On the Usefulness of Predicates

Per Austrin, Aalto University and KTH Royal Institute of Technology
Johan Håstad, KTH Royal Institute of Technology

Motivated by the pervasiveness of strong inapproximability results for Max-CSPs, we introduce a relaxed notion of an approximate solution of a Max-CSP. In this relaxed version, loosely speaking, the algorithm is allowed to replace the constraints of an instance by some other (possibly real-valued) constraints, and then only needs to satisfy as many of the new constraints as possible.

To be more precise, we introduce the following notion of a predicate P being *useful* for a (real-valued) objective Q: given an almost satisfiable Max-P instance, there is an algorithm that beats a random assignment on the corresponding Max-Q instance applied to the same sets of literals. The standard notion of a nontrivial approximation algorithm for a Max-CSP with predicate P is exactly the same as saying that P is useful for P itself.

We say that P is useless if it is not useful for any Q. This turns out to be equivalent to the following pseudo-randomness property: given an almost satisfiable instance of Max-P it is hard to find an assignment such that the induced distribution on k-bit strings defined by the instance is not essentially uniform.

Under the Unique Games Conjecture, we give a complete and simple characterization of useful Max-CSPs defined by a predicate: such a Max-CSP is useless if and only if there is a pairwise independent distribution supported on the satisfying assignments of the predicate. It is natural to also consider the case when no negations are allowed in the CSP instance, and we derive a similar complete characterization (under the UGC) there as well.

Finally, we also include some results and examples shedding additional light on the approximability of certain Max-CSPs.

Categories and Subject Descriptors: F.2 [**Theory of Computation**]: ANALYSIS OF ALGORITHMS AND PROBLEM COMPLEXITY

General Terms: Algorithms, Theory

Additional Key Words and Phrases: constraint satisfaction problem, usefulness, approximation resistance, unique games conjecture

ACM Reference Format:
ACM Trans. Comput. Theory V, N, Article A (January YYYY), 25 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

The motivation for this paper comes from the study of maximum constraint satisfaction problems (Max-CSPs). In a Max-CSP problem, we are given a sequence of constraints, each depending on a constant number of variables, and the goal is to find an assignment that maximizes the number of satisfied constraints. Essentially any such problem is NP-hard and a large number of papers have studied the question of approximability of this class of problems. The standard concept of approximability is that an

Work done while P.A. at the University of Toronto funded by NSERC. J.H. funded by ERC Advanced Investigator Grant 226203.

DOI:http://dx.doi.org/10.1145/0000000.0000000

algorithm is a C-approximation algorithm if, on any instance I, it outputs a number $A(I)$ such that $C \cdot O(I) \leq A(I) \leq O(I)$, where $O(I)$ is the optimum value of I.

There are finer measures of performance. For example, one can take C above to be a function of the optimum value $O(I)$. That is, for each fraction of constraints satisfied by the optimal solution, we try to determine the best solution that can be found efficiently. The only problem where this has been fully done explicitly is Max-Cut, where, assuming the unique games conjecture, O'Donnell and Wu [O'Donnell and Wu 2008] has found the entire curve of approximability. In a remarkable paper, Raghavendra [Raghavendra 2008] showed that, assuming the unique games conjecture, the best such approximation possible is the one given by a certain natural semidefinite programming-based (SDP) relaxation of the problem. However, understanding the performance of this SDP is difficult in general and in this paper we are interested in more explicit bounds.

Max-Cut turns out to be approximable in a very strong sense. To describe these results note that for Max-Cut it is the case that a random assignment satisfies half of the constraints on average. Whenever the optimum satisfies a fraction $1/2 + \epsilon$ of the constraints then it is possible to efficiently find an assignment that satisfies a fraction $1/2 + c(\epsilon)$ of the constraints where $c(\epsilon)$ is strictly positive, depending on ϵ [Charikar and Wirth 2004]. In other words, whenever the optimal solution satisfies a non-trivial fraction of the constraints it is possible to efficiently find an assignment that satisfies a smaller, but still non-trivial, fraction of the constraints.

In this paper the main focus is on the other end of the spectrum. Specifically we are interested in the following property: even if the optimal solution satisfies (almost) all the constraints it is still hard to find an assignment that satisfies a non-trivial fraction. This might sound like an unusual property, but evidence is mounting that most CSPs have this property. We say that such a CSP is *approximation resistant* (a formal definition appears in Section 2).

We shall focus on a special class of CSPs defined by a single predicate $P : \{-1, 1\}^k \rightarrow \{0, 1\}$ (throughout the paper we identify the Boolean value true with -1 and false with 1). Each constraint asserts that P applied to some k literals (each literal being either a variable or a negated variable) is true. We refer to this problem as Max-P, and say that P is approximation resistant if Max-P is.

Several predicates are proven to be approximation resistant in [Håstad 2001] and the most notable cases are when the predicate in question is the XOR, or the usual OR, of 3 literals. For the latter case, Max-3Sat, it is even the case that the hardness remains the same for satisfiable instances. This is clearly not the case for XOR since a satisfying assignment, if one exists, can be found by Gaussian elimination. Hast [Hast 2005] studied predicates of arity 4 and of the (exactly) 400 different predicates, 79 are proven to be approximation resistant, 275 are found to be non-trivially approximable while the status of the remaining 46 predicates was not determined. Some results exist also for larger arity predicates and we return to some of these results in Section 4. If one is willing to believe the unique games conjecture (UGC) of Khot [Khot 2002] then it was established in [Austrin and Håstad 2011] that an overwhelming majority of all predicates are approximation resistant. This paper relies on a result [Austrin and Mossel 2009] establishing that any predicate P such that the set of accepted strings $P^{-1}(1)$ supports a pairwise independent distribution is, assuming the UGC, approximation resistant.

In spite of all these impressive results we want to argue that approximation resistance is not the ultimate hardness condition for a CSP. Approximation can be viewed as relaxing the requirements: if there is an assignment that satisfies a large number, or almost all, of a given set of constraints, we are content in finding an assignment that satisfies a lower but still non-trivial number of the constraints. In some situations, instead of relaxing the number of constraints we want to satisfy it might make more sense to relax the constraints themselves.

Sometimes such relaxations are very natural, for instance if considering a threshold predicate we might want to lower the threshold in question. It also makes sense to have a real-valued measure of success. If we give the full reward for satisfying the original predicate we can have a decreasing reward depending on the distance to the closest satisfying assignment. This is clearly natural in the threshold predicate scenario but can also make good sense for other predicates.

It seems like the least we can ask for of a CSP is that when we are given an instance where we can satisfy (almost) all constraints under a predicate P then we can find an assignment that does something non-trivial for some, possibly real-valued, relaxation Q. This brings us to the key definition of our paper.

Definition 1.1. The predicate P is *useful* for the real-valued function $Q : \{-1, 1\}^k \to \mathbb{R}$, if and only if there is an $\epsilon > 0$ such that given an instance of Max-P where the optimal solution satisfies a fraction $\geq 1 - \epsilon$ of the constraints, there is a polynomial time algorithm to find an assignment x^0 such that

$$\frac{1}{m}\sum_{j=1}^{m} Q(\bar{x}^0_j) \geq \mathop{\mathbb{E}}_{x\in\{-1,1\}^k}[Q(x)] + \epsilon.$$

Here $\bar{x}^0_j$ denotes k-bit string giving the values of the k literals in the j'th constraint of P under the assignment x^0.

Given a notion of "useful" it is natural to define "useless". We say that P is useless for Q if, assuming P $\neq$ NP, it is not useful for Q. We choose to build the assumption P $\neq$ NP into the definition in order not to have to state it for every theorem – the assumption is in a sense without loss of generality since if P $=$ NP then uselessness is the same as a related notion we call information-theoretic uselessness which we briefly discuss in Section 3. Note that uselessness is a generalization of approximation resistance as that notion is the property that P is useless for P itself.

This observation implies that requiring a predicate to be useless for any relaxation is a strengthening of approximation resistance. The property of being a relaxation of a given predicate is somewhat in the eye of the beholder and hence we choose the following definition.

Definition 1.2. The predicate P is *(computationally) useless* if and only if it is useless for every $Q : \{-1, 1\}^k \to \mathbb{R}$.

As described in Section 4, it turns out that almost all approximation resistance proofs have indeed established uselessness. There is a natural reason that we get this stronger property. In a standard approximation resistance proof we design a probabilistically checkable proof (PCP) system where the acceptance criteria is given by P and the interesting step in the proof is to analyze the soundness of this PCP. In this analysis we use the Fourier expansion of P and it is proved that the only term that gives a significant contribution is the constant term. The fact that we are looking at the same P that was used to define the PCP is usually of no importance. It is thus equally easy to analyze what happens to any real-valued Q. In particular, it is straightforward to observe that the proof of [Håstad 2001] in fact establishes that parity of size at least 3 is useless. Similarly the proof of [Austrin and Mossel 2009], showing that any predicate that supports a pairwise independent measure is approximation resistant, also gives uselessness (but of course we still need to assume the unique games conjecture).

The possibly surprising, but technically not very difficult, result that we go on to establish (in Section 5) is that if the condition of [Austrin and Mossel 2009] is violated then we can find a real-valued function for which P is useful. Thus assuming the UGC we have a complete characterization of the property of being useless!

THEOREM 1.3. *Assuming the UGC, a predicate P is (computationally) useless if and only if there is a pairwise independent distribution supported on $P^{-1}(1)$.*

Disallowing negated variables. In Section 6 we briefly discuss what happens in the case when we do not allow negated variables, which in some cases may be more natural. In this situation we need to extend the notion of a trivial algorithm in that now it might make sense to give random but biased values to the variables. A simple example is when P accepts the all-one string in which case setting each variable to 1 with probability 1 causes us to satisfy all constraints (regardless of the instance), but probabilities strictly between 0 and 1 might also be optimal. Taking this into account our definitions extend.

