|_{S,X}(x) \leq 1$ for all $x \in S$, it follows that on its support, l has the same sign as f. Now apply 2.2(1).

For the reverse direction, we use a geometric argument to find the set S^* such that $f|_{X|^\perp_{S^*}} = f_S - f|_{S^*,X}$ has the same sign as f where ever non-zero.

Let C_f be a cone or orthant in $\mathcal{F}_{2^n}$ - viewed as $\mathbb{R}^{2^n}$ - and given by $\{g : sign(g) = sign(f)\}$. Notice that each facet of this cone is also a cone that contains exactly those functions that are 0 outside some subdomain S, and are either 0 or have the same sign as f on S. The entire cone C_f corresponds to $\bar{S}$ being empty. So we will denote the facet corresponding to subdomain S as $C_{f,S}$. Now $X^\perp$ is a subspace that satisfies at least one of the following properties.

(a) it completely contains some proper facet (of at least one lower dimension) $C_{f,S}$ of C_f,

(b) it cuts through a proper facet $C_{f,S}$ of C_f, or

(c) it is completely contained in the subspace formed by extending some proper facet $C_{f,S}$ of C_f to all orthants, i.e, the subspace containing exactly all functions that are 0 outside S.

In case (a), we simply choose $S^* = S$. Clearly, $f|_{S^*,X} = 0$ and we are done. In cases (b) and (c), we continue this process on a smaller cone $C_{f,S}$, starting with the function f_S instead of f, and the subspace $X^\perp_S$ instead of X. For the base case, when $|S| = 2$, in cases (b) and (c), it is easy to see that $f_S - f|_{S,X}$ does in fact have the same sign as f_S. □

Remark 4.7. *The use of subdomains as in the previous theorem is natural since in general, the nonexistence of a (close) uniform approximation can be established by showing nonexistence of a (close) uniform approximation over any distribution or subdomain $\mathcal{D}$. In particular, by Proposition 4.2, if a distribution $\mathcal{D}$ can be found such that B is a polynomially stable basis with respect to $<>_{\mathcal{D}}$, and and if $< f, h >_{\mathcal{D}}$ is small for every $h \in M$, then there is no approximation from $span(M)$ to f with the same sign. The results that follow are an application of this general idea.*

The next theorem considers a natural situation where B is an unstable family, and yet Proposition 4.2 can be applied to show nonexistence of a uniform approximation, and thereby provide a weighted threshold lower bound. Here, all the functions in B, when viewed as vectors in $\mathcal{F}_{2^n}$, form vector bundles, such that all the vectors in any one bundle are close to each other (have large scalar product), but any two bundles are nearly orthogonal to each other. The idea is to find a large enough subdomain where the vectors form a stable basis.

THEOREM 4.8. *If all pairs of functions g_i, g_j in the class B of Boolean functions satisfy either $| < g_i, g_j > | \leq \delta$ or $| < g_i, g_j > | \geq 1 - \delta$, for δ being typically significantly less than $1/2$, and furthermore, for a given Boolean function f, $| < f, g_i > | \leq \epsilon$, for all $g_i \in B$, then for $M \subseteq B$ with $|M| \leq min\{1/\delta^{1/3}, 1/\epsilon^{1/3}\}$, there is no $g \in span(M)$ with the same sign as f.*

PROOF. We first construct a subdomain/distribution $\mathcal{D}$ and show that

$$|| \sum_i < f, g_i >_{\mathcal{D}} g_{i,\mathcal{D}} ||_{\infty,\mathcal{D}} \leq 1.$$

Theorem 4.4 then completes the proof. The subdomain is constructed by first dividing M into bundles such that for pairs g_i, g_j in each bundle, $| < g_i, g_j > | \geq 1 - \delta$ and for g_i and g_j in different bundles, $| < g_i, g_j > | \leq \delta$. For each bundle M_k, we find a representative function $g_k \in M_k$ and remove from the $\mathcal{D}$ all points where $g_k \neq g_i$, for some $g_i \in M_k$. The subdomain $\mathcal{D}$ thus constructed is no less than $1 - |M|\delta$ of the entire domain; therefore, the values $| < f, g_i >_{\mathcal{D}} |$ are still no larger than $(\epsilon + |M|\delta)/(1 - |M|\delta)$, and the values $| < g_i, g_j >_{\mathcal{D}} |$ are either 1, i.e, $g_{i,\mathcal{D}} = g_{j,\mathcal{D}}$, or is at most $(\delta + |M|\delta)/(1 - |M|\delta)$, for $g_i, g_j \in M$. I.e, $g_{i,\mathcal{D}}$ and $g_{j,\mathcal{D}}$ are either identical or almost orthogonal. Furthermore, $||g_i||_{2,\mathcal{D}}$ is still 1, since the g_i are Boolean. Intuitively, the function $\sum_i < f, g_i >_{\mathcal{D}} g_{i,\mathcal{D}}$ is a reasonable approximation to the true projection $f|_{\mathcal{D},span(M)}$, since the g_i are almost orthonormal over $\mathcal{D}$. Furthermore,

