Electronic Colloquium on Computational Complexity, Report No. 132 (2010)

Approximating Linear Threshold Predicates*

Mahdi Cheraghchi † Johan Håstad ‡ Marcus Isaksson § Ola Svensson ¶

Abstract

We study constraint satisfaction problems on the domain $\{-1, 1\}$, where the given constraints are homogeneous linear threshold predicates. That is, predicates of the form $\operatorname{sgn}(w_1 x_1 + \cdots + w_n x_n)$ for some positive integer weights $w_1, \ldots, w_n$. Despite their simplicity, current techniques fall short of providing a classification of these predicates in terms of approximability. In fact, it is not easy to guess whether there exists a homogeneous linear threshold predicate that is approximation resistant or not.

The focus of this paper is to identify and study the approximation curve of a class of threshold predicates that allow for non-trivial approximation. Arguably the simplest such predicate is the majority predicate $\operatorname{sgn}(x_1 + \cdots + x_n)$, for which we obtain an almost complete understanding of the asymptotic approximation curve, assuming the Unique Games Conjecture. Our techniques extend to a more general class of "majority-like" predicates and we obtain parallel results for them. In order to classify these predicates, we introduce the notion of *Chow-robustness* that might be of independent interest.

1 Introduction

Constraint satisfaction problems or more succinctly CSPs are at the heart of theoretical computer science. In a CSP we are given a set of constraints, each putting some restriction on a constant size set of variables. The variables can take values in many different domains but in this paper we focus on the case of variables taking Boolean values. This is the most fundamental case and it has also attracted the most attention over the years. We also focus on the case where each condition is given by the same predicate, P, applied to a sequence of literals. The role of this predicate P is key in this paper and as it is more important for us than the number of variables, we reserve the letter n for the arity of this predicate while using N to be the number of variables in the instance. We also reserve m to denote the number of constraints.

Traditionally we ask for an assignment that satisfies all constraints and in this case it turns out that all Boolean CSPs are either NP-complete or belong to P and this classification was completed already in 1978 by Schaefer [15]. In this paper we study Max-CSPs which are optimization problems where we want to satisfy as many constraints as possible. Almost all Max-CSPs of interest turn out to be NP-hard and the main focus is that of efficient approximability.

The standard measure of approximability is given by a single number C and an algorithm is a C-approximation algorithm if it, on each input, finds an assignment with an objective value

*An extended abstract of this paper appeared in the proceedings of APPROX, 2010. This research is supported by the ERC Advanced investigator grant 226203. M. Cheraghchi is supported by the ERC Advanced investigator grant 228021.

†School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland. mahdi.cheraghchi@epfl.ch
‡KTH - Royal Institute of Technology, Stockholm, Sweden. johanh@kth.se
§Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden. maris@chalmers.se.
¶KTH - Royal Institute of Technology, Stockholm, Sweden. osven@kth.se

ISSN 1433-8092

that is at least C times the optimal value. Here we might allow randomization and be content if the assignment found satisfies these many constraints on average. A more refined question is to study the approximation curve where for each constant c, assuming that the optimal assignment satisfies cm constraints, we want to determine the maximal number of constraints that we can satisfy efficiently.

To get a starting point to discuss the quality of approximation algorithms it is useful to first consider the most simple algorithm that chooses the values of the variables randomly and uniformly from all values in $\{0,1\}^N$. If the predicate P is satisfied by t inputs in $\{0,1\}^n$ it is easy to see that this algorithm, on the average, satisfies $mt2^{-n}$ constraints. By using the method of conditional expectations it is also easy to deterministically find an assignment that satisfies this number of constraints.

A very strong type of hardness result possible for a Max-CSP is to prove that, even for instances where the optimal assignment satisfies all constraints, it is NP-hard to find an assignment that does significantly better (by a constant factor independent of N) than the above trivial algorithm. We call such a predicate "approximation resistant on satisfiable instances". A somewhat weaker, but still strong, negative result is to establish that the approximation ratio given by the trivial algorithm, namely $t2^{-n}$, is the best approximation ratio that can be obtained by an efficient algorithm. This is equivalent to saying that we cannot satisfy significantly more than $mt2^{-n}$ constraints when given an almost satisfiable instance. We call such a predicate "approximation resistant". It is well known that, unless P=NP, Max-3-Sat (i.e. when P is the disjunction of the three literals) is approximation resistant on satisfiable instances and Max-3-Lin (i.e. when P is the exclusive-or of three literals) is approximation resistant [8].

When it comes to positive results on approximability the most powerful technique is semi-definite programming introduced in this context in the classical paper by Goemans and Williamson [6] studying the approximability of Max-Cut, establishing the approximability constant $\alpha_{GW} \approx .878$. In particular, this result implies that Max-Cut is not approximation resistant. Somewhat surprisingly as proved by Khot et al. [12], this constant has turned out, assuming the Unique Games Conjecture, to be best possible. We note that these results have been extended in great generality and O'Donnell and Wu [14] determined the complete approximation curve of Max-Cut.

The general problem of determining which predicates are approximation resistant is still not resolved but as this is not the main theme of this paper let us cut this discussion short by mentioning a general result by Austrin and Mossel [2]. This paper relies on the Unique Games Conjecture by Khot [11] and proves that, under this conjecture, any predicate such that the set $P^{-1}(1)$ supports a pairwise independent measure is approximation resistant.

On the algorithmic side there is a general result by Hast, [7], that is somewhat complementary to the result of Austrin and Mossel. Hast considers the real valued function $P^{\leq 2}$ which is the sum of the linear and quadratic parts of the Fourier expansion of P. Oversimplifying slightly, the result by Hast says that if $P^{\leq 2}$ is positive on all inputs accepted by P then we can derive a non-trivial approximation algorithm and hence P is not approximation resistant.

To see the relationship between the results of Austrin and Mossel, and Hast, note that the condition of Austrin and Mossel is equivalent to saying that there is a probability distribution on inputs accepted by P such that the average of any unbiased quadratic function[1] is 0. In contrast, Hast needs that a particular unbiased quadratic function is positive on all inputs accepted by P. It is not difficult to come up with predicates that satisfies neither of these two conditions and hence we do not have a complete classification, even if we are willing to assume the Unique Games Conjecture. The combination of the two results, however, points to the class of predicates that can

[1]Throughout this work, we find it more convenient to represent Boolean values by {-1,+1} rather than {0,1}.

be written on the form

$$P(x) = \operatorname{sgn}(Q(x))$$

for a quadratic function Q as an interesting class of predicates to study and this finally brings us to the topic of this paper. We study this scenario in the simplest form by assuming that Q is in fact a linear function, $L : \{-1, 1\}^n \mapsto \{-1, 1\}$, without a constant term. In other words we have

$$P(x) = \operatorname{sgn}(L(x)) = \operatorname{sgn}\left(\sum_{i=1}^{n} w_i x_i\right),$$

for some, without loss of generality, positive integral weights $(w_i)_{i=1}^n$. Note that if we allow a constant term in L the situations is drastically different as for instance 3-SAT is the sign of linear form if we allow a non-zero constant term. One key difference is that a probability distribution supported on the set "$L(x) > 0$" cannot have even unbiased variables in the case when L is without constant term and thus hardness results such as the result by Austrin and Mossel do not apply.

To make life even simpler we make sure that L never takes the value 0 and as $L(-x) = -L(x)$, P accepts precisely half of the inputs and thus the number of constraints satisfied by a random assignment is, on the average, $m/2$.

The simplest such predicate is majority of an odd number of inputs. For this predicate it easy to see that Hast's condition is fulfilled and hence, for any odd value of n, his results imply that majority is not approximation resistant. This result generalizes to "majority-like" functions as follows. For a linear threshold functions, the Chow parameters, $\vec{P} = (\hat{P}(i))_{i=0}^n$, [3] are for, $i > 0$, defined to be the correlations between the output of the function and inputs x_i. We have that $\hat{P}(0)$ is the bias of the function and thus in our case this parameter is always equal to 0 and hence ignored.

Now if we order the weights $(w_i)_{i=1}^n$ in nondecreasing order then also the $\hat{P}(i)$'s are nondecreasing but in general quite different from the weights. It is well known that the Chow parameters determine the threshold function uniquely [3] but the computational problem of given $\vec{P}$, how to recover the weights, or even to compute P efficiently is an interesting problem and several heuristics have been proposed [10, 17, 9, 4] together with an empirical study that compares various methods [18]. More recently, the problem of finding an approximation of P given the Chow parameters has received increased attention, see e.g. [13] and [5]. The most naive method is to use $\vec{P}$ as weights. This does not work very well in general but this is a case of special interest to us as it is precisely when this method gives us back the original function that we can apply Hast's results directly. We call such a threshold function "Chow-robust" and we have not been able to find the characterization of this class of functions in the literature. If we ignore some error terms and technical conditions a sufficient condition to be Chow-robust is roughly that

$$\sum_{i=1}^{n} (w_i^3 - w_i) \leq 3 \sum_{i=1}^{n} w_i^2 \tag{1}$$

and thus it applies to functions with rather modest weights. We believe that this condition is not very far from necessary but we have not investigated this in detail.

Having established non-approximation resistance for such predicates we turn to study the full curve of approximability and, in asymptotic sense as a function of n, we get almost tight answers establishing both approximability results and hardness results. Our results do apply with degrading constants to more general threshold functions as our predicate P but let us here state them for majority. We have the following theorem.

Theorem 1.1. *(Informal) Given an instance of Max-Maj-n with n odd and m constraints, assume that the optimal assignment satisfies* $(1-\frac{\delta}{n+1})m$*, for* $\delta < 1$*. Then it is possible to efficiently find an assignment that satisfies* $\left(\frac{1}{2} + \Omega\left(\frac{(1-\delta)^{3/2}}{n^{1/2}}\right) - \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right)\right) m$ *constraints.*

Thus for large n we need almost satisfiable instances to get above the threshold $\frac{1}{2}$ obtained by a a random assignment. This might seem weak but we prove that this is probably the correct threshold.

Theorem 1.2. *(Informal) Assume the Unique Games Conjecture and let* $\epsilon > 0$ *be arbitrary. Then it is NP-hard to distinguish instances of Max-Maj-n where the optimal value is* $(1 - \frac{1}{n+1} - \epsilon)m$*, from those where the optimal value is* $(\frac{1}{2} + \epsilon)m$*.*

This proves that the range of instances to which Theorem 1.1 applies is essentially the correct one. A drawback of this theorem is that the error term in Theorem 1.1 dominates the systematic contribution of $(1-\delta)^{3/2}n^{-1/2}$ for δ very close to 1 and hence the threshold is not sharp. We are, however, able to sharply locate the threshold where something nontrivial can be done by combining our result with the general results by Hast. For details, see Section 3.

To see that the advantage obtained by the algorithm is also the correct order of magnitude we have the following theorem.

Theorem 1.3. *(Informal) Assume the Unique Games Conjecture and let* $\epsilon > 0$ *be arbitrary. Then there is an absolute constant c such that it is NP-hard to distinguish instances of Max-Maj-n where the optimal value is* $(1-\epsilon)m$*, from those where the optimal value is* $(\frac{1}{2} + \frac{c}{\sqrt{n}} + \epsilon)m$*.*

In summary, we get an almost complete understanding of the approximability curve of majority, at least in an asymptotic sense as a function of n. This complements the results for majority on three variables, for which there is a 2/3-approximation algorithm [19] and it is NP-hard to do substantially better [8].

The idea of the algorithm behind Theorem 1.1 is quite straightforward while its analysis gets rather involved. We set up a natural linear program which we solve and then use the obtained solution as biases in a randomized rounding. The key problem that arises is to carefully analyze the probability that a sum of biased Boolean variables is positive. In the case of majority-like variables we have the additional complication of the different weights. This problem is handled by writing the probability in question as a complex integral and then estimating this integral by the saddle-point method.

The hardness results given in Theorem 1.2 and Theorem 1.3 resort to the techniques of Austrin and Mossel [2]. The key to these results is to find suitable pairwise independent distributions relating to our predicate. In the case of majority it is easy to find such distributions explicitly, while in the case of more general weights the construction gets more involved.

In particular, we need to answer the following question: What is the minimal value of $\Pr[L(x) < 0]$ when x is chosen according to a pairwise independent distribution. This is a nice combinatorial question of independent interest.

An outline of the paper is as follows. In Section 2, we present notations and conventions used throught the paper, and also prove a result on weighted sums of balanced Bernoulli random variables that is used in the following sections. This is followed by the adaptation of Hast's algorithm for odd Chow-robust predicates and the proof that (essentially) the condition $\sum_{j=1}^{n} w_j^3 - w_j \leq 3\sum_{j=1}^{n} w_j^2$ on the weights is sufficient for a predicate to be Chow-robust. In Section 4, we present and analyze our main algorithm for Chow-robust predicates (Theorem 1.1 for the special case of majority). These positive results are then complemented in Section 5 where we show essentially tight hardness results assuming the increasingly prevalent Unique Games Conjecture. Finally, we discuss the obtained results together with interesting future directions (Section 6).

2 Preliminaries and Basic Technical Tools

In this section we introduce some notation and recall some results in complex analysis.

2.1 Notation

We consider the optimization problem Max-CSP(P) for homogeneous linear threshold predicates $P : \{-1,1\}^n \to \{-1,1\}$ of the form $P(x) = \text{sgn}(w_1 x_1 + \cdots + w_n x_n)$, where we assume that the weights are non-decreasing positive integers $1 \leq w_1 \leq \ldots \leq w_n$ such that $\sum_{j=1}^n w_j$ is odd and $w_{\max} := \max_j w_j = w_n$. The special case of equal weights, which requires n to be odd, is denoted by Maj_n, and we also write Max-Maj-n for Max-CSP(Maj_n). Using Fourier expansion, any such function can be written uniquely as

$$P(x) = \sum_{S \subseteq [n]} \hat{P}(S) \prod_{j \in S} x_j.$$

The Fourier coefficients are given by

$$\hat{P}(S) = \mathbb{E}[P(X) \prod_{j \in S} X_j],$$

where X is uniform on $\{-1,1\}^n$. Since all homogeneous linear threshold predicates are odd we have $\hat{P}(S) = 0$ when $|S|$ is even. We also write $\hat{P}(j) = \hat{P}(\{j\})$ for the first level Fourier coefficients (i.e. the Chow parameters) and let $P^{-1}(1)$ denote the set of assignments that satisfy P, i.e. $P^{-1}(1) = \{x : P(x) = 1\}$.