In the setting without negated variables it turns out that the unique games-based uselessness proof can be extended with slightly relaxed conditions with minor modifications. We are still interested in a distribution μ over $\{-1, 1\}^k$, supported on strings accepted by P, but we can allow two relaxations to the pairwise independence condition. The individual bits under μ need not be unbiased but each bit should have the same bias. Perhaps more interestingly, the bits need not be pairwise independent and we can allow positive (but for each pair of bits the same) correlations among the bits.

THEOREM 1.4 (INFORMAL). *When we do not allow negated variables, P is useless (assuming UGC) if and only if the accepting strings of P supports such a distribution.*

Note that this implies that any predicate that is useless when we allow negations is also useless when we do not allow negations while the converse is not true.

A basic computationally useless predicate in this setting is odd parity of an even number of variables (at least 4 variables). With even parity, or with odd parity of an odd number of variables, the predicate is also useless, but for the trivial reason that we can always satisfy all constraints (so the guarantee that we can satisfy most applications of P gives no extra information). Surprisingly we need the UGC to establish the result for odd parity of an even number of variables. As briefly discussed in Section 6 below it seems like new techniques are needed to establish NP-hardness results in this situation.

Adaptive uselessness and pseudorandomness. Our definition of uselessness is not the only possible choice. A stronger definition would be to let the algorithm choose a new objective Q based on the actual Max-P instance I, rather than just based on P. We refer to this as *adaptive* uselessness, and discuss it in Section 7. It turns out that in the settings discussed, adaptive uselessness is the same as non-adaptive uselessness.

In the adaptive setting when we allow negations clearly the task is to find an assignment such that the k-bit strings appearing in the constraints do not have the uniform distribution. This is the case as we can choose a Q which takes large values for the commonly appearing strings. Thus in this situation our results say that even given the promise that there is an assignment such that almost all resulting k-bit strings satisfy P, an efficient algorithm is unable to find any assignment for which the distribution on k-bit strings is not (almost) uniform.

Other results. When we come to investigating useful predicates and to determining pairs for which P is useful for Q it is of great value to have extensions of the result [Austrin and Mossel 2009]. These are along the lines of having distributions supported on the strings accepted by P where most pairs of variables are uncorrelated. Details of this can be found in Section 8.1.

Some examples relating to this theorem are given in Appendix B. In Appendix C we describe a predicate which is the sign of quadratic function but which is still approximation resistant. This shows that the condition of Austrin and Mossel of supporting a pairwise independent distribution is not necessary and sufficient when it comes to being approximation resistant.

Finally, in Appendix D, we take a brief look at the other end of the spectrum, and study CSPs which are highly approximable.

For the record let us point out that a preliminary version of this paper has appeared at the conference for Computational Complexity [Austrin and Håstad 2012].

2. PRELIMINARIES

We have a predicate P: $\{-1,1\}^k \rightarrow \{0,1\}$. The traditional approximation problem to study is Max-P in which an instance consists of m constraints over n variables. Each constraint is a k-tuple of literals, where a literal is either a variable or a negated variable. The goal is to find an assignment to the variables so as to maximize the number of resulting k-bit strings that satisfy the predicate P. To be more formal an instance is given by a set of indices $a_j^i \in [n]$ for $1 \leq i \leq k$ and $1 \leq j \leq m$ and complementations $b_j^i \in \{-1,1\}$. The jth k-tuple of literals contains the variables $(x_{a_j^i})_{i\in[k]}$ with $x_{a_j^i}$ negated iff $b_j^i = -1$. We use the short hand notation $P(x_{a_j}^{b_j})$ for the jth constraint.

We do not allow several occurrences of the same variable in one constraint. In other words, $a_j^i \neq a_j^{i'}$ for $i \neq i'$. The reason for this convention is that if the same variable appears twice we in fact have a different predicate on a smaller number of variables. This different predicate is of course somewhat related to P but does not share even basic properties such as the probability that it is satisfied by a random assignment. Thus allowing repeated variables would take us into a more complicated situation.

In this paper we assume that all constraints have the same weight but it is not hard to extend the results to the weighted case.

Definition 2.1. For $Q : \{-1,1\}^k \rightarrow \mathbb{R}$ define

$$E_Q = \mathop{\mathbb{E}}_{x\in\{-1,1\}^k} [Q(x)] .$$

Note that for a predicate P, an alternative definition of E_P is the probability that a uniformly random assignment satisfies P. It follows that the trivial algorithm that just picks a uniformly random assignment approximates Max-P within a factor E_P.

Definition 2.2. The predicate P is *approximation resistant* if and only if, for every $\epsilon > 0$, it is NP-hard to approximate Max-P within a factor $E_P + \epsilon$.

Another way to formulate this definition is that, again for any $\epsilon > 0$, it is NP-hard to distinguish instances for which the optimal solution satisfies a fraction $1 - \epsilon$ of the constraints from those where the optimal solution only satisfies a fraction $E_P + \epsilon$. One can ask for even more and we have the following definition.

Definition 2.3. The predicate P is *approximation resistant on satisfiable instances* if and only if, for any $\epsilon > 0$, it is NP-hard to distinguish instances of Max-P for which the optimal solution satisfies all the constraints from those instances where the optimal solution only satisfies a fraction $E_P + \epsilon$ of the constraints.

A phenomenon that often appears is that if P is approximation resistant then any predicate P' that accepts strictly more strings is also approximation resistant. Let us introduce a concept to capture this fact.

Definition 2.4. The predicate P is *hereditarily approximation resistant* if and only if, for any predicate P' implied by P (i.e., whenever $P(x)$ is true then so is $P'(x)$) is approximation resistant.

It turns out that 3-Lin, and indeed any parity of size at least three, is hereditarily approximation resistant. There are also analogous notions for satisfiable instances but as this is not the focus of the present paper we do not give the formal definition here. One of the few examples of a predicate that is approximation resistant but not hereditarily so is a predicate studied by Guruswami et al [Guruswami et al. 1998]. We discuss this predicate in more detail in Appendix B below.

Let us recall the definition of pairwise independence.

Definition 2.5. A distribution μ over $\{-1,1\}^k$ is *biased pairwise independent* if, for some $p \in [0,1]$, we have $\Pr_\mu[x_i = 1] = p$ for every $i \in [k]$ and $\Pr_\mu[x_i = 1 \wedge x_j = 1] = p^2$ for every $1 \leq i < j \leq k$ (i.e., if all two-dimensional marginal distributions are equal and product distributions).

We say that μ is *pairwise independent* if it is biased pairwise independent with $p = 1/2$ (i.e., if the marginal distribution on any pair of coordinates is uniform).

Finally we need a new definition of a distribution that we call uniformly positively correlated.

Definition 2.6. A distribution μ over $\{-1,1\}^k$ is *uniformly positively correlated* if, for some $p, \rho \in [0,1]$, with $\rho \geq p^2$, we have $\Pr_\mu[x_i = 1] = p$ for every $i \in [k]$ and $\Pr_\mu[x_i = 1 \wedge x_j = 1] = \rho$ for every $1 \leq i < j \leq k$ (i.e., if all two-dimensional marginal distributions are equal and the bits are positively correlated).

Note that we allow $\rho = p^2$ and thus any biased pairwise independent distribution is uniformly positively correlated.

3. INFORMATION-THEORETIC USEFULNESS

Clearly there must be some relation between P and Q for our notion to be interesting and let us discuss this briefly in the case when Q is a predicate.

If P and Q are not strongly related then it is possible to have instances where we can satisfy all constraints when applying P and only an E_Q fraction for Q. A trivial example would be if P is OR of three variables and Q is XOR. Then given the two constraints (x_1, x_2, x_3) and $(x_1, x_2, \bar{x}_3)$ it is easy to satisfy both constraints under P but clearly exactly one is always satisfied under Q. Thus we conclude that OR is not useful for XOR.

As another example let P be equality of two bits and Q non-equality and let the constraints be all pairs (x_i, x_j) for $1 \leq i < j \leq n$ (unnegated). It is possible to satisfy all constraints under P but it is not difficult to see that the maximal fraction goes to 1/2 under Q as n tends to infinity. We can note that the situation is the same for P being even parity and Q being odd parity if the size is even, while if the size of the parity is odd the situation is completely the opposite as negating a good assignment for P gives a good assignment for Q.

After these examples let us take a look in more detail at usefulness in an information-theoretic sense. It is not difficult to see that perfect and almost-perfect completeness are equivalent in this situation.

Definition 3.1. A predicate P is *information-theoretically useless* for Q if, for any $\epsilon > 0$ there is an instance such that

$$\max_x \frac{1}{m} \sum_{j=1}^{m} P(x_{a_j}^{b_j}) = 1$$

while

$$\max_x \frac{1}{m} \sum_{j=1}^{m} Q(x_{a_j}^{b_j}) \leq E_Q + \epsilon.$$

A trivial remark is that in the information-theoretic setting we cannot have total uselessness as P is always information-theoretically useful for itself or any predicate implied by P (unless P is trivial).

Let us analyze the above definition. Let μ be a probability measure on $\{-1, 1\}^k$ and let μ^p be the distribution obtained by first picking a string according to μ and then flipping each coordinate with probability p. Note that p need not be small and $p = 1$ is one interesting alternative as illustrated by the parity example above.

For a given μ let $Opt(Q, \mu)$ be the maximum over p of the expected value of $Q(x)$ when x is chosen according to μ^p. We have the following theorem.

THEOREM 3.2. *The predicate P is information-theoretically useless for Q if and only if there exists a measure μ supported on strings accepted by P such that $Opt(Q, \mu) = E_Q$.*

PROOF. Let us first see that if $Opt(Q, \mu) > E_Q$ for every μ then P is indeed useful for Q. Note that the space of measures on a finite set is compact and thus we have $Opt(Q, \mu) \geq E_Q + \delta$ for some fixed $\delta > 0$ for any measure μ.