$$|| \sum_i < f, g_i >_{\mathcal{D}} g_{i,\mathcal{D}} ||_{\infty,\mathcal{D}} \leq \sum_i | < f, g_i >_{\mathcal{D}} |$$

$$\leq \frac{(\epsilon + |M|\delta)}{(1 - |M|\delta)} |M|,$$

which is at most 1 provided $|M|$ is sufficiently small as in the statement of the theorem.

To find the true projection, we find orthonormal basis functions g_i^* for $span(M)|_{\mathcal{D}}$, from the functions $g_i|_{\mathcal{D}}$, using, for example, Gram-Schmidt orthonormalization. We omit the exact calculations. Basically, since $g_{i,\mathcal{D}}$ already forms a close-to-orthonormal basis, the orthonormalization does not blow-up either the ∞-norm of the functions g_i^*, or the values $| < f, g_i^* >_{\mathcal{D}} |$, and thus the projection $f|_{\mathcal{D},span(M)} = \sum_i < f, g_i^* >_{\mathcal{D}} g_i^*$ still continues to have a small ∞-norm provided $|M|$ is at most the bound given in the theorem. □

Next we carry further the idea of stable bases over subdomains/distributions and present a divide and conquer approach to showing that a function cannot be approximated from a space of functions.

Observation 4.9. *Let $\bigcup_i P_i \subseteq \{-1,1\}^n$, and for all i, let $P_i = P_0 \oplus s_i$, where '$\oplus$' stands for addition when $\{-1,1\}^n$ is viewed as $\mathbb{F}_2^n$, and s_i is the shift vector for P_i. In other words, the P_i's are shifts of each other. Given a function f on $\{-1,1\}^n$, we denote by f_{P_i} its restriction to P_i; furthermore, we shall view all of the functions f_{P_i} as being over P_0, by defining $f_{P_i}(x) := f(x \oplus s_i)$. Let f be Boolean, B be a set of Boolean functions, and M be a (typically small) subset of B.*

(i) If the functions f_{P_i} form an orthonormal basis for the space of functions from P_0 to the reals, and the functions $\{g_{P_i} : g \in M, i \in \mathbb{N}\}$ span a subspace X_M of dimension m, then there is no close approximation h to f from span(M), with $|h(x)| \geq \sqrt{m/|P_0|}$ for all $x \in P_0$.

(ii) If for some i, the set $\{g_{P_i} : g \in M\}$ forms an orthonormal set and $< f, g >_{P_i} < 1/|M|$ for all $g \in M$ then there is no approximation to f from span(M). This can be extended to the case where the set $\{g_{P_i} : g \in M\}$ forms orthonormal bundles as in Theorem 4.8.

Proof. For (i), we show that there is a P_i such that for the distribution $\mathcal{R}$ that is 1 on P_i and 0 elsewhere, $< f, g >_R \; < \; \sqrt{m/|P_0|}$, for all $g \in M$. Then the proof follows by 2.2(2). We in fact show something stronger. We show that $f|_{P_i}$ has no close 2-norm approximation. Since the functions $f_{P_i} : i \in \mathbb{N}$ form an orthonormal basis for a space of dimension $|P_0|$, and the set $\{g_{P_i} : g \in M, i \in \mathbb{N}\}$ only spans a subspace X_M of dimension m, it can be shown using simple linear algebra and geometry, that there must be at least one P_i such that the projection $f_{P_i}|_{X_M}$ has a 2-norm at most $\sqrt{m/|P_0|}$. In other words, we show that if, for all the P_i, the projections $f_{P_i}|_{X_M}$ had 2-norms exceeding δ, then the space spanned by the projections $f|_{P_i}|_{X_M}$, and therefore the space X_M would have dimension at least $|P_0|\delta^2$. Thus for some P_i, $< f_{P_i}, h^* >_{P_i} < \sqrt{m/|P_0|}$, for all $h^* \in X_M$, with 2-norm bounded by 1; and taking h^* to be any of the functions in $\{g_{P_i} : g \in M, i \in \mathbb{N}\}$, we have what we require. Notice that the above proof depends only on the dimension of the space X_M and goes through independent of the exact basis M.