For an instance $\mathcal{I} = (m, N, l, s)$ of Max-CSP(P) consisting of m constraints, N variables and matrices $l \in N^{m \times n}, s \in \{-1,1\}^{m \times n}$, the objective is to maximize the number of satisfied constraints or, equivalently, the average advantage

$$\mathsf{Adv}(x) := \frac{1}{m} \sum_{i=1}^{m} P(s_{i,1} x_{l_{i,1}}, \ldots, s_{i,n} x_{l_{i,n}})$$

subject to $x \in \{-1,1\}^N$.

2.2 Complex analysis background

We frequently use complex analysis to compute coefficients in series represented by generating functions. Recall that any complex function f which is analytic in a neighborhood, $0 < |z| < r_0$, of $z = 0$ can be represented as a Laurent series:

$$f(z) = \sum_{n=-\infty}^{n=\infty} a_n z^n \quad , \quad 0 < |z| < r_0.$$

The residue of f at $z = 0$,

$$\operatorname*{Res}_{z=0} f(z) = a_{-1},$$

can then be computed using Cauchy's Residue Theorem, which we state in a simplified form here:

Theorem 2.1 (Cauchy)**.** *Let C be a positively oriented simple closed contour containing the origin. If f is analytic inside and on C except at $z = 0$, then*

$$\operatorname*{Res}_{z=0} f(z) = \frac{1}{2\pi i} \oint_C f(z) dz$$

Thus, in order to compute the n'th coefficient b_n in a generating function

$$\sum_{n=0}^{\infty} b_n z^n = g(z) \quad , \quad |z| < r_0$$

we may apply Cauchy's theorem to $g(z)z^{-(n+1)}$ with a suitably selected contour.

2.3 Common Lemmas

We now present a technical lemma that is used in our calculations to bound integrands on the unit circle when we are not close to the point $z = 1$.

Lemma 2.2. *Suppose we are given real numbers p_j, $\frac{1}{4} \leq p_j \leq \frac{3}{4}$, $1 \leq j \leq n$ and positive integers $(w_j)_{j=1}^n$ such that $\sum_{j=1}^n w_j^3 < 100n$. Furthermore suppose that for at least t different values of j we have $w_j = 1$. Let $q_j = 1 - p_j$, then for any φ, $0 \leq \varphi \leq \pi$ we have*

$$|\prod_{j=1}^{n} (q_j + p_j e^{w_j i\varphi})| \leq e^{-0.01 \min(t, \varphi^2 n)}.$$

Proof. By multiplying by the conjugate we see that

$$|q_j + p_j e^{w_j i\varphi}|^2 = p_j^2 + q_j^2 + 2p_j q_j \cos(w_j\varphi) = 1 + 2p_j q_j (\cos(w_j\varphi) - 1). \tag{2}$$

Observe that for any φ, $\pi/8 \leq \varphi \leq \pi/2$ we have

$$1 + 2p_j q_j (\cos(\varphi) - 1) \leq 1 + 2 \cdot \frac{3}{16} (\cos(\pi/8) - 1),$$

which can be seen to be at most $e^{-0.02}$. As we have $w_j = 1$ for t different values of j, the product of the lemma is bounded by $e^{-t/100}$ for this range of φ and we turn to values $0 \leq \varphi \leq \pi/8$.

We claim that for any x, $0 \leq x \leq \pi/2$ we have $\cos(x) \leq 1 - \frac{4x^2}{\pi^2}$. To see this note that for $g(x) = 1 - \cos(x) - \frac{4x^2}{\pi^2}$ we have $g(0) = g(\pi/2) = 0$, $g'(0) = 0$, $g''(0) > 0$ and $g'''(x) \leq 0$ in the entire interval. It follows that each of $g''(x)$ and $g'(x)$ has a unique 0 in the interval $0 < x \leq \pi/2$ and g is unimodal.

As $p_j + q_j = 1$ and each of these numbers is at least $1/4$ it follows that (2), for $w_j \leq 4$ and the set of φ we are considering, is bounded by

$$1 - \frac{3}{8} \cdot \frac{4w_j^2\varphi^2}{\pi^2} \leq e^{-\varphi^2/20}.$$

By the condition on the sum of cubes we have at least $n/5$ different j with $w_j \leq 4$ and thus the lemma follows also in this case. □

To evaluate an integral with integrand $\varphi^k e^{-a\varphi^2}$, we use the following well known results.

Lemma 2.3 (Standard Integral)**.** *For $k > -1$ and $a > 0$,*

$$\int_0^\infty \varphi^k e^{-a\varphi^2}\, d\varphi = \frac{1}{2}\Gamma\left(\frac{k+1}{2}\right) a^{-\frac{k+1}{2}}.$$

Lemma 2.4 (Tail bound)**.** *For $k > -1$, and sequences $a(n) \geq 0$ and $\varphi_0(n) \geq 0$ such that $a(n) = \Omega(1)$ and $\varphi_0(n) = \mathcal{O}(1)$,*

$$\int_{\varphi_0(n)}^\infty \varphi^k e^{-a(n)\varphi^2}\, d\varphi = \mathcal{O}(e^{-a(n)\varphi_0^2(n)})$$

Proof. To ease notation we drop the explicit dependence on n. By a change of variables, and using $(x+y)^k \leq (2x)^k + (2y)^k$,

$$\begin{aligned} \int_{\varphi_0}^\infty \varphi^k e^{-a\varphi^2}\, d\varphi &= \int_0^\infty (\varphi+\varphi_0)^k e^{-a(\varphi+\varphi_0)^2}\, d\varphi \leq \int_0^\infty [(2\varphi)^k + \mathcal{O}(1)] e^{-a\varphi^2 - a\varphi_0^2}\, d\varphi \\ &= e^{-a\varphi_0^2} \int_0^\infty (2^k\varphi^k + \mathcal{O}(1)) e^{-a\varphi^2}\, d\varphi = e^{-a\varphi_0^2}\mathcal{O}(1) \end{aligned}$$

where the last inequality follows from Lemma 2.3. □

2.3.1 Balanced Bernoulli Random Variables

We say that a random variable X is a *balanced Bernoulli* random variable if $\Pr[X = 1] = \Pr[X = 0] = 1/2$. Here, we present two lemmas that are useful in Sections 3 and 5, where we analyze weighted sums of balanced Bernoulli random variables.

Lemma 2.5. *Suppose we are given n balanced Bernoulli random variables $Y_1, Y_2, \ldots, Y_n$ and positive integers $(w_j)_{j=1}^n$ such that $\sum_{j=1}^n w_j^3 < 100n$. Furthermore, suppose that for at least $\gamma \cdot \log n$ different values of j we have $w_j = 1$. Then for any $\varphi_0 \geq \frac{1}{10}\frac{\log n}{\sqrt{n}}$ and any integer t such that $\sum_{j=1}^n w_j + t = 0 \mod 2$,*

$$\Pr\left[\sum_{j=1}^n w_j Y_j = \frac{\sum_{j=1}^n w_j + t}{2}\right] = \frac{1}{\pi}\int_0^{\varphi_0} \cos(\varphi t/2) \prod_{j=1}^n \cos(w_j\varphi/2) + \mathcal{O}\left(\frac{1}{n^{0.01\gamma}}\right).$$

Proof. Let $Y = \sum_{j=1}^n w_j Y_j$. The probability generating function of Y is

$$g(z) = \prod_{j=1}^n \frac{1+z^{w_j}}{2}.$$

Letting $W = \sum_{j=1}^n w_j$ and

$$f(z) = \frac{g(z)}{z^{(W+t)/2+1}}$$

Cauchy's Residue Theorem gives

$$\Pr\left[\sum_{j=1}^n w_j Y_j = \frac{\sum_{j=1}^n w_j + t}{2}\right] = \frac{1}{2\pi i}\oint_C f(z)dz,$$

where we take C to be the unit circle,

$$C : z = e^{i\varphi} \qquad 0 \leq \varphi \leq 2\pi.$$

As $f(\overline{z}) = \overline{f(z)}$,

$$\frac{1}{2\pi i}\oint_C f(z)dz = \frac{1}{\pi}\int_0^\pi \mathrm{Re}[e^{i\varphi}f(e^{i\varphi})]d\varphi.$$

Expanding $f(e^{i\varphi})$ give us that

$$\begin{aligned} e^{i\varphi}f(e^{i\varphi}) &= g(e^{i\varphi})e^{-i\varphi(W+t)/2} \\ &= \left(\prod_{j=1}^n \frac{1+e^{iw_j\varphi}}{2}\right) e^{-i\varphi(W+t)/2} \end{aligned}$$

By Lemma 2.2, $\left|\prod_{j=1}^n \frac{1+e^{iw_j\varphi}}{2}\right| = \mathcal{O}(e^{-0.01\gamma\log n}) = \mathcal{O}(n^{-0.01\gamma})$ whenever $\varphi \geq \varphi_0$. Hence

$$\int_{\varphi_0}^\pi \left(\prod_{j=1}^n \frac{1+e^{iw_j\varphi}}{2}\right) e^{-i\varphi(W+t)/2}d\varphi = \mathcal{O}\left(\frac{1}{n^{0.01\gamma}}\right)$$

and we have thus that $\frac{1}{2\pi i}\oint_C f(z)dz = \frac{1}{\pi}\int_0^{\varphi_0}\mathrm{Re}[e^{i\varphi}f(e^{i\varphi})d\varphi] + \mathcal{O}\left(\frac{1}{n^{0.01\gamma}}\right)$. Multiplying each factor $\frac{1+e^{iw_j\varphi}}{2}$ of $g(e^{i\varphi}))$ by $e^{-i\varphi w_j/2}$ gives that

$$\begin{aligned} e^{i\varphi}f(e^{i\varphi}) &= \left(\prod_{j=1}^n \frac{1+e^{iw_j\varphi}}{2}\right) e^{-i\varphi(W+t)/2} \\ &= \left(\prod_{j=1}^n \frac{e^{-iw_j\varphi/2}+e^{iw_j\varphi/2}}{2}\right) e^{-i\varphi t/2}. \end{aligned}$$

The real part of this can thus be written as

$$\left(\prod_{j=1}^n \cos(w_j\varphi/2)\right)\cos\left(\varphi t/2\right),$$

which completes the proof of the lemma. □

Lemma 2.6. *Suppose we are given positive integers $(w_j)_{j=1}^n$ such that $\sum_{j=1}^n w_j^3 = O(n)$. Then for any $\varphi_0 : 10\frac{\log n}{\sqrt{n}} \geq \varphi_0 \geq \frac{1}{10}\frac{\log n}{\sqrt{n}}$ and any $k : -1 < k \leq 10$*

$$\int_0^{\varphi_0}\varphi^k\prod_{j=1}^n\cos(\varphi w_j/2)d\varphi = \left(1+\mathcal{O}\left(\frac{w_{\max}}{n}\right)\right)\frac{1}{2}\Gamma\left(\frac{k+1}{2}\right)\left(\frac{W^{(2)}}{8}\right)^{-\frac{k+1}{2}},$$

where $w_{\max} = \max_j w_j$ and $W^{(2)} = \sum_{j=1}^n w_j^2$.

Proof. Let $h(\varphi) = \prod_{j=1}^n \cos(\varphi w_j/2)$. We may use Taylor expansion to write

$$\cos(w_j\varphi/2) = 1 - \frac{w_j^2}{8}\varphi^2 + w_j^4\mathcal{O}(\varphi^4) = e^{-\frac{w_j^2}{8}\varphi^2 + w_j^4\mathcal{O}(\varphi^4)}$$

and hence

$$h(\varphi) = e^{-\frac{W^{(2)}}{8}\varphi^2 + W^{(4)}\mathcal{O}(\varphi^4)} = e^{-\frac{W^{(2)}}{8}\varphi^2}\left(1+W^{(4)}\mathcal{O}(\varphi^4)\right),$$

where $W^{(2)} = \sum_{j=1}^n w_j^2$ and $W^{(4)} = \sum_{j=1}^n w_j^4$. As $W^{(2)} \cdot \varphi_0^2 = \omega(\log n)$, Lemma 2.4 gives that

$$\int_{\varphi_0}^{\infty} \varphi^k e^{-\frac{W^{(2)}}{8}\varphi^2} \left(1 + W^{(4)}\mathcal{O}(\varphi^4)\right) d\varphi = e^{-\omega(\log n)}$$

and thus

$$\int_0^{\varphi_0} \varphi^k e^{-\frac{W^{(2)}}{8}\varphi^2} \left(1 + W^{(4)}\mathcal{O}(\varphi^4)\right) d\varphi = \int_0^{\infty} \varphi^k e^{-\frac{W^{(2)}}{8}\varphi^2} \left(1 + W^{(4)}\mathcal{O}(\varphi^4)\right) d\varphi + e^{-\omega(\log n)}.$$

Further, by Lemma 2.3

$$\int_0^{\infty} \varphi^k e^{-\frac{W^{(2)}}{8}\varphi^2} \left(1 + W^{(4)}\mathcal{O}(\varphi^4)\right) d\varphi = \frac{1}{2}\Gamma\left(\frac{k+1}{2}\right)\left(\frac{W^{(2)}}{8}\right)^{-\frac{k+1}{2}} + \mathcal{O}(W^{(4)})\left(W^{(2)}\right)^{-\frac{k+5}{2}}.$$

As $W^{(4)} \leq w_{\max} \sum_{j=1}^n w_j^3 = \mathcal{O}(w_{\max} n)$ and $W^{(2)} = \Omega(n)$, we have that $\mathcal{O}(W^{(4)} \cdot (W^{(2)})^{-2}) = \mathcal{O}\left(\frac{w_{\max}}{n}\right)$ and the lemma follows. □

3 Adaptation of the Algorithm by Hast

Using Fourier expansion we may write the advantage of an assignment to a Max-CSP(P) instance as

$$\mathsf{Adv}(x) = \frac{1}{m}\sum_{i=1}^m \operatorname{sgn}\left(\sum_{j=1}^n w_j s_{i,j} x_{l_{i,j}}\right) = \sum_{S\subseteq[N]:|S|\leq n} c_S \prod_{k\in S} x_k. \tag{3}$$

Hast [7] gives a general approximation algorithm for Max-CSP(P) that achieves a non-trivial approximation ratio whenever the linear part of the instance's objective function is large enough. We use his algorithm, but as our basic predicates are odd we have that $c_S = 0$ for any S of even size and we get slightly better bounds.