Consider any instance with

$$\max \frac{1}{m} \sum_{j=1}^{m} P(x_{a_j}^{b_j}) = 1$$

and let us consider the strings $(x_{a_j}^{b_j})_{j=1}^m$ when x is the optimal solution. These are all accepted by P and considering with which proportion each string appears we let this define a measure μ of strings accepted by P. By the definition of $Opt(Q, \mu)$, there is some p such that a random string from μ^p gives an expected value of at least $E_Q + \delta$ for $Q(x)$. It follows that flipping each bit in the optimal assignment for P with probability p we get an assignment such that

$$\mathbb{E}\left[\frac{1}{m} \sum_{j=1}^{m} Q(x_{a_j}^{b_j})\right] \geq E_Q + \delta$$

and thus P is information-theoretically useful for Q.

For the reverse conclusion we construct a random instance. Let μ be the measure guaranteed to exist by the assumption of the theorem.

We pick m random constraints independently as follows: $a_j \in [n]^k$ is a uniformly random set of indices from $[n]$ which are all different, and $b_j \in \{-1, 1\}^k$ is sampled according to μ. Note that the all-one solution satisfies all constraints since if $x_i = 1$ for all i then $x_{a_j}^{b_j}$ is distributed according to μ and therefore satisfies P.

Now we claim that, for an assignment with a fraction $1 - p$ variables set to 1, the expected value (over the choice of instance) of $\frac{1}{m} \sum_{j=1}^{m} Q(x_{a_j}^{b_j})$ is within an additive $O(\frac{1}{n})$ of $\mathbb{E}[Q(x)]$ when x is chosen according to μ^p. This is more or less immediate from the definition and the small error comes form the fact that we require the chosen variables to be different creating a small bias. Taking m sufficiently large compared to n the theorem now follows from standard large deviation estimates and an application of the union bound. □

Let us return to our main interest of studying usefulness in a computational context.

4. SOME EXAMPLES AND EASY THEOREMS

We have an almost immediate consequence of the definitions.

THEOREM 4.1. *If P is useless then P is hereditarily approximation resistant.*

PROOF. Let P' be any predicate implied by P. The fact that P is useless for P' states that it is hard to distinguish instances where we can satisfy P (and hence P') almost always from those where we can only satisfy P' on an $E_{P'} + \epsilon$ fraction of the constraints. The theorem follows. □

Clearly we have the similar theorem for satisfiable instances.

THEOREM 4.2. *If P is useless on satisfiable instances then P is hereditarily approximation resistant on satisfiable instances.*

The standard way to prove that a predicate P is approximation resistant is to design a Probabilistically Checkable Proof (PCP) where the acceptance criterion is given by P and to prove that we have almost perfect completeness (i.e., correct proofs of correct statements are accepted with probability $1 - \epsilon$) and soundness $E_P + \epsilon$. Usually it is easy to analyze the completeness and the main difficulty is the soundness. In this analysis of soundness, P is expanded using the discrete Fourier transform and the expectation of each term is analyzed separately.

The most robust way of making this analysis is to prove that each non-constant monomial has expectation at most ϵ. As any real-valued function can be expanded by the Fourier transform this argument actually shows that the predicate in question is computationally useless. To show the principle we give a proof of the following theorem in Appendix A.

THEOREM 4.3. *For any $k \geq 3$, parity of k variables is computationally useless.*

As stated above, most approximation resistance results turn out to give uselessness without any or only minor modifications of the proofs. In particular, if one is looking for sparse useless predicates, the recent result of Chan [Chan 2012] implies that there is such a predicate of arity d that accepts at most $2d$ strings.

Turning to satisfiable instances, for arity 3, the predicate “Not-Two” is computationally useless even on satisfiable instances [Håstad 2012]. Considering sparse predicates of larger arity, the predicates defined by Håstad and Khot [Håstad and Khot 2005] which accepts 2^{4k} inputs and have arity $4k + k^2$, have the same property. This paper presents two different predicates with these parameters and although it is likely that the result holds for both predicates we have only verified this for the “almost disjoint sets PCP” (Section 3.2.1 of [Håstad and Khot 2005]).

If we are willing to assume the unique games conjecture by Khot [Khot 2002] we can use the results of [Austrin and Mossel 2009] to get very strong results.

THEOREM 4.4. *Let P be a predicate such that the strings accepted by P supports a pairwise independent measure. Then, assuming the unique games conjecture, P is computationally useless.*

This follows immediately from the proof of [Austrin and Mossel 2009] as the proof shows that the expectation of each non-constant character is small.

As the unique games conjecture has imperfect completeness there is no natural way to use it to prove that certain predicates are computationally useless on satisfiable instances.

5. THE MAIN USEFULNESS RESULT

In this section we present our main algorithm showing that Theorem 4.4 is best possible in that any predicate that does not support a pairwise independent measure is in fact not computationally useless. We have the following result which is proved in [Austrin and Håstad 2011] but, as it is natural and not very difficult, we suspect that it is not original to that paper.

THEOREM 5.1. *Suppose that the set of inputs accepted by predicate P does not support a pairwise independent measure. Then there is a real-valued quadratic polynomial Q such that $Q(x) > E_Q$ for any $x \in P^{-1}(1)$.*

PROOF. The full proof appears in [Austrin and Håstad 2011] but let us give a sketch of the proof. For each $x \in \{-1, 1\}^k$ we can define a point $x^{(2)}$ in $k + \binom{k}{2}$ real dimensions where the coordinates are given by the coordinates of x as well as any pairwise product of coordinates $x_i x_j$. The statement that a set S supports a pairwise independent measure is equivalent with the origin being in the convex hull of the points $\{x^{(2)} \mid x \in S\}$. If the origin is not in the convex hull of these points then there is a separating hyperplane such that its normal vector $\vec{c} \in \mathbb{R}^{k+\binom{k}{2}}$ satisfies $(\vec{c}, x^{(2)}) > 0$ for all $x \in P^{-1}(1)$, and this hyperplane defines the quadratic function Q via $Q(x) = (\vec{c}, x^{(2)})$. □

We now have the following theorem.

THEOREM 5.2. *Let P be a predicate whose accepting inputs do not support a pairwise independent measure and let Q be the quadratic function proved to exist by Theorem 5.1. Then P is useful for Q.*

PROOF. To make the situation more symmetric let us introduce a variable x_0 which always takes the value 1 and replace each linear term x_i by $x_0 x_i$ and drop any constant term in Q. This makes Q homogeneous of degree 2. Note that negating all inputs does not change the value of Q and thus any solution with $x_0 = -1$ can be transformed to a legitimate solution by negating all variables. As each term is unbiased we have $E_Q = 0$ and thus the goal is to find an assignment that gives $\sum Q(x_{a_j}^{b_j}) \geq \delta m$ for some absolute constant δ. Now let

$$C = \max_{x \in \{-1,1\}^k} -Q(x) \qquad\qquad c = \min_{x \in P^{-1}(1)} Q(x).$$

By assumption we have that c and C are fixed constants where c is strictly larger than 0. Let D be the sum of the absolute values of all coefficients of Q.

Let us consider our objective function

$$F(x) = \sum_{i=1}^{m} Q(x_{a_j}^{b_j})$$

which is a quadratic polynomial with the sum of the absolute values of coefficients bounded by Dm. As we are guaranteed that we have an assignment that satisfies at least $(1-\epsilon)m$ clauses we know that the optimal value of F is at least $(1-\epsilon)cm - \epsilon Cm \geq cm - (c+C)\epsilon m$.

Consider the standard semidefinite relaxation where we replace each product $x_i x_j$ by an inner product (v_i, v_j) for unit length vectors v_i. This semidefinite program can be solved with arbitrary accuracy and let us for notational convenience assume that we have an optimal solution which, by assumption, has an objective value at least $cm - (c+C)\epsilon m$.

To round the vector-valued solution back to a Boolean valued solution we use the following rounding guided by a positive constant B.

(1) Pick a random vector r by picking each coordinate to be an independent normal variable with mean 0 and variance 1.
(2) For each i if $|(v_i, r)| \leq B$ set $p_i = \frac{B+(v_i,r)}{2B}$ and otherwise set $p_i = \frac{1}{2}$.
(3) Set $x_i = 1$ with probability p_i independently for each i and otherwise $x_i = -1$.

Remember that if x_0 gets the value -1 we negate all variables. The lemma below is the key to the analysis.

LEMMA 5.3. *We have*

$$\left|\mathbb{E}[x_i x_j] - \frac{1}{B^2}(v_i, v_j)\right| \leq be^{-B^2/2}.$$

for some absolute constant b.

PROOF. If $|(v_i, r)| \leq B$ and $|(v_j, r)| \leq B$ then $\mathbb{E}[x_i x_j] = \frac{1}{B^2}\mathbb{E}_r[(v_i, r)(v_j, r)]$. Now it is not difficult to see that $\mathbb{E}_r[(v_i, r)(v_j, r)] = (v_i, v_j)$ and thus using the fact that $Pr[|(v_i, r)| > B] \leq \frac{b}{2}e^{-B^2/2}$ for a suitable constant b, the lemma follows. □

Taking all the facts together we get that the obtained Boolean solution has expected value at least

$$\frac{1}{B^2}(cm - (c + C)\epsilon m) - be^{-B^2/2}Dm.$$

If we choose $\epsilon = \frac{c}{2(c+C)}$ and then B a sufficiently large constant we see that this expected value is at least δm for some absolute constant δ. The theorem follows.

6. THE CASE OF NO NEGATION

In our definition we are currently allowing negation for free. Traditionally this has not been the choice in most of the CSP-literature. Allowing negations does make many situations more smooth but both cases are of importance and let us here outline what happens in the absence of negation. We call the resulting class Max-P^+.

In this case the situation is different and small changes in P may result in large difference in the performance of "trivial" algorithms. In particular, if P accepts the all-zero or all-one string then it is trivial to satisfy all constraints by setting each variable to 0 in the first case and each variable to 1 in the second case.

