APPROXIMATION RESISTANCE AND LINEAR THRESHOLD FUNCTIONS

RIDWAN SYED

Abstract. In the boolean $Max-k-CSP(f)$ problem we are given a predicate $f:\{-1,1\}^k \to \{0,1\}$, a set of variables, and local constraints of the form $f(x_1,...,x_k)$, where each x_i is either a variable or negated variable. The goal is to assign values to variables as to maximize the fraction of constraints which evaluate to 1.

Many such problems are NP-Hard to solve exactly, and some are even NP-Hard to approximate better than a random assignment. Such problems, or rather their corresponding predicates f, are called approximation resistant.

Recently Khot et al. gave a characterization (assuming the Unique Games Conjecture) of a modified notion of approximation resistance in terms of the existence of a measure over a convex polytope corresponding to the predicate. In this paper we discuss this characterization as well as the approximability of linear threshold predicates (predicates of the form $sgn(w_1x_1+...+w_kx_k)$).

Contents

1. Constraint Satisfaction Problems and Hardness of Approximation

In this section we introduce the family of Constraint Satisfaction Problems, and discuss hardness of approximation. Throughout this paper we will focus on the $Max-k-CSP(f)$ problem, where f is a Boolean function. Informally, in the Boolean $Max-k-CSP(f)$ problem we are given a k-bit Boolean function $f:\{-1,1\}^k \to \{0,1\}$ (often called a predicate), a set of variables, and a set of local constraints of the form $f(x_1,...,x_k)$ where each of the x_i is a literal (variable or a negated variable). The objective is to assign values to the variables such that the fraction of constraints which simultaneously evaluate to 1 is maximized. We can formalize instances of this problem as follows.

Date: August 2014.

Definition 1.1. Fix $f : \{-1,1\}^k \to \{0,1\}$. An instance of $Max-k-CSP(f)$, $I = \{N, m, v, s\}$, consists of N variables, m constraints, a matrix of labels $v \in [N]^{m\times k}$, and a matrix of signs $s \in \{-1,1\}^{m\times k}$. For convenience the m constraints of I can be written as

$$f(s_{1,1}x_{v_{1,1}}, ..., s_{1,k}x_{v_{1,k}})$$
$$...$$
$$f(s_{m,1}x_{v_{m,1}}, ..., s_{m,k}x_{v_{m,k}})$$

The objective is to assign values to the N variables so as to maximize the following function:

$$Val(x) = \frac{1}{m} \cdot \sum_{i=1}^{m} f(s_{i,1}x_{v_{i,1}}, ..., s_{i,k}x_{v_{i,k}})$$

. We denote the optimal value for a given $Max - k - CSP(f)$ problem as,

$$Opt(I) = max_{x\in\{-1,1\}^N} Val(x)$$

Many well known combinatorial problems are in fact $Max - k - CSP(f)$ for some f and k.

Example 1.2. Let $f : \{-1,1\}^3 \to \{0,1\}$ be defined as $f(x_1, x_2, x_3) = 1$ if and only if at least one literal x_i is 1. This function is precisely the or of 3 bits, so on any instance I of $Max - 3 - CSP(f)$ we can write the ith constraint as

$$(s_{i,1}x_{v_{i,1}} \vee s_{i,2}x_{v_{i,2}} \vee s_{i,3}x_{v_{i,3}}).$$

This is precisely the problem of $Max - 3 - SAT$.

In particular, this problem (like many problems in this family) is NP-Hard to solve exactly. Assuming the widely believed conjecture that $P \neq NP$, this problem cannot be solved exactly by an efficient algorithm. Therefore, it is interesting to study the complexity of finding an approximate solution.

Definition 1.3. Let f be a k-bit Boolean function, and suppose that $f(x) = 1$ for exactly p inputs. Then, we say that f has density $\rho(f) = \frac{p}{2^k}$.

Definition 1.4. Fix a predicate f, and let $\alpha \leq 1$. If $P(I)$ is the output of an algorithm P, such that for all instances I of $Max - k - CSP(f)$, we have that

$$\alpha \cdot Opt(I) \leq P(I) \leq Opt(I),$$

we call P an α-approximation algorithm for f.

We call α the approximation ratio for the algorithm. We use a similar notion for randomized algorithms.

Definition 1.5. Fix a predicate f, and let $\alpha \leq 1$. If $P(I)$ is the output of a randomized algorithm P, such that for all instances I of $Max - k - CSP(f)$, we have that

$$\alpha \cdot Opt(I) \leq \mathbb{E}[P(I)],$$

we call P a randomized α-approximation algorithm for f.

We are naturally concerned with finding efficient approximation algorithms, and we present such an algorithm.

Proposition 1.6. *Fix a k-bit Boolean predicate f. There exists a randomized $\rho(f)$-approximation algorithm for $Max - k - CSP(f)$.*

Proof. Suppose we are given an instance of $Max - k - CSP(f)$, $I = \{N, m, v, s\}$. Independently assign each x_i to 1 with probability 1/2 and -1 with probability 1/2. Then, the probability that any particular constraint evaluates to 1 is $\rho(f)$. So the expected contribution of each constraint to the total number of constraints satisfied is $\rho(f)$. Linearity of expectation and the observation that no more than m constraints can be satisfied give the result. □

Although this algorithm achieves some positive approximation ratio, it is somewhat unsatisfying since the algorithm does not actually look at the instance of the problem or take advantage of any properties of f. For many predicates, one can achieve larger ratios using, for example semidefinite programming. However, for some predicates f the ratio given by $\rho(f)$ is in fact optimal i.e. it is NP-Hard to approximate $Max - k - CSP(f)$ by a larger ratio. We call such predicates approximation resistant. We present an equivalent definition of approximation resistance.