Theorem 3.1. *For any $\delta > 0$, there is a polynomial time algorithm which given an instance of Max-CSP(P) with objective function*

$$\mathsf{Adv}(x_1, \ldots, x_N) = \sum_{S\subseteq[N],|S|\leq n} c_S \prod_{k\in S} x_k$$

satisfying $\sum_{k=1}^N |c_{\{k\}}| \geq \delta$ and $c_S = 0$ for any set S of even cardinality, achieves $\mathbb{E}[\mathsf{Adv}(x)] \geq \frac{\delta^{3/2}}{8n^{3/4}}$.

Proof. Let $\epsilon > 0$ be a parameter to be determined. We set each x_i randomly and independently to one with probability $(1 + \operatorname{sgn}(c_{\{i\}})\epsilon)/2$. Clearly this implies that $\mathbb{E}[c_{\{i\}}x_i] = \epsilon|c_{\{i\}}|$ and that $|\mathbb{E}[\prod_{k\in S} x_k]| = \epsilon^{|S|}$.

By Cauchy Schwarz inequality and Parseval's identity we have that

$$\sum_{|T|=k} |\hat{P}(T)| \leq \binom{n}{k}^{1/2} \left(\sum_{|T|=k} \hat{P}^2(T)\right)^{1/2} \leq \binom{n}{k}^{1/2}$$

and hence

$$\sum_{|S|=k} |c_S| \leq \binom{n}{k}^{1/2}. \tag{4}$$

We conclude that the advantage of the given algorithm is, given that $c_S = 0$ for even cardinality S, at least

$$\epsilon \sum_{i=1}^{n} |c_i| - \sum_{|S|\geq 3} \epsilon^k |c_S| \geq \epsilon\delta - \sum_{k=3}^{n} \epsilon^k \binom{n}{k}^{1/2}. \tag{5}$$

The sum in (5) is, provided $\epsilon \leq (2\sqrt{n})^{-1}$, and using Cauchy-Schwarz bounded by

$$\left(\sum_{k=3}^{n} \left(\frac{1}{n}\right)^k \binom{n}{k}\right)^{1/2} \left(\sum_{k=3}^{n} (\epsilon^2 n)^k\right)^{1/2} \leq \left(1+\frac{1}{n}\right)^{n/2} (2\epsilon^6 n^3)^{1/2} \leq 3\epsilon^3 n^{3/2},$$

where we used $\sum_{k=0}^{n} \left(\frac{1}{n}\right)^k \binom{n}{k} = \left(1+\frac{1}{n}\right)^n$ and $\sum_{k=3}^{n} (\epsilon^2 n)^k \leq \epsilon^6 n^3 \sum_{k=0}^{\infty} \frac{1}{2^k}$ for the first inequality. Setting $\epsilon = \delta^{1/2}(2n^{3/4})^{-1}$, which is at most $(2\sqrt{n})^{-1}$ by (4), we see that the advantage of the algorithm is $\epsilon\delta - 3\epsilon^3 n^{3/2} = \frac{\delta^{3/2}}{8n^{3/4}}$ and the proof is complete. □

Let us see how to apply Theorem 3.1 in the case when P is majority of n variables. Suppose we are given an instance that is $1 - \frac{\delta}{n+1}$ satisfiable and let us consider

$$\sum_{i=1}^{N} c_{\{i\}} \alpha_i \tag{6}$$

where $x_i = \alpha_i$ is the optimal solution and prove that this is large. Any lower bound for this is clearly a lower bound for $\sum_{i=1}^{N} |c_{\{i\}}|$.

Let $\hat{P}_1$ be the value of any Fourier coefficient of a unit size set. Then any satisfied constraint contributes at least $\hat{P}_1$ to (6) while any other constraint contributes at least $-n\hat{P}_1$. We conclude that (6) is at least

$$\left(1 - \frac{\delta}{n+1}\right)\hat{P}_1 - \frac{\delta}{n+1} n\hat{P}_1 = (1-\delta)\hat{P}_1.$$

Using Theorem 3.1 and the fact that $\hat{P}_1 = \Theta(n^{-1/2})$ we get the following corollary.

Theorem 3.2. *Suppose we are given an instance of Max-Maj-n which is $(1 - \frac{\delta}{n+1})$-satisfiable. Then it is possible, in probabilistic polynomial time, to find an assignment that satisfies a fraction $\frac{1}{2} + \Omega((1-\delta)^{3/2} n^{-3/2})$ of the constraints.*

Let us sketch how to generalize this theorem to predicates other than majority. Clearly the key property is to establish that the sum (6) is large when most constraints can be simultaneously satisfied. In order to have any possibility for this to be true it must be that whenever a constraint is satisfied, then the contribution to (6) is positive and this is exactly being "Chow-robust" as discussed in the introduction. Furthermore, to get a quantitative result we must also make sure that it is positive by some fixed amount. Let us turn to a formal definition.

Recall that the Chow parameters of a predicate P is its degree-0 and degree-1 Fourier coefficients, i.e., $\hat{P}(0), \hat{P}(1), \ldots, \hat{P}(n)$ for $i = 1, 2, \ldots, n$. As we are here dealing with an odd predicate, $\hat{P}(0) = 0$. If it holds for all $x \in \{-1,1\}^n$ that $P(x) = \mathrm{sgn}(\hat{P}(1)x_1 + \hat{P}(2)x_2 + \cdots + \hat{P}(n)x_n)$, we say that the predicate is *Chow-robust* and it is *γ-Chow-robust* iff

$$0 < \gamma \leq \min_{x:P(x)=1} \left(\sum_{j=1}^{n} \hat{P}(j)x_j\right).$$

Let us state our extension of Theorem 3.2 in the present context.

Theorem 3.3. *Let* $P(x) = \mathrm{sgn}(w_1x_1 + w_2x_2 + \cdots + w_nx_n)$ *be a* γ*-Chow-robust predicate and suppose that* $\mathcal{I}$ *is a* $1 - \frac{\delta\gamma}{\gamma+\sum_{j=1}^n \hat{P}(j)}$ *satisfiable instance of Max-CSP*(P) *where* $\delta < 1$*. Then there is a polynomial time algorithm that achieves* $\mathbb{E}[\mathsf{Adv}(x)] = \frac{(1-\delta)^{3/2}\gamma^{3/2}}{8n^{3/4}}$.

Proof. The linear part of the advantage of an assignment can be written as

$$\sum_{k=1}^{N} c_{\{k\}} x_k = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n} \hat{P}(j) s_{i,j} x_{l_{i,j}}.$$

For any assignment x we have thus

$$\sum_{k=1}^{N} |c_{\{k\}}| \geq \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n} \hat{P}(j) s_{i,j} x_{l_{i,j}}.$$

Now since $\mathcal{I}$ is $1 - \frac{\delta\gamma}{\gamma+\sum_{j=1}^n \hat{P}(j)}$ satisfiable there is an assignment x such that for at least a $1 - \frac{\delta\gamma}{\gamma+\sum_{j=1}^n \hat{P}(j)}$ fraction of the constraints

$$P(s_{i,1}x_{\ell_{i,1}}, \ldots, s_{i,n}x_{\ell_{i,n}}) = 1 \text{ and thus } \sum_{j=1}^{n} \hat{P}(j) s_{i,j} x_{\ell_{i,j}} \geq \gamma$$

using that P is γ-Chow-robust. As the linear part of the remaining constraints is greater than $-\sum_{j=1}^n \hat{P}(j)$, we have that

$$\begin{aligned} \sum_{k=1}^{N} |c_{\{k\}}| &\geq \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{n} \hat{P}(j) s_{i,j} x_{l_{i,j}} \\ &\geq \left(1 - \frac{\delta\gamma}{\gamma + \sum_{j=1}^n \hat{P}(j)}\right)\gamma - \frac{\delta\gamma}{\gamma + \sum_{j=1}^n \hat{P}(j)} \sum_{j=1}^{n} \hat{P}(j) \\ &= (1-\delta)\gamma. \end{aligned}$$

Theorem 3.1 now gives the result. □

Given Theorem 3.3 it is interesting to discuss sufficient conditions for P to be Chow-robust and we have the following theorem.

Theorem 3.4. *Suppose we are given positive integers* $(w_j)_{j=1}^n$ *such that*

$$\beta(w) := 1 - \frac{\sum_{j=1}^n (w_j^3 - w_j)}{3\sum_{j=1}^n w_j^2} > 0. \tag{7}$$

Further, suppose that for at least $400 \log n$ *different values of* j*, say* $1, 2, \ldots, n_1$*, we have* $w_j = 1$*. Then the predicate* $P(x) = \mathrm{sgn}(x_1 + \cdots + x_{n_1} + w_{n_1+1}x_{n_1+1} + \cdots + w_nx_n)$ *is* γ*-Chow-robust with* $\gamma = \left(\beta(w) - \mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right)\hat{P}(1)$*, provided that* n *is large enough so that this is positive.*

Before presenting the proof of the above theorem, let us comment on the condition on the $\Omega(\log n)$ weights that we require to be one. This should be viewed as a technical condition and we could have chosen other similar conditions. In particular, we have made no effort to optimize the constant 400. In our calculations this condition is used to bound the integrand of a complex integral on the unit circle when we are not close to the point $z = 1$ and this could be done in many ways. We would like to point out that although there are choices for the technical condition, some condition is needed. The condition should imply some mathematical form of "when z on the unit circle is far from 1 then many numbers of the form z^{w_j} are not close to 1". Sets of weights violating such conditions are cases when almost all weights have a common factor. An interesting example is the function which, for odd n, has $n-4$ weights equal to 3 and 4 weights equal to 1. This function is not Chow-robust for any value of n. The above example shows that there are functions with weights of at most 3 that are not Chow-robust. This is a tight bound as the techniques used in the proof of Theorem 3.4 can be used to show that a function with all weights equal to 1 or 2 is Chow-robust.

Proof of Theorem 3.4

Let us start with a lemma that can be used to bound higher moments of the weights when $\beta(w) > 0$.

Lemma 3.5. *Let $w = (w_1, \ldots, w_n)$ be positive integers such that $\beta(w) > 0$. Then $\sum_{j=1}^n w_j^3 \leq 64n$ and $w_{\max} \leq 4n^{1/3}$.*

Proof. Since $\beta(w) > 0$ we have $\sum_{j=1}^n (w_j^3 - w_j) < 3\sum_{j=1}^n w_j^2$. Hence, by Hölder's inequality,

$$\sum_{j=1}^n w_j^3 \leq 3\sum_{j=1}^n w_j^2 + \sum_{j=1}^n w_j \leq 4\sum_{j=1}^n w_j^2 \leq 4\left(\sum_{j=1}^n w_j^3\right)^{2/3}\left(\sum_{j=1}^n 1\right)^{1/3}$$

and the bounds $\sum_{j=1}^n w_j^3 \leq 64n$ and $w_{\max} \leq 4n^{1/3}$ follow. □

We proceed by analyzing the linear threshold predicate where the linear Fourier coefficients are used as weights. Let $P_C(x) = \operatorname{sgn}(\hat{P}(1)x_1 + \hat{P}(2)x_2 + \cdots + \hat{P}(n)x_n)$. Since $\hat{P}(1) = \hat{P}(2) = \cdots = \hat{P}(n_1)$,

$$P_C(x) = \operatorname{sgn}\left(x_1 + \cdots + x_{n_1} + \frac{\hat{P}(n_1+1)}{\hat{P}(1)}x_{n_1+1} + \cdots + \frac{\hat{P}(n)}{\hat{P}(1)}x_n\right).$$

A sufficient condition for P to be γ-Chow-robust is then

$$\sum_{j=n_1+1}^n \left|\frac{\hat{P}(j)}{\hat{P}(1)} - w_j\right| < 1 - \frac{\gamma}{\hat{P}(1)}. \tag{8}$$

To see this, consider an x such that $P(x) = 1$ and hence $\sum_{j=1}^n w_j x_j \geq 1$. The above condition implies that $\sum_{j=1}^n \frac{\hat{P}(j)}{\hat{P}(1)}x_j = \sum_{j=1}^n \left(\frac{\hat{P}(j)}{\hat{P}(1)} - w_j\right)x_j + \sum_{j=1}^n w_j x_j \geq \frac{\gamma}{\hat{P}(1)}$ and we have as required that $\sum_{j=1}^n \hat{P}(j)x_j \geq \gamma$.