We propose to extend the set of trivial algorithms to allow the algorithm to find a bias $r \in [-1, 1]$ and then set all variables randomly with expectation r, independently. The algorithm to outperform is then the algorithm with the optimal value of r. Note that this algorithm is still oblivious to the instance as the optimal r depends solely on P. We extend the definition of E_Q for this setting.

Definition 6.1. For Q: $\{-1, 1\}^k \mapsto \mathbb{R}$ and $r \in [-1, 1]$, define

$$E_Q(r) = \mathop{\mathbb{E}}_{x \in \{-1,1\}^k_{(r)}} Q(x), \qquad E_Q^+ = \max_{r \in [-1,1]} E_Q(r),$$

where $\{-1, 1\}^k_{(r)}$ denotes the r-biased hypercube.

Using this definition we now get extensions of the definitions of approximation resistance and uselessness of Max-P^+, and we say that P is *positively approximation resistant* or *positively useless*.

6.1. Positive usefulness in the information theoretic setting

The results of Section 3 are not difficult to extend and we only give an outline. The main new component to address is the fact that 0 and 1 are not symmetric any longer.

As before let μ be a probability measure and let $\mu^{p,q}$ be the distribution obtained by first picking a string according to μ and then flipping each coordinate that is one to a zero with with probability p and each coordinate that is zero to one with probability q (of course all independently). For a given μ let $Opt^+(Q,\mu)$ be the maximum over p and q of the expected value of $Q(x)$ when x is chosen according to $\mu^{p,q}$. We have the following theorem.

THEOREM 6.2. *The predicate P is positively information-theoretically useless for Q if and only if there exists a measure supported on strings accepted by P such that $Opt^+(Q,\mu) = E_Q^+$.*

PROOF. The proof follows the proof of Theorem 3.2, and we leave the easy modifications to the reader. □

Let us return to the more interesting case of studying positive uselessness in the computational setting.

6.2. Positive usefulness in the computational setting

Also in this situation we can extend the result from the situation allowing negations by using very similar techniques. We first extend the hardness result Theorem 4.4 based on pairwise independence to this setting and we can now even allow a uniformly positively correlated distribution.

THEOREM 6.3. *Let P be a predicate such that the strings accepted by P supports a uniformly positively correlated distribution. Then, assuming the unique games conjecture, P is positively useless.*

A similar theorem was noted in [Austrin 2010], but that theorem only applied for pairwise independent distributions. The relaxed condition that the distribution only needs to be positively correlated is crucial to us as it allows us to get a tight characterization. As the proof of Theorem 6.3 has much in common with the proof of Theorem 8.3 stated below we give the proofs of both theorems in Section 9.

Let us turn to establishing the converse of Theorem 6.3. We start by extending Theorem 5.1.

THEOREM 6.4. *Suppose that the set of inputs accepted by predicate P does not support a uniformly positively correlated measure. Then there is a real-valued quadratic polynomial Q such that $Q(x) > E_Q^+$ for any $x \in P^{-1}(1)$. Furthermore, Q can be chosen such that the optimal bias r giving the value E_Q^+ satisfies $|r| < 1$.*

PROOF. As in the proof of Theorem 5.1 for each $x \in \{-1,1\}^k$ we can define a point $x^{(2)}$ in $k + \binom{k}{2}$ real dimensions where the coordinates are given by the coordinates of x as well as any pairwise product of coordinates $x_i x_j$. We consider two convex bodies, K_1 and K_2 where K_1 is the same body we saw in the proof of Theorem 5.1 – the convex hull of $x^{(2)}$ for all x accepted by P.

For each $b \in [-1,1]$ we have a point y^b with the first k coordinates equal to b and the rest of the coordinates equal to b^2. We let K_2 be the convex hull of all these points.

The hypothesis of the theorem is now equivalent to the fact that K_1 and K_2 are disjoint. Any hyperplane separating these two convex sets would be sufficient for the first part of the theorem but to make sure that the optimal r satisfies $|r| < 1$ we need to consider how to find this hyperplane more carefully.

Suppose p_2 is a point in K_2 such that $d(p_2, K_1)$, i.e., the distance from p_2 to K_1, is minimal. Furthermore let p_1 be the point in K_1 minimizing $d(p_1, p_2)$. One choice for the separating hyperplane is the hyperplane which is orthogonal to the line through

p_1 and p_2 and which intersects this line at the midpoint between p_1 and p_2. As in Theorem 5.1 we get a corresponding quadratic form, Q, and it is not difficult to see that the maximum of Q over K_2 is taken at p_2 (and possibly at some other points). Thus if we can make sure that p_2 does not equal $(1^k, 1^{\binom{k}{2}})$ or $(-1^k, 1^{\binom{k}{2}})$ we are done.

We make sure that this is the case by first applying a linear transformation to the space. Note that applying a linear transformation does not change the property that K_1 and K_2 are non-intersecting convex bodies but it does change the identity of the points p_1 and p_2.

As P does not support a uniformly positively correlated measure it does not accept either of the points 1^k or -1^k as a measure concentrated on such a point is uniformly positively correlated. This implies that K_1 is contained in the strip

$$\left|\sum_{i=1}^{k} y_i\right| \le k-2.$$

We also have that K_2 is contained in the strip

$$\left|\sum_{i=1}^{k} y_i\right| \le k,$$

and that it contains points with the given sum taking any value in the above interval. Furthermore the points we want avoid satisfy $|\sum_{i=1}^{k} y_i| = k$. Now apply a linear transformation that stretches space by a large factor in the direction of the vector $(1^k, 0^{\binom{k}{2}})$ while preserving the space in any direction orthogonal to this vector. It is easy to see that for a large enough stretch factor, none of the points $(1^k, 1^{\binom{k}{2}})$ or $(-1^k, 1^{\binom{k}{2}})$ can be the point in K_2 that is closest to K_1. The theorem follows. □

Given Theorem 5.2 the next theorem should be no surprise.

THEOREM 6.5. *Let P be a predicate whose set of accepting inputs does not support a uniformly positively correlated measure and Q be the quadratic function proved to exist by Theorem 6.4. Then P is positively useful for Q.*

PROOF. The proof is a small modification of the proof of Theorem 5.2 and let us only outline these modifications.

Let r be the optimal bias of the inputs to get the best expectation of Q and let us consider the expected value of $\frac{1}{m}\sum Q(x_{a_j}^{b_j})$ given that we set x_i to one with probability $(1 + r + y_i)/2$. This probability can be written a quadratic form in y_i and we want to optimize this quadratic form under the conditions that $|r + y_i| \le 1$ for any i. Note that the constant term is E_Q^+ and if we introduce a new variable y_0 that always takes the value 1 we can write the resulting expectation as

$$E_Q^+ + \sum_{i \ne j} c_{ij} y_i y_j, \tag{1}$$

for some real coefficients c_{ij}. As before we relax this to a semi-definite program by replacing the products $y_i y_j$ in (1) by inner products (v_i, v_j) and relaxing the constraints to

$$\|r v_0 + v_i\| \le 1,$$

for any $i \ge 1$ and $\|v_0\| = 1$. Solving this semi-definite program we are now in essentially the same situation as in the proof of Theorem 5.2. The fact that $|r| < 1$ ensures that a sufficiently large scaling of the inner products results in probabilities in the interval $[0, 1]$. We omit the details. □

Theorem 6.3 proves that having odd parity on four variables is positively useless but assumes the UGC. It seems natural to hope that this theorem could be establish based solely on $NP \neq P$, but we have been unable to do so.

Let us briefly outline the problems encountered. The natural attempt is to try a long-code based proof for label cover instance similar to the proof [Håstad 2001]. A major problem is that all currently known such proofs read two bits from the same long code. Considering functions Q that benefit from two such bits being equal gives us trouble through incorrect proofs where each individual long code is constant. For instance we currently do not know how to show that odd parity is not useful for the "exactly three" function based only on NP≠P.

7. ADAPTIVE USELESSNESS AND PSEUDORANDOMNESS

We now discuss the adaptive setting, when we allow the algorithm to choose the new objective function Q based on the Max-P instance. Formally, we make the following definition.

Definition 7.1. The predicate P is *adaptively useful*, if and only if there is an $\epsilon > 0$ such that there is a polynomial time algorithm which given a Max-P instance with value $1 - \epsilon$ finds an objective function $Q : \{-1, 1\}^k \to [0, 1]$ and an assignment x such that

$$\frac{1}{m}\sum_{j=1}^{m} Q(x_{a_j}^{b_j}) \geq \mathop{\mathbb{E}}_{x \in \{-1,1\}^k}[Q(x)] + \epsilon.$$

Note that we need to require Q to be bounded since otherwise the algorithm can win by simply scaling Q by a huge constant. Alternatively, it is easy to see that in the presence of negations, adaptive usefulness is equivalent with requiring that the algorithm finds an assignment x such that the distribution of the k-tuples $\{x_{a_j}^{b_j}\}_{j \in [m]}$ is ϵ-far in statistical distance from uniform for some $\epsilon > 0$ (not the same ϵ as above). In fact, since k is constant it is even equivalent with requiring that the min-entropy is bounded away from k, in particular that there is some $\alpha \in \{-1, 1\}^k$ and $\epsilon > 0$ such that at least a $2^{-k} + \epsilon$ fraction of the $x_{a_j}^{b_j}$'s attain the string α.

Adaptive uselessness trivially implies non-adaptive uselessness. In the other direction, with the interpretation of deviating from the uniform distribution over $\{-1, 1\}^k$, it is easy to see that the proof of the hardness result based on pairwise independence from the non-adaptive setting works also for adaptive uselessness.

This result can, by a slightly more careful argument, be extended also to the case without negations. The characterization is then that the algorithm is attempting to produce a distribution on k-tuples that is far from being uniformly positively correlated. In this setting, it does not seem meaningful to think of adaptive uselessness as a pseudorandomness property.

8. A NEW APPROXIMATION RESISTANCE RESULT

In this section we describe how the pairwise independence condition of [Austrin and Mossel 2009] can be relaxed somewhat to give approximation resistance for a wider range of predicates. Some examples illuminating this theorem are given in Appendix B.