For (ii), since the g_{P_i} form an orthonormal set for some P_i, by 4.2, any approximation g from their span to f is a high energy approximation with $< f, g >_{P_i} \geq 1/|M|$. The result follows. □

○ This general method has been used in several papers, although not stated as such, for example [22], [20], [11]: in particular, (i) above is the crux of the "geometric or variation rank" method used in [22] to show that DIP_2 is not closely approximable by the span of few symmetric functions. Using traditional terminology, (ii) represents a combination of "divide-and-conquer," or the "spectral method."

5. Examples of stable basis families.

In this section, we show several examples of Boolean basis families B consisting of functions over discrete and continuous domains that are *poly* stable. By the results of the earlier sections, it then follows that showing weighted threshold lower bounds involving functions from such basis families B can be proven by scalar product arguments. One of these examples provides a method for showing lower bounds for an algebraic circuit model of computation.

The first two examples are straightforward, and hence we omit the proofs.

Theorem 5.1. *(1) Let B be a set of And functions that are monotone or anti-monotone with respect to the same sets of variables. Then B is stable: for any $M \subseteq B$, $||\sum\limits_{h\in M} a_h h||_2^2 \geq 1/\sqrt{|M|}\sum\limits_{h\in M} a_h^2$.*

(2) Let B be a set of symmetric threshold functions t over $\{-1,1\}^n$, defined as $t(x) := sign(\sum\limits_i x_i - \alpha)$. Then B is stable: for any $M \subseteq B$, $||\sum\limits_{h\in M} a_h h||_2^2 \geq 1/|M|\sum\limits_{h\in M} a_h^2$.

Conjecture 5.2. *It is easy to see that general And functions do not form a stable basis, since some of them can be obtained as large coefficient linear combinations of others. However, we conjecture that they form a quasistable basis. In fact, it seems likely that all of the known* weighted *threshold lower bounds have been proven for quasistable basis families. If this is the case, although their current proofs seem to use the strong first method in Proposition 3.5, by Proposition 4.2, they can, in fact, be proven using the weaker Method 3, or Theorem 4.4, i.e, using scalar product estimates that have already commonly been used for proving non-existence of close approximations and thereby unweighted threshold lower bounds.*

Observe that the above conjecture does not contradict the result in [20] that there exist functions, which, when expressed as thresholds of *And* functions, *need* large integer weights: these functions have no *close* uniform approximations from *And* functions, but do have uniform approximations. This invalidates the use of Method (3) in Proposition 3.5 as a way of proving nonexistence of uniform approximations. This occurence is common: we will see examples in the next section of basis families that are stable, even orthogonal, but which require large integer weights for approximating certain functions, i.e, these functions do provide uniform approximations for certain functions, but do not provide close approximations. *However, since the basis is stable*, irrespective of the weights the existence of uniform approximation does imply the existence of high energy approximations from these basis families. Therefore, potentially, the scalar product estimates of Method 2 can still be used to show nonexistence of uniform approximations or weighted threshold lower bounds for these basis families.

Next, we show stability of a class of weakly symmetric threshold functions over the continuous domain $(0,1)^n$. These functions form a non-trivial family: for instance, the discrete analog of these functions are unstable as will be shown in the next section. Then, using Proposition 4.2, any function that has a small scalar product with such thresholds cannot be computed as a weighted threshold of such thresholds.

Open Question 5.3. *The algebraic model of computation over $\mathbb{R}^n$ or $(0,1)^n$ consisting of threshold circuits with gates over $(0,1)^n$ at the bottom level, and the usual threshold gates over $\{-1,1\}^n$ at the higher levels is related to arithmetic circuits, (see [7] and see [23]), and linear (algebraic) decision trees studied in, for example, [29] and [6], although the exact relationship is unknown. It would be useful to investigate this relationship since that would help to transfer lower bounds such as those to be described below to the decision tree model and viceversa. In particular, a threshold of m threshold gates over the reals can be simulated by a linear decision tree of depth m, but clearly, there are functions computable by linear decision trees of depth m that are not computable by a threshold of m thresholds.*

NOTE: For convenience, we assume that the range of the functions is $\{0,1\}$ instead of the usual $\{-1,1\}$.