Definition 1.7. Fix a k-bit Boolean predicate f. If for all positive ϵ it is NP-hard to distinguish between instances of $Max - k - CSP(f)$ which are $1 - \epsilon$ satisfiable and instances which are not satisfiable by more than $\rho(f) + \epsilon$ fraction of constraints, we say that f is approximation resistant.

2. Basic Fourier Analysis of Boolean Functions

One of the most useful tools in the analysis of Boolean functions is the discrete Fourier Transform, which we discuss in this section. We will show that any Boolean function can be expressed uniquely as a multilinear polynomial, and we will prove some useful facts about this representation.

Now, we consider the vector space of functions of the form $f : \{-1, 1\}^n \to \mathbb{R}$. Observe that we can represent each f as a column vector by stacking its 2^n values in lexicographic order of the corresponding input. In other words, the space of k-bit Boolean functions is 2^n-dimensional.

Definition 2.1. Let $f, g : \{-1, 1\}^n \to \mathbb{R}$. We define the inner product $< \cdot\, , \,\cdot >$ as

$$< f, g > = 2^{-n} \cdot \sum_{x \in \{-1,1\}^k} (f \cdot g)(x) = \mathbb{E}_{x \sim \{-1,1\}^n}[f(x) \cdot g(x)]$$

where $\cdot$ is the usual product over reals.

Next, we define the class of parity functions, which as we shall see forms an orthonormal basis for the space of n-bit Boolean functions.

Definition 2.2. Let $S \subseteq [n]$. Then define the S-th parity function $\chi_S : \{-1, 1\}^n \to \{-1, 1\}$ to be

$$\chi_S(x) = \prod_{i \in S} x_i.$$

For the special case where $S = \emptyset$, $\chi_S(x)$ is defined to be identically 1.

The name parity functions comes from the fact that the value of $\chi_S(x)$ is determined by the parity of -1s appearing in S. Observe that there are 2^n such parity functions. We now show that these functions form an orthonormal set.

Proposition 2.3. *The set of functions* $\{\chi_S(x) \mid S \subseteq [n]\}$ *is orthonormal. Let* $S, T \subseteq [n]$*. Then,*

$$< \chi_S, \chi_T >= \begin{cases} 1 & if\ S = T \\ 0 & otherwise \end{cases}$$

Proof. First, since the parity functions have range $\{-1, 1\}$ we have that

$$\begin{aligned} \chi_S(x) \cdot \chi_T(x) &= \prod_{i \in S} x_i \cdot \prod_{i \in T} x_i \\ &= \prod_{i \in S \Delta T} x_i \cdot \prod_{i \in S \cap T} x_i^2 \\ &= \prod_{i \in S \Delta T} x_i \\ &= \chi_{S \Delta T}(x). \end{aligned}$$

If S and T are equal, we have

$$\begin{aligned} < \chi_S, \chi_T > &= \mathbb{E}_{x \sim \{-1,1\}^n}[\chi_S(x) \cdot \chi_T(x)] \\ &= \mathbb{E}_{x \sim \{-1,1\}^n}[\chi_{S \Delta T}(x)] \\ &= \mathbb{E}_{x \sim \{-1,1\}^n}[\chi_{\emptyset}(x)] \\ &= 1. \end{aligned}$$

Otherwise, we have,

$$\begin{aligned} < \chi_S, \chi_T > &= \mathbb{E}_{x \sim \{-1,1\}^n}[\chi_S(x) \cdot \chi_T(x)] \\ &= \mathbb{E}_{x \sim \{-1,1\}^n}[\chi_{S \Delta T}(x)] \\ &= \mathbb{E}_{x \sim \{-1,1\}^n}[\prod_{i \in S \Delta T} x_i] \\ &= \prod_{i \in S \Delta T} \mathbb{E}_{x \sim \{-1,1\}^n}[x_i] \\ &= 0 \end{aligned}$$

where the fourth equality follows from independence, and the last equality follows from the fact that $x_i = -1$ and $x_i = 1$ are equally likely. □

Hence the 2^n parity functions form an orthonormal set. From orthogonlity we have that the parity functions are in fact linearly independent. Since the space of n-bit functions is 2^n dimensional, it follows that the parity functions span the space. Hence we can write any Boolean function as a multilinear polynomial:

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x) = \sum_{S \subseteq [n]} \hat{f}(S) \prod_{i \in S} x_i$$

where each $\hat{f}(S)$ is a real constant. We refer to $\hat{f}(S)$ as the Fourier coefficient of f on S. The following is an explicit formula for computing the Fourier coefficients of f.

