Discrete Mathematics and Theoretical Computer Science DMTCS vol. **14:2**, 2012, 205–228

On Quadratic Threshold CSPs†

Per Austrin‡ Siavosh Benabbas§ Avner Magen

Department of Computer Science, University of Toronto, Canada

received 24th February 2012, revised 18th October 2012, accepted 2nd November 2012.

A predicate $P : \{-1, 1\}^k \to \{0, 1\}$ can be associated with a constraint satisfaction problem $\textsc{Max}\ \mathrm{CSP}(P)$. P is called "approximation resistant" if $\textsc{Max}\ \mathrm{CSP}(P)$ cannot be approximated better than the approximation obtained by choosing a random assignment, and "approximable" otherwise. This classification of predicates has proved to be an important and challenging open problem. Motivated by a recent result of Austrin and Mossel (Computational Complexity, 2009), we consider a natural subclass of predicates defined by signs of quadratic polynomials, including the special case of predicates defined by signs of linear forms, and supply algorithms to approximate them as follows.

In the quadratic case we prove that every *symmetric* predicate is approximable. We introduce a new rounding algorithm for the standard semidefinite programming relaxation of $\textsc{Max}\ \mathrm{CSP}(P)$ for any predicate $P : \{-1, 1\}^k \to \{0, 1\}$ and analyze its approximation ratio. Our rounding scheme operates by first manipulating the optimal SDP solution so that all the vectors are nearly perpendicular and then applying a form of hyperplane rounding to obtain an integral solution. The advantage of this method is that we are able to analyze the behaviour of a set of k rounded variables together as opposed to just a pair of rounded variables in most previous methods.

In the linear case we prove that a predicate called "Monarchy" is approximable. This predicate is not amenable to our algorithm for the quadratic case, nor to other LP/SDP-based approaches we are aware of.

Keywords: Combinatorial Optimization, Approximation Algorithms, Constraint Satisfaction Problems

1 Introduction

This paper studies the approximability of *constraint satisfaction problems* (CSPs). Given a predicate $P : \{-1, 1\}^k \to \{0, 1\}$, the $\textsc{Max}\ \mathrm{CSP}(P)$ problem is defined as follows. An instance is given by a list of k-tuples (clauses) of literals over some set of variables $x_1, \dots, x_n$, where a literal is either a variable or its negation. A clause is satisfied by an assignment to the variables if P is satisfied when applied to the value assigned to the k literals of the clause. The goal is then to find an assignment to the variables that maximizes the number of satisfied clauses. Our specific interest is predicates of the form $P(x) = \frac{1+\mathrm{sign}(Q(x))}{2}$ where $Q : \mathbb{R}^k \to \mathbb{R}$ is a quadratic polynomial with no constant term, i.e.,

†A preliminary version of this work appeared as [ABM10]

‡Email: `austrin@cs.toronto.edu`. Work done while the author was at KTH, Stockholm, supported by ERC advanced investigator grant 226203.

§Email: `siavosh@cs.toronto.edu`

$Q(x) = \sum_{i=1}^k a_i x_i + \sum_{i \neq j} b_{ij} x_i x_j$ for some set of coefficients $a_1, \ldots, a_n$ and $b_{11}, \ldots, b_{nn}$. While this special case is arguably very rich and interesting in its own right, we give some further motivations below. But first, we give some background to the study of MAX CSP(P) problems in general.

A canonical example of a MAX CSP(P) problem is when $P(x_1, x_2, x_3) = x_1 \vee x_2 \vee x_3$ is a disjunction of three variables, in which case MAX CSP(P) is the classic MAX 3-SAT problem. Another well-known example is the MAX 2-LIN(2) problem in which $P(x_1, x_2) = x_1 \oplus x_2$. As MAX CSP$(P)$ is NP-hard for almost all choices of P much effort has been put into understanding the best possible approximation ratio achievable in polynomial time.(i) A (randomized) algorithm is said to have approximation ratio $\alpha \leq 1$ if, given an instance with optimal value Opt, it produces an assignment with (expected) value at least $\alpha \cdot \mathrm{Opt}$.

The arguably simplest approximation algorithm is to pick a uniformly random assignment. As this algorithm satisfies each constraint with probability $\frac{|P^{-1}(1)|}{2^k}$ it follows that it gives an approximation ratio of $\frac{|P^{-1}(1)|}{2^k}$. In their classic paper [GW95], Goemans and Williamson used semidefinite programming to obtain improved approximation algorithms for predicates on two variables. For instance, for MAX 2-LIN(2) they gave an algorithm with approximation ratio $\alpha_{GW} \approx 0.878$. Following [GW95], many new approximation algorithms were found for various specific predicates, improving upon the random assignment algorithm. However, for some cases, perhaps most prominently the MAX 3-SAT problem, no such progress was made. Then, in another classic paper Håstad [Hås01] proved that MAX 3-SAT is in fact NP-hard to approximate within $7/8 + \epsilon$ (for any constant $\epsilon > 0$), showing that a random assignment in fact gives the best possible worst-case approximation that can be obtained in polynomial time.

Predicates which exhibit this behaviour are called *approximation resistant*. One of the main open questions along this line of research is to characterize which predicates admit a non-trivial approximation algorithm, and which predicates are approximation resistant. For predicates on three variables, the work of Håstad together with work of Zwick [Zwi98] shows that a predicate is resistant iff it is implied by an XOR of the three variables, or the negation thereof, where a predicate P is said to imply a predicate P' if $P(x) = 1 \Rightarrow P'(x) = 1$. For four variables, Hast [Has05] made an extensive classification leaving open the status of 46 out of 400 different predicates. It is worthwhile to note that in a celebrated result Raghavendra [Rag08] presents an approximation algorithm that (assuming the Unique Games Conjecture) achieves the best approximation factor for MAX CSP(P) for any predicate P; however, the approximation factor of this algorithm for any particular P is essentially impossible to understand and [Rag08] only proves that no other algorithm can do better (assuming the Unique Games Conjecture).(ii) Thus, this algorithm is not useful in determining which predicates are approximation resistant.

There have been several papers [ST00, EH08, ST09], mainly motivated by the soundness-query tradeoff for PCPs, giving increasingly general conditions under which predicates are approximation resistant. In a recent paper [AM09], the first author and Mossel proved that, if there exists an unbiased pairwise independent distribution on $\{-1, 1\}^k$ whose support is contained in $P^{-1}(1)$, then P is approximation resistant under the Unique Games Conjecture [Kho02]. This condition is very general and turned out to give many new cases of resistant predicates [AH11]. A related result by Benabbas et al. [BGMT12] that is independent of complexity assumptions, shows that under the same condition on P, the so-called Sherali-

(i) It follows from the dichotomy theorem of [Cre95] that (for any boolean predicate P) MAX CSP(P) is NP-hard unless P depends on at most 1 variable.

(ii) [Rag08] also presents a (exponential time) algorithm to compute the approximation factor for a specific predicate P up to any required precision.

Adams SDP hierarchy—which is a strong version of the Semidefinite Programming approach—does not beat a random assignment. Indeed, when it comes to algorithms, there are very few systematic results that give algorithms for large classes of predicates. One such result can be found in [Has05]. Given the result of [AM09], assuming the Unique Games Conjecture, such systematic results can only work for predicates that do not support pairwise independence. A very natural subclass of these predicates are those of the form $\frac{1+\mathrm{sign}(Q)}{2}$ for Q a quadratic polynomial as described above. To be more precise, the following fact from [AH11] is our main motivation for studying this type of predicates.

Fact 1. *A predicate P does not support pairwise independence if and only if there exists a quadratic polynomial $Q : \{-1,1\}^k \to \mathbb{R}$ with no constant term that is positive on all of $P^{-1}(1)$ (in other words, P implies a predicate of the form $\frac{1+\mathrm{sign}(Q)}{2}$).*

Given that the main tool for approximation algorithms—semidefinite programming—works by optimizing quadratic forms, it seemed natural and intuitive to hope that predicates of this form are always approximable. This however turns out to be false—in [AH12], a predicate is constructed that is the sign of a quadratic polynomial and still approximation resistant assuming the Unique Games Conjecture. Loosely speaking, the main crux is that semidefinite programming is good for optimizing the degree-2 part of the Fourier expansion of a predicate, which unfortunately can behave very differently from P itself or the quadratic polynomial used to define P (we elaborate on this below.) However, it turns out that when we restrict our attention to the special case of signs of symmetric polynomials (i.e., polynomials which are invariant under any permutation of their variables), this can not happen, and we can obtain an approximation algorithm, which is our first result.

Theorem 1. *Let $P : \{-1,1\}^k \to \{0,1\}$ be a predicate that is of the form $P(x) = \frac{1+\mathrm{sign}(Q(x))}{2}$ where Q is a symmetric quadratic polynomial with no constant term. Then P is not approximation resistant.*

A very natural special case of the signs of (not necessarily symmetric) quadratic polynomials is the case when $P(x)$ is simply the sign of a linear form, i.e., a linear threshold function. While we cannot prove that linear threshold predicates are approximable in general, we do believe this is the case, and make the following conjecture.

Conjecture 1. *Let $P : \{-1,1\}^k \to \{0,1\}$ be a predicate that is a sign of a linear form with no constant term. Then P is not approximation resistant.*

We view the resolution of this conjecture as a very natural and interesting open problem. As in the quadratic case, the difficulty stems from the fact that the low-degree part of P can be unrelated to the linear form used to define P. Specifically, it can be the case that the low-degree part of the arithmetization of P vanishes or becomes negative for some inputs where the linear/quadratic polynomial is positive (i.e. accepting inputs), and unfortunately this seems to make the standard SDP approach fail. The perhaps most extreme case of this phenomenon is exhibited by the predicate $\mathrm{Monarchy} : \{-1,1\}^k \to \{0,1\}$ suggested by Håstad [Hås09], in which the first variable (the "monarch") decides the outcome, unless all the other variables unite against it. In other words,

$$\mathrm{Monarchy}(x) = \frac{1 + \mathrm{sign}((k-2)x_1 + \sum_{i=2}^{k} x_i)}{2}.$$

Now, for the input $x_1 = -1$, $x_2 = \ldots = x_k = 1$, the linear part of the Fourier expansion of $\mathrm{Monarchy}$ takes value $-1 + o_k(1)$, whereas the linear form used to define monarchy is positive on this input, hence the value of the predicate is 1. Again, we stress that this means that known algorithms and techniques to obtain explicit bounds do not apply (while the results of Raghavendra [Rag08] in theory allow us to pin down the approximability to within any desired precision, they are not practical enough to let us determine whether a given predicate is approximation resistant[iii]). However, in this case we are still able to achieve an approximation algorithm, which is our second result.

Theorem 2. *The predicate* Monarchy *is* not *approximation resistant.*

This shows that there is some hope in overcoming the apparent barriers to proving Conjecture 1.