Theorem 5.4. *Let B be a set of threshold functions over $(0,1)^n$ that are symmetric on k disjoint sets of indices $u_1,\dots,u_k$. I.e, each such function t has the form $sign(\sum_i a_i(\sum_{j\in u_i} x_j) + a_0)$. Denoting $\sum_{j\in u_i} x_j$ simply as u_i, and appropriately* *normalizing, t can be viewed as a threshold function on $(0,1)^k$. Clearly B is a stable basis family if and only if the corresponding basis family over $(0,1)^k$ thus obtained is also a stable basis family. If, for any two such functions t_i and t_j, over $(0,1)^k$, a separation condition holds: $||t_i - t_j||_2 \geq 1/p(n)$, for some polynomial p, then B is a poly stable basis, for some polynomial $poly(n)$ that depends on k and p.*

Proof. The proof proceeds by developing each t_i as a Fourier series, studying the coefficients in this expansion and showing that for some frequency we have the desired estimate needed for stability. We indicate how this works for two dimensions on $(0,1)^2$. Consider a threshold function t on the unit square, whose defining line, say $ax_1 + x_2 = c$ intersects the square on the lines $x_1 = 0$ and $x_1 = 1$. The other types of threshold functions behave similarly. We develop t as a Fourier series with coefficients $\hat{t}(u_1, u_2)$, where recall that

$$\hat{t}(u_1,u_2) = \int_0^1 \int_0^1 t(x_1,x_2) e^{2\pi i(u_1x_1+u_2x_2)} dx_1 dx_2$$

$$= \int_0^1 dx_1 \int_{c-ax_1}^1 e^{2\pi i(u_1x_1+u_2x_2)} dx_2$$

$$= \int_0^1 e^{2\pi i(u_1x_1)} [1/(2\pi i u_2) e^{2\pi i(u_2x_2)}]_{c-ax_1}^1 dx_1$$

$$= \int_0^1 1/(2\pi i u_2) e^{2\pi i(u_1x_1)} dx_1$$

$$- \int_0^1 1/(2\pi i u_2) e^{2\pi i(u_1x_1+u_2(c-ax_1))} dx_1$$

$$= \frac{1}{(4\pi)^2(u_2(u_1-au_2)} \Big(e^{2\pi i(c-a)u_2} - e^{2\pi i(c)u_1} \Big).$$

Now if we take a linear combination of threshold functions, say $\sum\limits_j^m b_j t_j$, then using the obvious notation, the $(u_1, u_2)^{th}$ Fourier coefficient of this linear combination is

$$= \sum_j^m b_j \frac{1}{(4\pi)^2(u_2(u_1 - a_j u_2))} \Big(e^{2\pi i(c_j - a_j)u_2} - e^{2\pi i(c_j)u_1} \Big). \qquad A$$

We now analyze A as a function of u_1 and u_2, for $u_1 \leq q$ and $u_2 \leq q$, where q is some polynomial depending on the polynomial p, and the number of threshold functions m. We use the fact that the functions $2^{2\pi i \beta_j u_2}$ are almost orthogonal in the interval $-q \leq u_2 \leq q$, for the different β_j in A, and use it to obtain the polynomial *poly* required to show that these threshold functions form a *poly* stable basis. The proof extends to arbitrary, fixed dimensions k.

We show that the modulus of A is large for some u_1, u_2 and to do so, we observe the following:

i) There are at most $p_1(n)$ threshold functions which are pairwise separated as in the assumption of the theorem. Here, p_1 depends on p and the dimension k (recall that we have assumed that $k = 2$ for this proof).

ii) There exists $p_2(n)$ depending on $p(n)$ s.t. if $|a_i - a_j| < \frac{1}{p_2(n)}$ then $|c_i - c_j| > \frac{2}{p_2(n)}$ and so $|(c_i - a_i) - (c_j - a_j)| > \frac{1}{p_2(n)}$

(ii) follows immediately from the separation assumption. We now consider the b_{j_o} of maximal modulus. We will assume $|a_{j_o}| \geq \frac{1}{p_2(n)}$. The case $|a_{j_o}| < \frac{1}{p_2(n)}$ is similar.

$$\ldots (1)$$

We will show that after averaging over appropriate intervals of u_1 and u_2, the sum of those terms in A for which $|a_i - a_{j_o}| \leq \frac{1}{p_3(n)}$, where p_3 is appropriately chosen, is much bigger than the sum of those terms for which $|a_i - a_{j_o}| > \frac{1}{p_3(n)}$. To do that, we consider the factor $\frac{1}{u_2(u_1 - a_{j_o} u_2)}$ appearing in (A). We see that, given k_o, there exists $p_3(n)$ and intervals of integers

$$\begin{aligned} I_1 &= (c_1 n^{k_1}, c_1 n^{k_1} + n^{k_2}) \quad \text{and} \\ I_2 &= (c_2 n^{k_3}, c_2 n^{k_3} + n^{k_2}) \quad \text{where} \\ & \quad n^{k_2} > n^{k_o} p_2(n) \end{aligned}$$

such that the following holds:

If $u_1, v_1, \in I_1$ and $u_2, v_2 \in I_2$, and if $|a_i - a_{j_o}| > \frac{1}{p_3(n)}$ we have $\left|\frac{1}{(v_1 - a_{j_o} v_2)}\right| \leq \frac{1}{n^{k_o}} \left|\frac{1}{u_2(u_1 - a_{j_o} - u_2)}\right|$

$$\ldots (2)$$

In view of (2) it is enough to show that

$$\sum_{\substack{u_2 \in I_1 \\ u_2 \in I_2}} \left| \sum_{\substack{j: \\ |a_j - a_{j_o}| < \frac{1}{p_3(n)}}} b_j \frac{1}{u_2(u_1 - a_j u_2)} \left[e^{2\pi i (c_j - a_j) u_2} - e^{2\pi i c_j u_1} \right] \right|^2$$

is sufficiently by

$$\ldots (3)$$

By expanding (3) we obtain

$$\textstyle\sum_{u_1 \in I_1} \sum_{I_2 \in I_2} \sum_{i,j} b_i \bar{b}_j \frac{1}{u_2(u_1 - a_i u_2)} \frac{1}{u_2(u_j - a_j u_2)} \\ \left[e^{2\pi i [(c_i - a_i) - (c_j - a_j)] u_2} + e^{2\pi j (c_i - c_j) u_1} - e^{2\pi i (c_i - a_i) u_2 - c_j u_1} - e^{2\pi i (c_j - a_j) u_2 - c_i u_1} \right]$$

$$\ldots (4)$$

Now (ii) and (2) imply that in (4), the sum of the terms for which $i \neq j$ is much smaller than the sum of those terms for which $i = j$. And the sum of the terms for which $i = j$ is $\geq |I_1||I_2| \cdot (\sum_j (b_j)^2) \cdot \frac{1}{p_4(n)}$, where $p_4(n)$, $|I_1|$ and $|I_2|$ depend only on p and the dimension. This completes the proof of the theorem. □

6. Unstable Boolean bases and examples.

In this section, we give a method for establishing instability of bases and show that there exist Boolean bases that are unstable. In fact, we show that a highly restricted family of threshold functions is already unstable. The first proposition gives straightforward but strong conditions that imply stability of basis families, in the case that the domain of the functions S is a discrete set, i.e, the space of functions under consideration is a finite dimensional space. These conditions, theoretically, provide a means to construct unstable Boolean bases.

Proposition 6.1. *A basis M of independent functions over a finite, discrete set S can be represented as a matrix with $|S|$ rows and $|M|$ columns, with each column h representing a basis function, and each row x representing a point in S, so that the entry $M(h, x)$ is nothing but the value $h(x)$. Clearly, M is a polynomially stable basis if, for all column vectors a with $|M|$ entries,* *it holds that $||Ma||_2 \geq 1/poly(|M|)||a||_2$. Now consider a decomposition of M in to $|M| \times |M|$ submatrices $M_i : i \leq |S|/|M|$. For some such decomposition, if at least $1/poly(|M|)$ of these square matrices M_i are nonsingular and have inverses whose 2-norms are bounded by $poly(|M|)$, then M is a polynomially stable basis.*

The condition in the above proposition is usually too strong to be useful for establishing *stability* of basis families. Besides the fact that the submatrices M_i could be nonsingular, it could also be that although *all* the submatrices have large inverses, the sets of vectors a for which each submatrix behaves unstably are disjoint sets, and therefore, for any single vector a, a sufficiently large number of submatrices behave stably. However, the proposition above gives an easy way to construct *unstable* bases M, by using only submatrices M_i whose inverses have large 2-norms.

Do there exist unstable *Boolean* bases M? In other words, do there exist square Boolean matrices whose inverses have large 2-norms (since M can be constructed using several copies of such square matrices as described in Proposition 6.1)?

A straightforward upper bound of $n^{n/2}$ exists on the (∞-norm of the) inverse of $\{-1, 1\}$-valued matrices, since the determinant is given as $n!$ times the volume of the simplex formed by the rows. In addition, matrices with fairly large 2-norms for their inverses ($\Omega(2^{n/2})$) can be easily constructed from simple arguments [13], and using cyclic codes. The next result of [1] and [28] shows that, in fact, the upper bound of $n^{n/2}$ is quite tight, and [28] gives a construction of ill-conditioned matrices that meet this bound for the inverse 2-norm.