Proposition 2.4. *Let $f : \{-1,1\}^n \to \mathbb{R}$ and let $S \subseteq [n]$. Then,*

$$\hat{f}(S) =< f, \chi_S >= \mathbb{E}_{x\sim\{-1,1\}^n}[f(x) \cdot \chi_S].$$

Proof. We can represent $f(x)$ as the polynomial $\sum_{T\subseteq[n]} \hat{f}(T)\chi_T(x)$. Then,

$$\begin{aligned} < f, \chi_S > &=< \sum_{T\subseteq[n]} \hat{f}(T)\chi_T, \chi_S > \\ &= \sum_{T\subseteq[n]} \hat{f}(T) < \chi_T, \chi_S > \\ &= \hat{f}(S)\cdot < \chi_S, \chi_S > \;+\; \sum_{T\neq S} \hat{f}(T) < \chi_T, \chi_S > \\ &= \hat{f}(S) \end{aligned}$$

where the second equality follows from linearity of the inner product and the last equality follows from orthonormality. □

Now, we prove a useful formula for computing the dot product of two arbitrary Boolean functions. We refer to this formula as Plancherel's Theorem.

Theorem 2.5. *Let $f, g : \{-1,1\}^n \to \mathbb{R}$. Then,*

$$< f, g >= \sum_{S\subseteq[n]} \hat{f}(S)\hat{g}(S)$$

.

Proof. We proceed as follows:

$$\begin{aligned} < f, g > &=< \sum_{S\subseteq[n]} \hat{f}(S)\chi_S, \sum_{T\subseteq[n]} \hat{f}(T)\chi_T > \\ &= \sum_{S,T\subseteq[n]} \hat{f}(S)\hat{g}(T) < \chi_S, \chi_T > \\ &= \sum_{S=T\subseteq[n]} \hat{f}(S)\hat{g}(T) < \chi_S, \chi_T > \;+\; \sum_{S\neq T\subseteq[n]} \hat{f}(S)\hat{g}(T) < \chi_S, \chi_T > \\ &= \sum_{S\subseteq[n]} \hat{f}(S)\hat{g}(S) \end{aligned}$$

where the second equality follows from linearity of inner product and the last equality follows from orthonormality. □

A useful corrolary follows:

Corollary 2.6. *Let $f : \{-1,1\}^n \to \mathbb{R}$. Then,*

$$< f, f > \;=\; \mathbb{E}_{x\sim\{-1,1\}^n}[f^2(x)] \;=\; \sum_{S\subseteq[n]} \hat{f}^2(S).$$

Moreover, if f takes values in $\{-1,1\}$ we have that $\sum_{S\subseteq[n]} \hat{f}^2(S) = 1$.

This formula is referred to as Parseval's Theorem.

3. Unique Games and (Strong) Approximation Resistance

In this section we discuss the Unique Games Conjecture due to Khot, and some of its connections to the theory of approximation resistance.

Definition 3.1. An instance of Unique Games, $\mathcal{I} = (G(V,E),[n],\{\pi_e|e \in E\})$, consists of a directed graph G, a set of labels $[n]$, and a set of permutations (also called constraints) $\{\pi_e|e \in E\}$ of $[n]$. A labelling of the graph is a function $L : [V] \to [n]$. We say that an edge $e = (u,v) \in E$ is satisfied by the labelling L if $(\pi_e \circ L)(u) = L(v)$. The goal is to produce a labelling function L so as to maximize the following objective function:

$$\frac{1}{|E|}|\{e = (u,v) \in E \ : \ (\pi_e \circ L)(u) = L(v)\}|$$

. We denote the maximum value of the above objective function as $OPT(\mathcal{I})$.

Conjecture 3.2. *For all $0 < \delta \leq 1$, there exists N, such that for instances of Unique Games $\mathcal{I} = (G(V,E),[N],\{\pi_e|e \in E\})$ it is NP-Hard to distinguish between the cases where $OPT(\mathcal{I}) \geq 1-\delta$ and $\delta \geq OPT(\mathcal{I})$.*

Remark 3.3. If a decision problem is NP-Hard assuming the Unique Games Conjecture, we say that the problem is UG-Hard

Recall that a predicate f is approximation resistant if it is NP-Hard to distinguish the case where an instance of $Max-k-CSP(f)$ is more than $1-\epsilon$ satisfiable from the case of being no more than $\rho(f)+\epsilon$ satisfiable. We can obtain many strong results for approximation resistance by relaxing the hardness condition to UG-Hardness. The following result of Austrin and Mossel gives a sufficient condition for approximation resistance under the Unique Games Conjecture.

Theorem 3.4. *Let f be a predicate on k variables, and suppose there exists a probability distribution μ over $\{-1,1\}^k$ such that,*

$$Pr_{x\sim\mu}[f(x) = 1] = 1,$$

for all $i \in [k]$,

$$\mathbb{E}_{x\sim\mu}[x_i] = 0,$$

and for all $i, j \in [k]$ where $i \neq j$,

$$\mathbb{E}_{x\sim\mu}[x_i x_j] = \mathbb{E}_{x\sim\mu}[x_i]\mathbb{E}_{x\sim\mu}[x_j].$$

Then, the Unique Games Conjecture implies that f is approximation resistant.

Thus, a sufficient condition for approximation resistance is the existence of a balanced, pairwise independent distribution over satisfying assignments. It would be interesting to obtain a necessary and sufficient condition for approximation resistance. A recent work of Khot, Tulsiani, and Worah gives a necessary and sufficient condition for a special case of approximation resistance called strong approximation resistance.