8.1. Relaxed Pairwise Independence Conditions

We first define the specific kind of distributions whose existence give our hardness result.

Definition 8.1. A distribution μ over $\{-1,1\}^k$ *covers* $S \subseteq [k]$ if there is an $i \in S$ such that $\mathbb{E}_\mu[x_i] = 0$ and $\mathbb{E}_\mu[x_i x_j] = 0$ for every $j \in S \setminus \{i\}$.

Definition 8.2. Fix a function $Q : \{-1,1\}^k \to \mathbb{R}$ and a pair of coordinates $\{i,j\} \subseteq [k]$. We say that a distribution μ over $\{-1,1\}^k$ is $\{i,j\}$-negative with respect to Q if $\mathbb{E}_\mu[x_i] = \mathbb{E}_\mu[x_j] = 0$ and $\mathrm{Cov}_\mu[x_i, x_j]\hat{Q}(\{i,j\}) \le 0$.

Our most general condition for approximation resistance (generalizing both Theorem B.3 in the appendix and [Austrin and Mossel 2009]) is as follows.

THEOREM 8.3. *Let $P : \{-1,1\}^k \to \{-1,1\}$ be a predicate and let $Q : \{-1,1\}^k \to \mathbb{R}$ be a real valued function. Suppose there is a probability distribution μ supported on P with the following properties:*

- *For each pair $\{i,j\} \subseteq [k]$, it holds that μ is $\{i,j\}$-negative with respect to Q*
- *For each $S \neq \emptyset, |S| \neq 2$ such that $\hat{Q}(S) \neq 0$, it holds that μ covers S*

Then P is not useful for Q, assuming the Unique Games Conjecture. In particular, if the conditions are true for $Q = P$, then P is approximation resistant.

We are not aware of any approximation resistant predicates that do not satisfy the conditions given in Theorem 8.3. On the other hand we see no reason to believe that it is tight.

9. PROOFS OF UG-HARDNESS

In this section we give the proofs of the extensions Theorems 6.3 and 8.3 of [Austrin and Mossel 2009]. It is well-known that the key part in deriving UG-hardness for a CSP is to design *dictatorship tests* with appropriate properties — see e.g. [Raghavendra 2008] for details.

9.1. Background: Polynomials, Quasirandomness and Invariance

To set up the dictatorship test we need to mention some background material.

For $b \in [-1,1]$, we use $\{-1,1\}^n_{(b)}$ to denote the n-dimensional Boolean hypercube with the b-biased product distribution, i.e., if x is a sample from $\{-1,1\}^n_{(b)}$ then the expectation of i'th coordinate is $\mathbb{E}[x_i] = b$ (equivalently, $x_i = 1$ with probability $(1+b)/2$), independently for each $i \in [n]$). Whenever we have a function $f : \{-1,1\}^n_{(b)} \to \mathbb{R}$ we think of it as a random variable and hence expressions like $\mathbb{E}[f]$, $\mathrm{Var}[f]$, etc, are interpreted as being with respect to the b-biased distribution. We equip $L^2(\{-1,1\}^n_{(b)})$ with the inner product $\langle f, g\rangle = \mathbb{E}[f \cdot g]$ for $f, g : \{-1,1\}^n_{(b)} \to \mathbb{R}$.

For $S \subseteq [n]$ define $\chi_S : \{-1,1\}^n_{(b)} \to \mathbb{R}$ by

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

where $\chi : \{-1,1\}_{(b)} \to \mathbb{R}$ is defined by

$$\chi(x_i) = \frac{x_i - \mathbb{E}[x_i]}{\sqrt{\mathrm{Var}[x_i]}} = \begin{cases} -\sqrt{\frac{1+b}{1-b}} & \text{if } x_i = -1 \\ \sqrt{\frac{1-b}{1+b}} & \text{if } x_i = 1 \end{cases}.$$

The functions $\{\chi_S\}_{S\subseteq[n]}$ form an orthonormal basis with respect to the inner product $\langle\cdot,\cdot\rangle$ on $L^2(\{-1,1\}^n_{(b)})$ and thus any function $f:\{-1,1\}^n_{(b)}\to\mathbb{R}$ can be written as

$$f(x)=\sum_{S\subseteq[n]}\hat{f}(S;b)\chi_S(x),$$

where $\hat{f}(S;b)$ are the Fourier coefficients of f (with respect to the b-biased distribution).

With this view in mind it is convenient to think of functions f in $L^2(\{-1,1\}^n_{(b)})$ as multi-linear polynomials $F:\mathbb{R}^n\to\mathbb{R}$ in the random variables $X_i=\chi(x_i)$, viz.,

$$F(X)=\sum_{S\subseteq[n]}\hat{f}(S;b)\prod_{i\in S}X_i.$$

We say that such a polynomial is (d,τ)-*quasirandom* if for every $i\in[n]$ it holds that

$$\sum_{\substack{i\in S\subseteq[n]\\|S|\le d}}\hat{f}(S;b)^2\le\tau.$$

A function $f:\{-1,1\}^n_{(b)}\to\mathbb{R}$ is said to a be a *dictator* if $f(x)=x_i$ for some $i\in[n]$, i.e., f simply returns the i'th coordinate. The polynomial F corresponding to a dictator is $F(X)=b+\sqrt{1-b^2}X_i$. Note that a dictator is in some sense the extreme opposite of a (d,τ)-quasirandom function as a dictator is not $(1,\tau)$-quasirandom for $\tau<1-b^2$.

We are interested in distributions μ over $\{-1,1\}^k$. In a typical situation we pick n independent samples of μ, resulting in k strings $\vec{x}_1,\dots,\vec{x}_k$ of length n, and to each such string we apply some function $f:\{-1,1\}^n\to\{-1,1\}$. With this in mind, define the following $k\times n$ matrix X of random variables. The j'th column which we denote by X^j has the distribution obtained by picking a sample $x\in\{-1,1\}^k$ from μ and letting $X^j_i=\chi(x_i)$, independently for each $j\in[n]$. Then, the distribution of $(f(\vec{x}_1),\dots,f(\vec{x}_k))$ is the same as the distribution of $(F(X_1),\dots,F(X_k))$, where X_i denotes the i'th row of X.

Now, we are ready to state the version of the invariance principle [Mossel et al. 2010; Mossel 2010] that we need.

THEOREM 9.1. *For any $\alpha>0,\epsilon>0,b\in[-1,1],k\in\mathbb{N}$ there are $d,\tau>0$ such that the following holds. Let μ be any distribution over $\{-1,1\}^k$ satisfying:*

(*1*) $\mathbb{E}_{x\sim\mu}[x_i]=b$ *for every $i\in[k]$ (i.e., all biases are identical).*
(*2*) $\mu(x)\ge\alpha$ *for every $x\in\{-1,1\}^k$ (i.e., μ has full support).*

Let X be the $k\times n$ matrix defined above, and let Y be a $k\times n$ matrix of standard jointly Gaussian variables with the same covariances as X. Then, for any (d,τ)-quasirandom multi-linear polynomial $F:\mathbb{R}^n\to\mathbb{R}$, it holds that

$$\left|\mathbb{E}\left[\prod_{i=1}^k F(X)\right]-\mathbb{E}\left[\prod_{i=1}^k F(Y)\right]\right|\le\epsilon.$$

9.2. The Dictatorship Test

The dictatorship tests we use to prove Theorems 6.3 and 8.3 are both instantiations of the test used in [Austrin and Mossel 2009], with slightly different analyses, so we start by recalling how this test works.

In what follows we extend the domain of our predicate $P:\{-1,1\}^k\to\{0,1\}$ to $[-1,1]^k$ multi-linearly. Thus, we have $P:[-1,1]^k\to[0,1]$.

To prove hardness for Max-P, one analyzes the performance of the dictatorship test in Figure 1, which uses a distribution μ over $\{-1,1\}^k$ that we assume is supported on $P^{-1}(1)$ and satisfies condition (1) of Theorem 9.1, which is the case in both Theorems 6.3 and 8.3.

```
    INPUT: A function $f : \{-1,1\}^L \to [-1,1]$
    OUTPUT: Accept/Reject
(1) Let $\mu_\epsilon = (1-\epsilon)\mu + \epsilon U_b$, where $U_b$ denotes the product distribution over $\{-1,1\}^k$
    where each bit has bias $b$.
(2) Pick $L$ independent samples from $\mu_\epsilon$, giving $k$ vectors $\vec{x}_1, \ldots, \vec{x}_k \in \{-1,1\}^L$.
(3) Accept with probability $P(f(\vec{x}_1), \ldots, f(\vec{x}_k))$.
```

Fig. 1. Dictatorship test

The completeness property of the test is easy to establish (and only depends on μ being supported on strings accepted by P).

LEMMA 9.2. *If f is a dictatorship function then A accepts with probability $\geq 1-\epsilon$.*

For the soundness analysis, the arguments are going to be slightly different for the two Theorems 6.3 and 8.3. It is convenient to view f in its multi-linear form as described in the previous section. Thus, instead of looking at $f(\vec{x}_1), \ldots, f(\vec{x}_k)$ we look at $F(X_1), \ldots, F(X_k)$. In both cases, the goal is to prove that there are d and τ such that if F is (d,τ)-quasirandom then the expectation of $Q(F(X_1), \ldots, F(X_k))$ is small (at most $E_Q + \epsilon$ for Theorem 8.3 and at most $E_Q^+ + \epsilon$ for Theorem 6.3).

In general, it is also convenient to apply the additional guarantee that F is balanced (i.e., satisfying $\mathbb{E}[F(X)] = 0$). This can be achieved by the well-known trick of folding, and is precisely what causes the resulting Max-P instances to have negated literals. In other words, when we prove Theorem 6.3 on the hardness of Max-P^+, we are not going to be able to assume this.

Theorem 8.3: Relaxed Approximation Resistance Conditions. The precise soundness property for Theorem 8.3 is as follows.