FACT 6.2. *([1] and [28]) Take the columns of a matrix M to be the linear monomials x_i over $\{-1, 1\}^{|M|}$. Then there exists an $|M| \times |M|$ submatrix whose inverse has 2-norm at least $\Omega((n/c)^{n/2})$ for some constant c.*

PROOF. We sketch the proofs for completeness. The proof uses a result of [14] that constructs a threshold function which does not use the constant term, and which requires large (integer) weights, or in other words, when these weights are normalized, there is a large difference between the smallest and the largest weights. The result acheives this by showing that there is a function f such that for each a, for which $sign(Ma) = sign(f)$, there are 2 entries i, j in a such that $|a_i|$ is almost 1, $|a_j|$ is very small, and all entries of Ma have absolute value at least $|a_j|$. It follows by a familiar result on linear polytopes that there is at least one nonsingular square submatrix of M, say M^*, such that all entries of M^*a have absolute value *exactly* $|a_j|$. But since $|a_i|$ is much larger than $|a_j|$, this must mean that $(M^*)^{-1}$ has a large ∞-norm. It follows that $(M^*)^{-1}$ has a large 2-norm since M^* is just an $n \times n$ matrix. □

REMARK 6.3. *The result seems counter-intuitive, since the linear monomials forming M are a stable, even orthonormal basis. However, as remarked after 5.2, this does not contradict the above fact, since even if many submatrices in M have inverses with large 2-norms, on any given vector a, significantly many of them behaves stably. In other words, as mentioned in the discussion following Conjecture 5.2, the result in [14] shows that while there is no close approximation from the linear monomials to a certain function f, a uniform approximation exists. Hence non-existence of uniform approximations from linear monomials cannot be established using method 4 of Proposition 3.5. But for such functions f, since the basis is stable, 4.2 shows that a high-energy approximation exists, and thus conversely, to show non-existence of uniform approximation from the linear monomials, method 3 and Theorem 4.4 can still be used.*

While the above result shows that unstable Boolean bases exist, these bases are quite artificial since they are constructed using several copies of matrices M_i whose inverses have large 2-norms, and it is not clear that any natural family of Boolean functions behave in this manner.

Next we give a natural family of threshold functions that forms an unstable basis. Recall Theorem 5.1 that symmetric threshold functions form a stable basis. Next we show that weakening this symmetry even slightly creates an unstable basis.

DEFINITION 6.4. *A threshold function $t_{u,v}$ is* **2-symmetric** *if it is of the form*

$$t_{u,v}(x) := sign(a_1 \sum_{i \in u} x_i + a_2 \sum_{i \in v} x_i - a_0),$$

where u and v are disjoint sets of indices. Thus $t_{u,v}$ can be viewed as a threshold function in 2 variables, over $\{0, \dots, n\}^2$, given as

$$t_{u,v}(u_1, u_2) := sign(a_1 u_1 + a_2 u_2 - a_0).$$

We refer to the latter as **2-threshold** *functions.*

The following is straightforward and was assumed earlier for the case of threshold functions over continuous domains.

FACT 6.5. *The set of 2-symmetric threshold functions over $\{-1,1\}^n$ is a polynomially stable basis if and only if the set of $\{0,1\}$-valued 2-threshold functions over $\{0,\ldots,n\}^2$ is also a polynomially stable basis.*

We use an $n \times n$ $\{0,1\}$ matrix M^* given by [13] that has a cyclic structure, whose inverse has a 2-norm close to $2^{n/2}$, to prove the following theorem. We describe the matrix M^* here. The rows r_i of M^* all have 3 non-zero entries, and have the following cyclic pattern:

$$r_1 = (10110\ldots0); r_2 = (01110\ldots0);$$

$$r_3 = (001011\ldots0); r_4 = (000111\ldots0); \text{etc.}$$

giving nonsingularity. Now, we get

$$(r_1 + r_2) - 2(r_3 + r_4) + 4(r_5 + r_6)\ldots + 2^{n/2}(r_{n-1} + r_n)$$

$$= (110\ldots0),$$

thereby showing that M^* has a large 2-norm for its inverse.

THEOREM 6.6. *The set of 2-symmetric threshold functions $t_{u,v}$ for any fixed set of indices u and v is not polynomially stable.*

PROOF. To prove the theorem, we show that there is a linearly independent set of at most $N \leq 10n$ 2-threshold functions t_j, and coefficients a_j such that at least one of the $|a_j| = 1$, but $||\sum_j^N a_j t_j||_2 \leq 1/2^{n/2}$. To achieve this, we build an $n \times n^2$ matrix M to satisfy the following properties:
(i) the columns of M correspond to the domain points in $\{1,\ldots,n\}^2$, and whose rows correspond to n Boolean functions h_i;
(ii) M will embed the $n \times n$ submatrix M^* (described before the theorem) as a submatrix; All the other entries in M will be 0.
(iii) Each of the functions (rows) h_i will be a linear combination of at most 10 threshold functions t_j;
(iv) all of the $N \leq 10n$ threshold functions thus used will be linearly independent over the entire domain i.e, over all the columns of M.