Definition 3.5. Fix a k-bit Boolean predicate f and let $\epsilon > 0$. If it is NP-hard to distinguish between the case where an instance of $Max-k-CSP(f)$ is $1-\epsilon$ satisfiable from the case where for all assignments x, $|\rho(f) - val(x)| \leq \epsilon$ we say f is strongly approximation resistant.

Where approximation resistance corresponds to intractibility of beating a random assignment, strong approximation resistance corresponds to the intractibility of differing from a random assignment in either direction. Moreover, strong approximation resistance implies approximation resistance. In addition, for odd predicates, i.e. those predicates for which $f(-x) = 1 - f(x)$, the two notions are in fact equivalent. This follows from the observation that any assignment which satisfies fewer than $\frac{1}{2} - \epsilon$ constraints can be negated (by switching the signs of each coordinate) to yield an assignment which satisfies more than $\frac{1}{2} + \epsilon$ constraints.

Before we can state the Khot et al. characterization of strong approximation resistance some definitions are in order.

Definition 3.6. Let μ be a probability distribution over $\{-1,1\}^k$. We define the symmetric matrix of first and second moments $\zeta(\mu) \in \mathbb{R}^{(k+1)\times(k+1)}$ as follows:

$$\text{for } i \in [k],\ \zeta(\mu)_{0,i} = \mathbb{E}_{x\sim\mu}[x_i]$$

$$\text{for distinct } i,j \in [k],\ \zeta(\mu)_{i,j} = \mathbb{E}_{x\sim\mu}[x_i x_j]$$

and all diagonal entries are 1.

Definition 3.7. Let f be a predicate on k variables. Define $\mathcal{C}(f) = \{\zeta(\mu) : Pr_{x\sim\mu}[f(x) = 1] = 1\}$.

Definition 3.8. Let $S \subset [k]$. Then define ζ^S to be the projection of ζ to the coordinates indexed by $\{0\} \cup S$. Let $\pi : S \to S$ be a permutation. Define $\zeta^{S,\pi}$ be ζ^S with rows and columns permuted by π. Finally, let $b \in \{-1,1\}^{|S|}$ be a vector of signs. Define $\zeta^{S,\pi,b} = \zeta^{S,\pi} \circ ((1\ b)(1\ b)^T)$ where $\circ$ is the entrywise product of two matrices.

Definition 3.9. Let Λ be a probability measure on $k+1 \times k+1$ real matrices, $S \subset [k]$, $\pi : S \to S$ be a permutation, and $b \in \{-1,1\}^{|S|}$. Denote by $\Lambda^{S,\pi,b}$ the measure on $|S|+1 \times |S|+1$ matrices obtained by randomly sampling a matrix ζ according to the measure Λ and then selecting $\zeta^{S,\pi,b}$.

Recall that any predicate $f : \{-1,1\}^k \to \{0,1\}$ can be written as the multilinear polynomial $\sum_{S\subset[k]} \hat{f}(S) \prod_{i\in S} x_i$. We are now ready to state the characterization.

Theorem 3.10. *Let f be a predicate on k variables. Then, f is strongly approximation resistant if and only if there exists a probability measure Λ supported on $\mathcal{C}(f)$ such that for all $t \in [k]$ the following function vanishes identically*

$$\Lambda^t = \sum_{|S|=t} \sum_{\pi:S\to S} \sum_{b\in\{-1,1\}^t} (\hat{f}(S)\Lambda^{S,\pi,b} \prod_{i=1}^{t} b_i).$$

Notice that this characterization is in some sense a generalization of the sufficient condition obtained by Austrin and Mossel. The condition that satisfying assignments of f support a balanced pairwise independent distribution is equivalent to $I \in \mathcal{C}(f)$. Then we can simply take $\Lambda(I) = 1$. There are several open questions regarding this characterization.

Question 3.11. *Is the characterization recursive?*

Question 3.12. *When a Λ satisfying the characterization exists, is it always finitely supported?*

Question 3.13. *Can the characterization take simpler form for explicit families of predicates?*

More concretely, it would be interesting to apply the characterization on balanced linear threshold functions. It is currently open whether or not there exists an approximation resistant linear threshold predicate. Perhaps the characterization takes simpler form for these predicates.

4. Linear Threshold Functions

We now turn to majority and majority-like predicates, the so-called linear threshold functions. We will explore approximability results for a particular class of such functions, and discuss a possible candidate for approximation resistance. Unless stated otherwise, we will consider functions $f : \{-1, 1\}^k \to \{-1, 1\}$.

Definition 4.1. Define $L : \{-1, 1\}^k \to \mathbb{R}$ as $L(x) = w_0 + w_1x_1 + w_2x_2 + ... + w_kx_k$, where each $w_i \in \mathbb{R}$. We shall often refer to these w_i's as weights. Let $f : \{-1, 1\}^k \to \{-1, 1\}$ be such that $f(x) = sgn(L(x))$. Then we say that f is a linear threshold function. For definiteness, in the case that $L(x) = 0$, we define $f(x) = 1$.