A recent related work of Cheraghchi et al. [CHIS12] also studies the approximability of predicates defined by linear forms. However, their focus is on establishing precise quantitative bounds on the approximability, rather than the more qualitative distinction between approximable and approximation resistant predicates. As such, their results only apply to predicates which were already known to be approximable. Specifically, they consider "Majority-like" predicates P where the linear part of the Fourier expansion of P behaves similarly to P itself (in the sense explained above).

Techniques: Our starting point in both our algorithms is the standard SDP relaxation of $\textsc{Max}\ \mathrm{CSP}(P)$. The main difficulty in rounding the solutions of these SDPs is that current rounding algorithms offer no analysis of the joint distribution of the outcome of the rounding for k variables, when $k > 2$. (Interestingly, when some modest value $k = 3$ is used, often some numerical methods are employed to complete the analysis [KZ97, Zwi98].) Unfortunately, such analysis seems essential to understanding the performance of the algorithm for $\textsc{Max}\ \mathrm{CSP}(P)$ as each constraint depends on k variables. Indeed, even a local argument would have to argue about the outcome of the rounding algorithm for k variables together.

For Theorem 1, we give a new and more direct proof of a theorem by Hast [Has05], giving a general condition on the low-degree part of the Fourier Expansion which guarantees a predicate is approximable (Theorem 4). We then show that this condition holds for predicates which are defined by symmetric quadratic polynomials. The basic idea behind our new algorithm is as follows. First, observe that the SDP solution in which all vectors are perpendicular is easy to analyze when the usual hyperplane rounding is employed, as in this case the obtained integral values are distributed uniformly. This motivates the following approach: start with the perpendicular configuration and then perturb the vectors in the direction of the optimal SDP solution. This perturbation acts as a differentiation operator, and as such allows for a "linear snapshot" of what is typically a complicated system. For each clause we analyze the probability that hyperplane rounding outputs a satisfying assignment, as a function of the inner products of vectors involved. Now, the object of interest is the gradient of this function at "zero". The hope is that since the optimal SDP solution (almost) satisfies this clause, it has a positive inner product with the gradient, and so can act as a global recipe that works for all clauses. It is important to stress that since we are only concerned with getting an approximation algorithm that works slightly better than random we can get away with this linear simplification. We show that this condition on the gradient translates into a condition on the low-degree part of the Fourier expansion of the predicate.

As it turns out, the predicate Monarchy which we tackle in Theorem 2 does not exhibit the aforementioned desirable property. In other words, the gradient above does not generally have a positive inner product with an optimal SDP solution. Instead, we show that when all vectors are sufficiently far from $\pm\mathbf{v}_0$ it is possible to get a similar guarantee on the gradient using high (but not too high) moments of the

[iii] In fact it is not even known whether, given a predicate P, it is decidable to check if P is approximation resistant.

vectors. We can then handle vectors which are very close to $\pm\mathbf{v}_0$ separately by rounding them deterministically to ± 1.

Organization: The rest of the paper is organized as follows. First, we introduce some definitions and preliminaries including the standard SDP relaxation of $\textsc{Max}\ \mathrm{CSP}(P)$ in Section 2. Then, in Section 3 we give our new algorithm for this SDP relaxation and characterize the predicates for which it gives an approximation ratio better than a random assignment. We then take a closer look at signs of symmetric quadratic forms in Section 4 and show that these satisfy the condition of the previous section, proving Theorem 1. In Section 5 we give the approximation algorithm for the $\mathrm{Monarchy}$ predicate and its somewhat tedious analysis. Finally, we give a discussion and some directions for future work in Section 6.

2 Preliminaries

In what follows $\mathbb{E}$ stands for expectation. For any positive integer n we use the notation $[n]$ for the set $\{1,\ldots,n\}$. For a finite set S (often a subset of $[n]$) we use the notation $\{-1,1\}^S$ for the set of all $-1,1$ vectors indexed by elements of S. For example, $|\{-1,1\}^S| = 2^{|S|}$. When $x \in \{-1,1\}^S$ and $y \in \{-1,1\}^{S'}$ are two vectors indexed by disjoint sets, i.e. $S \cap S' = \emptyset$, we use $x \circ y \in \{-1,1\}^{S \cup S'}$ to denote their natural concatenation.

We use φ and Φ for the probability density function and the cumulative distribution function of a standard normal random variable, respectively. We use the notation $\mathbb{S}^n$ for the $n+1$ dimensional sphere, i.e. , the set of unit vectors in $\mathbb{R}^{n+1}$.

Throughout the paper, we use $\mathrm{sign}(x)$ for the sign function defined as,

$$\mathrm{sign}(x) = \begin{cases} 1 & x > 0, \\ -1 & x \le 0. \end{cases}$$

Note that $\mathrm{sign}(0) = -1$.

2.1 Fourier Representation

Consider the set of real functions with domain $\{-1,1\}^k$ as a vector space. It is well known that the following set of functions called the characters form a complete basis for this space, $\chi_S(x) \stackrel{\text{def}}{=} \prod_{i\in S} x_i$. In fact if we define inner products of functions as $f \cdot g \stackrel{\text{def}}{=} \mathbb{E}_x\left[f(x)g(x)\right]$ this basis will be orthonormal and every function will have a unique *Fourier expansion* when written in this basis,

$$f = \sum_{S \subseteq [k]} \widehat{f}(S)\chi_S, \qquad\qquad \widehat{f}(S) \stackrel{\text{def}}{=} f \cdot \chi_S.$$

The values $\left\{\widehat{f}(S)\right\}_{S\subseteq[n]}$ are often called the *Fourier coefficients* of f. We write $f^{=d}$ for the part of the function that is of degree d, i.e.,

$$f^{=d}(x) = \sum_{|S|=d} \widehat{f}(\{S\})\chi_S(x).$$

Fig. 1: Standard SDP relaxation of MAX CSP(P).

$$
\begin{array}{lcll}
\text{Maximize} & & \displaystyle\frac{1}{m}\sum_{i=1}^{m}\sum_{\omega\in\{-1,1\}^{T_i}} C_i(\omega)\ I_{T_i,\omega} & \\
\text{Where,} & \forall S\subset[n], |S|\le k, \omega\in\{-1,1\}^S & I_{S,\omega}\in[0,1] & \\
 & \forall i\in[n] & \mathbf{v}_i\in S^n & \\
 & & \mathbf{v}_0\in S^n & \\
\text{Subject to} & \forall S\subset[n], |S|\le k & \displaystyle\sum_{\omega\in\{-1,1\}^S} I_{S,\omega}=1 & (1)\\
 & \forall S\subset S'\subset[n], |S'|\le k, \omega\in\{-1,1\}^S & \displaystyle\sum_{\omega'\in\{-1,1\}^{S'\setminus S}} I_{S',\omega'\circ\omega}=I_{S,\omega} & (2)\\
 & \forall i\in[n] & I_{\{i\},(1)}-I_{\{i\},(-1)}=\mathbf{v}_0\cdot\mathbf{v}_i & (3)\\
 & \forall i,j\in[n] & I_{\{i,j\},(1,1)}-I_{\{i,j\},(-1,1)} & \\
 & & -I_{\{i,j\},(1,-1)}+I_{\{i,j\},(-1,-1)}=\mathbf{v}_i\cdot\mathbf{v}_j & (4)
\end{array}
$$

It is easy to see that whenever $f:\{-1,1\}^k\to\mathbb{R}$ is odd (i.e. $f(x)=-f(-x)$) and $S\subseteq[k]$ is an even sized set $\widehat{f}(S)=0$:

$$\widehat{f}(S)=\mathop{\mathbb{E}}_x[f(x)\chi_S(x)]=-\mathop{\mathbb{E}}_x[f(-x)\chi_S(x)]=-\mathop{\mathbb{E}}_x[f(-x)\chi_S(-x)]=-\mathop{\mathbb{E}}_y[f(y)\chi_S(y)]=-\widehat{f}(S),$$

where we have used the fact that χ_S is an even function for even $|S|$.

2.2 Semidefinite Relaxation

For any fixed P, MAX CSP(P) has a natural SDP relaxation that can be seen in Figure 1. The essence of this relaxation is that each $I_{S,*}$ is a distribution, often called a *local distribution*, over all possible assignments to the variables in set S as enforced by (1). Whenever, S_1 and S_2 intersect (2) guarantees that their marginal distributions on the intersection agree. Also, (3) and (4) ensure that $\mathbf{v}_0\cdot\mathbf{v}_i$ and $\mathbf{v}_i\cdot\mathbf{v}_j$ are equal to the bias of variable x_i and the correlation of the variables x_i and x_j in the local distributions respectively. The clauses of the instance are $C_1,\ldots,C_m$, with C_i being an application of P (possibly with some variables negated) on the set of variables T_i. The objective function is the fraction of the clauses that are satisfied.

Observe that the reason this SDP is not an exact formulation but a relaxation is that these distributions are defined only on sets of size up to k. It is worth mentioning that this program is weaker than the kth round of the Lasserre hierarchy for this problem while stronger than the kth round of the Sherali-Adams hierarchy. From here on the only things we use in the rounding algorithms are the vectors $\mathbf{v}_0,\ldots,\mathbf{v}_n$ and the existence of the local distributions.

3 (ϵ, η)-hyperplane rounding

In this section we define a rounding scheme for the semidefinite program of MAX CSP(P) and proceed to analyze its performance. The rounding scheme is based on the usual hyperplane rounding but is more flexible in that it uses two parameters $\epsilon \geq 0$ and η where ϵ is a sufficiently small constant and η is an arbitrary real number. We will then formalize a (sufficient) condition involving P and η under which our approximation algorithm has approximation factor better than that of a random assignment. In the next section we show that this condition is satisfied (for some η) by signs of symmetric quadratic polynomials.

Given an instance of MAX CSP(P), our algorithm first solves the standard SDP relaxation of the problem (Figure 1.) It then employs the rounding scheme outlined in Figure 2 to get an integral solution.

Fig. 2: (ϵ, η)-Hyperplane Rounding

INPUT: $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{S}^n$.

OUTPUT: $x_1, \ldots, x_n \in \{-1, 1\}$.

1. Define unit vectors $\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_n \in \mathbb{S}^n$ such that for all $0 \leq i < j$,
$$\mathbf{w}_i \cdot \mathbf{w}_j = \epsilon(\mathbf{v}_i \cdot \mathbf{v}_j),$$
2. Let $\mathbf{g} \in \mathbb{R}^{n+1}$ be a random $(n+1)$-dimensional Gaussian.
3. Assign each x_i as,
$$x_i = \begin{cases} 1 & \text{if } \mathbf{w}_i \cdot \mathbf{g} > -\eta(\mathbf{w}_0 \cdot \mathbf{w}_i), \\ -1 & \text{otherwise.} \end{cases}$$

Note that when $\epsilon = 0$ the rounding scheme above simplifies to assigning all x_i's uniformly and independently at random which satisfies $\frac{|P^{-1}(1)|}{2^k}$ fraction of all clauses in expectation. For non-zero ϵ, η will determine how much weight is given to the position of $\mathbf{v}_0$ compared to the correlation of the variables.