Since M^{*-1} is known to have a large 2-norm, it will then follow that there is a $1 \times n$ coefficient vector a such that $||aM||_2 \leq 1/2^{n/2}||a||_2$. Furthermore, since each of the rows of of M is a linear combination of at most 10 threshold functions, the result follows.

We now describe the construction of M that satisfies (i)-(iv). We mimic the structure of the matrix M^* described above on a zigzagging subset S consisting of n of the domain lattice points $\{0,\ldots,n\}^2$. The set S will correspond to the columns of M^*. Each row r_i of M^* will obtained as the support of a linear combination of at most 10 threshold functions.

It is easy to see that we can obtain a row r_i of M (and M^*) of the form $(00\ldots111\ldots0)$ as a function supported at 3 adjacent lattice points such as

$$\begin{array}{ccccc}
0 & \ldots & 0 & \ldots & 0 \\
0 & \ldots & 0 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
0 & \ldots & 111 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
0 & \ldots & 0 & \ldots & 0
\end{array}$$

on the domain, and we can obtain the next row r_{i+1} of the form $(00\ldots1101\ldots0)$ as a function supported at 3 lattice points such as

$$\begin{array}{ccccc}
0 & \ldots & 0 & \ldots & 0 \\
0 & \ldots & 0 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
0 & \ldots & 110 & \ldots & 0 \\
0 & \ldots & 001 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
0 & \ldots & 0 & \ldots & 0
\end{array}$$

on the domain. Thus the set S of n points corresponding to the columns of the matrix M^* are as given below. The numbers refer to the rows of M and M^*.

$$\begin{array}{ccccccc}
0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & \ldots & 0 & 0 & (5,6,8) & \ldots & 0 \\
0 & \ldots & 0 & 0 & (5,6,7) & \ldots & 0 \\
0 & \ldots & (1,2,3) & (3,4,5) & (3,4,6) & \ldots & 0 \\
0 & \ldots & (1,2,4) & 0 & 0 & \ldots & 0 \\
0 & \ldots & (1) & (2) & 0 & \ldots & 0 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0
\end{array}$$

Notice that each row or function h_i (which is 0 everywhere else on the $n \times n$ lattice except at the 3 specific points) can be obtained as a simple linear combination of atmost 10 threshold functions, while making sure that all the threshold functions thus used are linearly independent. Thus M satisfies all the conditions (i)-(iv), whereby the proof is complete. □

OPEN QUESTION 6.7. *It is an interesting open problem whether the better lower bounds of [1] and [28] on the 2-norms of the inverses of Boolean matrices can be mimicked by threshold functions (or any other natural family of Boolean functions) as in the above proof. This would show that the stability of families of threshold functions is even worse than that indicated by the above theorem. This provides an intuition as to why nonexistience of uniform approximations from threshold functions (or lower bounds for a weighted threshold of thresholds is so hard to prove: recall the notorious question as to whether LT_2 is different from NP). However, it is still possible that threshold function families satisfy some other notion of stability such as quasistability in the weak sense, because that would be sufficient to permit the use of Methods 2 or 3 of Proposition 3.5 or Theorem 4.4 to show such lower bounds.*

Acknowledgements

The work of Per Enflo was partially supported by NSF grant CCR 440103. The work of Meera Sitharam was partially supported by NSF grant CCR 9409809. We are grateful to Victor Grinberg for interesting discussions.

References

[1] N. Alon, “Threshold gates, coin weighing, and indecomposable hypergraphs,” *Manuscript*, 1996.

[2] J. Aspnes, R. Beigel, M. Furst, S. Rudich, “The expressive power of voting polynomials,” *Combinatorica* 14, pp. 1-14, 1994.

[3] L. Babai, N. Nisan, M. Szegedy, “Multiparty protocols and pseudorandom sequences,” *Proc.* 21^{st} *Ann. ACM Symp. on Theory of Computing,* pp. 1-11, 1989.

[4] J. Bruck, “Harmonic analysis of polynomial threshold functions,” *SIAM Journal of Discrete Mathematics,* 3 (2), pp. 168-177, 1990.

[5] J. Bruck, R. Smolensky, “Polynomial threshold functions, AC^0 functions, and spectral norms,” 31^{st} *Ann. IEEE Symp. Foundations of CS,* pp. 632-641, 1990.