Example 4.2. Let k be odd, and let $L(x)$ be such that $w_0 = 0$, and $w_1 = w_2 = ... = w_k = 1$. Then, $f(x) = sgn(L(x))$ is 1 if and only if a majority of the coordinates of x are equal to 1. For this reason this linear threshold function is called the majority on k variables and is denoted as $Maj_k(x)$. Notice that it is not necessary for the weights to all be equal to 1. In fact, any positive constant will suffice.

Remark 4.3. A linear threshold function $f(x)$ can be represented as the sign of infinitely many different functions, $L(x)$. This follows from the observation that scaling the weights of $L(x)$ by a positive real constant will not change the value of $f(x)$ on any of the inputs. Furthermore, observe that we can add small constants to each weight without changing the value of $f(x)$. Consequently given a linear threshold function $f(x) = sgn(L(x))$, we can find $L'(x)$ with all integer weights such that $f(x) = sgn(L(x))$.

Remark 4.4. In our discussion, we will only consider predicates with integer weights. For additional simplicity, we consider functions with non-negative weights, and in particular we require that $w_0 = 0$. Additionally, we will restrict our predicates to those for which $L(x) \neq 0$ for all inputs. Now, assuming these restrictions, observe that for every input x, $L(-x) = -L(x)$. So, for all predicates we consider, $f(-x) = -f(x)$, or in other words, f is odd. We also adopt the convention of writing the weights in non-decreasing order.

Recall that any boolean function $f(x)$ can be represented as a multilinear polynomial of the form, $f(x) = \sum_{S\subseteq[n]} \hat{f}(S) \cdot \prod_{i\in S} x_i$. The following theorem, due to Chow, states that a linear threshold function is completely determined by its Fourier coefficients on the empty and unit sets.

Theorem 4.5. *Let $f(x) : \{-1, 1\}^k \to \{-1, 1\}$ be a linear threshold function, and let $g(x) : \{-1, 1\}^k \to \{-1, 1\}$ be such that $\hat{f}(S) = \hat{g}(S)$ for $|S| < 2$. Then for all x, $f(x) = g(x)$.*

Proof. Let $f(x) = sgn(L(x))$. Observe that we can perturb the weights of $L(x)$ such that $L(x) \neq 0$ on all inputs, without changing the value of $f(x)$. So without loss of generality we assume $L(x)$ is always non-zero. Since the Fourier coeficients of $L(x)$ are nonzero only on the empty and unit size sets, by Plancherel's Theorem we have that

$$\sum_{|S|<2} \hat{f}(S)\hat{L}(S) = \mathbb{E}_{x\sim\{-1,1\}^k}[f(x)L(x)].$$

Similarly, we have that

$$\sum_{|S|<2} \hat{g}(S)\hat{L}(S) = \mathbb{E}_{x\sim\{-1,1\}^k}[g(x)L(x)].$$

Since, $\hat{f}(S) = \hat{g}(S)$ on empty and unit sets, we have that $\mathbb{E}_{x\sim\{-1,1\}^k}[f(x)L(x)] = \mathbb{E}_{x\sim\{-1,1\}^k}[g(x)L(x)]$. Now, since $f(x) = sgn(L(x))$, $f(x)L(x) = |L(x)|$, and since $g(x)$ is boolean valued, we have that $f(x)L(x) \geq g(x)L(x)$. But since these two functions are equal in expectation and non-zero, we have that $f(x) = g(x)$. □

The Fourier coefficients of a function on empty and unit sets are often called the function's Chow parameters. It is currently unknown how to construct a linear threshold function $f(x)$ exactly from its Chow parameters. A naive approach to this problem would be to simply use the Chow parameters themselves as weights. Indeed this method does work for several functions, including $Maj_k(x)$. It is not hard to show that this does not work in general.

Example 4.6. Consider the linear threshold function on 5 variables, $f(x) = sgn(x_1+x_2+x_3+x_4+3x_5)$. The Chow parameters of this function are $\hat{f}(\{i\}) = \frac{1}{8}$ for $1 \leq i \leq 4$ and $\hat{f}(\{5\}) = \frac{7}{8}$. The function we obtain by plugging in these parameters as weights (after scaling) is $f'(x) = sgn(x_1 + x_2 + x_3 + x_4 + 7x_5)$. However, $f(1,1,1,1,-1) = 1$ while $f'(1,1,1,1,-1) = -1$.

The functions for which the Chow parameters yield the correct weights, referred to in the literature as Chow-robust functions, present an interesting class of functions for which to explore approximability. As we shall see, the optimization problem $Max - CSP(f)$ for Chow-robust f is in fact nontrivially approximable. Before proceeding we note (but do not prove) a useful fact about the Chow parameters of Majority functions.

Fact 4.7. *Let* $f : \{-1,1\}^k \to \{-1,1\}$ *be the Majority function on* k *variables. Then, for each* $i \in [k]$, $\hat{f}(\{i\}) = 2^{1-k} \cdot \binom{k-1}{\frac{k-1}{2}} = \Theta(\frac{1}{\sqrt{k}})$.