Notice that in the pursuit of a rounding algorithm that has approximation ratio better than $\frac{|P^{-1}(1)|}{2^k}$ it is possible to assume that the optimal integral solution is arbitrary close to 1 as otherwise random assignment already delivers an approximation factor better than $\frac{|P^{-1}(1)|}{2^k}$. In particular, the optimal vector solution can be assumed to be that good. This observation is in fact essential to our analysis. Let us for the sake of simplicity first consider the case where the value of the vector solution is precisely 1. Fix a clause, say $P(x_1, x_2, \ldots, x_k)$. (In general, without loss of generality we can assume that the current clause is on k *variables* as opposed to k *literals*. This is simply because one can assume that $\neg x_i$ is a separate variable from x_i with SDP vector $-\mathbf{v}_i$.) Since the SDP value is 1, every clause (and this clause in particular) is completely satisfied by the SDP, hence the local distribution $I_{[k],*}$ is supported on the set of satisfying assignments of P. The hope now is that when ϵ increases from zero to some small positive value this distribution helps to boost the probability of satisfying the clause (a little) beyond $\frac{|P^{-1}(1)|}{2^k}$. This becomes a question of differentiation. Specifically, consider the probability of satisfying the clause at hand as a function of ϵ. We want to show that for some $\epsilon > 0$ the value of this function is bigger than its value at zero. This is closely related to the derivative of the function at zero and in fact the first step is to analyze this derivative. The following Theorem relates the value of this function and its derivative at zero to the predicate P and the vectors $\mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_k$.

Theorem 3. *For any fixed η, the probability that $P(x_1, \ldots, x_k)$ is satisfied by the assignment behaves as follows at $\epsilon = 0$:*

$$\Pr\left[(x_1, \ldots, x_k) \in P^{-1}(1)\right] = \frac{|P^{-1}(1)|}{2^k}$$

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} \Pr\left[(x_1, \ldots, x_k) \in P^{-1}(1)\right] = \frac{2\eta}{\sqrt{2\pi}} \sum_{i=1}^{k} \widehat{P}(\{i\})\mathbf{v}_0 \cdot \mathbf{v}_i + \frac{2}{\pi} \sum_{i<j} \widehat{P}(\{i,j\})\mathbf{v}_i \cdot \mathbf{v}_j. \tag{5}$$

Proof: The first claim follows from the fact that for $\epsilon = 0$, $x_1, \ldots, x_k$ are uniform and independent. To prove (5) we first introduce some notation. For an assignment $\omega \in \{-1, 1\}^k$ define the function p_ω on $(k+1) \times (k+1)$ semidefinite matrices as follows. For a positive semidefinite matrix $A_{(k+1)\times(k+1)}$, consider a set of vectors $\mathbf{w}'_0, \ldots, \mathbf{w}'_k$ whose Gram matrix is A, and consider running steps 2 and 3 of the rounding procedure on $\mathbf{w}'_i$'s instead of $\mathbf{w}_i$'s. Define $p_\omega(A)$ as the probability that the rounding procedure outputs ω. In what follows, we will use $A^* = A^*(\epsilon)$ to denote the Gram matrix of the set of vectors $\mathbf{w}_1, \ldots, \mathbf{w}_n$ used by the algorithm (which depends on ϵ). Clearly, we have that

$$\Pr\left[(x_1, \ldots, x_k) \in P^{-1}(1)\right] = \sum_{\omega \in P^{-1}(1)} \Pr\left[(x_1, \ldots, x_k) = \omega\right] = \sum_{\omega \in P^{-1}(1)} p_\omega(A^*).$$

We start by computing $\frac{\mathrm{d}}{\mathrm{d}\epsilon} p_\omega(A^*)$ using the chain rule.

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} p_\omega(A^*(\epsilon)) = \sum_{i \le j} \left(\left.\frac{\mathrm{d}}{\mathrm{d}\epsilon} a^*_{ij}\right|_{\epsilon=0} \cdot \left.\frac{\partial}{\partial a_{ij}} p_\omega(A)\right|_{A=I} \right) = \sum_{i<j} \mathbf{v}_i \cdot \mathbf{v}_j \left.\frac{\partial}{\partial a_{ij}} p_\omega(A)\right|_{A=I}, \tag{6}$$

where when we talk about $\frac{\partial}{\partial a_{ij}} p_\omega(A)$ we consider A a *symmetric* positive semidefinite matrix so a_{ji} changes with a_{ij}. Now to compute $\left.\frac{\partial}{\partial a_{ij}} p_\omega(A)\right|_{A=I}$ we compute a formula for $p_\omega(A)$ where A is equal to I in every entry except the ij and ji entries where it is a_{ij}. Define J_{ij} as the matrix which is zero on every coordinate except coordinates ij and ji where it is 1. Then,

$$\left.\frac{\partial}{\partial a_{ij}} p_\omega(A)\right|_{A=I} = \left.\frac{\mathrm{d}}{\mathrm{d}t} p_\omega(I + tJ_{ij})\right|_{t=0}.$$

Note that for $t \in [-1, 1]$, $I+tJ$ is positive semidefinite, so $p_\omega(I+tJ)$ is well-defined. Now observe that the geometric realization of $\mathbf{w}'_0, \ldots, \mathbf{w}'_k$ defining $p_\omega(I + tJ_{ij})$ is simple; in particular, all the vectors are perpendicular except the pair $\mathbf{w}'_i$ and $\mathbf{w}'_j$, so $p_\omega(I + tJ_{ij})$ can be readily computed using a case analysis, with two cases depending on whether i or j equals zero, as follows.

First, if $i = 0$, the value of all variables are going to be assigned independently, and all but x_j are going to be assigned values uniformly. It is easy to see that for any $j \in [k]$ and any $\omega \in \{-1,1\}^k$,

$$\begin{aligned}
p_\omega(I + tJ_{0j}) &= 2^{-(k-1)} \cdot \begin{cases} \Pr\left[\mathbf{w}'_j \cdot g \geq -\eta t\right] & \text{if } \omega_j = 1 \\ \Pr\left[\mathbf{w}'_j \cdot g \leq -\eta t\right] & \text{if } \omega_j = -1 \end{cases} \\
&= 2^{-(k-1)} \cdot \begin{cases} 1 - \Phi(-\eta t) & \text{if } \omega_j = 1 \\ \Phi(-\eta t) & \text{if } \omega_j = -1 \end{cases} \\
&= 2^{-(k-1)} \cdot \left(\frac{1+\omega_j}{2} - \omega_j \Phi(-\eta t)\right).
\end{aligned}$$

Differentiating, we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} p_\omega(I + tJ_{0j}) &= -2^{-(k-1)} \omega_j \frac{\mathrm{d}}{\mathrm{d}t} \Phi(-\eta t) = 2^{-(k-1)} \eta \omega_j \varphi(-\eta t) \\
&= \frac{2^{-(k-1)} \eta \omega_j e^{-\eta^2 t^2/2}}{\sqrt{2\pi}},
\end{aligned}$$

from which we get the identity

$$\left.\frac{\partial}{\partial a_{0j}} p_\omega(A)\right|_{A=I} = \frac{2^{-(k-1)} \eta \omega_j}{\sqrt{2\pi}}. \tag{7}$$

Let us then consider the case where both i and j are non-zero. In this case, each variable is going to be assigned a value in $\{-1,1\}$ uniformly at random, and all these assignments are independent except the assignments of the ith and the jth variable. So we can imagine that x_l for all $l \notin \{i,j\}$ are assigned a uniformly independent value in $\{-1,1\}$ and then x_i and x_j are assigned random values depending only on g projected to the linear subspace spanned by $\mathbf{w}'_i$ and $\mathbf{w}'_j$; we will use $\tilde{\mathbf{g}}$ for this projection. The joint distribution of x_i and x_j is then not hard to understand: x_i is uniformly random and $x_j \neq x_i$ if and only $\tilde{\mathbf{g}}$ lies in one of two segment of the unit circle (in this linear subspace) shown in Figure 3. We note that this analysis is identical to the one used by [GW95]. We have,

$$\begin{aligned}
p_\omega(I + tJ_{ij}) &= 2^{-(k-1)} \cdot \begin{cases} \frac{1}{\pi} \arccos t & \text{if } \omega_i \neq \omega_j \\ 1 - \frac{1}{\pi} \arccos t & \text{if } \omega_i = \omega_j \end{cases} \\
&= 2^{-(k-1)} \left(\frac{1+\omega_i\omega_j}{2} - \frac{1}{\pi} \omega_i \omega_j \arccos t\right)
\end{aligned}$$

Differentiating this expression, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} p_\omega(I + tJ_{ij}) = -\frac{2^{-(k-1)} \omega_i \omega_j}{\pi} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \arccos t = \frac{2^{-(k-1)} \omega_i \omega_j}{\pi \sqrt{1-t^2}},$$

and we can conclude that

$$\left.\frac{\partial}{\partial a_{ij}} p_\omega(A)\right|_{A=I} = 2^{-(k-1)} \omega_i \omega_j / \pi. \tag{8}$$

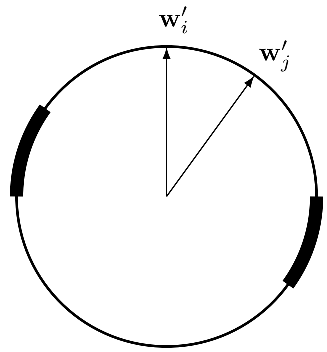

Fig. 3: The joint distribution of x_i and x_j: If $\tilde{\mathbf{g}}/||\tilde{\mathbf{g}}||_2$ is in the thickly drawn arcs then $x_i \neq x_j$.