We continue by analyzing the quotient $\frac{\hat{P}(j_0)}{\hat{P}(1)}$ for a fixed $j_0 : n_1+1 \leq j_0 \leq n$. As P is a monotone function, a degree-1 Fourier coefficient equals that coordinates influence, i.e.,

$$\begin{aligned}\hat{P}(j_0) = \mathrm{Inf}_{j_0}(P) &= \Pr_x[P(x_1, \ldots, x_{j_0}, \ldots, x_k) \neq P(x_1, \ldots, -x_{j_0}, \ldots, x_k)] \\ &= \Pr[\frac{W+1}{2} - w_{j_0} \leq X^{(j_0)} \leq \frac{W+1}{2} - 1],\end{aligned}$$

where $W = \sum_{j=1}^{n} w_j$ and $X^{(j_0)} = \sum_{j=1, j\neq j_0}^{n} w_j X_j$ is a weighted sum of $n-1$ balanced independent Bernoulli random variables. By Lemma 2.5 and the assumption that $|\{j : w_j = 1\}| \geq 400 \log n$, we have

$$\Pr_x \left[X^{(j_0)} = \frac{W + 1 - 2t}{2} \right] = \int_0^{\varphi_0} \cos\left(\varphi \frac{w_{j_0} + 1 - 2t}{2}\right) \prod_{j=1, j\neq j_0}^{n} \cos(w_j \varphi/2) d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right),$$

where $\varphi_0 = \frac{\log n}{\sqrt{n}}$. We can thus write $\hat{P}(j_0)$ as

$$\int_0^{\varphi_0} \sum_{k=1}^{w_{j_0}} \cos\left(\varphi \frac{w_{j_0} + 1 - 2k}{2}\right) \prod_{j=1, j\neq j_0}^{n} \cos(w_j \varphi/2)\, d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right).$$

Similarly, we can write $\hat{P}(1)$ as

$$\int_0^{\varphi_0} \sum_{k=1}^{w_1} \cos\left(\varphi \frac{w_1 + 1 - 2k}{2}\right) \prod_{j=2}^{n} \cos(w_j \varphi/2)\, d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right),$$

which equals (since $w_1 = 1$)

$$\int_0^{\varphi_0} \prod_{j=2}^{n} \cos(w_j \varphi/2)\, d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right),$$

Letting $h(\varphi) = \prod_{j=2, j\neq j_0}^{n} \cos(w_j \varphi/2)$,

$$\frac{\hat{P}(j_0)}{\hat{P}(1)} = \frac{\int_0^{\varphi_0} \sum_{k=1}^{w_{j_0}} \cos(\varphi/2) \cos\left(\varphi \frac{w_{j_0}+1-2k}{2}\right) h(\varphi)\, d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right)}{\int_0^{\varphi_0} \cos(w_{j_0}\varphi/2) h(\varphi)\, d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right)} \tag{9}$$

Using Taylor expansion we may write

$$\cos(\varphi w_{j_0}/2) \;=\; 1 - \frac{w_{j_0}^2}{8}\varphi^2 + w_{j_0}^4 \mathcal{O}(\varphi^4)$$

and

$$\begin{aligned} \sum_{k=1}^{w_{j_0}} \cos(\varphi/2) \cos\left(\varphi \frac{w_{j_0} + 1 - 2k}{2}\right) &= \sum_{k=1}^{w_{j_0}} \left(1 - \frac{1 + (w_{j_0} + 1 - 2k)^2}{8}\varphi^2 + w_{j_0}^4 \mathcal{O}(\varphi^4)\right) \\ &= w_{j_0} - \frac{2w_{j_0} + w_{j_0}^3}{24}\varphi^2 + w_{j_0}^5 \mathcal{O}(\varphi^4), \end{aligned}$$

where the last equality follows from the identities $\sum_{k=1}^{w_{j_0}} k = w_{j_0}(w_{j_0}+1)/2$ and $\sum_{k=1}^{w_{j_0}} k^2 = w_{j_0}(w_{j_0}+1)(2w_{j_0}+1)/6$. By the above calculations, the numerator of (9) equals

$$w_{j_0} \int_0^{\varphi_0} \left(1 - \frac{2 + w_{j_0}^2}{24}\varphi^2 + w_{j_0}^4 \mathcal{O}(\varphi^4)\right) h(\varphi)\, d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right)$$

and its denominator equals

$$\int_0^{\varphi_0} \left(1 - \frac{w_{j_0}^2}{8}\varphi^2 + w_{j_0}^4 \mathcal{O}(\varphi^4)\right) h(\varphi)\, d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right).$$

Simplifications then give us that

$$\frac{\hat{P}(j_0)}{\hat{P}(1)} - w_{j_0} = \frac{w_{j_0}\int_0^{\varphi_0}\left(\frac{w_{j_0}^2-1}{12}\varphi^2 + w_{j_0}^4\,\mathcal{O}(\varphi^4)\right)h(\varphi)\,d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right)}{\int_0^{\varphi_0}\left(1-\frac{w_{j_0}^2}{8}\varphi^2 + w_{j_0}^4\,\mathcal{O}(\varphi^4)\right)h(\varphi)\,d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right)} \tag{10}$$

To estimate this expression we will use Lemma 2.6. Letting $S_2 = \sum_{j=2,j\neq j_0}^n w_j^2$, the numerator of (10) then equals

$$\left(1+\mathcal{O}\left(\frac{w_{\max}}{n}\right)\right)\left(\frac{w_{j_0}^3 - w_{j_0}}{12}\frac{\sqrt{\pi}}{4}\left(\frac{S_2}{8}\right)^{-\frac{3}{2}} + \mathcal{O}(w_{j_0}^5)\left(\frac{S_2}{8}\right)^{-\frac{5}{2}}\right),$$

which can be simplified, by using that $S_2 = \Omega(n)$ and $2 \le w_{j_0} \le w_{\max} \le 4n^{1/3}$, to

$$\left(1+\mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right)\frac{w_{j_0}^3 - w_{j_0}}{12}\frac{\sqrt{\pi}}{4}\left(\frac{S_2}{8}\right)^{-\frac{3}{2}}.$$

Similarly, by Lemma 2.6, (10)'s denominator can be written as

$$\left(1+\mathcal{O}\left(\frac{w_{\max}}{n}\right)\right)\left(\frac{\sqrt{\pi}}{2}\left(\frac{S_2}{8}\right)^{-\frac{1}{2}} + \mathcal{O}(w_{j_0}^2)\left(\frac{S_2}{8}\right)^{-\frac{3}{2}}\right)$$

and simplified to

$$\left(1+\mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right)\frac{\sqrt{\pi}}{2}\left(\frac{S_2}{8}\right)^{-\frac{1}{2}}$$

Substituting in these evaluations, we obtain that

$$\begin{aligned}\frac{\hat{P}(j_0)}{\hat{P}(1)} - w_{j_0} &= \frac{w_{j_0}^3 - w_{j_0}}{24}\left(\frac{S_2}{8}\right)^{-1}\left(1+\mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right)\\ &= \frac{w_{j_0}^3 - w_{j_0}}{3\sum_{j=2,j\neq j_0}^n w_j^2}\left(1+\mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right)\\ &= \frac{w_{j_0}^3 - w_{j_0}}{3\sum_{j=1}^n w_j^2}\left(1+\mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right),\end{aligned}$$

where we used that $\sum_{j=2,j\neq j_0}^n w_j^2 = \left(1+\mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right)\sum_{j=1}^n w_j^2$ for the last equality.

We now conclude the proof of the theorem by observing that the sufficient condition (8) for P to be γ-Chow-robust is satisfied if

$$\left(1+\mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right)\sum_{j=1}^n (w_j^3 - w_j) < \left(1 - \frac{\gamma}{\hat{P}(1)}\right)3\sum_{j=1}^n w_j^2.$$

Indeed, then

$$\sum_{j=n_1+1}^n \left|\frac{\hat{P}(j)}{\hat{P}(1)} - w_j\right| \le \left(1+\mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right)\frac{\sum_{j=1}^n w_j^3 - w_j}{3\sum_{j=1}^n w_j^2} \le 1 - \frac{\gamma}{\hat{P}(1)}.$$

The statement now follows from observing that we can select γ to be

$$\left(1 - \left(1+\mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right)\frac{\sum_{j=1}^n w_j^3 - w_j}{3\sum_{j=1}^n w_j^2}\right)\hat{P}(1) = \left(\beta(w) + \mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right)\hat{P}(1),$$

where we used that $\sum_{j=1}^n w_j^2 = \Omega(n)$ and $\sum_{j=1}^n w_j^3 = \mathcal{O}(n)$.

4 Our Main Algorithm

We now give an improved algorithm for Max-CSP(P) for homogeneous linear threshold predicates. Recall that we write the i'th constraint as $P(s_{i,1}x_{l_{i,1}}, \ldots, s_{i,n}x_{l_{i,n}}) = \operatorname{sgn}(L_i(x))$,where $L_i(x) = \sum_{j=1}^n w_j s_{i,j} x_{l_{i,j}}$, and let $W := \sum_{j=1}^n w_j$. The algorithm which is parametrized by a noise parameter $0 < \epsilon < 1$ is described as follows:

Algorithm $A_{\mathbf{LP},\epsilon}$

1. Let x^*, Δ^* be the optimal solution to the following linear program

$$\begin{array}{ll} \text{maximize} & \frac{1}{m}\sum_{i=1}^m \Delta_i \\ \text{subject to} & L_i(x) \geq \Delta_i, \forall i \in [m] \\ & x \in [-1,1]^N, \Delta \in [-W, 1]^m \end{array}$$

2. Pick $X_1, \ldots, X_N \in \{-1, 1\}$ independently with bias $\mathbb{E}[X_i] = \epsilon x_i^*$ and return this assignment.

As in Theorem 3.4, we now define $\beta(w)$ for a set of weights $w = (w_1, \ldots, w_n)$ as

$$\beta(w) = 1 - \frac{\sum_{j=1}^n (w_j^3 - w_j)}{3\sum_{j=1}^n w_j^2}.$$

Note that $\beta \leq 1$ for any set of weights, while for majority $\beta = 1$. Further, if $\beta(w) > 0$, then Theorem 3.4 showed that P is γ-Chow-robust provided that n is large enough.

In Section 4.1 we show that on $1 - \frac{\delta}{1+W}$ satisfiable instances, where $\delta < \beta$, the above algorithm achieves an advantage of $\Omega(\frac{1}{\sqrt{n}})$ for large enough n. In particular, we will prove the following theorem:

Theorem 4.1. *Fix any homogeneous threshold predicate $P(x) = \operatorname{sgn}(w_1x_1 + \cdots + w_nx_n)$ having $w_j = 1$ for at least $200\log n$ different values of j and satisfying $\beta := \beta(w) > 0$. Then, for any $1 - \frac{\delta}{1+W}$ satisfiable instance $\mathcal{I}$ of Max-CSP(P), where $\delta < \beta$, we have*

$$\mathbb{E}[\mathsf{Adv}(A_{LP,\epsilon}(\mathcal{I}))] = (\beta - \delta)^{3/2}\,\Omega\left(\frac{1}{\sqrt{n}}\right) - \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right), \tag{11}$$

where $\epsilon = (\beta - \delta)^{1/2}\epsilon_0$ and $\epsilon_0 > 0$ is an absolute constant.

Thus, for δ bounded away from β, and large enough n, this algorithm is an improvement over the algorithm of Theorem 3.3. We may also note that both the algorithm $A_{\mathrm{LP},\epsilon}$ and the algorithm of Theorem 3.3 can be de-randomized using the method of conditional expectation.

4.1 Analysis of the Algorithm (Proof of Theorem 4.1)

The crucial point for Theorem 4.1 is that on $1 - \frac{\delta}{1+W}$ satisfiable instances, $\bar{\Delta}^* := \frac{1}{m}\sum_{i=1}^m \Delta_i^* \geq 1 - \delta > 1 - \beta$. Too see this, take any assignment $x \in \{-1,1\}^N$ which satisfies a $1 - \frac{\delta}{1+W}$ fraction of all constraints. Let $\Delta_i = 1$ on all satisfied constraints and $\Delta_i = -W$ on the constraints which are not satisfied. This is a feasible solution to the linear program since clearly $L_i(x) \geq \Delta_i$, $\forall i$. Hence,

$$\bar{\Delta}^* = \frac{1}{m}\sum_{i=1}^m \Delta_i^* \geq \frac{1}{m}\sum_{i=1}^m \Delta_i = \left(1 - \frac{\delta}{1+W}\right) + \frac{\delta}{1+W}(-W) = 1 - \delta. \tag{12}$$

The next lemma shows that for large enough n we can use this advantage whenever no $|\Delta_i(x^*)|$ is too large, provided that we pick ϵ small enough.

Lemma 4.2. *Fix $0 < \epsilon < 1/2$ and let $X = w_1X_1 + \cdots + w_nX_n$ be a sum of n independent Bernoulli random variables where $w_1, \ldots, w_n$ are positive integer weights such that $w_j = 1$ for at least $200 \log n$ different j's and $\beta := \beta(w) > 0$ given by (7). Further, let $\Pr(X_j = 1) = \frac{1+\epsilon x_j}{2}$ where $-1 \le x_j \le 1$, and let $\sigma^2 = \operatorname{Var} X$ and $\Delta = \sum_{j=1}^n w_j x_j$. Then, if $|\Delta| \le n^{1/3}$,*

$$\Pr\left(X \ge \frac{W+1}{2}\right) \ge \frac{1}{2} + \frac{\epsilon(\Delta - 1 + \beta)}{\sigma\sqrt{8\pi}} + \epsilon^3 \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) + \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right).$$

Proof. As before we let $\epsilon_j = \epsilon x_j$, $p_j = \frac{1+\epsilon_j}{2}$ and $q_j = 1 - p_j$. Further, we let $W^{(t)} = \sum_{j=1}^n w_j^t$ and $w_{\max} = \max_j w_j$ while noting that by Lemma 3.5 the assumption $\beta > 0$ implies $W^{(3)} \le 64n$ and $w_{\max} \le 4n^{1/3}$. Hence we also have $W^{(4)} = \mathcal{O}(n^{4/3})$ and $W^{(5)} = \mathcal{O}(n^{5/3})$.

Now, X has the probability generating function

$$g(z) = \sum_{j=0}^{n} \Pr(X = j) z^j = \prod_{j=1}^{n} [q_j + p_j z^{w_j}].$$

Hence, the series $(\Pr(X \le i))_{i=0}^{\infty}$ has the generating function $\frac{g(z)}{1-z}$. Letting

$$f(z) = \frac{g(z)}{(1-z)z^{(W+1)/2}},$$

Cauchy's Residue Theorem gives

$$\Pr\left(X \le \frac{W-1}{2}\right) = \frac{1}{2\pi i}\oint_C f(z)dz = \frac{1}{2\pi i}\int_{C_1} f(z)dz + \frac{1}{2\pi i}\int_{C_2} f(z)dz,$$

where the contour C is the concatenation of the following two arcs, enclosing the pole $z = 0$ but not $z = 1$ (see Figure 1),

$$\begin{aligned} &C_1\colon z = e^{i\varphi}, && 2a \le \varphi \le 2\pi - 2a \\ &C_2\colon z = 1 + re^{-i\varphi}, && \pi/2 + a \le \varphi \le 3\pi/2 - a \end{aligned}$$

where $a(r) = \arcsin(\frac{r}{2})$ and $r > 0$ is a small parameter that we will later let go to 0.