Recall that for an instance of $Max - CSP(f)$ on k variables, m constraints, and with matrices $v \in [N]^{m\times k}$, $s \in \{-1,1\}^{m\times k}$, the objective function we wish to maximize is the advantage of an assignment x which we can write as:

$$Adv(x) = \frac{1}{m}\sum_{i=1}^{m} f(s_{i,1}x_{v_{i,1}}, ..., s_{i,k}x_{v_{i,k}}).$$

We call this function $Adv(x)$ rather than $Val(x)$ to emphasize that we've switched notation to the range of f being in $\{-1,1\}$. We have restricted ourselves to odd predicates, so for a random assignment any constraint is satisfied with probability $\frac{1}{2}$. By linearity of expectation, we can thus expect to satisfy half of all constraints. In other words, $\mathbb{E}_{x\sim\{-1,1\}^k}[Adv(x)] = 0$. Using Fourier expansion we can re-write the advantage as:

$$Adv(x) = \sum_{S\subseteq[N],|S|\leq k} c(S) \cdot \prod_{i\in S} x_i,$$

where $c(S)$ is simply $\hat{f}(S)$ scaled to reflect the signs of s and the exact choice of variables in each constraint in the instance. In 2005 Hast gave a general algorithm which achieves a positive advantage (and thus a non-trivial approximation) whenever the coefficients of linear terms of the above objective function are non-zero. A recent paper by Cheraghchi et al. gave a modified version of the algorithm for odd predicates, which we present below. First we prove a technical lemma bounding the sum of the coefficients $c(S)$.

Lemma 4.8. *Let $f : \{-1,1\}^k \to \{-1,1\}$, and let I be an instance of $Max - CSP(f)$ with objective function*

$$Adv(x) = \sum_{S \subseteq [N], |S| \leq n} c(S) \cdot \prod_{i \in S} x_i$$

Then $\sum_{|S|=n} |c(S)| \leq \binom{k}{n}^{1/2}$.

Proof. Let $n \leq k$. Clearly, $|c(S)| \leq |\hat{f}(S)|$. Thus, we have

$$\begin{aligned} \sum_{|S|=n} |c(S)| &\leq \sum_{|S|=n} |\hat{f}(S)| \\ &\leq (\sum_{n=1}^{\binom{k}{n}} 1)^{1/2} (\sum_{|S|=n} \hat{f}^2(S))^{1/2} \\ &\leq (\sum_{n=1}^{\binom{k}{n}} 1)^{1/2} \\ &= \binom{k}{n}^{1/2}, \end{aligned}$$

where the second inequlity follows from Cauchy-Schwarz and the third inequality follws from Parseval's theorem. □

Theorem 4.9. *Let $f : \{-1,1\}^k \to \{-1,1\}$ be an odd function, and let I be an instance of $Max - k - CSP(f)$ with objective function*

$$Adv(x) = \sum_{S \subseteq [N], |S| \leq n} c(S) \cdot \prod_{i \in S} x_i.$$

If $\sum_{i=1}^{k} |c(\{i\})| \geq \alpha$, then there exists a polynomial time algorithm which outputs an assignment x such that $\mathbb{E}[Adv(x)] \geq \frac{\alpha^{3/2}}{8k^{3/4}}$.

Proof. Let $\beta = \alpha^{1/2}(2k^{3/4})^{-1}$. Set each x_i independently and randomly such that $Pr[x_i = 1] = \frac{1}{2} + \beta \frac{sgn(c(\{i\}))}{2}$. Clearly, $\mathbb{E}[c(\{i\})x_i] = \beta |c(\{i\})|$ and $|\mathbb{E}[\prod_{i \in S} x_i]| =$

$\beta^{|S|}$. Thus we have that

$$\begin{aligned}\mathbb{E}[Adv(x)] &= \mathbb{E}[\sum_{S\subseteq[N],|S|\leq n} c(S)\cdot\prod_{i\in S} x_i] \\ &= \mathbb{E}[\sum_{i=1}^{k} c(\{i\})x_i] + \mathbb{E}[\sum_{|S|\geq 3} c(S)\prod_{i\in S} x_i] \\ &\geq \beta\sum_{i=1}^{k}|c(\{i\})| - \sum_{|S|\geq 3}\beta^{|S|}|c(S)| \\ &\geq \beta\alpha - \sum_{n=3}^{k}\beta^n\binom{k}{n}^{1/2},\end{aligned}$$

where the first inequality follows from the fact that $c(S) = 0$ on sets of size 2. Now, since $\alpha \leq 1$, $\beta \leq (2\sqrt{k})^{-1}$. Thus, $\sum_{n=3}^{k}(\beta^2 k)^n \leq \beta^6 k^3 \sum_{n=0}^{\infty}\frac{1}{2^n} \leq 2\beta^6 k^3$. Then by Cauchy-Schwarz,

$$\begin{aligned}\sum_{n=3}^{k}\beta^n\binom{k}{n}^{1/2} &\leq (\sum_{n=3}^{k}\frac{1}{k}^n\binom{k}{n})^{1/2}(\sum_{n=3}^{k}(\beta^2 k)^n)^{1/2} \\ &\leq (1+\frac{1}{k})^{n/2}(2\beta^6 k^3)^{1/2} \\ &\leq 3\beta^3 k^{3/2}\end{aligned}$$

where the second inequality follows from $\sum_{n=0}^{k}\frac{1}{k}^n\binom{k}{n} = (1+\frac{1}{k})^k$. Hence, $\mathbb{E}[Adv(x)] \geq \beta\alpha - 3\beta^3 k^{3/2} = \frac{\alpha^{3/2}}{8k^{3/4}}$. □

We can apply this theorem to the majority function as follows.