Now combining (6) with (7) and (8) we get,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Pr\left[(x_1,\ldots,x_k)\in P^{-1}(1)\right] &= \sum_{\omega\in P^{-1}(1)}\sum_{i<j}\mathbf{v}_i\cdot\mathbf{v}_j\,\frac{\partial}{\partial a_{ij}}p_\omega(A)\Big|_{A=I}\\
&= \sum_{1\le i\le k}\mathbf{v}_0\cdot\mathbf{v}_i\frac{2^{-(k-1)}\eta}{\sqrt{2\pi}}\sum_{\omega\in P^{-1}(1)}\omega_i\\
&\quad+\sum_{1\le i<j\le k}\mathbf{v}_i\cdot\mathbf{v}_j\frac{2^{-(k-1)}}{\pi}\sum_{\omega\in P^{-1}(1)}\omega_i\omega_j\\
&= \sum_{1\le i\le k}\mathbf{v}_0\cdot\mathbf{v}_i\frac{2\eta}{\sqrt{2\pi}}\mathop{\mathbb{E}}_{\omega}\left[\omega_i P(\omega)\right]+\sum_{1\le i<j\le k}\mathbf{v}_i\cdot\mathbf{v}_j\frac{2}{\pi}\mathop{\mathbb{E}}_{\omega}\left[\omega_i\omega_j P(\omega)\right]\\
&= \frac{2\eta}{\sqrt{2\pi}}\sum_{i=1}^{k}\widehat{P}(\{i\})\mathbf{v}_0\cdot\mathbf{v}_i+\frac{2}{\pi}\sum_{i<j}\widehat{P}(\{i,j\})\mathbf{v}_i\cdot\mathbf{v}_j.
\end{aligned}$$

Which completes the proof. $\square$

Now, the inner products $\mathbf{v}_i\cdot\mathbf{v}_j$ are equal to the moments of the local distributions $I_{\{i,j\},*}$, which in turn agree with those of the local distribution $I_{[k],*}$. It follows that,

$$\frac{2\eta}{\sqrt{2\pi}}\sum_{i=1}^{k}\widehat{P}(\{i\})\mathbf{v}_0\cdot\mathbf{v}_i+\frac{2}{\pi}\sum_{i<j}\widehat{P}(\{i,j\})\mathbf{v}_i\cdot\mathbf{v}_j=\mathop{\mathbb{E}}_{\omega\sim I_{[k],*}}\left[\frac{2\eta}{\sqrt{2\pi}}P^{=1}(\omega)+\frac{2}{\pi}P^{=2}(\omega)\right]. \quad (9)$$

Thus, in order for the derivative in (5) to be positive for all possible values of the $\mathbf{v}_i$'s (that have SDP objective value 1), it is necessary and sufficient that $\frac{2\eta}{\sqrt{2\pi}}P^{=1}(\omega)+\frac{2}{\pi}P^{=2}(\omega)$ is positive for every $\omega\in P^{-1}(1)$. This leads us to the following Theorem formulating a condition under which our rounding algorithm works.

Theorem 4. *Suppose that there exists an $\eta \in \mathbb{R}$ such that*

$$\frac{2\eta}{\sqrt{2\pi}} P^{=1}(\omega) + \frac{2}{\pi} P^{=2}(\omega) > 0 \tag{10}$$

for every $\omega \in P^{-1}(1)$. Then P is approximable.

As mentioned in the Techniques section, this theorem is not new. It was previously found by Hast [Has05]. However, his algorithm and analysis are completely different from ours (using different algorithms to optimize the linear and quadratic parts of the predicate, and case analysis depending on the behaviour of the integral solution). We believe that our algorithm is simpler and considerably more direct.

The general strategy for the proof, which can be found below, is as follows. We will concentrate on a clause that is almost satisfied by the SDP solution. By Equation 10 and Theorem 3 the first derivative of the probability that this clause is satisfied by the rounded solution is at least some positive global constant (say δ) at $\epsilon = 0$. We will then show that for small enough ϵ the second derivative of this probability is bounded in absolute value by, say, Γ at any point in $[0, \epsilon]$. Now we can apply Taylor's theorem to show that if ϵ is small enough the probability of success is at least $\frac{|P^{-1}(1)|}{2^k} + \delta\epsilon - \Gamma\epsilon^2/2$ which for $\epsilon = \delta/\Gamma$ is at least $\frac{|P^{-1}(1)|}{2^k} + \delta^2/2\Gamma$.

Proof of Theorem 4: Consider the optimal vector solution $\mathbf{v}_1, \ldots, \mathbf{v}_n \in S^n$. Note that the optimal *integral* solution will have objective value no more than that of $\mathbf{v}_0, \ldots, \mathbf{v}_n$. So if we fix a constant δ_2, we can always assume that the vector solution achieves objective value at least $1 - \delta_2$. Otherwise, a random assignment to the variables will achieve an objective value of $\frac{|P^{-1}(1)|}{2^k}$ and approximation factor $\frac{|P^{-1}(1)|}{2^k}/(1-\delta_2) > \frac{|P^{-1}(1)|}{2^k}(1+\delta_2)$. So, in this case even a random assignment shows that the predicate is not approximation resistant. From here on we assume that the vector solution achieves an objective value of $1 - \delta_2$, where δ_2 is some constant to be set later. Now, applying a simple Markov type inequality one can see that at least a $1 - \sqrt{\delta_2}$ fraction of the clauses must have SDP value at least $1 - \sqrt{\delta_2}$. Consider one such clause, say $P(x_1, \ldots, x_k)$. We will show that the probability that this clause is satisfied by the rounded solution is slightly more than $\frac{|P^{-1}(1)|}{2^k}$.

Let $\delta_1 > 0$ be the minimum value of the left hand side of (10) over all $\omega \in P^{-1}(1)$, and let s denote its minimum over all $\omega \in \{-1, 1\}^k$. By Theorem 3 and (9), we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\epsilon} \Pr\left[(x_1, \ldots, x_k) \in P^{-1}(1)\right] &= \frac{2\eta}{\sqrt{2\pi}} \sum_{j>0} \widehat{P}(\{j\}) \mathbf{v}_j \cdot \mathbf{v}_0 + \frac{2}{\pi} \sum_{0<i<j} \widehat{P}(\{i,j\}) \mathbf{v}_i \cdot \mathbf{v}_j \\
&= \mathop{\mathbb{E}}_{\omega \sim I_{[k],*}} \left[\frac{2\eta}{\sqrt{2\pi}} P^{=1}(\omega) + \frac{2}{\pi} P^{=2}(\omega)\right] \\
&\geq (1 - \sqrt{\delta_2})\delta_1 + \sqrt{\delta_2} s \geq \delta_1/2
\end{aligned}$$

where the last step holds provided δ_2 is sufficiently small[(iv)] compared to δ_1 and s. This shows that the first derivative (at $\epsilon = 0$) of the probability that P is satisfied by the rounded solution is bounded from below by the constant $\delta_1/2$.

(iv) In particular, if $\delta_2 \leq \delta_1^2/4(\delta_1 - \min(s, 0))^2$

All that remains is to show that the second derivative of this probability can not be too large in absolute value. We will need the following lemma about the second derivative of the orthant probability of normal random variables.

Lemma 5. *For fixed k, define the function* $\mathrm{ort}(\nu, \Sigma)$ *as the* orthant probability *of the multivariate normal distribution with mean ν and covariance matrix Σ, where $\nu \in \mathbb{R}^k$ and $\Sigma_{k\times k}$ is a positive semidefinite matrix. That is,*

$$\mathrm{ort}(\nu, \Sigma) \stackrel{\text{def}}{=} \Pr_{\mathbf{x}\sim N(\nu,\Sigma)} [\mathbf{x} \geq 0] .$$

There exists a global constant Γ that upper bounds all the second partial derivatives of $\mathrm{ort}()$ *when Σ is close to I. In particular, for all k, there exist $\kappa > 0$ and Γ, such that for all $i_1, j_1, i_2, j_2 \in [k]$, all vectors $\nu \in \mathbb{R}^k$ and all positive definite matrices $\Sigma_{k\times k}$ satisfying*

$$|I - \Sigma|_\infty, |\nu|_\infty < \kappa,$$

we have,

$$\left|\frac{\partial^2}{\partial\Sigma_{i_1 j_1}\partial\Sigma_{i_2 j_2}} \mathrm{ort}(\nu, \Sigma)\right| < \Gamma,$$

$$\left|\frac{\partial^2}{\partial\Sigma_{i_1 j_1}\partial\nu_{i_2}} \mathrm{ort}(\nu, \Sigma)\right| < \Gamma,$$

$$\left|\frac{\partial^2}{\partial\nu_{i_1}\partial\nu_{i_2}} \mathrm{ort}(\nu, \Sigma)\right| < \Gamma.$$

The proof of this lemma is rather technical, but the general outline is as follows. First we write down the orthant probability as an integral of the probability density function over the positive orthant. Then we observe that each of the inner integrals as well as the probability density function and their partial derivatives are continuous, so we can apply Leibniz's integral rule iteratively to move the differentiation under the integral. We then differentiate the probability density function and the result will be in the form of the expectation of a degree 2 expression in x_i's under the same distribution. We can then bound these expression in terms of the means and correlations of the variables. For the interested reader a full proof is presented in an appendix.

Now, similar to the proof of Theorem 3 we can write,

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} \Pr\left[(x_1, \ldots, x_k) \in P^{-1}(1)\right] &= \frac{\mathrm{d}}{\mathrm{d}\epsilon} \sum_{\omega\in P^{-1}(1)} \sum_{0\leq i<j\leq k} \mathbf{v}_i \cdot \mathbf{v}_j \frac{\partial}{\partial a_{ij}} p_\omega(A) \\ &= \sum_{\omega\in P^{-1}(1)} \sum_{0\leq i<j\leq k} \mathbf{v}_i \cdot \mathbf{v}_j \sum_{0\leq i'<j'\leq k} \mathbf{v}_{i'} \cdot \mathbf{v}_{j'} \frac{\partial^2}{\partial a_{ij}\partial a_{i'j'}} p_\omega(A).\end{aligned} \tag{11}$$

One can think of $\mathbf{g} \cdot \mathbf{w}_1 + \eta\mathbf{w}_0 \cdot \mathbf{w}_1, \mathbf{g} \cdot \mathbf{w}_1 + \eta\mathbf{w}_0 \cdot \mathbf{w}_2, \ldots$ as a set of joint Gaussian random variables. In particular for a fixed ω define $\nu \in \mathbb{R}^n$ and a positive definite matrix $\Sigma_{k\times k}$ as,

$$\begin{aligned} &\forall 1 \leq i \leq k && \nu_i = \eta A_{0i}\omega_i = \epsilon\, \eta\omega_i \mathbf{v}_0 \cdot \mathbf{v}_i \\ &\forall 1 \leq i \leq k && \Sigma_{ii} = 1 \\ &\forall 1 \leq i < j \leq k && \Sigma_{ij} = \Sigma_{ji} = A_{ij}\omega_i\omega_j = \epsilon\, \omega_i\omega_j \mathbf{v}_i \cdot \mathbf{v}_j \end{aligned}$$