LEMMA 9.3. *Suppose μ is a distribution as in the statement of Theorem 8.3 and that the function F is folded. Then for every $\epsilon > 0$ there are d, τ such that whenever F is (d,τ)-quasirandom then*

$$\mathbb{E}[Q(F(X_1), \ldots, F(X_k))] \leq E_Q + \epsilon.$$

PROOF. We write $Q(x) = \sum_{S \subseteq [k]} \hat{Q}(S) \prod_{i \in S} x_i$, where $\hat{Q}(S)$ are the Fourier coefficients of Q with respect to the uniform distribution, and $\hat{Q}(\emptyset) = E_Q$.

We analyze the expectation of Q term by term. Fix a set $\emptyset \neq S \subseteq [k]$ and let us analyze $\mathbb{E}[\prod_{i \in S} F(X_i)]$. Let d, τ be the values given by Theorem 9.1, when applied with ϵ chosen as $\epsilon/2^k$ and the α given by the distribution μ_ϵ (note that this distribution satisfies the conditions of Theorem 9.1). There are two cases.

Case 1: $|S| = 2$. Let $S = \{i,j\}$. The conditions on μ guarantee that μ is $\{i,j\}$-negative with respect to Q, i.e., for any column a we have $\mathbb{E}[X_i^a] = \mathbb{E}[X_j^a] = 0$ and $\hat{Q}(S)\,\mathbb{E}[X_i^a X_j^a] \leq 0$. Let $\rho = \mathbb{E}[X_i^a X_j^a]$ (as each column a is identically distributed this value does not depend on a). Then we have

$$\hat{Q}(S)\,\mathbb{E}[F(X_i)F(X_j)] = \hat{Q}(S)\,\mathbb{S}_\rho(F)$$

where $\mathbb{S}_\rho(F)$ is the noise stability of F at ρ. Moreover, since the function F is folded it is an odd function, which in particular implies that $\mathbb{S}_\rho(F)$ is odd as well so that $\operatorname{sgn}(\mathbb{S}_\rho(F)) = \operatorname{sgn}(\rho)$. Since $\hat{Q}(S)\rho \le 0$ it follows that $\hat{Q}(S)\,\mathbb{S}_\rho(F) \le 0$ as well, so S can not even give a positive contribution to the acceptance probability.

Case 2: $S \neq \emptyset, |S| \neq 2$, $\hat{Q}(S) \neq 0$. This is the more interesting case. The conditions on μ guarantee that μ covers S, i.e., there is an $i^* \in S$ such that $\mathbb{E}[X^a_{i^*}] = 0$ and $\mathbb{E}[X^a_{i^*} X^a_j] = 0$ for all $j \in S \setminus \{i^*\}$. By Theorem 9.1, we know that if F is (d, τ)-quasirandom we have

$$\left| \mathbb{E}\left[\prod_{i \in S} F(X_i)\right] - \mathbb{E}\left[\prod_{i \in S} F(Y_i)\right] \right| \le \epsilon/2^k,$$

where Y is a jointly Gaussian matrix with the same first and second moments as X. But then, for every column a, the conditions on the second moments imply that $Y^a_{i^*}$ is a standard Gaussian completely independent from all other entries of Y. This implies that

$$\mathbb{E}\left[\prod_{i \in S} F(Y_{i^*})\right] = \mathbb{E}\left[F(Y_{i^*})\right] \cdot \mathbb{E}\left[\prod_{i \in S \setminus \{i^*\}} F(Y_i)\right] = 0,$$

where the second equality is by the assumption that F is folded. This implies that S has at most a negligible contribution of $\epsilon/2^k$ to the acceptance probability of the test.

Combining the two cases, it immediately follows that

$$\mathbb{E}[Q(F(X_1), \ldots, F(X_k))] = \sum_{S \subseteq [k]} \hat{Q}(S)\,\mathbb{E}\left[\prod_{i \in S} F(X_i)\right] \le E_Q + \epsilon.$$

□

Theorem 6.3: No Negations. As mentioned earlier, in the case when negated literals are not allowed, we can no longer assume that F is folded. Furthermore, the distribution μ over $\{-1,1\}^k$ used is only assumed to be pairwise uniformly correlated. The precise soundness is as follows.

LEMMA 9.4. *Suppose μ is a uniformly positively correlated distribution. Then for every $Q : [-1,1]^k \to [-1,1]$ and $\epsilon > 0$ there are d, τ such that whenever F is (d, τ)-quasirandom then*

$$\mathbb{E}[Q(F(X_1), \ldots, F(X_k))] \le E^+_Q + \epsilon.$$

PROOF. Similarly to the previous lemma, we are going to take d, τ to be the values given by Theorem 9.1 with ϵ chosen as $\frac{\epsilon}{2 \cdot 2^k}$.

Note that since μ_ϵ is a combination of μ and U_b, both being uniformly positively correlated, μ_ϵ is also uniformly positively correlated.

Let $b = \mathbb{E}_{x \sim \mu_\epsilon}[x_i]$ and $\rho = \mathbb{E}_{x \sim \mu_\epsilon}[x_i x_j] \ge b^2$ be the bias and correlation of μ_ϵ, respectively. Define a new distribution η over $\{-1,1\}^k$ as

$$\eta = c U_{\sqrt{\rho}} + (1-c) U_{-\sqrt{\rho}},$$

where $c = \frac{b + \sqrt{\rho}}{2\sqrt{\rho}} \in [0,1]$ (recall that $U_{\sqrt{\rho}}$ denotes the product distribution where all biases are $\sqrt{\rho}$).

Then η has the same first and second moments as μ_ϵ and therefore, writing Z for the corresponding matrix from η, we can apply Theorem 9.1 twice and see that for every $S \subseteq [k]$

$$\left| \mathbb{E}\left[\prod_{i \in S} F(X_i)\right] - \mathbb{E}\left[\prod_{i \in S} F(Z_i)\right] \right| \leq \epsilon/2^k,$$

implying

$$|\mathbb{E}[Q(F(X_1), \ldots, F(X_k))] - \mathbb{E}[Q(F(Z_1), \ldots, F(Z_k))]| \leq \epsilon.$$

Now, the column Z^1 of Z can be written as a convex combination of two product distributions R^+ and R^- over $\mathbb{R}^k$ (resulting from applying the character χ to $U_{\sqrt{\rho}}$ and $U_{-\sqrt{\rho}}$, respectively). By linearity of expectation, we can replace Z^1 with one of R^+ and R^- without decreasing the expectation of $Q(F(\cdot), \ldots, F(\cdot))$. Repeating this for all columns, we end up with a random matrix W, each column of which is either distributed like R^+ or like R^-, and satisfying

$$\mathbb{E}[Q(F(W_1), \ldots, F(W_k))] \geq \mathbb{E}[Q(F(Z_1), \ldots, F(Z_k))].$$

But now since each column of W is distributed according to a product distribution (with identical marginals), the rows of W are independent and identically distributed, implying that

$$\mathbb{E}[Q(F(W_1), \ldots, F(W_k))] \leq E_Q^+.$$

Combining all our inequalities, we end up with

$$\mathbb{E}[Q(F(X_1), \ldots, F(X_k))] \leq E_Q^+ + \epsilon,$$

as desired. □

10. CONCLUDING REMARKS

We have introduced a notion of (computational) uselessness of constraint satisfaction problems, and showed that, assuming the unique games conjecture, this notion admits a very clean and nice characterization. This is in contrast to the related and more well-studied notion of approximation resistance, where the indications are that a characterization, if there is a reasonable one, should be more complicated.

Our inability to obtain any non-trivial NP-hardness results for positive uselessness, instead of Unique Games-based hardness is frustrating. While [Håstad 2001] proves odd parity of four variables to be positively approximation resistant, obtaining positive uselessness by the same method appears challenging.

Another direction of future research is understanding uselessness in the completely satisfiable case.

We have focused on CSPs defined by a single predicate P (with or without negated literals). It would be interesting to generalize the notion of usefulness to a general CSP (defined by a family of predicates). Indeed, it is not even clear what the correct definition is in this setting, and we leave this as a potential avenue for future work. Another possible direction is to consider an analogous notion for the decision version of a CSP rather than the optimization version.

Acknowledgement. We are grateful to a number of anonymous referees for useful comments on the presentation of this paper.

REFERENCES

Per Austrin. 2010. Improved Inapproximability for Submodular Maximization. In *APPROX-RANDOM*. 12–24.

P. Austrin and J. Håstad. 2011. Randomly supported independence and resistance. *SIAM J. Comput.* 40 (2011), 1–27.

P. Austrin and J. Håstad. 2012. On the Usefulness of Predicates. In *Proceeings of 27th Annual IEEE Conference on Computational Complexity*. 53–63.

P. Austrin and E. Mossel. 2009. Approximation Resistant Predicates from Pairwise Independence. *Computational Complexity* 18 (2009), 249–271.

S. O. Chan. 2012. Approximation Resistance from Pairwise Independent Subgroups. (2012). ECCC techical report, 2012, No 110.

M. Charikar and A. Wirth. 2004. Maximizing quadratic programs: extending Grothendieck's inequality. In *Proceedings of 45th Annual IEEE Symposium of Foundations of Computer Science*. 54–60.

V. Guruswami, D. Lewin, M. Sudan, and L. Trevisan. 1998. A tight characterization of NP with 3 query PCPs.. In *Proceedings of 39th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, Palo Alto, 8–17.

G. Hast. 2005. *Beating a random assignment*. KTH, Stockholm. Ph.D Thesis.

J. Håstad. 2001. Some optimal inapproximability results. *Journal of ACM* 48 (2001), 798–859.

J. Håstad. 2007. On the efficient approximability of constraint satisfaction problems. In *Surveys in Combinatorics 2007*, A. Hilton and J.Talbot (Eds.), Vol. 346. Cambridge University Press, 201–222.

J. Håstad. 2012. On the NP-hardness of Max-Not-2. In *Proceedings of Approx 2012, Springer Lecture Notes in Computer Science, Vol 7408*. 170–181.