Corollary 4.10. *Let $f(x) = Maj_k(x)$. Suppose I is an instance of $Max-k-CSP(f)$ which is $(1-\frac{\delta}{k+1})$-satisfiable. Then in polynomial time, we can produce an assignment $x \in \{-1,1\}^k$ for which $\mathbb{E}[Adv(x)] = \Omega(\frac{1-\delta}{k}^{3/2})$.*

Remark 4.11. Note that we could have assumed I is $1-\epsilon$ satisfiable, but we use the notation above for analytic convenience.

Proof. Let $x*$ be the assignment which maximizes the number of satisfied constraints. Then clearly, $\sum_{i=1}^{k} c(\{i\})x*_i \leq \sum_{i=1}^{k}|c(\{i\})|$. Note that all the non-zero Chow parameters are equal to some constant $\hat{f}(\{i\})$. Any satisfied constraint must add at least $\hat{f}(\{i\})$ to $\sum_{i=1}^{k}|c(\{i\})|$, while any unsatisfiable constraint subtracts at most $n\hat{f}(\{i\})$. Hence,

$$\sum_{i=1}^{k}|c(\{i\})| \geq (1-\frac{\delta}{k+1})\hat{f}(\{i\}) - \frac{\delta}{k+1}k\hat{f}(\{i\}) = (1-\delta)\hat{f}(\{i\})$$

. The result now follows from Theorem 1.8 and the fact that $\hat{f}(\{i\}) = \Theta(\frac{1}{\sqrt{k}})$. □

We can also extend this result to the previously mentioned Chow-robust predicates. We refine our definition of Chow-robustness with an explicit bound.

Definition 4.12. Let $f(x)$ be an odd linear threshold function, and let $\gamma > 0$. We say that $f(x)$ is γ-robust if for all satisfying assignments x, $\gamma \geq \sum_{i=1}^{k}\hat{f}(\{i\})x_i$.

Clearly a function is Chow-robust if and only if there exists positive γ such that $f(x)$ is γ-robust. We obtain the following approximation for Chow-robust functions.

Corollary 4.13. *Let $f(x)$ be γ-robust, and let I be an instance of $Max-k-CSP(f)$ which is $1-\frac{\delta\cdot\gamma}{\gamma+\sum_{j=1}^{k}\hat{f}(\{j\})}$ satisfiable, with $\delta<1$. Then in polynomial time, we can produce an assignment $x\in\{-1,1\}^k$ for which $\mathbb{E}[Adv(x)]=\frac{(\gamma(1-\delta))^{3/2}}{8k^{3/4}}$.*

Proof. For any assignment x, we have that

$$\sum_{n=1}^{k}|c_n|\geq\sum_{n=1}^{k}c_nx_n=\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k}\hat{f}(\{j\})s_{i,j}x_{l_{i,j}}$$

. Since I is $1-\frac{\delta\cdot\gamma}{\gamma+\sum_{j=1}^{k}\hat{f}(\{j\})}$-satisfiable, for the optimal assignment x^*, we have that $\sum_{j=1}^{k}\hat{f}(\{j\})s_{i,j}x^*_{l_{i,j}}\geq\gamma$ for at least $1-\frac{\delta\cdot\gamma}{\gamma+\sum_{j=1}^{k}\hat{f}(\{j\})}$ fraction of constraints. For the unsatisfied constraints, $\sum_{j=1}^{k}\hat{f}(\{j\})s_{i,j}x^*_{l_{i,j}}\geq-\sum_{j=1}^{k}\hat{f}(\{j\})$. Thus we have that,

$$\begin{aligned}\sum_{n=1}^{k}|c_n|&\geq\sum_{n=1}^{k}c_nx_n\\&=\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{k}\hat{f}(\{j\})s_{i,j}x_{l_{i,j}}\\&\geq(1-\frac{\delta\cdot\gamma}{\gamma+\sum_{j=1}^{k}\hat{f}(\{j\})})\gamma-\frac{\delta\cdot\gamma}{\gamma+\sum_{j=1}^{k}\hat{f}(\{j\})}\sum_{j=1}^{k}\frac{f}{(}\{j\})\\&=(1-\delta)\gamma.\end{aligned}$$

We apply Theorem 4.8 to finish the proof. □

Thus, we have that the family of Chow-robust linear threshold functions is approximable. It would be interesting to find a linear threshold function which is in fact approximation resistant, and if they exist, it would be of interesting to find distributions on weights for which there is a high probability of approximation resistance.

Definition 4.14. Let $W\subseteq\mathbb{N}$ be finite, and let μ be a probability distribution over W. Let $w_1,...,w_k$ be chosen independently from W according to μ. Then, we say $f(x)=sgn(w_1x_1+...+w_kx_k)$ is a μ-random linear threshold function.

The simplest such distribution to study is when the weights are chosen from the uniform distribution over $[m]$, where m is an integer. In our discussion, for convenience, we will restrict m to be odd. Before determining whether linear threshold functions sampled from such a distribution are with high probability approximation resistant, it is useful to check that we are unlikely to get good approximations using Hast's algorithm. The only known predicates which can be nontrivially approximated using Hast's algorithm are the Chow robust predicates. Therefore we would like to show that for large values of γ, we are unlikely to sample a γ-robust predicate. Towards that end we prove Proposition 4.15, which gives a tradeoff between γ and the probability of sampling a γ-robust predicate.