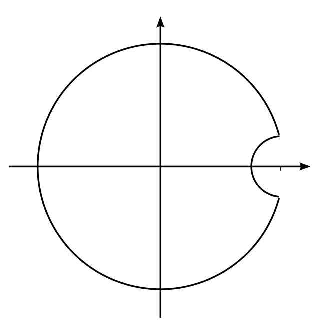

Figure 1: The contour

The second integral is

$$\frac{1}{2\pi i}\int_{C_2} f(z)dz = \frac{-1}{2\pi}\int_{\pi/2+a}^{3\pi/2-a} re^{-i\varphi}f(1+re^{-i\varphi})d\varphi =$$
$$= \frac{1}{2\pi}\int_{\pi/2+a}^{3\pi/2-a} \frac{\prod_{j=1}^{n}[q_j+p_j(1+re^{-i\varphi})^{w_j}]}{(1+re^{-i\varphi})^{(W+1)/2}}d\varphi \to \frac{1}{2} \text{ as } r\to 0,$$

since the integrand converges uniformly to 1. Let us now concentrate on the first integral. As $f(\overline{z}) = \overline{f(z)}$,

$$\frac{1}{2\pi i}\int_{C_1} f(z)dz = \frac{1}{\pi}\int_{2a}^{\pi} \mathrm{Re}[e^{i\varphi}f(e^{i\varphi})]d\varphi. \tag{13}$$

Expanding $f(e^{i\varphi})$ gives us that

$$e^{i\varphi}f(e^{i\varphi}) = e^{i\varphi/2}\frac{\prod_{j=1}^{n}[q_j+p_je^{iw_j\varphi}]}{(1-e^{i\varphi})e^{i\varphi W/2}} = \frac{\prod_{j=1}^{n}[q_j+p_je^{iw_j\varphi}]e^{-i\varphi W/2}}{-2i\sin(\varphi/2)} =$$
$$= \frac{\prod_{j=1}^{n}[q_je^{-iw_jp_j\varphi}+p_je^{iw_jq_j\varphi}]}{-2i\sin(\varphi/2)}e^{i\varphi\epsilon\Delta/2},$$

where the last equality follows from $W/2+\epsilon\Delta/2 = \sum_{j=1}^{n} w_jp_j$.

To analyze the integral as $r\to 0$ (and thus $a\to 0$), we first use Lemma 2.2 to argue that the integral between $\varphi_0 := \frac{\log n}{\sqrt{n}}$ and π is small. Indeed, since we have at least $200\log n$ weights which are 1, Lemma 2.2 together with the bound $\sin(x)\geq \frac{2}{\pi}x$ for $0\leq x\leq\frac{\pi}{2}$ implies that

$$\left|\frac{1}{\pi}\int_{\varphi_0}^{\pi}\mathrm{Re}[e^{i\varphi}f(e^{i\varphi})]d\varphi\right| \leq \frac{1}{2\pi\sin(\varphi_0/2)}\int_{\varphi_0}^{\pi}\prod_{j=1}^{n}\left|q_j+p_je^{iw_j\varphi}\right|d\varphi \leq \frac{\sqrt{n}}{2\log n}\mathcal{O}\left(\frac{1}{n^2}\right) = \mathcal{O}\left(\frac{1}{n}\right).$$

We now consider the integral between 0 and φ_0. We start with the following identity of the integrand.

Claim 4.3. *For $0\leq\varphi<\varphi_0$, we have*

$$\mathrm{Re}[e^{i\varphi}f(e^{i\varphi})] = e^{-\sigma^2\frac{\varphi^2}{2}}\left(-\frac{\epsilon\Delta}{2} - S_3\frac{\varphi^2}{6} + \epsilon^3\mathcal{O}(n\varphi^2) + \mathcal{O}(n^{5/3}\varphi^4)\right)\left(1+\mathcal{O}\left(\frac{\log^4 n}{n^{2/3}}\right)\right),$$

where $S_3 = \sum_{j=1}^{n} w_j^3q_jp_j(p_j^2-q_j^2)$.

Proof of Claim. Using Taylor expansion we may write

$$q_je^{-iw_jp_j\varphi}+p_je^{iw_jq_j\varphi} = 1 - w_j^2q_jp_j\frac{\varphi^2}{2} + iw_j^3q_jp_j(p_j^2-q_j^2)\frac{\varphi^3}{6} + w_j^4\mathcal{O}(\varphi^4) + iw_j^5\mathcal{O}(\varphi^5),$$

where we have separated the real and complex errors. Letting $\mathrm{Log}\, z = \log|z| + i\,\mathrm{Arg}\, z$ denote the principal logarithm ($-\pi < \mathrm{Arg}\, z\leq\pi$) which is analytic except on the non-positive part of the real axis, we can use another Taylor expansion to write (note that $\mathrm{Re}(q_je^{-iw_jp_j\varphi}+p_je^{iw_jq_j\varphi})>0$ for $\varphi<\varphi_0$ when n is large enough, since $w_j\leq 4n^{1/3}$)

$$q_je^{-iw_jp_j\varphi}+p_je^{iw_jq_j\varphi} = e^{\mathrm{Log}(q_je^{-iw_jp_j\varphi}+p_je^{iw_jq_j\varphi})}$$
$$= e^{-w_j^2q_jp_j\frac{\varphi^2}{2}+iw_j^3q_jp_j(p_j^2-q_j^2)\frac{\varphi^3}{6}+w_j^4\mathcal{O}(\varphi^4)+iw_j^5\mathcal{O}(\varphi^5)}.$$

Using $\sigma^2 = \sum_{j=1}^n w_j^2 q_j p_j$ and the expression for S_3 stated in the claim we have

$$e^{i\varphi} f(e^{i\varphi}) \quad = \quad e^{-\sigma^2 \frac{\varphi^2}{2}} \frac{e^{i\varphi\epsilon\frac{\Delta}{2} + iS_3\frac{\varphi^3}{6} + W^{(4)}\mathcal{O}(\varphi^4) + iW^{(5)}\mathcal{O}(\varphi^5)}}{-2i\sin(\varphi/2)}.$$

Since $W^{(4)} = \mathcal{O}(n^{4/3})$ and $W^{(5)} = \mathcal{O}(n^{5/3})$, we may write the real part of this as

$$\mathrm{Re}[e^{i\varphi} f(e^{i\varphi})] = -e^{-\sigma^2\frac{\varphi^2}{2}} \frac{\sin(h(\varphi))}{\varphi} \frac{\varphi/2}{\sin(\varphi/2)} e^{\mathcal{O}(n^{4/3}\varphi^4)},$$

where

$$h(\varphi) = \epsilon\frac{\Delta}{2}\varphi + S_3\frac{\varphi^3}{6} + \mathcal{O}(n^{5/3}\varphi^5). \tag{14}$$

Taylor expansions of e^x, $\sin(x)$ and $\frac{x}{\sin(x)} = \frac{1}{1+\mathcal{O}(x^2)} = 1 + \mathcal{O}(x^2)$ gives

$$\mathrm{Re}[e^{i\varphi} f(e^{i\varphi})] = e^{-\sigma^2\frac{\varphi^2}{2}} \frac{h(\varphi) + \mathcal{O}([h(\varphi)]^3)}{\varphi} (1 + \mathcal{O}(\varphi^2))(1 + \mathcal{O}(n^{4/3}\varphi^4)).$$

The product of the last two factors is $1 + \mathcal{O}\left(\frac{\log^4 n}{n^{2/3}}\right)$ since $\varphi \leq \log(n)/\sqrt{n}$. For the second factor, first note that since $|S_3| = \mathcal{O}(n)$ we have $h(\varphi) = \epsilon\frac{\Delta}{2}\varphi + \mathcal{O}(n\varphi^3)$ and thus, since $|\Delta| \leq n^{1/3}$ and $\varphi \leq \log(n)/\sqrt{n}$,

$$[h(\varphi)]^3 = \frac{\epsilon^3\Delta^3}{8}\varphi^3 + \Delta^2 n\mathcal{O}(\varphi^5) + \Delta n^2\mathcal{O}(\varphi^7) + n^3\mathcal{O}(\varphi^9) = \epsilon^3 n\mathcal{O}(\varphi^3) + n^{5/3}\mathcal{O}(\varphi^5). \tag{15}$$

Combining (14) and (15), the second factor becomes

$$\frac{h(\varphi) + \mathcal{O}([h(\varphi)]^3)}{\varphi} = \epsilon\frac{\Delta}{2} + S_3\frac{\varphi^2}{6} + \epsilon^3\mathcal{O}(n\varphi^2) + \mathcal{O}(n^{5/3}\varphi^4),$$

proving the claim. □

We now compute the part of the integral (13) from $\varphi = 0$ to φ_0:

Claim 4.4. *We have that*

$$I_2 := \frac{1}{\pi}\int_0^{\varphi_0} \mathrm{Re}[e^{i\varphi} f(e^{i\varphi})] d\varphi \leq -\frac{\epsilon(\Delta - 1 + \beta)}{\sigma\sqrt{8\pi}} + \epsilon^3\mathcal{O}\left(\frac{1}{\sqrt{n}}\right) + \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right).$$

Proof of Claim. By the previous claim,

$$I_2 = \frac{1}{\pi}\int_0^{\varphi_0} e^{-\sigma^2\frac{\varphi^2}{2}} \left(-\frac{\epsilon\Delta}{2} - S_3\frac{\varphi^2}{6} + \epsilon^3\mathcal{O}(n\varphi^2) + \mathcal{O}(n^{5/3}\varphi^4)\right)\left(1 + \mathcal{O}\left(\frac{\log^4 n}{n^{2/3}}\right)\right) d\varphi.$$

Using Lemma 2.3 and 2.4 and noting that $e^{-\frac{\sigma^2}{2}\varphi_0^2} = e^{-\Omega(\log^2(n))} = \frac{1}{n^{\Omega(\log n)}}$ since $\sigma^2 = \Theta(n)$, this is

$$I_2 = \left(-\frac{\epsilon\Delta}{2\sigma\sqrt{2\pi}} - \frac{S_3}{6\sigma^3\sqrt{2\pi}} + \frac{\epsilon^3\mathcal{O}(n)}{\sigma^3} + \frac{\mathcal{O}(n^{5/3})}{\sigma^5}\right)\left(1 + \mathcal{O}\left(\frac{\log^4 n}{n^{2/3}}\right)\right) + \frac{1}{n^{\Omega(\log n)}}.$$

Now,

$$
\begin{aligned}
|S_3| &= \left|\sum_{j=1}^n w_j^3 q_j p_j (p_j^2 - q_j^2)\right| = \left|\sum_{j=1}^n w_j^3 \frac{1-\epsilon_j^2}{4}\epsilon_j\right| = \left|\sum_{j=1}^n \frac{w_j^3\epsilon_j}{4} - \sum_{j=1}^n \frac{w_j^3\epsilon_j^3}{4}\right| = \\
&= \left|\sum_{j=1}^n \frac{w_j^3\epsilon_j}{4}\right| + \epsilon^3\mathcal{O}(n) = \left|\sum_{j=1}^n \frac{(w_j^3 - w_j)\epsilon_j + \epsilon\Delta}{4}\right| + \epsilon^3\mathcal{O}(n) \le \epsilon\sum_{j=1}^n \frac{w_j^3 - w_j}{4} + \left|\frac{\epsilon\Delta}{4}\right| + \epsilon^3\mathcal{O}(n).
\end{aligned}
$$

Recall that $1-\beta = \frac{\sum_{j=1}^n (w_j^3 - w_j)}{3\sum_{j=1}^n w_j^2}$. Together with the bound $\sigma^2 = \sum_{j=1}^n w_j^2 \frac{1-\epsilon_j^2}{4} \ge \frac{1-\epsilon^2}{4}\sum_{j=1}^n w_j^2$ this gives

$$
\epsilon\sum_{j=1}^n \frac{w_j^3 - w_j}{4} = \epsilon(1-\beta)\frac{3}{4}\sum_{j=1}^n w_j^2 \le 3\epsilon(1-\beta)\frac{\sigma^2}{1-\epsilon^2} = 3\epsilon(1-\beta)\sigma^2 + \epsilon^3\mathcal{O}(n).
$$

Thus $|S_3| \le 3\epsilon(1-\beta)\sigma^2 + \left|\frac{\epsilon\Delta}{4}\right| + \epsilon^3\mathcal{O}(n)$, which together with $\sigma^2 = \Theta(n)$ implies that

$$
I_2 \le -\frac{\epsilon(\Delta - 1 + \beta)}{\sigma\sqrt{8\pi}}\left(1 + \mathcal{O}\left(\frac{\log^4 n}{n^{2/3}}\right)\right) + \epsilon^3\mathcal{O}\left(\frac{1}{\sqrt{n}}\right) + \mathcal{O}\left(\frac{1}{n^{5/6}}\right).
$$

The claim now follows from $|\Delta| \le n^{1/3}$, which gives

$$
\frac{\epsilon(\Delta - 1 + \beta)}{\sigma\sqrt{8\pi}}\mathcal{O}\left(\frac{\log^4 n}{n^{2/3}}\right) = \mathcal{O}\left(\frac{n^{1/3}\log^4 n}{n^{1/2}n^{2/3}}\right) = \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right)
$$

□

Letting $a \to 0$ and summing all integrals we get

$$
\begin{aligned}
\Pr\left(X \le \frac{n-1}{2}\right) &\le \frac{1}{2} + \mathcal{O}\left(\frac{1}{n}\right) - \frac{\epsilon(\Delta - 1 + \beta)}{\sigma\sqrt{8\pi}} + \epsilon^3\mathcal{O}\left(\frac{1}{\sqrt{n}}\right) + \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right) \\
&= \frac{1}{2} - \frac{\epsilon(\Delta - 1 + \beta)}{\sigma\sqrt{8\pi}} + \epsilon^3\mathcal{O}\left(\frac{1}{\sqrt{n}}\right) + \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right).
\end{aligned}
$$

□

In order to extend this bound to larger (negative) Δ's we use the central limit theorem with explicit error bounds:

Theorem 4.5 (Berry-Esseen)**.** *Let $Y = Y_1 + \cdots + Y_n$ be a sum of independent random variables satisfying $\mathbb{E}[Y_i] = 0$ for all i, $\sqrt{\sum \mathbb{E}[Y_i^2]} = \sigma$, and $\sum \mathbb{E}[|Y_i|^3] = \rho_3$. Then*

$$
\sup_x |\Pr[Y \le x] - \Phi(x/\sigma)| \le C\rho_3/\sigma^3,
$$

where Φ is the cdf of a standard Gaussian random variable, and C is an absolute constant. It has been shown that one can take $C = 0.7915$ [16].