It is easy to verify that $p_\omega(A)$ is indeed the orthant probability of Gaussian distribution with mean ν and correlation matrix Σ. So according to Lemma 5 and (11) for $\epsilon \le \min(\kappa, \kappa/|\eta|)$,

$$\left| \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} \Pr\left[(x_1, \dots, x_k) \in P^{-1}(1)\right] \right| \le 2^k k^4 \Gamma,$$

where $2^k k^4$ is a bound on the number of terms in (11) and κ and Γ are constants only depending on k. Now for every such ϵ_0 according to Taylor's theorem for some $0 \le \epsilon' \le \epsilon_0$,

$$\begin{aligned} \Pr\left[(x_1, \dots, x_k) \in P^{-1}(1)\right]\Big|_{\epsilon=\epsilon_0} &= \frac{|P^{-1}(1)|}{2^k} + \epsilon_0 \frac{\mathrm{d}}{\mathrm{d}\epsilon} \Pr\left[(x_1, \dots, x_k) \in P^{-1}(1)\right]\Big|_{\epsilon=0} \\ &\quad + \frac{\epsilon_0^2}{2} \frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2} \Pr\left[(x_1, \dots, x_k) \in P^{-1}(1)\right]\Big|_{\epsilon=\epsilon'} \\ &\ge \frac{|P^{-1}(1)|}{2^k} + \epsilon_0 \delta_1/2 - \epsilon_0^2 2^k k^4 \Gamma/2. \end{aligned}$$

Setting ϵ_0 appropriately[(v)], this is at least $\frac{|P^{-1}(1)|}{2^k} + \delta_3$ for $\delta_3 = \epsilon_0\delta_1/4$ (which crucially does *not* depend on δ_2). This shows that each clause for which the vector solution gets a value of $(1 - \sqrt{\delta_2})$ is going to be satisfied by the rounded solution with probability at least $\frac{|P^{-1}(1)|}{2^k} + \delta_3$. As these constitute a $1 - \sqrt{\delta_2}$ fraction of all clauses, the overall expected value of the objective function for the rounded solution is at least

$$\left(1 - \sqrt{\delta_2}\right)\left(\frac{|P^{-1}(1)|}{2^k} + \delta_3\right) \ge \frac{|P^{-1}(1)|}{2^k} + \delta_3 - \sqrt{\delta_2}.$$

If we set $\delta_2 < (\delta_3/2)^2$, this is at least $\frac{|P^{-1}(1)|}{2^k} + \delta_3/2$, which provides a lower bound on the approximation ratio of the algorithm on instances with optimal value at least $1 - \delta_2$. This completes the proof. $\square$

4 Signs of Symmetric Quadratic Polynomials

In this section we study signs of symmetric quadratic polynomials, and give a proof of Theorem 1. Consider a predicate $P : \{-1, 1\}^k \to \{0, 1\}$ that is the sign of a symmetric quadratic polynomial with no constant term, i.e.,

$$P(x) = \frac{1 + \mathrm{sign}\left(\alpha \sum_i x_i + \beta \sum_{i<j} x_i x_j\right)}{2}$$

for some constants α and β. We would like to apply the (ϵ, η)-rounding scheme to MAX CSP(P), which in turn requires us to understand the low-degree Fourier coefficients of P. Note that because of symmetry, the value of a Fourier coefficient $\widehat{P}(S)$ depends only on $|S|$.

We will prove that "morally" β has the same sign as the degree-2 Fourier coefficient of P and that if one of them is 0 then so is the other. This statement is not quite true (consider for instance the predicate $P(x_1, x_2) = \frac{1+\mathrm{sign}(x_1+x_2)}{2} = \frac{1+x_1+x_2+x_1x_2}{4}$), however it is always true that by slightly adjusting β (without changing P), we can assure that this is the case.

(v) $\epsilon_0 = \min(\kappa, \kappa/|\eta|, 2^{-k}k^{-2}\delta_1/2\Gamma)$ will do

Theorem 6. *For any P of the above form, there exists β' with the property that $\beta' \cdot \widehat{P}(\{1,2\}) \geq 0$ and $\beta' = 0$ iff $\widehat{P}(\{1,2\}) = 0$, satisfying*

$$P(x) = \frac{1 + \operatorname{sign}\left(\alpha \sum x_i + \beta' \sum x_i x_j\right)}{2}.$$

Proof: Let us define

$$P_\beta(x) = \frac{1 + \operatorname{sign}(\alpha \sum x_i + \beta \sum x_i x_j)}{2}$$

where we consider α fixed and β as a variable. First, we have the following claim:

Claim 1. *$\widehat{P_\beta}(\{1,2\})$ is a monotonically non-decreasing function in β. Furthermore, if $P_{\beta_1} \neq P_{\beta_2}$ then $\widehat{P_{\beta_1}}(\{1,2\}) \neq \widehat{P_{\beta_2}}(\{1,2\})$.*

Proof: Fix two arbitrary values $\beta_1 < \beta_2$ of β, and let $\Delta P : \{-1,1\}^k \to \{-1,0,1\}$ be the difference $\Delta P = P_{\beta_2} - P_{\beta_1}$. Consider an input $x \in \{-1,1\}^k$. It follows from the definition of P_β that if $\Delta P(x) > 0$ then $\sum_{i<j} x_i x_j > 0$, and similarly if $\Delta P(x) < 0$ then $\sum_{i<j} x_i x_j < 0$. Now since ΔP is symmetric, the level-2 Fourier coefficient of ΔP equals ,

$$\widehat{\Delta P}(\{1,2\}) = \frac{1}{\binom{n}{2}} \sum_{i<j} \widehat{\Delta P}(\{i,j\}) = \frac{1}{\binom{n}{2}} \mathop{\mathbb{E}}_x \left[\Delta P(x) \sum_{i<j} x_i x_j \right] \geq 0,$$

with equality holding only if ΔP is zero everywhere, i.e., if $P_{\beta_1} = P_{\beta_2}$. This completes the proof of the Claim. □

Suppose first that either $\alpha = 0$ or k is odd. It is easy to check that in these two cases, $\widehat{P_0}(\{1,2\}) = 0$ (if $\alpha = 0$ the function P_0 is constant, and if $\alpha \neq 0$ but k is odd the function P_0 is odd so its Fourier coefficients of size 2 are zero). Consider the set of values $B = \{\, \beta \mid \widehat{P_\beta}(\{1,2\}) = 0 \,\}$. From Claim 1 it follows that B is an interval (though it is possible for this interval to consist of a single point) and that the function P_β is the same for all $\beta \in B$. For $\beta < 0$, $\beta \notin B$, Claim 1 shows that $\widehat{P_\beta}(\{1,2\}) < 0$ and so has the same sign as β, and similarly for $\beta > 0$, $\beta \notin B$. For $\beta \in B$ we see that $P_\beta = P_0$ and thus we can set $\beta' = 0$.

The remaining case is that of even k, and $\alpha \neq 0$. Notice that if $|\beta|$ is sufficiently small compared to $|\alpha|$, say, $|\beta| \leq |\alpha|/k^2$, then $P_\beta(x)$ only differs from $P_0(x)$ on balanced inputs (i.e., having $\sum x_i = 0$). Let B be the set of all such sufficiently small β's. For $\beta \in B$, the only contribution to $\widehat{P_\beta}(\{1,2\})$ comes from points x that are balanced (i.e., having $\sum x_i = 0$). The reason is that for all other x, the contribution $\operatorname{sign}(\alpha \sum x_i) x_1 x_2$ to $\widehat{P_0}(\{1,2\})$ is cancelled by the contribution from the point $-x$.

For balanced points, we have $\sum_{i<j} x_i x_j = 2\binom{n/2}{2} - (n/2)^2 = -\frac{n}{2} < 0$ and therefore $\operatorname{sign}(\alpha \sum x_i + \beta \sum x_i x_j) = \operatorname{sign}(-\beta)$, implying

$$\sum_{i<j} \widehat{P_\beta}(\{i,j\}) = 2^{-k} \sum_{x : \sum x_i = 0} \sum_{i<j} \operatorname{sign}(-\beta) x_i x_j,$$

which has the same sign as $-\operatorname{sign}(-\beta)$. Thus we see that if $\beta > 0$, $\beta \in B$ we have $\widehat{P_\beta}(\{1,2\}) > 0$, and if $\beta \leq 0$, $\beta \in B$ we have $\widehat{P_\beta}(\{1,2\}) < 0$.

From Claim 1 it follows that whenever $\beta \neq 0$ (not necessarily in B) we can simply take $\beta' = \beta$, and that when $\beta = 0$ we can take β' to be a negative value close to 0 (e.g., $\beta' = -|\alpha|/k^2$). □

We are now ready to prove Theorem 1.

Theorem 1 (restated)**.** *Let $P : \{-1,1\}^k \to \{0,1\}$ be a predicate that is of the form $P(x) = \frac{1+\mathrm{sign}(Q(x))}{2}$ where Q is a symmetric quadratic polynomial with no constant term. Then P is not approximation resistant.*

Proof: Without loss of generality, we can take $Q(x) = \alpha \sum x_i + \beta \sum x_i x_j$ where β satisfies the property of β' in Theorem 6.

If $\widehat{P}(\{1,2\}) = \beta = 0$, we set $\eta = \alpha/\widehat{P}(\{1\})$ (note that in this case we can assume that α, and hence also $\widehat{P}(\{1\})$ is non-zero as otherwise P is the trivial predicate that is always false). We then have, for every $x \in P^{-1}(1)$,

$$\frac{2\eta}{\sqrt{2\pi}} P^{=1}(x) + \frac{2}{\pi} P^{=2}(x) = \frac{2\alpha}{\sqrt{2\pi}} \sum x_i,$$

which is positive by the definition of P. If $\widehat{P}(\{1,2\}) \neq 0$, we set $\eta = \sqrt{\frac{2}{\pi}} \frac{\alpha}{\widehat{P}(\{1\})} \cdot \frac{\widehat{P}(\{1,2\})}{\beta}$. In this case for every $x \in P^{-1}(1)$,

$$\frac{2\eta}{\sqrt{2\pi}} P^{=1}(x) + \frac{2}{\pi} P^{=2}(x) = \frac{2\widehat{P}(\{1,2\})}{\pi\beta} \left(\alpha \sum x_i + \beta \sum x_i x_j\right) > 0,$$

since β agrees with $\widehat{P}(\{1,2\})$ in sign and $Q(x) > 0$. In either cases, using Theorem 4 and the respective choices of η we conclude that P is approximable. □

5 Monarchy

In this section we prove that for $k > 4$ the $\mathrm{Monarchy}$ predicate is not approximation resistant. Notice that $\mathrm{Monarchy}$ is defined only for $k > 2$, and that the case $k = 3$ coincides with the predicate $\mathrm{majority}$ that is known not to be approximation resistant. Further, the case $k = 4$ is handled by [Has05].(vi)

Just like the algorithm of Theorem 4 we first solve the natural semidefinite program of $\mathrm{Monarchy}$, and then use a rounding algorithm to construct an integral solution out of the vectors. The rounding algorithm, which is given in Figure 4, has two parameters $\epsilon > 0$ and an odd positive integer ℓ, both depending on k. These will be fixed in the proof.