J. Håstad and S. Khot. 2005. Query efficient PCPs with perfect completeness. *Theory of Computing* 1 (2005), 119–149.

S. Khot. 2002. On the power of unique 2-Prover 1-Round games. In *Proceedings of 34th ACM Symposium on Theory of Computating*. 767–775.

E. Mossel. 2010. Gaussian Bounds for Noise Correlation of Functions. *GAFA* 19 (2010), 1713–1756.

E. Mossel, R. O'Donnell, and K. Oleszkiewicz. 2010. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics* 171, 1 (2010), 295–341. http://front.math.ucdavis.edu/0503.5503

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

Prasad Raghavendra. 2008. Optimal algorithms and inapproximability results for every CSP. In *In Proc. 40 th ACM STOC*. 245–254.

U. Zwick. 1998. Approximation algorithms for constraint satisfaction problems involving at most three variables per constraint. In *Proceedings 9th Annual ACM-SIAM Symposium on Discrete Algorithms*. ACM, 201–210.

A. PROOF OF THEOREM 4.3

THEOREM 4.3 RESTATED. *For any $k \geq 3$, parity of k variables is computationally useless.*

PROOF. To avoid cumbersome notation let us only give the proof in the case $k = 3$. We assume the reader is familiar with the PCP defined for this case in [Håstad 2001] to prove that Max-3-Lin is approximation resistant. We claim that the same instances show that 3-Lin is computationally useless.

Indeed consider an arbitrary $Q : \{-1, 1\}^3 \to \mathbb{R}$ and consider its Fourier-expansion

$$Q(x) = \sum_{S \subseteq [3]} \hat{Q}_S \chi_S(x). \tag{2}$$

Now we need to consider $\frac{1}{m} \sum_{i=1}^{m} Q(x_{a_j}^{b_j})$ and we can expand each term using (2) and switch the the order of summation. Remember that $\hat{Q}_\emptyset = E_Q$ and thus we need to make sure that

$$\frac{1}{m} \sum_{i=1}^{m} \chi_S(x_{a_j}^{b_j}) \tag{3}$$

is small for any non-empty S (unless there is a good strategy for the provers in the underlying two-prover game). This is done in [Håstad 2001] for $S = \{1, 2, 3\}$ as this is the only Fourier coefficient that appears in the expansion of parity itself.

For smaller, non-empty, S, it is easy to see that (3) equals 0. Bits read corresponding $A(f)$ and $B(g_i)$ are pairwise independent and pairing terms for f and $-f$ proves that $\mathbb{E}[B(g_1)B(g_2)] = 0$.

The result follows in the case of parity of 3 variables and the extension to the general case is straightforward and left to the reader. □

B. SOME EXAMPLES

Let us first recall one of the few known examples of a predicate that is approximation resistant but not hereditarily approximation resistant.

Example B.1. Consider the predicate $GLST : \{-1, 1\}^4 \to \{0, 1\}$ defined by

$$GLST(x_1, x_2, x_3, x_4) = \begin{cases} x_2 \neq x_3 & \text{if } x_1 = -1 \\ x_2 \neq x_4 & \text{if } x_1 = 1 \end{cases}.$$

This predicate was shown to be approximation resistant by Guruswami et al. [Guruswami et al. 1998], but there is no pairwise independent distribution supported on its accepting assignments – indeed it is not difficult to check that $x_2x_3 + x_2x_4 + x_3x_4 < 0$ for all accepting inputs. This predicate also implies $NAE(x_2, x_3, x_4)$, the not-all-equal predicate and this is known to be non-trivially approximable [Zwick 1998].

When the predicate $GLST$ is proved to be approximation resistant in [Guruswami et al. 1998] the crucial fact is that not all terms appear in the Fourier expansion of P. We have

$$GLST(x_1, x_2, x_3, x_4) = \frac{1}{2} - \frac{x_2x_3}{4} - \frac{x_2x_4}{4} + \frac{x_1x_2x_3}{4} - \frac{x_1x_2x_4}{4}.$$

The key is that no term in the expansion contains both of the variables x_3 and x_4, corresponding to two questions in the PCP that are very correlated and hence giving terms that are hard to control.

In other words, when proving approximation resistance it suffices to only analyze those terms appearing in the Fourier expansion of a predicate P. In the context of the pairwise independent condition (which only gives UG-hardness, not NP-hardness), this means that it suffices to find a distribution which is pairwise independent on those pairs of variables that appear together in some term.

However, these are not the only situations where we can deduce that P is approximation resistant.

Example B.2. Consider the predicate

$$P(x_1, x_2, x_3, x_4) = GLST(x_1, x_2, x_3, x_4) \vee (x_1 = x_2 = x_3 = x_4 = 1),$$

which is the $GLST$ predicate with the all-ones string as an additional accepting assignment. One can check that there is no pairwise independent distribution supported on $P^{-1}(1)$, and since P has an odd number of accepting assignments, all its Fourier coefficients are non-zero. However, Max-P is known to be approximation resistant [Hast 2005].

The result of [Hast 2005] proving that this predicate is resistant is somewhat more general. In particular, it says the following.

THEOREM B.3 ([HAST 2005], THEOREM 6.5). *Let $P : \{-1, 1\}^4 \to \{0, 1\}$ be a predicate on 4 bits. Suppose $\hat{P}(\{3, 4\}) \geq 0$ and that P accepts all strings $x_1x_2x_3x_4$ with $\prod_{i=1}^{3} x_i = -1$ and $x_3 = -x_4$. Then P is approximation resistant.*

The statement of [Hast 2005], Theorem 6.5 is slightly different. The above statement is obtained by flipping the sign of x_3 in the statement of [Hast 2005]. If is not difficult to see that Theorem 8.3 extends Theorem B.3 to a much larger class of predicates (but giving UG-hardness, whereas [Hast 2005] gives NP-hardness).

Example B.4 (*Example B.2 continued*). Consider the distribution μ used to prove approximation resistance for $GLST$, i.e., the uniform distribution over the four strings $x_1x_2x_3x_4$ satisfying $x_1x_2x_3 = -1$ and $x_3 = -x_4$ (note that the condition of Theorem B.3 is precisely that P should accept these inputs). First, it satisfies

$$\hat{P}(\{3,4\}) \mathop{\mathbb{E}}_{\mu}[x_3x_4] = \frac{1}{16} \cdot (-1) < 0,$$

and all other pairwise correlations are 0, so μ satisfies the $\{i,j\}$-negativity condition of Theorem 8.3. Further, for $|S| > 2$ it holds that either x_1 or x_2 is in S. Since $\mathbb{E}_\mu[x_1] = 0$ and $\mathbb{E}_\mu[x_1x_j] = 0$ for all $j \neq 1$ (and similarly for x_2), this shows that any $|S| > 2$ is covered by μ. Finally since all $\mathbb{E}_\mu[x_i] = 0$, all four singleton S are also covered by μ. Hence Theorem 8.3 implies that Max-P is resistant (under the UGC).

Example B.5. Consider the predicate $P(x) = x_1 \oplus ((x_2 \oplus x_3) \vee x_4)$. This predicate is known to be approximation resistant [Hast 2005]. Let us see how to derive this conclusion using Theorem 8.3 (albeit only under the UGC). The Fourier expansion of P is

$$P(x) = \frac{1}{2} + \frac{x_1}{4} - \frac{x_1x_4}{4} - \frac{x_1x_2x_3}{4} - \frac{x_1x_2x_3x_4}{4},$$

and the distribution we use is uniform over:

$$\{\, x \in \{-1,1\}^4 \mid x_1x_2x_3 = -1, x_4 = 1\}.$$

Each of x_1, x_2, x_3 are unbiased, and x_4 is completely biased but as it does not appear as a singleton in the expansion of P this is not an issue. Further, all pairwise correlations are 0, and it is easy to check that this is sufficient for Theorem 8.3 to apply.

We only used Theorem 8.3 to get approximation resistance in a few examples. It can also be used to give interesting examples of P and Q such that P is not useful for Q but we leave the creation of such examples to the reader.

C. RESISTANT SIGNS OF QUADRATIC FORMS

In this section we consider a slightly different example from the ones considered in Section B. Suppose that in the definition of uselessness we only considered predicates Q rather than arbitrary real-valued functions. Would we get the same set of useless predicates?

The answer to this question is not obvious. Any real-valued function Q can be written in the form

$$Q(x) = q_0 + \sum_i c_i P_i(x)$$

where the sum is over different predicates and each coefficient c_i is non-negative. This implies that if P is useful for a real-valued function Q then there is a collection of predicates $\{P_i\}$ such that on any instance we can do better than random on one of these predicates. However, it does not imply that there is a single predicate P' for which P is useful. On the other hand it does imply that the standard proofs of uselessness for P can not work since these show that P is useless for all P' on the same instance.

It is natural for any P that does not support a pairwise independent distribution to try to find a predicate P' such that P is useful for P'. Given the discussion of Section 5 a very natural candidate is,

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

where Q is the quadratic form guaranteed by Theorem 5.1. Note that it may or may not be the case that $P = P'$. We now present an example to show that this choice does not always work.

THEOREM C.1. *There is a predicate, P, of the form $\operatorname{sgn}(Q(x))$ where Q is a quadratic function without a constant term that is approximation resistant (assuming the UGC).*

PROOF. Let L_1 and L_2 be two linear forms with integer coefficients which only assume odd values and only depends on variables x_i for $i \geq 3$. Define

$$Q(x) = 10(L_1(x) + x_1)(L_2(x) + x_2) + x_1 L_2(x) + 2x_2 L_1(x), \tag{4}$$

and let $P(x) = \operatorname{sgn}(Q(x))$. We establish the following properties of P.

(1) For all α such that $\{1, 2\} \subseteq \alpha$ we have $\hat{P}_\alpha = 0$.
(2) There is a probability distribution μ supported on strings accepted by P such that $\mathbb{E}_\mu[x_i] = 0$ for all i and $\mathbb{E}_\mu[x_i x_j] = 0$ for all $i < j$ with $(i, j) \neq (1, 2)$.