Lemma 4.6. *Fix $0 < \epsilon < 1$. Let $X = w_1X_1 + \cdots + w_nX_n$ be a sum of n independent Bernoulli random variables with positive integer weights of total sum $W = \sum_{j=1}^n w_j$, and suppose $\Pr(X_j = 1) = \frac{1+\epsilon x_j}{2}$ where $-1 \le x_j \le 1$. Further, let $\sigma^2 = \operatorname{Var} X$ and $\Delta = \sum_{j=1}^n w_j x_j$. Then, if $\Delta \le 0$,*

$$
\Pr\left(X \ge \frac{W+1}{2}\right) \ge \frac{1}{2} + \frac{\epsilon\Delta}{\sigma\sqrt{8\pi}} - \frac{C}{(1-\epsilon^2)^{3/2}}\frac{\sum_{j=1}^n w_j^3}{n^{3/2}}.
$$

Proof. As before we let $\epsilon_j = \epsilon x_j$, $p_j = \frac{1+\epsilon_j}{2}$ and $q_j = 1 - p_j$. Now, let $Y_j = w_j(X_j - p_j)$ and $Y = Y_1 + \cdots + Y_n$. Then $\mathbb{E}[Y_j] = 0$,

$$\sum_{j=1}^n \mathbb{E}[Y_j^2] = \sigma^2 = \sum_{j=1}^n w_j^2 \operatorname{Var} X_j = \sum_{j=1}^n w_j^2 \frac{1-\epsilon_j^2}{4} \geq n\frac{1-\epsilon^2}{4}$$

and

$$\sum_{j=1}^n \mathbb{E}[|Y_j|^3] = \sum_{j=1}^n w_j^3(p_j q_j^3 + q_j p_j^3) = \sum_{j=1}^n w_j^3 \frac{1-\epsilon_j^4}{8} \leq \sum_{j=1}^n \frac{w_j^3}{8}.$$

Further, we have

$$\Pr\left(X \leq \frac{W}{2}\right) = \Pr\left(Y \leq \frac{W}{2} - \sum_{j=1}^n w_j p_j\right) = \Pr\left(Y \leq -\frac{\epsilon\Delta}{2}\right).$$

Applying Berry-Esseen and using $\Phi(x) \leq \frac{1}{2} + \frac{x}{\sqrt{2\pi}}$ for $x \geq 0$ this is at most

$$\Phi\left(\frac{-\epsilon\Delta}{2\sigma}\right) + \frac{C}{(1-\epsilon^2)^{3/2}} \frac{\sum_{j=1}^n w_j^3}{n^{3/2}} \leq \frac{1}{2} + \frac{-\epsilon\Delta}{\sigma\sqrt{8\pi}} + \frac{C}{(1-\epsilon^2)^{3/2}} \frac{\sum_{j=1}^n w_j^3}{n^{3/2}}.$$

□

We are now ready to prove Theorem 4.1 by combining the bounds for small and large Δ.

Proof of Theorem 4.1. Let x^* be the optimal solution to the linear program in $A_{LP,\epsilon}$ for some $\epsilon > 0$ that we will specify later. By Lemma 4.2 and 4.6 we have the following lower bounds on the probability that the i'th constraint $\operatorname{sgn}(L_i(X))$ with bias $\mathbb{E}(L_i(X)) = \epsilon L_i(x^*) \geq \epsilon\Delta_i^*$ is satisfied:

$$\begin{cases} \frac{1}{2} + \frac{\epsilon(\Delta_i^* - 1 + \beta)}{\sigma\sqrt{8\pi}} + \epsilon^3 \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) + \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right), & \text{if } |\Delta_i^*| \leq n^{1/3} \\ \frac{1}{2} + \frac{\epsilon\Delta_i^*}{\sigma\sqrt{8\pi}} - \frac{C}{(1-\epsilon^2)^{3/2}} \frac{\sum_{j=1}^n w_j^3}{n^{3/2}}, & \text{if } \Delta_i^* \leq 0 \end{cases}$$

Here we have used the fact that having a larger expected value of the linear form $L_i(x^*)$ than its lower bound Δ_i^* can only increase the probability of a constraint being satisfied. Since by (12), $\bar{\Delta}^* \geq 1 - \delta > 0$, and further $\Delta_i^* \leq 1$ for all i, we must have $\Delta_i^* \geq -n^{1/3}$ for at least a fraction $1 - \frac{1}{n^{1/3}}$ of the constraints. Thus, the expected fraction of satisfied constraints is

$$\begin{aligned} \mathbb{E}\left[\frac{\mathsf{Adv}(A_{LP,\epsilon}(\mathcal{I})) + 1}{2}\right] \geq \frac{1}{2} \quad &+ \quad \frac{\epsilon(\bar{\Delta}^* - 1 + \beta)}{\sigma\sqrt{8\pi}} + \epsilon^3 \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) + \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right) \\ &- \quad \frac{1}{n^{1/3}} \frac{C}{(1-\epsilon^2)^{3/2}} \frac{\sum_{j=1}^n w_j^3}{n^{3/2}}, \end{aligned}$$

where we used that $\beta \leq 1$. Now, by Lemma 3.5, we have $\sum_{j=1}^n w_j^3 \leq 64n$ and also $\sigma = \Theta(\sqrt{n})$. This together with $\bar{\Delta}^* \geq 1 - \delta$ implies that

$$\mathbb{E}[\mathsf{Adv}(A_{LP,\epsilon}(\mathcal{I}))] \geq \epsilon(\beta - \delta)\Omega\left(\frac{1}{\sqrt{n}}\right) + \epsilon^3 \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) + \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right).$$

Now, letting $\epsilon = (\beta - \delta)^{1/2}\epsilon_0$ for some absolute constant $\epsilon_0 > 0$ small enough so that the first term dominates the second by a constant factor for any n, we have

$$\mathbb{E}[\mathsf{Adv}(A_{LP,\epsilon}(\mathcal{I}))] \geq (\beta - \delta)^{3/2}\Omega\left(\frac{1}{\sqrt{n}}\right) + \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right).$$

□

4.2 Application to Majority

As $\beta = 1$ for Maj_n the following result follows directly from Theorem 3.3 (by also adjusting the hidden constants in (11) so that this expression becomes non-positive when $200 \log n > n$):

Corollary 4.7. (Formal statement of Theorem 1.1) *For all $1 - \frac{\delta}{n+1}$ satisfiable instances $\mathcal{I}$ of Max-Maj-n, where $\delta < 1$, we have*

$$\mathbb{E}[\mathsf{Adv}(A_{LP,\epsilon}(\mathcal{I}))] = (1-\delta)^{3/2}\Omega\left(\frac{1}{\sqrt{n}}\right) - \mathcal{O}\left(\frac{\log^4 n}{n^{5/6}}\right),$$

where $\epsilon = (1-\delta)^{1/2}\epsilon_0$ and $\epsilon_0 > 0$ is an absolute constant.

5 Unique Games Hardness

In this section, we show hardness of approximation results for majority-like predicates under the Unique Games Conjecture. This complements our algorithmic results obtained in Sections 3 and 4.

5.1 The Basic Tool

The hardness results in this section are under the increasingly prevalent assumption that the Unique Games Conjecture (UGC) holds (see Appendix A for a definition). The basic tool that we use is the result by Austrin and Mossel [2], which states that the UGC implies that a predicate is approximation resistant if it supports a uniform pairwise independent distribution, and hard to approximate if it "almost" supports a uniform pairwise independent distribution. We now state their result in a simplified form tailored for the application at hand:

Theorem 5.1 ([2])**.** *Let $P\colon \{-1,1\}^n \to \{-1,1\}$ be a n-ary predicate and let μ be a balanced pairwise independent distribution over $\{-1,1\}^n$. Then, for any $\epsilon > 0$, the UGC implies that it is NP-hard to distinguish between those instances of Max-CSP(P)*

- *that have an assignment satisfying at least a fraction $\Pr_{x\in(\{-1,1\}^n,\mu)}[P(x) = 1] - \epsilon$ of the constraints;*
- *and those for which any assignment satisfies at most a fraction $|P^{-1}(1)|/2^n + \epsilon$ of the constraints.*

5.2 Application to the Majority Predicate

We now give a fairly easy application of Theorem 5.1 to the predicate Maj_n. Later, we generalize this approach to more general homogeneous linear threshold predicates.

Theorem 5.2. (Formal statement of Theorem 1.2) *For any $\epsilon > 0$ the UGC implies that it is NP-hard to distinguish between those instances of Max-Maj-n*

- *that have an assignment satisfying at least a fraction $1 - \frac{1}{n+1} - \epsilon$ of the constraints;*
- *and those for which any assignment satisfies at most a fraction $1/2 + \epsilon$ of the constraints.*

Proof. Consider the following distribution μ over $\{-1,+1\}^n$: with probability $\frac{1}{n+1}$, all the bits in μ are fixed to -1, and with probability $\frac{n}{n+1}$, μ samples a vector with $(n+1)/2$ ones, chosen uniformly at random among all possibilities. To see that this gives a pairwise independent distribution let $X = (X_1, \ldots, X_n)$ be drawn from μ. Then $\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n+1} \cdot (-n) + \frac{n}{n+1} \cdot 1 = 0$ and

$\mathbb{E}\left[\sum_{\substack{i,j=1\\i\neq j}}^n X_iX_j\right] = \mathbb{E}\left[\left(\sum_{i=1}^n X_i\right)^2\right] - n = \frac{1}{n+1}\cdot(n^2) + \frac{n}{n+1}\cdot 1 - n = 0$. Because of the symmetry of the coordinates, it follows that for all i, $\mathbb{E}[X_i] = 0$ and for every $i \neq j$, $\mathbb{E}[X_iX_j] = 0$. Therefore, the distribution μ is balanced pairwise independent. Theorem 5.1 now gives the result. □

For predicate Maj_n, we can also obtain a hardness result for almost satisfiable instances:

Theorem 5.3. (Formal statement of Theorem 1.3) *For any $\epsilon > 0$ the UGC implies that it is NP-hard to distinguish between those instances of Max-Maj-n*

- *that have an assignment satisfying at least a fraction $1-\epsilon$ of the constraints;*
- *and those for which any assignment satisfies at most a fraction $\frac{1}{2}+c_n\frac{1}{\sqrt{n}}+\epsilon$ of the constraints, where*

$$c_n = \frac{\sqrt{n}}{2^{n-2}}\binom{n-2}{\frac{n-1}{2}} \approx \sqrt{\frac{2}{\pi}}.$$

Proof. Let $k = n-2$ and consider the predicate $P\colon \{-1,1\}^k \to \{-1,1\}$ defined as $P(x) = \operatorname{sgn}(x_1 + \cdots + x_k + 2)$. Our interest in P stems from the fact that Max-Maj-n is at least as hard to approximate as Max-CSP(P). Indeed, given an instance of Max-CSP(P), we can construct an instance of Max-Maj-n by letting each constraint $P(l_1, \ldots l_k)$ equal $\mathsf{Maj}_n(y_1, y_2, l_1, \ldots, l_k)$ for two new variables y_1 and y_2, that are the same in all constraints and always appear in the positive form. As any good solution to the instance of Max-Maj-n sets both y_1 and y_2 to one, we can conclude that any optimal assignments to the two instances satisfy the same fraction of constraints.

Now consider the following distribution μ over $\{-1,1\}^k$: with probability $\frac{1}{k+1}$, all the bits in μ are fixed to ones, and with probability $\frac{k}{k+1}$, μ samples a vector with $(k+1)/2$ minus ones, chosen uniformly at random among all possibilities. The same argument as in the proof of Theorem 5.2 shows that the distribution μ is uniform and pairwise independent. Theorem 5.1 now gives that for any $\epsilon > 0$ the UGC implies that it is NP-hard to distinguish between those instances of Max-CSP(P) that have an assignment satisfying a fraction $1-\epsilon$ of the constraints; and those for which any assignment satisfies at most a fraction

$$\frac{|P^{-1}(1)|}{2^k} + \epsilon = \frac{1}{2^k}\sum_{j=0}^{\frac{k+1}{2}}\binom{k}{j} + \epsilon = \frac{1}{2} + \frac{\binom{k}{\frac{k+1}{2}}}{2^k} + \epsilon = \frac{1}{2} + \sqrt{\frac{2}{\pi k}} + o(1/k) + \epsilon.$$

The result now follows from the observation above that we can construct an instance of Max-Maj-n from an instance of Max-CSP(P) such that optimal assignments to the two instances satisfy the same fraction of the constraints. □

Taking the convex combination of the results in Theorems 5.2 and 5.3 yields:

Corollary 5.4. *For any $\delta : 0 \leq \delta \leq 1$ and any $\epsilon > 0$, the UGC implies that it is NP-hard to find an assignment x to a given $1 - \frac{\delta}{n+1} - \epsilon$ satisfiable instance of Max-Maj-n achieving*

$$\mathsf{Adv}(x) \geq (1-\delta)c_n\frac{1}{\sqrt{n}} + \epsilon,$$

where c_n is the constant defined in Theorem 5.3.

5.3 Hardness for More General Predicates

We will now prove hardness of approximation for more general predicates than majority. Let us first recall the main idea for proving hardness of Maj_n. Since all weights are one, we have $w_{\max} = 1$. One can now observe that the constructed balanced pairwise independent distribution μ over $\{-1,1\}^n$ in Theorem 5.2 can be defined as follows. With probability $\frac{1}{n+1}$ sample a vector where the jth bit is set to 1 with probability $0 = \frac{1-\frac{w_j}{w_{\max}}}{2}$ independent of the other bits, and with probability $1 - \frac{1}{n+1}$ sample a vector X such that $\sum_{j=1}^n X_j = 1$, or equivalently $\sum_{j=1}^n w_j X_j = w_{\max}$, chosen uniformly at random among all possibilities.

We will prove (in Theorem 5.6) that a distribution μ essentially defined as above is an almost balanced pairwise distribution for homogeneous linear threshold predicates of the form $\operatorname{sgn}(w_1x_1 + w_2x_2 + \cdots + w_nx_n)$ with $400\log n$ unit weights and $\sum_{j=1}^n w_j^3 < 100n$. We then, using a general result show that such a distribution can be slightly adjusted to obtain a perfect balanced pairwise distribution. These two results are then combined, in Theorem 5.5 below (proved in Section 5.3.3), to obtain the desired hardness results.