Remark 1. *As the reader may have observed, the "geometric" power of SDP is not used in the rounding scheme in Figure 4, and indeed a linear programming relaxation of the problem would suffice for the algorithm we propose. However, in the interest of consistency and being able to describe the techniques in a language comparable to Theorem 1 we elected to use the SDP framework.*

(vi) In the notation of [Has05], $\mathrm{Monarchy}$ on 4 variables is the predicate `0000000101111111`, which is listed as approximable in Table 6.6. We remark that this is not surprising since $\mathrm{Monarchy}$ in this case is simply a majority in which x_1 serves as a tie-breaker variable.

Fig. 4: Rounding SDP solutions for Monarchy

INPUT: "biases" $b_1 = \mathbf{v}_0 \cdot \mathbf{v}_1, \ldots, b_n = \mathbf{v}_0 \cdot \mathbf{v}_n$.

PARAMETERS: An odd integer ℓ and $\epsilon \in [0, 1]$.

OUTPUT: $x_1, \ldots, x_n \in \{-1, 1\}$.

1. Choose a parameter $\tau \in [1/(2k^2), 1/k^2]$ uniformly at random.
2. For all i,
   (a) If $b_i > 1 - \tau$ or $b_i < -1 + \tau$, set x_i to 1 or -1 respectively.
   (b) Otherwise, set x_i (independent of all other x_j's), randomly to -1 or 1 such that $\mathbb{E}[x_i] = \epsilon b_i^\ell$. In particular, set $x_i = 1$ with probability $(1 + \epsilon b_i^\ell)/2$ and $x_i = -1$ with probability $(1 - \epsilon b_i^\ell)/2$.

We will first discuss the intuition behind the analysis of the algorithm ignoring, for now, the greedy ingredient (2a above). Notice that for $\epsilon = 0$ the rounding gives a uniform assignment to the variables, hence the expected value of the obtained solution is $1/2$. As long as $\epsilon > 0$ is small enough, the probability of success for a clause is essentially only affected by the degree-one Fourier coefficients of $\mathrm{Monarchy}$. Now, fix a clause and assume that the SDP solution completely satisfies it. Specifically, consider the clause $\mathrm{Monarchy}(x_1, \ldots, x_k)$, and define $b_1, \ldots, b_k$ as the corresponding biases. As the analysis will show, the rounding scheme above satisfies $\mathrm{Monarchy}(x_1, \ldots, x_k)$ with a probability that is essentially $1/2$ plus some *positive linear combination* of the ϵb_i^ℓ. Our objective is then to fix ℓ that would make the value of this combination positive (and independent from n). It turns out that the maximal b_i in magnitude (call it b_j) is always positive in this case. Oversimplifying, imagine that $|b_j| \geq |b_i| + \xi$ for all i different than j where ξ is some positive constant. In this setting it is easy to take ℓ (a function of k) that makes the effect of all b_i other than b_j vanish, ensuring a positive addition to the probability as desired so that overall the expected fraction of satisfied clauses is more than 1/2.

More realistically, the above slack ξ does not generally exist. However, we can show that a similar condition holds provided that the $|b_i|$ are bounded away from 1. This condition suffices to prove that the rounding algorithm works for clauses that do not have any variables with bias very close to ± 1. The case where there are b_i that are very close to 1 in magnitude is where the greedy ingredient of the algorithm (2a) is used, and it can be shown that when τ is roughly $1/k^2$, this ingredient works. In particular, we can show that for each clause, if rule (2a) is used to round one of the variables, it is used to round essentially every variable in the clause. Also, if this happens, the clause is going to be satisfied with high probability by the rounded solution.

The last complication stems from the fact that the clauses are generally not completely satisfied by the SDP solution. However, a standard averaging argument implies that it is enough to deal with clauses that are *almost* satisfied by the SDP solution. For any such clause the SDP induces a probability distribution on the variables that is mostly supported on satisfying assignments, compared to *only* on satisfying assignments in the above ideal setting. As such, the corresponding b_i's can be thought of as a perturbed version of the biases in that ideal setting. Unfortunately, the greedy ingredient of the algorithm is very sensitive to such small perturbations. In particular, if the biases are very close to the set threshold, τ, a small perturbation can break the method. To avoid this, we choose the actual threshold randomly, and we manage to argue that only a small fraction of the clauses end up in such unfortunate configurations.

This completes the high level description of the proof of our second result.

Theorem 2 (restated)**.** *The predicate* Monarchy *is* not *approximation resistant.*

Proof: As in Theorem 4 we can assume that the objective value of the SDP solution is at least $1-\delta$ for a fixed constant δ to be set later, and we can focus on clauses with SDP value at least $1-\sqrt{\delta}$. Again, as in Theorem 4 we consider one of these constraints and without loss of generality assume that this constraint is on $x_1, \ldots, x_k$. Remember that the variables $I_{[k],*}$ define a distribution, say μ, on $\{-1,1\}^k$ such that

$$\Pr_{y\sim\mu}[\mathrm{Monarchy}(y)=1] \geq 1-\sqrt{\delta}, \qquad\qquad \forall i \qquad b_i = \mathop{\mathbb{E}}_{y\sim\mu}[y_i].$$

Given that we choose τ uniformly at random in an interval of length $1/(2k^2)$, for any particular clause the probability that $|b_1|$ is of distance less than $2\sqrt{\delta}$ from $1-\tau$ is at most $8\sqrt{\delta}k^2$ and in particular, there are in expectation no more than a $8\sqrt{\delta}k^2$ fraction of clauses for which the following does not hold.

$$|b_1| \notin [1-\tau-2\sqrt{\delta}, 1-\tau+2\sqrt{\delta}]. \tag{12}$$

We will assume that (12) holds for our clause.

Given that μ is almost completely supported on points that satisfy Monarchy, we will first prove a few properties of distributions supported on satisfying points of Monarchy.

Lemma 7. *For a distribution ν on $\{-1,1\}^k$, completely supported on satisfying points of monarchy, i.e. , $\Pr_{y\sim\nu}[\mathrm{Monarchy}(y)=1]=1$, let*

$$\forall i \qquad b_i \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{y\sim\nu}[y_i].$$

Then,

$$\forall i>1 \qquad b_i \geq -b_1, \tag{13}$$

$$\frac{1}{k-1}\sum_{i>1} b_i \geq -b_1 + (1+b_1)/(k-1). \tag{14}$$

Proof: These two properties are linear conditions on the distribution, so if we check them for all points satisfying Monarchy they will follow for every distribution by convexity. They are two types of points in $\mathrm{Monarchy}^{-1}(1)$. There is the point $(-1,1,\ldots,1)$, and there are points of the form $(1, z_2, \ldots, z_n)$ where not all z_i's are -1. One can check that (13-14) hold for both kinds of points. □

We can write our distribution μ as $\mu = (1-\sqrt{\delta})\mu_0 + \sqrt{\delta}\mu_1$ where μ_0 is a distribution completely supported on $\mathrm{Monarchy}^{-1}(1)$ and μ_1 is a general distribution on $\{-1,1\}^k$. Notice that the biases of μ_0 satisfy the equations (13) and (14) while the biases of μ_1 do not falsify them with a margin bigger than 2, i.e. the left hand side will be no less than the right hand side minus 2. Lemma 7 then immediately implies that for our μ and $b_1, \ldots, b_n$,

$$\forall i>1 \qquad b_i \geq -b_1 - 2\sqrt{\delta}, \tag{15}$$

$$\frac{1}{k-1}\sum_{i>1} b_i \geq -b_1 + (1+b_1)/(k-1) - 2\sqrt{\delta}. \tag{16}$$

We are now ready to prove the following lemma. It essentially shows that the deterministic rounding of the variables with big bias does not interfere with the randomized rounding of the rest of the variables.

Lemma 8. *For any clause for which the SDP value is at least $1-\sqrt{\delta}$, and b_1 satisfies the range requirement of (12), one of the following happens,*

1. *x_1 is deterministically set to -1 and all the rest of x_i's are deterministically set to 1.*
2. *x_1 is deterministically set to 1, and at least two of the other x_i's are not deterministically set to -1.*
3. *x_1 is deterministically set to 1, and for some $i>1$, $b_i \geq 1-3/2(k-2)$.*
4. *x_1 is not set deterministically, and no other x_i is deterministically set to -1.*

Proof: The proof is by case analysis based on how x_1 is rounded.

First, assume that x_1 is deterministically set to -1. It follows from (15) that all the other b_i's are at least $-b_1-2\sqrt{\delta}$ so by the assumption in (12) we know that we are in case 1 of the lemma and we are done.

Now, assume that x_1 is deterministically set to 1. If for two distinct i's $b_i > -1+\tau$ we are in case 2 of the lemma and we are done. Otherwise, assume that b_j is the biggest of b_i's, and in particular all other b_i's are at most $-1+\tau$, we have,

$$
\begin{aligned}
\frac{1}{k-1}\sum_{i>1} b_i &\geq -b_1+(1+b_1)/(k-1)-2\sqrt{\delta} && \text{by (16)}\\
\frac{1}{k-1}\sum_{i>1} b_i &\leq \frac{b_j}{k-1}+\frac{k-2}{k-1}(-1+\tau) && \text{by assumption}\\
\Rightarrow b_j &\geq -(k-2)b_1+1-2\sqrt{\delta}(k-1)-(k-2)(-1+\tau)\\
&= 1-(k-2)(b_1-1+\tau)-2\sqrt{\delta}(k-1)\\
&\geq 1-(k-2)\tau-2\sqrt{\delta}(k-1)\\
&> 1-1/(k-2)-2\sqrt{\delta}(k-1) && \text{by } \tau\leq 1/k^2\\
&\geq 1-3/2(k-2) && \text{if } \delta < 1/16(k-1)^2(k-2)^2.
\end{aligned}
$$

This shows that we are in the third case of the lemma and we are done.

Finally, assume that x_1 is not deterministically rounded, i.e. , $-1+\tau\leq b_1\leq 1-\tau$. It follows from (12) that in fact, $b_1 < 1-\tau-2\sqrt{\delta}$. So, one can use (15) to deduce that for all $i>1$,

$$b_i \geq -b_1-2\sqrt{\delta} > -1+\tau+2\sqrt{\delta}-2\sqrt{\delta} = -1+\tau.$$

So, we are in case 4 of the lemma and we are done. □

We can now look at different cases and show that the parameters ϵ and ℓ can be set appropriately such that in all the cases the rounded $x_1,\ldots,x_k$ satisfy the predicate with probability at least $1/2+\gamma$ for some constant γ. Intuitively, in the first three cases the clause is deterministically rounded and has a high probability of being satisfied, while the fourth case is when the analysis of the clause needs arguments about the absolute values of the biases, and the clause is satisfied with probability slightly above $1/2$. We will handle the first three cases first.