These two conditions clearly make it possible to apply Theorem 8.3. Loosely speaking the second condition makes it possible to construct at PCP such that we can control sums over all nontrivial characters except those that contain both 1 and 2. The first conditions implies that these troublesome terms do not appear in the expansion of P and hence we can complete the analysis.

We claim that property 1 is equivalent to the statement that every setting of the variables x_i for $i \geq 3$ results in a function on x_1 and x_2 that has the Fourier coefficient of size 2 equal to 0. In other words it should be a constant, one of the variables x_1 or x_2 or the negation of such a variable. Let us check that this is the case for Q defined by (4).

Fix any value of x_i, $i \geq 3$ and we have the following cases.

(1) $|L_1| \geq 3$ and $|L_2| \geq 3$.
(2) $|L_1| = 1$ and $|L_2| \neq 1$.
(3) $|L_2| = 1$.

In first case clearly the first term determines the sign of Q and $P = \operatorname{sgn}(L_1(x)L_2(x))$ and in particular the sub-function is independent of x_1 and x_2.

The second case is almost equally straightforward. When $x_1 = L_1(x)$ then the first term dominates and the answer is $\operatorname{sgn}(x_1 L_2(x))$. When $x_1 = -L_1(x)$ the first term is 0 and as $|L_2(x)x_1| \geq 3$ while $|L_1(x)x_2| = 1$ the answer also in this case is $\operatorname{sgn}(x_1 L_2(x))$.

Finally let us consider the third case. Then if $x_2 = L_2(x)$ our function Q reduces to

$$20(L_1(x) + x_1)x_2 + x_2(2L_1(x) + x_1)$$

and any nonzero term of this sum has sign $\operatorname{sgn}(x_2 L_1(x))$. Finally if $x_2 = -L_2(x)$ we get

$$x_2(2L_1(x) - x_1)$$

and again the sign is that of $\operatorname{sgn}(x_2 L_1(x))$. We conclude that in each case we have one of the desired functions and property 1 follows.

We establish property 2 in the case when each L_i is the sum of 5 variables not occurring in the other linear form. Thus for example we might take

$$L_1(x) = x_3 + x_4 + x_5 + x_6 + x_7$$

and

$$L_2(x) = x_8 + x_9 + x_{10} + x_{11} + x_{12}.$$

We describe the distribution μ in a rather indirect way to later be able to analyze it. Let $c = \frac{7-\sqrt{41}}{8} \approx .0746$.

(1) Fix $|L_1(x)|$ to 1,3 or 5 with probabilities $\frac{1}{2}+2c$, $\frac{1}{2}-3c$, and c, respectively.
(2) Fix $|L_2(x)|$ to 1 or 3 each with a probability $\frac{1}{2}$.
(3) Pick a random $b \in \{-1,1\}$ taking each value with probability $\frac{1}{2}$.
(4) Suppose $|L_1(x)| \geq 3$ and $|L_2(x)| \geq 3$. Set $\operatorname{sgn}(L_1(x)) = \operatorname{sgn}(L_2(x)) = b$ and $x_1 = x_2 = -b$.
(5) Suppose $|L_2| \neq 1$ and $|L_1| = 1$. Set $\operatorname{sgn}(L_1(x)) = -\operatorname{sgn}(L_2(x)) = b$ and $x_1 = x_2 = -b$.
(6) Suppose $|L_2| = 1$. Set $\operatorname{sgn}(L_1(x)) = x_1 = x_2 = b$ and $\operatorname{sgn}(L_2(x))) = -b$ with probability $\frac{1+12c}{2}$ and $\operatorname{sgn}(L_2(x)) = b$ with probability $\frac{1-12c}{2}$.
(7) Choose x_i for $i \geq 3$ uniformly at random given the values $L_1(x)$ and $L_2(x)$.

Now first note that by the analysis in establishing property 1 we always pick an assignment such that $Q(x) > 0$. This follows as in the three cases the output of the function is $\operatorname{sgn}(L_1(x)L_2(x))$, $\operatorname{sgn}(x_1 L_2(x))$, and $\operatorname{sgn}(L_1(x)x_2)$, respectively and they are all chosen to be b^2.

As b is a random bit, it is easy to see that $\mathbb{E}[x_i] = 0$ for any i and we need to analyze $\mathbb{E}[x_i x_j]$ for $(i,j) \neq (1,2)$. In our distribution we always have $x_1 = x_2$ and the variables in L_1 and L_2 are treated symmetrically and hence it is sufficient to establish the following five facts.

(1) $\mathbb{E}[L_1^2(x)] = 5$.
(2) $\mathbb{E}[L_2^2(x)] = 5$.
(3) $\mathbb{E}[x_1 L_1(x)] = 0$.
(4) $\mathbb{E}[x_1 L_2(x)] = 0$.
(5) $\mathbb{E}[L_1(x)L_2(x)] = 0$.

The first expected value equals

$$\left(\frac{1}{2}+2c\right)\cdot 1 + \left(\frac{1}{2}-3c\right)\cdot 9 + 25\cdot c = 5$$

while the second equals

$$\frac{1\cdot 1 + 1\cdot 9}{2} = 5.$$

For the third expected value note that $x_1 L_1(x) = -|L_1(x)|$ when $L_2(x) = 3$ while it equals $x_1 L_1(x) = |L_1(x)|$ when $L_2(x) = 1$. The two cases happens each with probability $1/2$ and as $|L_1(x)|$ is independent of $|L_2(x)|$ the equality follows.

To analyze the fourth value, first observe that conditioned on $|L_2(x)| = 1$ we have $\mathbb{E}[x_1 L_2(x)] = -12c$. On the other hand when $|L_2(x)| = 3$ we have

$$\mathbb{E}[x_1 L_2(x)] = 3(\frac{1}{2}+2c) - 3(\frac{1}{2}-3c) - 3c = 12c,$$

giving the result in this case. Finally, conditioned on $|L_2(x)| = 1$ we have

$$\mathbb{E}[L_1(x)L_2(x)] = -12c\left((\frac{1}{2}+2c) + 3(\frac{1}{2}-3c) + 5c\right) = -(24c - 24c^2)$$

and conditioned on $|L_2(x)| = 3$ we have

$$\mathbb{E}[L_1(x)L_2(x)] = -3(\frac{1}{2}+2c) + 9(\frac{1}{2}-3c) + 15c = 3 - 18c$$

giving a total expected value of

$$\frac{1}{2}(24c^2 - 24c) + \frac{1}{2}(3 - 18c) = 3/2 + 12c^2 - 21c$$

and c was chosen carefully to make this quantity 0. □

D. ONE RESULT AT THE OTHER END OF THE SPECTRUM

We have focused on computationally useless predicates that do not enable us to do essentially anything. Knowing that there is an assignment that satisfies almost all the constraints does not enable us to do better for any function.

At the other end of the spectrum we could hope for predicates where even more moderate promises can be sufficient to find useful assignments efficiently.

One possibility is to ask for a predicate that is useful for all functions Q. This is too much to ask for, as discussed in Section 3, if P and Q are sufficiently unrelated it might be the case that there are instances where we can satisfy P on all constraints while the best assignment when we consider condition Q only satisfies essentially a fraction E_Q. One possible definition is to say that P should be useful for any Q which is not excluded by this information theoretic argument. This is a potential avenue of research which we have not explored and hence we have no strong feeling about what to expect. One complication here is of course that the characterization of Theorem 3.2 is not very explicit and hence might be difficult to work with.

The payoff Q we must always consider is the traditional question of approximability namely $Q = P$ but let us weaken the promise from the optimum being almost one to being just slightly above the random threshold.

Definition D.1. A predicate P is *fully approximable* if for any $\epsilon > 0$ there is a $\delta > 0$ such that if the optimal value of a Max-P instance is $E_P + \epsilon$ then one can efficiently find an assignment that satisfies an $(E_P + \delta)$-fraction of the constraints.

First note that the most famous example of a fully approximable predicate is Max-Cut and in fact any predicate of arity two is fully approximable. This definition has been explored previously in [Håstad 2007] but given that this is not a standard venue for results on Max-CSPs, let us restate the following theorem of [Håstad 2007].

THEOREM D.2. *[Håstad 2007] A predicate P is fully approximable if and only if the Fourier expansion of P contains no term of degree at least 3.*

We refer to [Håstad 2007] for the not too difficult proof. It is not difficult to find the complete list of such predicates. A predicate on three variables is fully approximable iff it accepts equally many even and odd strings. Up to negations and permutations of variables, the only predicate that depends genuinely on four variables with this property is

$$P(x) = \frac{2 + x_1x_3 + x_1x_4 + x_2x_3 - x_2x_4}{4}.$$

Let us sketch the argument why this is the case. We are looking for a degree-two polynomial that takes values in $\{0, 1\}$. If is has any linear term, introduce a new variable x_0 and replace x_i by x_0x_i. This is a polynomial that depends on one more variable and still takes values in $\{0, 1\}$. This follows as when $x_0 = 1$ is the old polynomial and the fact that $P(-x) = P(x)$ establishes the same property when $x_0 = -1$. In view of this, we can assume that the polynomial only has terms of degree 0 and 2. Let us assume that the monomial x_1x_2 appears with a nonzero coefficient, which we may assume, by negation is positive. Then as fixing the other variables and considering all predicates on two variables, we see that this coefficient must by $1/2$ or $1/4$. In the former case we see that the obtained function must always be the parity of x_1 and x_2 on thus the polynomial only depends on two variables. In the latter case x_1 must appear in one other term which we may assume is $x_1x_3/4$. Now when $x_2 = x_3$, it is not difficult to see that the induced function must be the parity of x_1 and x_2 and thus the polynomial must be of the form $c + x_1(x_2 + x_3)/4 + (x_2 - x_3)L(x)$ for some linear function L. It is not difficult to see that we must have $c = 1/2$ and that L can only contain one new variable.