[6] A. Björner, L. Lovász, A.C. Yao, "Linear decision trees, volume estimates and topological bounds," 24th *Ann. ACM Symp. Theory of Comp,* pp. 170-177, 1992.

[7] J. von zur Gathen, "Algebraic complexity theory," *Ann. rev. comp. sci* 3, 317-347, 1988.

[8] M. Goldman, "On the power of a threshold gate at the top," *Manuscript*, 1995.

[9] M. Goldman, J. Håstad, "On the power of small depth threshold circuits," 31st *Ann. IEEE Symp. Foundations of CS*, pp. 610-618, 1990.

[10] M. Goldman, J. Håstad, "A simple lower bound for monotone clique using a communication game," *Information Processing Letters* 41, 221-226, 1991.

[11] M. Goldman, J. Håstad, A.A. Razborov, "Majority gates vs. general weighted threshold gates," 32nd *Ann. IEEE Symp. Foundations of CS*, 1991.

[12] M. Grigni, M. Sipser, "Monotone separation of logspace from NC^1," *Proc.* 6th *IEEE Conf. on Structure in Complexity Theory,* pp. 294-298, 1991.

[13] V. Grinberg, *Personal communication.*

[14] J. Håstad, "On the size of weights for threshold gates," *SIAM J. Disc. Math,* pp. 484-492, 1994.

[15] A. Hajnal, W. Maass, G. Turán, "On the communication complexity of graph properties," 20th *Ann. ACM Symp. on Theory of Computing*, pp. 186-191, 1988.

[16] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, G. Turán, "Threshold circuits of bounded depth," 28th *Ann. IEEE Symp. Foundations of CS*, pp. 99-110, 1987.

[17] J. Jackson, "An efficient membership query al gorithm for learning DNF with respect to the uniform distribution, " *Proc.* 35th *Ann. IEEE Symp. Foundations of CS,* pp. 42-53, 1994.

[18] M. Karchmer, A. Wigderson, "Monotone circuits for connectivity require super-logarithmic depth," 21st *Ann. ACM Symp. on Theory of Computing*, pp. 539-550, 1988, *SIAM J. Disc. Math* 3, 255-265, 1990.

[19] M. Krause, "Geometric arguments yield better bounds for threshold circuits and distributed computing," 6th *IEEE conf. on Struct. in Compl. Theory,* pp. 314-322, 1991.

[20] M. Krause, P. Pudlák "On the power of depth 2 circuits with threshold and modulo gates," 26th *Ann. Symp. Theory of Comput.*, pp. 48-58, 1994.

[21] M. Krause, P. Pudlák "On Computing Boolean functions by sparse real polynomials," 36^{th} *Ann.IEEE Symp. Foundations of CS*, pp. 682-691, 1995.

[22] M. Krause, S. Waack, "Variation ranks of communication matrices and lower bounds for depth two circuits having symmetric gates and unbounded fan-in," 32^{nd} *Ann. IEEE Symp. Foundations of CS*, pp. 777-782, 1991.

[23] N. Nisan, A.W. Widgerson, "Lower bounds on arithmetic circuits via partial derivatives," 36^{th} *Ann. IEEE Symp. Foundations of CS,* pp. 16-25, 1995.

[24] R. Raz, A. Wigderson, "Monotone circuits for matching require linear depth," *Proc.* 22^{nd} *Ann. ACM Symp. Theory of Computing,* pp. 287-292, 1990.

[25] M. Sitharam. "Evaluating spectral norms for constant depth circuits with symmetric gates," *J. computational complexity,* 6, pp. 167-189, 1995.

[26] M. Sitharam. "Approximation from linear spaces and applications to complexity," *Manuscript* submitted to Electronic colloquium on computational complexity and this journal.

[27] K.I. Siu, J. Bruck, "On the power of threshold circuits with small weights," *SIAM Journal of Discrete Mathematics,* 4 (3), pp. 423-435, 1991.

[28] V.H. Vu, "On ill-conditioned Boolean matrices and their applications," *Manuscript,* 1996.

[29] A.C. Yao, "Algebraic decision trees and the Euler characteristic," 33^{rd} *Ann. IEEE Symp. Foundations of CS,* pp. 268-278, 1992.

Manuscript received July, 2001

Per Enflo
Department of Mathematics and Computer Sciences
Kent State University
Kent, OH 44240, USA
`enflo@mcs.kent.edu`

Meera Sitharam
Computer and Information Science and Engineering
University of Florida
Gainesville, FL 32611, USA
`sitharam@cise.ufl.edu`