Lemma 9. *If one of the three first cases of Lemma 8 happen for a clause then it is satisfied in the rounded solution with probability at least* 1, $1-(1+\epsilon)^2/4$, *and* $1/2+\epsilon 2^{-2\ell-1}$ *respectively. In particular, if* ϵ *and* ℓ *are constants (only depending on* k*) and* $\epsilon < \sqrt{2}-1$ *the clause is satisfied with probability at least* $1/2+\gamma$ *for some constant* γ *independent of* n*.*

Proof: The first two cases are easy: in the first case the clause is always satisfied by $x_1, \ldots, x_k$ while in the second case it is satisfied if and only if at least one $i > 1$, x_i is set to $+1$. Given that at least two of these x_i's are not deterministically rounded to -1, the clause is unsatisfied with probability at most $(1+\epsilon)^2/4$.

Now, assume that a clause is in the third case. Then, we know that for some $i > 1$, $b_i \geq 1-3/2(k-2)$, so the clause is satisfied with probability at least $\frac{1+\epsilon(1-3/2(k-2))^\ell}{2} \geq \frac{1}{2}+\epsilon 2^{-2\ell-1}$, where we have used $k \geq 4$. □

All that remains is to show that in the fourth case the clause is satisfied with some probability greater than $1/2$ and to find the suitable values of ϵ and ℓ in the process. This is formulated as the next lemma.

Lemma 10. *There are constants* $\epsilon = \epsilon(k)$, $\ell = \ell(k)$, *and* $\gamma = \gamma(k)$, *such that, for any small enough* δ, *any clause for which the fourth case of Lemma 8 holds is satisfied with probability at least* $1/2+\gamma$, *if we round with parameters* ϵ *and* ℓ*.*

The idea of the proof is to look at such a clause and inspect the objective value of the rounded solution using Fourier analysis. It is not hard to see that for small enough ϵ only the linear part of the Fourier expansion of $\mathrm{Monarchy}$ (i.e. $\mathrm{Monarchy}^{=1}$) "matters". So if we can prove that the linear part of the Fourier expansion contributes something positive we can set ϵ small enough and ignore the higher degree part. To do so, we will use (16) to prove that if ℓ is chosen big enough, this linear part is dominated by a positive term.

Proof of Lemma 10: We can assume that none of the x_i's are deterministically rounded as being rounded to 1 only helps us. We consider two cases: either $b_1 \leq 0$, or $b_1 \geq 0$. But first let us write the probability that this clause is satisfied in terms of the Fourier coefficients.

$$\begin{aligned}
\Pr\left[\mathrm{Monarchy}(x)=1\right] &= \widehat{\mathrm{Monarchy}}(\emptyset) + \sum_{i=1}^{k} \widehat{\mathrm{Monarchy}}(\{i\})\epsilon b_i^\ell + \sum_{i<j} \widehat{\mathrm{Monarchy}}(\{i,j\})\epsilon^2 b_i^\ell b_j^\ell + \cdots \\
&\geq \widehat{\mathrm{Monarchy}}(\emptyset) + \epsilon \sum_{i=1}^{k} \widehat{\mathrm{Monarchy}}(\{i\}) b_i^\ell - \epsilon^2 2^k \max_{S:|S|>1} |\widehat{\mathrm{Monarchy}}(S)| \\
&= 1/2 + \epsilon \sum_{i=1}^{k} \widehat{\mathrm{Monarchy}}(\{i\}) b_i^\ell - \epsilon^2 2^k. \qquad (17)
\end{aligned}$$

So all we have to do is to find a value of ℓ and a positive lower bound for $\sum_{i=1}^{k} \widehat{\mathrm{Monarchy}}(\{i\}) b_i^\ell$ that holds for all valid b_i's. It is easy to see that

$$\widehat{\mathrm{Monarchy}}(\{1\}) = 1/2 - 2^{1-k} \qquad \forall_{i>1} \quad \widehat{\mathrm{Monarchy}}(\{i\}) = 2^{1-k}.$$

Define,

$$C \stackrel{\text{def}}{=} \widehat{\text{Monarchy}}(\{1\})/\widehat{\text{Monarchy}}(\{2\}) \approx 2^{k-2}.$$

$$f(b_1,\ldots,b_k) \stackrel{\text{def}}{=} \frac{1}{\widehat{\text{Monarchy}}(\{2\})}\sum_{i=1}^{k}\widehat{\text{Monarchy}}(\{i\})b_i^\ell = Cb_1^\ell + \sum_{i>1} b_i^\ell.$$

In other words, $f(b_1,\ldots,b_k)$ is the part of (17) that we want to obtain a positive lower bound for, scaled by $\widehat{\text{Monarchy}}(\{2\})$. As $\widehat{\text{Monarchy}}(\{2\})$ is a constant it is then sufficient to lower bound $f(b_1,\ldots,b_k)$.

First, assume $b_1 \le (2k-4)^{-3}/2$. We know that,

$$\begin{aligned}
f(b_1,\ldots,b_k) &= Cb_1^\ell + \sum_{i>1} b_i^\ell \ge Cb_1^\ell + (k-1)(\frac{1}{k-1}\sum_i b_i)^\ell && \text{by concavity of } x^\ell,\\
&\ge Cb_1^\ell + (k-1)(-b_1 + (1+b_1)/(k-1) - 4\sqrt{\delta})^\ell && \text{by (16),}\\
&\ge Cb_1^\ell + (k-1)(-b_1 + \tau/(k-1) - 4\sqrt{\delta})^\ell && \text{as we are in case 4,}\\
&\ge Cb_1^\ell + (k-1)(-b_1 + 1/2(k-2)^2(k-1) - 4\sqrt{\delta})^\ell && \text{by choice of } \tau,\\
&\ge Cb_1^\ell + (k-1)(-b_1 + 1/4(k-2)^3 - 4\sqrt{\delta})^\ell &&\\
&> Cb_1^\ell + (k-1)(-b_1 + 1/8(k-2)^3)^\ell && \text{assuming } \delta < 2^{-5}(k-2)^{-6}.
\end{aligned}$$

Now if $b_1 \ge 0$ we are done as $f(b_1,\ldots,b_k)$ would be at least $(k-1)(2^k-4)^{-3\ell}2^{-\ell}$ and any constant ℓ will do the job. Otherwise, note that the expression inside the parenthesis is at least $(2k-4)^{-3}$ bigger than b_1 in absolute value. So, if we take ℓ to be big enough the second expression is going to dominate the expression. Specifically, first assume $|b_1| \ge (4k-8)^{-3}$,

$$\begin{aligned}
f(b_1,\ldots,b_k) &> Cb_1^\ell + (k-1)(-b_1 + (2k-4)^{-3})^\ell\\
&\ge Cb_1^\ell - b_1^\ell(k-1) + (k-1)\ell b_1^{\ell-1}(2k-4)^{-3}\\
&= ((C-k+1)b_1 + \ell(k-1)(2k-4)^{-3})b_1^{\ell-1}\\
&\ge ((C-k+1)b_1 + \ell(k-1)(2k-4)^{-3})(4k-8)^{-3\ell+3} && \text{as } \ell-1 \text{ is even}\\
&> (-(C-k+1) + \ell(k-1)(2k-4)^{-3})(4k-8)^{-3\ell+3}, && \text{by } b_1 < 1
\end{aligned}$$

which clearly has a constant lower bound if $\ell \ge (C-k+1)(2k-4)^3$. Now if $|b_1| < (4k-8)^{-3}$ we can write,

$$\begin{aligned}
f(b_1,\ldots,b_k) &> Cb_1^\ell + (k-1)(-b_1 + (2k-4)^{-3})^\ell\\
&> Cb_1^\ell + (k-1)(2k-4)^{-3\ell}\\
&> -C(4k-8)^{-3\ell} + (k-1)(2k-4)^{-3\ell} && \text{by } |b_1| < (4k-8)^{-3}\\
&= (2k-4)^{-3\ell}(-C2^{-3\ell} + (k-1)),\\
&> (2k-4)^{-3\ell}(k-2), && \text{as } 3\ell > \log_2 C
\end{aligned}$$

which is a constant as long as ℓ is some constant. This completes the case of $b_1 \leq (2k-4)^3/2$. Note that so far we have assumed that ℓ is some constant and at least $\max(\log_2 C/3, (C-k+1)(2k-4)^3)$.

Lets assume $b_1 > (2k-4)^{-3}/2$. One can write,

$$
\begin{aligned}
f(b_1, \ldots, b_k) &= Cb_1^\ell + \sum_{i>1} b_i^\ell \geq Cb_1^\ell + (k-1)(-b_1 - 4\sqrt{\delta})^\ell && \text{by (15),} \\
&= b_1^\ell \left(C - (k-1)(1 + 4\sqrt{\delta}/b_1)^\ell \right) \\
&> b_1^\ell \left(C - (k-1)(1 + 8(2k-4)^3\sqrt{\delta})^\ell \right) && \text{by } b_1 > (2k-4)^{-3}/2 \\
&> b_1^\ell \left(C - (k-1)\exp(8(2k-4)^3\sqrt{\delta}\ell) \right) && \text{by } e^x > 1 + x \\
&\geq b_1^\ell \geq (2k-4)^{-3\ell} 2^{-\ell},
\end{aligned}
$$

where to get the last line we have assumed that

$$
\begin{aligned}
\delta &\leq (4k-8)^{-6}\ell^{-2}(\ln(C-1) - \ln(k-1))^2 \\
\sqrt{\delta} &\leq (4k-8)^{-3}\ell^{-1}(\ln(C-1) - \ln(k-1)) \\
8(2k-4)^3\sqrt{\delta}\ell &\leq \ln(C-1) - \ln(k-1) \\
(k-1)\exp(8(2k-4)^3\sqrt{\delta}\ell) &\leq C-1.
\end{aligned}
$$

So, as long as $C - 1 > k - 1$ we can set $\ell = \max(\log_2 C/3, (C-k+1)(2k-4)^3)$ and prove that $f(b_1, \ldots, b_k)$ has a constant lower bound depending on k, provided that we assume $\delta < \delta_0$ where δ_0 is another function of k. One can check that $C > k - 2$ whenever $k > 4$. This completes the proof. $\square$

We can now finish the proof of Theorem 2. In particular, solve the SDP relaxation of $\mathrm{Monarchy}$, if the objective value is smaller than $1 - \delta$ output a uniformly random solution, and if it is bigger apply the rounding algorithm in Figure 4. In the first case the expected objective of the output is $1/2$, while the optimal solution can not have objective value more than $1 - \delta$, giving rise to approximation ratio $1/2(1-\delta) > (1+\delta)/2$. In the second case, at least a $(1 - \sqrt{\delta})$ fraction of the clauses have objective $1 - \sqrt{\delta}$ and from these in expectation at least a $1 - 16\sqrt{\delta}(k-2)^2$ fraction satisfy (12). We can apply Lemma 9 and Lemma 10 to these. So, the objective function is at least,

$$(1 - \sqrt{\delta})(1 - 16\sqrt{\delta}(k-2)^2)(1/2 + \gamma) \geq 1/2 + \gamma - 17\sqrt{\delta}(k-2)^2.$$

This is clearly more than $1/2 + \gamma/2$ for small enough δ, which finishes the proof of the Theorem. $\square$

6 Discussion

We have given algorithms for two cases of MAX CSP(P) problems not previously known to be approximable. The first case, signs of symmetric quadratic forms, follows from the condition that the low-degree part of the Fourier expansion behaves "roughly" like the predicate in the sense of Theorem 4. The second case, Monarchy, is interesting since it does not satisfy the condition of Theorem 4. As far as we are aware, this is the first example of a predicate which does not satisfy this property but is still approximable. Monarchy is of course only a very special case of Conjecture 1, and we leave the general form open.