Lemma 4.15. *Let μ be the uniform distribution over $[m]$ for some even positive integer m. If f is μ-random,*

$$Pr[\forall i,\ \hat{f}(\{i\}) \leq 2^{2-(k-s\sqrt{k})/2}\binom{(k-s\sqrt{k}-2)/2}{(k-s\sqrt{k}-2)/4}] \geq 1-ke^{-t^2/2}$$

where $0 \leq s \leq \sqrt{k}$.

Proof. First, we write $f(x) = sgn(w_1x_1 + ... + w_kx_k)$, where the weights are written in nondecreasing order. We can upperbound $\hat{f}(\{k\})$ as follows. Let $g(x) = sgn(w_jx_j + w_{j+1}x_{j+1} + ...w_kx_k)$, where the weights $Z = \{w_j, ..., w_k\}$ are precisely those which are greater than $\frac{m}{2}$. Clearly $\hat{g}(\{k\}) \geq \hat{f}(\{k\})$, since the fraction of weights on x_k is higher in g than in f. The influence of x_k in g, is maximized when $w_j = ... = w_{k-1} = \frac{m}{2}+1$ and $w_k = m$. Clearly such a predicate is equivalent to $h(x) = sgn(x_j + ... + x_{k-1} + 2x_k)$, and moreover the influence of x_k in h is

$$2M\hat{aj}_{|Z|}(\{k\}) = 2^{2-(|Z|-s\sqrt{|Z|})/2}\binom{(|Z|-s\sqrt{|Z|}-2)/2}{(|Z|-s\sqrt{|Z|}-2)/4}.$$

Hence

$$\hat{f}(\{k\}) \leq 2M\hat{aj}_{|Z|}(\{k\}).$$

Now, since the probability that any given weight is in Z is 1/2, $\mathbb{E}[|Z|] = \frac{k}{2}$. Then by Chernoff's inequality we have that

$$Pr[|Z| \leq (k-s\sqrt{k})/2] \leq e^{-s^2/2}$$

where $0 \leq s \leq \sqrt{k}$. Hence,

$$Pr[\hat{f}(\{k\}) \geq 2^{2-(k-s\sqrt{k})/2}\binom{(k-s\sqrt{k}-2)/2}{(k-s\sqrt{k}-2)/4}] = Pr[|Z| \leq (k-s\sqrt{k})/2]$$
$$\leq e^{-s^2/2}.$$

Since all other weights w_i are smaller than w_k, the above inequalities hold for all $\hat{f}(\{i\})$. Finally, applying the union bound inequality gives the result. □

Now, we give the main result.

Proposition 4.16. *Let μ be the uniform distribution over $[m]$ for some even positive integer m. Fix $\gamma > 2^{2-(k-s\sqrt{k})/2}\binom{(k-s\sqrt{k}-2)/2}{(k-s\sqrt{k}-2)/4}$. Then, if f is μ-random, $Pr[f \text{ is not } \gamma-\text{robust}] \geq 1-ke^{-s^2/2}$, where $0 \leq s \leq \sqrt{k}$.*

Proof. Recall that if f were γ-robust, on satisfying assignments $\sum_{i=1}^k \hat{f}(\{i\})x_i \geq \gamma$, and on all others $\sum_{i=1}^k \hat{f}(\{i\})x_i \leq \gamma$.
Now, consider the assignment of x which sets all coordinates to 1. On this assignment the sum $\sum_{i=1}^k \hat{f}(\{i\})x_i$ is positive. We can negate the assignments of each coordinate in some arbitrary order, and at each step we will reduce the value of this sum. After negating some coordinate j, the sum becomes negative. However, for f to be γ-robust negating such a coordinate must reduce the sum from greater than or equal to γ to less than or equal to γ. In this case, $\hat{f}(\{j\}) \geq \gamma$. However, by Lemma 4.14, the probability that $\hat{f}(\{j\}) < \gamma$ is at least $1-ke^{-s^2/2}$. Hence $Pr[f \text{ is not } \gamma-\text{robust}] \geq 1-ke^{-s^2/2}$. □

Thus, as we increase the size of γ it becomes exponentially less likely that we sample a γ-robust predicate. In fact, in practice it is unlikely that we draw a γ-robust predicate for any size γ. We take this as a "sanity-check" that it is not obvious that randomly sampled predicates yield large (or any) approximation.

Acknowledgments. The author thanks Professor Madhur Tulsiani, for his guidance throughout the project.

References

[1] Per Austrin and Elchanan Mossel (2009). Approximation Resistant Predicates From Pairwise Independence
[2] Mahdi Cheraghchi, Johan Håstad, Marcus Isaksson, and Ola Svensson (2011). Approximating Linear Threshold Predicates
[3] Gustav Hast (2005). Beating a Random Assignment
[4] Johan Håstad (2007). On the efficient approximability of constraint satisfaction problems
[5] Subhash Khot (2010). On the Unique Games Conjecture
[6] Subhash Khot, Madhur Tulsiani, Pratik Worah (2014). A Characterization of Strong Approximation Resistance
[7] Ryan O'Donnell (2014). Analysis of Boolean Functions