Theorem 5.5. *Suppose we are given positive integers $(w_j)_{j=1}^n$ such that $\sum_{j=1}^n w_j^3 < 100n$ and $\sum_{j=1}^n w_j$ is odd. Further, suppose that for at least $400\log n$ different values of j we have $w_j = 1$. Let $P(x) = \operatorname{sgn}(w_1x_1 + \cdots + w_nx_n)$, then, for any $\epsilon > 0$, the UGC implies that it is NP-hard to distinguish between those instances of Max-CSP(P)*

- *that have an assignment satisfying at least a fraction $1 - \mathcal{O}\left(\frac{w_{\max}^4}{n}\right) - \epsilon$ of the constraints;*
- *and those for which any assignment satisfies at most a fraction $1/2 + \epsilon$ of the constraints.*

5.3.1 Almost Balanced Pairwise Distribution

In this section we prove the following:

Theorem 5.6. *Suppose we are given positive integers $(w_j)_{j=1}^n$ such that $\sum_{j=1}^n w_j^3 < 100n$ and $w_j = 1$ for at least $400\log n$ different values of j. Then there is a distribution μ over $\{-1,1\}^n$ satisfying*

$$\begin{aligned}
\Pr[\textstyle\sum_{j=1}^n w_j X_j > 0] &= 1 - \mathcal{O}\left(\frac{w_{\max}^2}{n}\right) \\
\mathbb{E}[X_i] &= 0 && \text{for } i = 1, \ldots, n \\
\mathbb{E}[X_i X_j] &= \mathcal{O}\left(\frac{w_{\max}^4}{n^2}\right) && \text{for } i, j : 1 \le i < j \le n.
\end{aligned}$$

The technical part of proving this theorem is captured by the following lemma.

Lemma 5.7. *Suppose we are given positive integers $(w_j)_{j=1}^n$ satisfying the conditions of Theorem 5.6. Let $\tilde{w}_{\max}$ be the smallest integer such that $\tilde{w}_{\max} \ge \max_j w_j$ and $\sum_{j=1}^n w_j + \tilde{w}_{\max} = 0 \mod 2$, then there is a distribution ν over $\{-1,1\}^n$ with support $\{x : \sum_{j=1}^n w_j x_j = \tilde{w}_{\max}\}$ satisfying*

$$\begin{aligned}
\mathbb{E}[X_j] &= w_j \frac{\tilde{w}_{\max}}{W^{(2)}} + \mathcal{O}\left(\frac{\tilde{w}_{\max}^4}{n^2}\right) && \text{for } j = 1, \ldots, n \\
\mathbb{E}[X_i X_j] &= -\frac{w_i w_j}{W^{(2)}} + \mathcal{O}\left(\frac{\tilde{w}_{\max}^4}{n^2}\right) && \text{for } 1 \le i < j \le n,
\end{aligned}$$

where $W^{(2)} = \sum_{j=1}^n w_j^2$.

Before giving the proof, let us see how this lemma implies Theorem 5.6. The Distribution ν is such that $\mathbb{E}[X_j] = w_j \frac{\tilde{w}_{\max}}{\sum_{j=1}^n w_j^2} + \beta_j$ where $\beta_j = \mathcal{O}(\frac{\tilde{w}_{\max}^4}{n^2})$ for $j = 1, \ldots, n$. Let $\beta_{\max} = \max_j |\beta_j|$ and let $T = \tilde{w}_{\max}^2 + (1+\beta_{\max})W^{(2)}$. Now define the distribution μ over $\{-1,1\}^n$ as follows:

with probability $\frac{\tilde{w}_{\max}^2}{T}$: sample a vector where the j:th bit is set to 1 with probability $\frac{1-\frac{w_j}{\tilde{w}_{\max}}}{2}$ independent of the other bits;

with probability $\frac{W^{(2)}}{T}$: sample a vector from ν; and

with probability $\frac{W^{(2)}\beta_{\max}}{T}$: sample a vector where the j:th bit is set to 1 with probability $\frac{1-\frac{\beta_j}{\beta_{\max}}}{2}$.

To verify that μ satisfies $\mathbb{E}[X_j] = 0$ and $\mathbb{E}[X_i X_j] = \mathcal{O}(\frac{w_{\max}^4}{n^2})$ is now an easy task and left to the reader. Furthermore, μ satisfies (as required)

$$\begin{aligned}\Pr\left[\sum_{j=1}^n w_j X_j > 0\right] &\geq \frac{W^{(2)}}{T} = \frac{W^{(2)}}{\tilde{w}_{\max}^2 + (1+\beta_{\max})W^{(2)}} = 1 - \frac{\beta_{\max}W^{(2)} + \tilde{w}_{\max}^2}{\tilde{w}_{\max}^2 + (1+\beta_{\max})W^{(2)}} = \\ &= 1 - \mathcal{O}\left(\frac{\tilde{w}_{\max}^2}{n}\right) = 1 - \mathcal{O}\left(\frac{w_{\max}^2}{n}\right),\end{aligned}$$

where we used that $\tilde{w}_{\max} \leq w_{\max} + 1$. Let us remark that the $\mathcal{O}(\cdot)$ terms of Lemma 5.7 arise when we estimate probabilities using complex integrals. If we omit those terms then the distribution μ, defined as above with $\beta_{max} = 0$, would be balanced pairwise independent and satisfy $\Pr[\sum_{j=1}^n w_j X_j] \geq \frac{W^{(2)}}{T} = 1 - \frac{w_{\max}^2}{w_{\max}^2 + \sum_{j=1}^n w_j^2}$, i.e., essentially matching the bound of Theorem 5.10.

We proceed with the proof of Lemma 5.7.

Proof. That the set $\{X : \sum_{j=1}^n w_j X_j = \tilde{w}_{\max}\}$ is non-empty follows from that $\sum_{j=1}^n w_j + \tilde{w}_{\max} = 0$ mod 2 and that the assumptions on the weights $(w_j)_{j=1}^n$ imply that $w_{\max} < 5n^{1/3}$, $w_j < 5$ for at least $n/5$ different values of j, and $w_j = 1$ for at least $400 \log n$ different values of j. Now let ν be the distribution over $\{-1,1\}^n$ that samples uniformly at random among all possibilities a vector in $\{X : \sum_{j=1}^n w_j X_j = \tilde{w}_{\max}\}$.

By definition, we have that the support of ν is $\{x : \sum_{j=1}^n w_j x_j = \tilde{w}_{\max}\}$. We proceed by analyzing the expectation of X_{j_0} with respect to ν for a fixed $j_0 : 1 \leq j_0 \leq n$. Let $W = \sum_{j=1}^n w_j$ and let $Y = \sum_{j=1, j\neq j_0}^n w_j Y_j$ be the weighted sum of $n-1$ balanced Bernoulli random variables. Further, let A and B be the events that $Y = \frac{W + \tilde{w}_{\max} - 2w_{j_0}}{2}$ and $Y = \frac{W+\tilde{w}_{\max}}{2}$, respectively. With this notation, the expectation $\mathbb{E}[X_{j_0}]$ can be written as

$$\frac{\Pr[A] - \Pr[B]}{\Pr[A] + \Pr[B]}.$$

Applying Lemma 2.5 with $\varphi_0 = \frac{\log n}{\sqrt{n}}$ and using the assumption $|\{j : w_j = 1\}| \geq 400 \log n$, gives

$$\Pr[A] = \int_0^{\varphi_0} \cos\left(\varphi \frac{\tilde{w}_{\max} - w_{j_0}}{2}\right) h(\varphi)\, d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right)$$

and

$$\Pr[B] = \int_0^{\varphi_0} \cos\left(\varphi \frac{\tilde{w}_{\max} + w_{j_0}}{2}\right) h(\varphi)\, d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right),$$

where $h(\varphi) = \prod_{j=1, j\neq j_0}^n \cos(w_j\varphi/2)$. Using Taylor expansion, we may write

$$\cos\left(\varphi\frac{\tilde{w}_{\max} - w_{j_0}}{2}\right) + \cos\left(\varphi\frac{\tilde{w}_{\max} + w_{j_0}}{2}\right) = 2 + \mathcal{O}(\tilde{w}_{\max}^2\varphi^2)$$

and

$$\cos\left(\varphi\frac{\tilde{w}_{\max} - w_{j_0}}{2}\right) - \cos\left(\varphi\frac{\tilde{w}_{\max} + w_{j_0}}{2}\right) = \tilde{w}_{\max}w_{j_0}\varphi^2/2 + \mathcal{O}(\tilde{w}_{\max}^3 w_{j_0}\varphi^4).$$

Substituting in these expansions give us that

$$\mathbb{E}[X_{j_0}] \quad = \quad \frac{\int_0^{\varphi_0} h(\varphi)(\frac{\tilde{w}_{\max}w_{j_0}}{2}\varphi^2 + \mathcal{O}(\tilde{w}_{\max}^3 w_{j_0}\varphi^4))d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right)}{\int_0^{\varphi_0} h(\varphi)(2 + \mathcal{O}(\varphi^2\tilde{w}_{\max}^2))d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right)}$$

We now use Lemma 2.6 to evaluate this expression. Letting $S_2 = \sum_{j=1, j\neq j_0}^n w_j^2$, we can thus write $\mathbb{E}[X_{j_0}]$ as

$$\frac{\frac{\sqrt{\pi}}{8}\tilde{w}_{\max}w_{j_0}\left(\frac{S_2}{8}\right)^{-3/2} + \mathcal{O}(\tilde{w}_{\max}^3 w_{j_0})\left(\frac{S_2}{8}\right)^{-5/2}}{\sqrt{\pi}\left(\frac{S_2}{8}\right)^{-1/2} + \mathcal{O}(\tilde{w}_{\max}^2)\left(\frac{S_2}{8}\right)^{-3/2}}(1 + \mathcal{O}(\frac{\tilde{w}_{\max}}{n})).$$

As $S_2 = \Omega(n)$ and $w_{\max} = \mathcal{O}(n^{1/3})$, this can be simplified to

$$\frac{(1 + \mathcal{O}(\frac{\tilde{w}_{\max}^2}{n}))\frac{\sqrt{\pi}}{8}\tilde{w}_{\max}w_{j_0}\left(\frac{S_2}{8}\right)^{-1}}{(1 + \mathcal{O}(\frac{\tilde{w}_{\max}^2}{n}))\sqrt{\pi}}(1 + \mathcal{O}(\frac{\tilde{w}_{\max}}{n})) = \frac{\tilde{w}_{\max}w_{j_0}}{S_2}(1 + \mathcal{O}(\frac{\tilde{w}_{\max}^2}{n})),$$

which in turn (since $\sum_{j=1}^n w_j^2 = (1 + \mathcal{O}(\frac{\tilde{w}_{\max}^2}{n}))\sum_{j=1, j\neq j_0}^n w_j^2$) equals

$$\frac{\tilde{w}_{\max}w_{j_0}}{\sum_{j=1}^n w_j^2}(1 + \mathcal{O}(\frac{\tilde{w}_{\max}^2}{n})) = \frac{\tilde{w}_{\max}w_{j_0}}{\sum_{j=1}^n w_j^2} + \mathcal{O}(\frac{\tilde{w}_{\max}^4}{n^2}).$$

We complete the proof by analyzing the expectation of $X_{i_0}X_{j_0}$ with respect to ν for fixed $i_0, j_0 : 1 \leq i_0 < j_0 \leq n$. The arguments are similar to the ones used above for $\mathbb{E}[X_j]$ and sketched in the following. Let now $Y = \sum_{j=1, j\notin\{i_0,j_0\}}^n w_j Y_j$ be the weighted sum of $n-2$ balanced Bernoulli random variables. Further, let A, B, C, and D be the events that $Y = \frac{W + \tilde{w}_{\max} - 2w_{i_0} - 2w_{j_0}}{2}$, $Y = \frac{W + \tilde{w}_{\max}}{2}$, $Y = \frac{W + \tilde{w}_{\max} - 2w_{i_0}}{2}$, and $Y = \frac{W + \tilde{w}_{\max} - 2w_{j_0}}{2}$, respectively. With this notation, the expectation $\mathbb{E}[X_{i_0}X_{j_0}]$ can be written as

$$\frac{\Pr[A] + \Pr[B] - \Pr[C] - \Pr[D]}{\Pr[A] + \Pr[B] + \Pr[C] + \Pr[D]}.$$

Applying Lemma 2.5, we obtain for $E \in \{A, B, C, D\}$

$$\Pr[E] = \int_0^{\varphi_0} \cos\left(\varphi\frac{\tilde{w}_{\max} + g(E)}{2}\right) h(\varphi)\,d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right)$$

where $h(\varphi)$ now is $\prod_{j=1, j\notin\{i_0,j_0\}}^n \cos(w_j\varphi/2)$ and $g(A) = -(w_{i_0} + w_{j_0}), g(B) = w_{i_0} + w_{j_0}, g(C) = -(w_{i_0} - w_{j_0})$, and $g(D) = w_{i_0} - w_{j_0}$. Similar to before, we use Taylor expansions to obtain

$$\mathbb{E}[X_{i_0}X_{j_0}] \quad = \quad -\frac{\int_0^{\varphi_0} h(\varphi)(w_{i_0}w_{j_0}\varphi^2 + \mathcal{O}(\tilde{w}_{\max}^2 w_{i_0}w_{j_0}\varphi^4))d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right)}{\int_0^{\varphi_0} h(\varphi)(4 + \mathcal{O}(\varphi^2\tilde{w}_{\max}^2))d\varphi + \mathcal{O}\left(\frac{1}{n^4}\right)}$$

and then evaluate the expression using Lemma 2.6 to get that this equals

$$-\frac{\frac{\sqrt{\pi}}{4}w_{i_0}w_{j_0}\left(\frac{S_2}{8}\right)^{-3/2}+\mathcal{O}(\tilde{w}_{\max}^2 w_{i_0}w_{j_0})\left(\frac{S_2}{8}\right)^{-5/2}}{2\sqrt{\pi}\left(\frac{S_2}{8}\right)^{-1/2}+\mathcal{O}(\tilde{w}_{\max}^2)\left(\frac{S_2}{8}\right)^{-3/2}}(1+\mathcal{O}(\frac{\tilde{w}_{\max}}{n}),$$

where $S_2 = \sum_{j=1, j\notin\{i_0,j_0\}}^n w_j^2$. Similar simplifications as before now yields the desired result. □

In the next section we discuss how to find a distribution with a prescribed set of correlations. We later use this to make our almost pairwise independent distribution perfectly pairwise independent.