A further interesting special case of the conjecture is a generalization of $\mathrm{Monarchy}$ called "republic", defined as $\mathrm{sign}(\frac{k}{2}x_1 + \sum_{i=2}^{k} x_i)$. In this case the x_1 variable needs to get a $1/4$ fraction of the other variables on its side. We do not even know how to handle this seemingly innocuous example.

It is interesting that the condition on P for our (ϵ, η)-rounding to succeed turned out to be precisely the same as the condition previously found by Hast [Has05], with a completely different algorithm. It would be interesting to know whether this is a coincidence or there is a larger picture that we can not yet see.

As we mentioned in the introduction, there are very few results which give approximation algorithms for large classes of predicates, and it would be very interesting if new such algorithms could be devised.

References

[ABM10] Per Austrin, Siavosh Benabbas, and Avner Magen. On quadratic threshold CSPs. In *LATIN'10*, pages 332–343, 2010.

[AH11] Per Austrin and Johan Håstad. Randomly supported independence and resistance. *SIAM Journal on Computing*, 40(1):1–27, 2011.

[AH12] Per Austrin and Johan Håstad. On the Usefulness of Predicates. In *CCC'12*, pages 53–63, 2012.

[AM09] Per Austrin and Elchanan Mossel. Approximation Resistant Predicates from Pairwise Independence. *Computational Complexity*, 18(2):249–271, 2009.

[BGMT12] Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8(1):269–289, 2012.

[CHIS12] Mahdi Cheraghchi, Johan Håstad, Marcus Isaksson, and Ola Svensson. Approximating linear threshold predicates. *ACM Trans. Comput. Theory*, 4(1):2:1–2:31, March 2012.

[Cre95] Nadia Creignou. A dichotomy theorem for maximum generalized satisfiability problems. *J. Comput. Syst. Sci.*, 51:511–522, December 1995.

[EH08] Lars Engebretsen and Jonas Holmerin. More efficient queries in pcps for np and improved approximation hardness of maximum csp. *Random Struct. Algorithms*, 33(4):497–514, 2008.

[GW95] Michel X. Goemans and David P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *Journal of the ACM*, 42:1115–1145, 1995.

[Hås01] Johan Håstad. Some Optimal Inapproximability Results. *Journal of the ACM*, 48(4):798–859, 2001.

[Has05] Gustav Hast. *Beating a Random Assignment – Approximating Constraint Satisfaction Problems*. PhD thesis, KTH – Royal Institute of Technology, 2005.

[Hås09] Johan Håstad. Personal communication, 2009.

[Kho02] Subhash Khot. On the Power of Unique 2-prover 1-round Games. In *STOC'02*, pages 767–775, 2002.

[KZ97] Howard Karloff and Uri Zwick. A 7/8-Approximation Algorithm for MAX 3SAT? In *FOCS'97*, pages 406–415, 1997.

[Rag08] Prasad Raghavendra. Optimal Algorithms and Inapproximability Results For Every CSP? In *STOC'08*, pages 245–254, 2008.

[ST00] Alex Samorodnitsky and Luca Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *STOC'00*, pages 191–199, 2000.

[ST09] Alex Samorodnitsky and Luca Trevisan. Gowers uniformity, influence of variables, and pcps. *SIAM Journal on Computing*, 39(1):323–360, 2009.

[Woo26] Frederick S. Woods. *Advanced Calculus: A Course Arranged with Special Reference to the Needs of Students of Applied Mathematics*. Ginn, Boston, MA, USA, 1926.

[Zwi98] Uri Zwick. Approximation Algorithms for Constraint Satisfaction Problems Involving at Most Three Variables Per Constraint. In *SODA'98*, pages 201–210, 1998.

Appendix: Proof of Lemma 5

In this section we present the proof Lemma 5 restated here for convenience.

Lemma 5 (restated)**.** *For fixed k, define the function* $\mathrm{ort}(\nu, \Sigma)$ *as the orthant probability of the multivariate normal distribution with mean ν and covariance matrix Σ, where $\nu \in \mathbb{R}^k$ and $\Sigma_{k\times k}$ is a positive semidefinite matrix. That is,*

$$\mathrm{ort}(\nu, \Sigma) \stackrel{\text{def}}{=} \Pr_{x\sim N(\nu,\Sigma)} [x \geq 0] .$$

There exists a global constant Γ that upper bounds all the second partial derivatives of $\mathrm{ort}()$ *when Σ is close to I. In particular, for all k, there exist $\kappa > 0$ and Γ, such that for all $i_1, j_1, i_2, j_2 \in [k]$, all vectors $\nu \in \mathbb{R}^k$ and all positive definite matrices $\Sigma_{k\times k}$,*

$$\begin{aligned}
|I - \Sigma|_\infty, |\nu|_\infty < \kappa \Rightarrow \left|\frac{\partial^2}{\partial\Sigma_{i_1j_1}\partial\Sigma_{i_2j_2}} \mathrm{ort}(\nu, \Sigma)\right| &< \Gamma, \\
\left|\frac{\partial^2}{\partial\Sigma_{i_1j_1}\partial\nu_{i_2}} \mathrm{ort}(\nu, \Sigma)\right| &< \Gamma, \\
\left|\frac{\partial^2}{\partial\nu_{i_1}\partial\nu_{i_2}} \mathrm{ort}(\nu, \Sigma)\right| &< \Gamma.
\end{aligned}$$

Proof: We know that whenever Σ is a (*full rank*) positive definite matrix,

$$\begin{aligned}
\mathrm{ort}(\nu, \Sigma) &= \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \int_{x_1=0}^{+\infty} \cdots \int_{x_k=0}^{+\infty} \phi(x, \nu, \Sigma)\mathrm{d}x_1 \cdots \mathrm{d}x_k \\
\phi(x, \nu, \Sigma) &= \exp\left(-\frac{1}{2}(x-\nu)^\top \Sigma^{-1}(x-\nu)\right) \\
&= \exp\left(-\frac{1}{2}\sum_{l,m\in[k]} (x_l - \nu_l)(x_m - \nu_m)(-1)^{l+m}|\Sigma^{ml}|/|\Sigma|\right)
\end{aligned}$$

where ϕ is a normalization of the probability density function of the multivariate normal distribution and Σ^{ml} is the minor of Σ obtained by removing row m and column l. More abstractly, we can write

$$\phi(x, \nu, \Sigma) = \exp\left(\sum_{l,m\in[k]} (x_l - \nu_l)(x_m - \nu_m)p_{lm}(\Sigma)/q(\Sigma)\right),$$

where p_{lm} and q are polynomials of degree $\leq k$ in Σ (depending only on k), and where $q(\Sigma)$ is bounded away from 0 in a region around Σ. Let $y_i = x_i - \nu_i$. We can then write

$$\frac{\partial^2}{\partial\Sigma_{ij}\partial\Sigma_{i'j'}}\phi(x, \nu, \Sigma) = \frac{\phi(x, \nu, \Sigma)}{q(\Sigma)^4} \sum_{l,m} y_l y_m (A_{lm,ij,i'j'}(\Sigma) + y_l y_m B_{lm,ij,i'j'}(\Sigma))$$

where $A_{lm,ij,i'j'}$ and $B_{lm,ij,i'j'}$ are polynomials depending only on p_{lm}, q, i, j, i' and j'. Thus, for all Σ in a neighbourhood around I, we have

$$\begin{aligned}\left|\frac{\partial^2}{\partial\Sigma_{ij}\partial\Sigma_{i'j'}}\phi(x,\nu,\Sigma)\right| &\le C\sum_{l,m}\left(|y_l y_m| + y_l^2 y_m^2\right)\phi(x,\nu,\Sigma)\\ &\le C\sum_{l,m}\left(\frac{1}{2}+\frac{3}{2}y_l^2 y_m^2\right)\phi(x,\nu,\Sigma)\end{aligned}$$

for a constant C (depending only on k). By an iterative application of Leibniz Integral Rule (Theorem 11, below) we can bound the second derivative of $\mathrm{ort}(\nu,\Sigma)$ as

$$\begin{aligned}\left|\frac{\partial^2}{\partial\Sigma_{ij}\partial\Sigma_{i'j'}}\,\mathrm{ort}(\nu,\Sigma)\right| &\le \sum_{l,m\in[k]}\frac{C}{\sqrt{(2\pi)^k|\Sigma|}}\int_{x_1=0}^{+\infty}\cdots\int_{x_k=0}^{+\infty}\big(\frac{1}{2}+\frac{3}{2}(x_l-\nu_l)^2(x_m-\nu_m)^2\big)\phi(x,\nu,\Sigma)\\ &\le \sum_{l,m\in[k]}\frac{C}{\sqrt{(2\pi)^k|\Sigma|}}\int_{x_1=-\infty}^{+\infty}\cdots\int_{x_k=-\infty}^{+\infty}\left(\frac{1}{2}+\frac{3}{2}(x_l-\nu_l)^2(x_m-\nu_m)^2\right)\phi(x,\nu,\Sigma)\\ &= C\sum_{l,m\in[k]}\big(\frac{1}{2}+\frac{3}{2}\mathop{\mathbb{E}}_{x\sim N(\nu,\Sigma)}\left[(x_l-\nu_l)^2(x_m-\nu_m)^2\right]\big)\\ &= C\sum_{l,m\in[k]}\big(\frac{1}{2}+\frac{3}{2}\Sigma_{ll}\Sigma_{mm}+3\Sigma_{lm}^2\big)\le 19k^2C,\end{aligned}$$

where we have assumed each element of Σ is at most 2 in absolute value. The cases of partial derivatives with respect to ν_i's follows similarly. □

Theorem 11 (Leibniz Integral Rule, see "Differentiation of a Definite Integral." §60 in [Woo26])**.** *Consider a real function of two variables $f(x,y)$ and assume that for some $x_0, x_1, y_0, y_1 \in \mathbb{R}$ both f and $\frac{\partial}{\partial x}f$ are continuous in the region $[x_0,x_1]\times[y_0,y_1]$. Then for any $x\in(x_0,x_1)$,*

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_{y_0}^{y_1} f(x,y)\,\mathrm{d}y = \int_{y_0}^{y_1}\frac{\partial f(x,y)}{\partial x}\,\mathrm{d}y$$