5.3.2 Existance of Correlated Random Bits

Given a matrix $\alpha \in \mathbb{R}^{n\times n}$, a necessary and sufficient condition for the existance of a random real vector with covariance matrix α is that α is symmetric and positive semidefinite. For sufficiency it is enough to consider a (possibly degenerate) normal distribution. However, for a random vector of bits $X \in \{-1,1\}^n$ this is not sufficient even if we require $\alpha_{i,i} = 1$ for all i. An example is $n = 3$ and $\alpha_{i,j} = -\frac{1}{2}$ for $i \neq j$. Although α is positive semidefinite no such distribution on bits exist since for bits $|X_1 + X_2 + X_3| \geq 1$, but such an α we imply

$$\mathbb{E}[(X_1 + X_2 + X_3)^2] = \sum_{i,j} \alpha_{i,j} = 3 - \frac{6}{2} = 0$$

We will give a sufficient condition for random bits, but first we start with a classical lemma:

Lemma 5.8. *Let $Z_1, Z_2 \sim \mathrm{Norm}(0,1)$ be standard normals with covariance $\mathbb{E}[Z_1Z_2] = \rho$. Then $\Pr(\mathrm{sgn}(Z_1) \neq \mathrm{sgn}(Z_2)) = \frac{\arccos(\rho)}{\pi}$.*

Proof. Let $Y \sim \mathrm{Norm}(0,1)$ be a standard normal variable independent of Z_1 and Z_2, and let

$$Z_2' = \rho Z_1 + \sqrt{1-\rho^2}Y = \cos(\varphi)Z_1 + \sin(\varphi)Y$$

where $\varphi = \arccos(\rho)$. Then (Z_1, Z_2) and (Z_1, Z_2') are identically distributed. Further Z_2' is the first coordinate of the random vector (Z_1, Y) rotated by an angle φ. But since the distribution of (Z_1, Y) is rotationally symmetric, the probability that the sign of the first coordinate changes under such a rotation is $\frac{\varphi}{\pi}$, i.e.

$$\Pr(\mathrm{sgn}(Z_1) \neq \mathrm{sgn}(Z_2)) = \Pr(\mathrm{sgn}(Z_1) \neq \mathrm{sgn}(Z_2')) = \frac{\varphi}{\pi} = \frac{\arccos(\rho)}{\pi}.$$

□

Using this lemma we now show that for the existance of random bits it is sufficient to have pairwise covariance bounded by $\frac{2}{\pi n}$.

Lemma 5.9. *Let $\alpha \in \mathbb{R}^{n\times n}$ be a symmetric matrix with $\alpha_{i,i} = 1$ and $|\alpha_{i,j}| \leq \frac{2}{\pi n}$ for all $i \neq j$. Then there exist a distribution on random bits $X_1, \ldots, X_n$ taking values in $\{-1,1\}$ such that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_iX_j] = \alpha_{i,j}$, for all $i, j \in [n]$.*

Proof. Let $S \in \mathbb{R}^{n\times n}$ be defined by

$$S_{i,j} = \sin\left(\frac{\pi}{2}\alpha_{i,j}\right) = \cos\left(\frac{\pi}{2}(1-\alpha_{i,j})\right)$$

and note that $S_{i,i} = 1$ while $|S_{i,j}| \leq \frac{\pi}{2}|\alpha_{i,j}| \leq \frac{1}{n}$ for all $i \neq j$. Furthermore, S is positive semidefinite since for any $x \in \mathbb{R}^n \setminus \{0\}$ we have

$$x^T Sx \geq \sum_{i=1}^n x_i^2 - \sum_{i,j\in[n]:i\neq j} \frac{1}{n} x_i x_j \geq \sum_{i=1}^n x_i^2 - \sum_{i=1}^n \frac{1}{\sqrt{n}}|x_i| \sum_{j=1}^n \frac{1}{\sqrt{n}}|x_j| \geq 0,$$

where the last inequality follows from Cauchy-Schwarz. Since S is positive semidefinite we can let $Z \sim \mathrm{Norm}(0, S)$ be a (possibly degenerate) n-dimensional normal with mean 0 and covariance matrix S. We will take $X_i = \mathrm{sgn}(Z_i)$, for $i = 1 \ldots n$. Clearly $\mathbb{E}[X_i] = 0$, and further for any i, j we have

$$\mathbb{E}[X_i X_j] = 1 - 2\Pr(\mathrm{sgn}(Z_i) \neq \mathrm{sgn}(Z_j)) = 1 - 2\frac{\arccos(S_{i,j})}{\pi} = \alpha_{i,j},$$

where we have used Lemma 5.8. □

5.3.3 Proof of Theorem 5.5

By Theorem 5.6, there is a distribution μ over $\{-1,1\}^n$ satisfying $\Pr[P(X) = 1] = 1 + \mathcal{O}\left(\frac{w_{\max}^2}{n}\right)$, $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i X_j] = \beta_{ij}$, where $\beta_{ij} = \mathcal{O}\left(\frac{w_{\max}^4}{n^2}\right)$. Let $\beta_{\max} = \max_{i,j} |\beta_{i,j}|$ and let $\alpha_{i,j} = -\frac{\beta_{ij}}{\beta_{\max}} \frac{2}{\pi n}$ for all $i \neq j$. Let now ν denote the distribution over $\{-1,1\}^n$ satisfying $\mathbb{E}[X_i] = 0$ and $\mathbb{E}[X_i X_j] = \alpha_{i,j}$. Such a distribution is guaranteed to exist by the sufficient condition shown in Lemma 5.9 (since $|\alpha_{i,j}| \leq \frac{2}{\pi n}$ for all $i \neq j$). Let $p = \frac{2}{\beta_{\max}\pi n + 2}$ and consider the distribution D: with probability p sample from μ and with probability $1 - p$ sample from ν. It is easy to see that this distribution is balanced pairwise independent. Furthermore,

$$\begin{aligned}
\Pr_{X\in(\{-1,1\}^n,D)}[P(X) = 1] &\geq p \Pr_{X\in(\{-1,1\}^n,\mu)}[P(X) = 1] \\
&= \left(1 - \frac{\beta_{\max}\pi n}{\beta_{\max}\pi n + 2}\right)\left(1 + \mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right) \\
&= \left(1 - \mathcal{O}\left(\frac{w_{\max}^4}{n}\right)\right)\left(1 + \mathcal{O}\left(\frac{w_{\max}^2}{n}\right)\right) \\
&= 1 - \mathcal{O}\left(\frac{w_{\max}^4}{n}\right).
\end{aligned}$$

Theorem 5.1 now gives the result.

5.4 A Limitation of Our Technique

Finally, we give a limitation on the technique of proving hardness by constructing balanced pairwise distributions.

Theorem 5.10. *Let $P(x) = \mathrm{sgn}(w_1 x_1 + \cdots + w_n x_n)$ be a homogeneous linear threshold predicate. For any balanced pairwise independent distribution μ over $\{-1,1\}^n$,*

$$\Pr_{x\in(\{-1,1\}^n,\mu)}[P(x) = 1] \leq 1 - \frac{1}{4}\frac{w_{\max}^2}{\sum_{j=1}^n w_j^2},$$

where $w_{\max} = \max_j w_j$.

Proof. Let $X = \sum_{j=1}^{n} w_j X_j$ and let μ be a pairwise independent distribution over $\{-1,1\}^n$. Throughout the proof all expectations and probabilities are taken with respect to the distribution μ. Since μ is balanced and pairwise independent

$$\mathrm{Var}[X] = \mathbb{E}[X^2] = \sum_{j=1}^{n} w_j^2.$$

Now let $p = \Pr[X \le 0 | X_n = -1]$. By pairwise independence $\mathbb{E}[X|X_n = -1] = -w_n = -w_{\max}$. Hence,

$$\mathbb{E}[X|X \le 0, X_n = -1] \le -\frac{w_{\max}}{p} \quad \text{and} \quad \mathbb{E}[X^2|X \le 0, X_n = -1] \ge \frac{w_{\max}^2}{p^2}.$$

To summarize, we have that

$$\sum_{j=1}^{n} w_j^2 = \mathbb{E}[X^2] \ge \frac{p}{2}\frac{w_{\max}^2}{p^2} = \frac{w_{\max}^2}{2p}$$

and since $\Pr[X \le 0] \ge p/2$ the statement follows. □

By the above theorem, a homogeneous linear threshold precidate can only partially support any balanced pairwise independent distribution, and the gap shown by the theorem directly affects the degree of satisfiability that one can expect from the almost satisfiable instances in Theorem 5.5.

6 Conclusions

We have studied, and obtained rather tight bounds for the approximability curve of "majority-like" predicates. There are still many questions to be addressed and let us mention a few.

This work has been in the context of predicates given by Chow-robust threshold functions. Within this class we already knew, by the results of Hast [7], that no such predicate can be approximation resistant and our contribution is to obtain sharp bounds on the nature of how approximable these predicates are. It is a very nice open question whether there are any approximation resistant predicates given as thresholds of balanced linear functions. It is not easy to guess the answer to this question.

Looking at our results from a different angle one has to agree that the approximation algorithm we obtain is rather weak. For large values of n we only manage to do something useful on almost satisfiable instances and in this case we beat the random assignment by a rather slim margin. On the other hand we also prove that this is the best we can do. One could ask the question whether there is any other predicate that genuinely depends on n variables, accepts about half the inputs and which is easier to approximate than majority. It is not easy to guess what such a predicate would be but there is also very little information to support the guess that majority is the easiest predicate to approximate.

Using the results of Austrin and Mossel, Austrin and Håstad [1] proved that almost all predicates are approximation resistant. One way to interpret the results of this paper is that it indicates that the following statement might be true. For the few predicates of large arity where we can get some nontrivial approximation, we should not hope for too strong positive results.

References

[1] P. Austrin and J. Håstad. Randomly supported independence and resistance. In *38th Annual ACM Symposium on Theory of Computation*, pages 483–492, 2009.

[2] P. Austrin and E. Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009.

[3] C. K. Chow. On the characterization of threshold functions. In *Proceedings of the 2nd Annual Symposium on Switching Circuit Theory and Logical Design*, pages 34–38, 1961.

[4] M. Dertouzos. *Threshold logic: a synthesis approach.* MIT Press, Cambridge, MA, 1965.

[5] I. Diakonikolas and R. A. Servedio. Improved approximation of linear threshold functions. In *IEEE Conference on Computational Complexity*, pages 161–172, 2009.

[6] M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995.

[7] G. Hast. Beating a random assignment. In *Proceedings of APPROX-RANDOM*, pages 134–145, 2005.

[8] J. Håstad. Some optimal inapproximability results. *Journal of ACM*, 48:798–859, 2001.

[9] K. R. Kaplan and R. O. Winder. Chebyshev approximation and threshold functions. *IEEE Transactions on Electronic Computers*, EC-14(2):250–252, 1965.

[10] P. Kaszerman. A geometric test-synthesis procedure for a threshold device. *Information and Control*, 6(4):381–398, 1963.

[11] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of 34th ACM Symposium on Theory of Computating*, pages 767–775, 2002.

[12] S. Khot, E. Mossel G. Kindler, and R. O'Donnell. Optimal inapproximability results for Max-Cut and other 2-variable CSPs? In *Proceedings of 45th Annual IEEE Symposium of Foundations of Computer Science*, pages 146–154, 2004.

[13] R. O'Donnell and R. A. Servedio. The Chow parameters problem. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 517–526, 2008.

[14] R. O'Donnell and Y. Wu. An optimal SDP algorithm for Max-Cut and equally optimal long code tests. In *Proceedings of 40th ACM Symposium on Theory of Computating*, pages 335–344, 2008.

[15] T. Schaefer. The complexity of satisfiability problems. In *Conference record of the Tenth annual ACM Symposium on Theory of Computing*, pages 216–226, 1978.

[16] I. S. Shiganov. Refinement of the upper bound of a constant in the remainder term of the central limit theorem. *Journal of Soviet mathematics*, 35:109–115, 1986.

[17] R. O. Winder. Threshold logic in artificial intelligence. *Artificial Intelligence, IEEE Publication*, S-142:107–128, 1963.

[18] R. O. Winder. Threshold gate approximations based on Chow parameters. *IEEE Transactions on Computers*, 18(4):372–375, 1969.

[19] U. Zwick. Approximation algorithms for constraint satisfaction problems involving at most three variables per constraint. In *Proceedings of the ninth annual ACM-SIAM Symposium on Discrete Algorithms*, pages 201–210, 1998.

A The Unique Games Conjecture

Although we do not directly use the Unique Games Conjecture (UGC), we define it here for the sake of completeness. An instance of Unique Games $\mathcal{L} = (G(V, W, E), [L], \{\pi_{v,w}\}_{(v,w)})$ consists of a regular bipartite graph $G(V, W, E)$ and a set $[L]$ of labels. For each edge $(v, w) \in E$ there is a constraint specified by a permutation $\pi_{v,w} : [L] \mapsto [L]$. The goal is to find a labeling $\ell : (V \cup W) \mapsto [L]$ so as to maximize $val(\ell) := \Pr_{e \in E}[\ell \text{ satisfies } e]$, where a labeling ℓ is said to satisfy an edge $e = (v, w)$ if $\ell(v) = \pi_{v,w}(\ell(w))$. For a Unique Game instance $\mathcal{L}$, we let $OPT(\mathcal{L}) = \max_{\ell: V \cup W \mapsto [L]} val(\ell)$. The now famous UGC that has been extensively used to prove strong hardness of approximation results can be stated as follows.

Conjecture A.1 ([11]). *For any constants $\zeta, \gamma > 0$, there is a sufficiently large constant $L = L(\zeta, \gamma)$ such that, for Unique Game instances $\mathcal{L}$ with label set $[L]$, it is NP-hard to distinguish between $OPT(\mathcal{L}) \geq 1 - \zeta$ and $OPT(\mathcal{L}) \leq \gamma$.*

ECCC ISSN 1433-8092

http://eccc.hpi-web.de