Satisfying degree d equations over $GF[2]^n$

Johan Håstad
johanh@kth.se
Royal Institute of Technology
Stockholm, Sweden

April 14, 2011

Abstract

We study the problem where we are given a system of polynomial equations defined by multivariate polynomials over $GF[2]$ of fixed constant degree $d > 1$ and the aim is to satisfy the maximal number of equations. A random assignment approximates this problem within a factor 2^{-d} and we prove that for any $\epsilon > 0$, it is NP-hard to obtain a ratio $2^{-d} + \epsilon$. When considering instances that are perfectly satisfiable we give a probabilistic polynomial time algorithm that, with high probability, satisfies a fraction $2^{1-d} - 2^{1-2d}$ and we prove that it is NP-hard to do better by an arbitrarily small constant. The hardness results are proved in the form of inapproximability results of Max-CSPs where the predicate in question has the desired form and we give some immediate results on approximation resistance of some predicates.

1 Introduction

The study of polynomial equations is a basic question of mathematics. In this paper we study a problem we call Max-d-Eq where we are given a system of m equations of degree d in n variables over $GF[2]$. As we consider the case of d constant all polynomials are given in the dense representation. Many problems can be coded as polynomial equations and in particular it is easy to code 3-Sat as equations of degree 3 and thus determining whether we can simultaneously satisfy all equations is NP-complete. It is hence natural to study the question of satisfying the maximal number of equations and our interests turn to approximation algorithms. We say that an algorithm is a C-approximation algorithm if it always returns a solution which satisfies at least $C \cdot \mathrm{OPT}$ equations where OPT is the number of equations satisfied by the optimal solution. The PCP-theorem [2, 1] shows that it is NP-hard to approximate the Max-d-Eq within some constant $C < 1$ and from the results of [6] it is not difficult to get an explicit constant of inapproximability. Given the importance of the problem it is, however, natural to try to determine the exact approximability of the problem and this is the purpose of this paper.

The result of [6] proves that the optimal approximability constant for linear equations ($d = 1$) is $\frac{1}{2}$. This approximability is obtained by simply picking a random assignment independently of the equations at hand. To prove tightness it is established that for any $\epsilon > 0$ it is NP-hard to approximate the answer better than within a factor $\frac{1}{2} + \epsilon$. This is proved by constructing a suitable Probabilistic Checkable Proof (PCP). It turns out that these results extend almost immediately to the higher degree case giving the optimal constant 2^{-d} for degree-d equations. We proceed to study the case when all equations can be simultaneously satisfied.

In the case of linear equations, it follows by Gaussian elimination that once it is possible to satisfy all equations one can efficiently find such a solution. The situation for higher degree equations turns out to be more interesting. Any implied affine condition can be used to eliminate a variable but this turns out to be the limit of what can be achieved. To be more precise, from a characterization of low weight code words from Reed-Muller codes [7] it follows that any equation satisfied by a fraction lower than $2^{1-d} - 2^{1-2d}$ must imply an affine condition. This number turns out to be the sharp threshold of approximability for satisfiable instances.

The upper bounds is obtained by using implied affine conditions and then choosing an assignment at random. For $d \geq 3$ we are not able to derandomize this algorithm and thus in general this is a probabilistic algorithm.

The lower bound is proved by constructing a PCP very much inspired by [6] and indeed nothing in the current paper uses facts not known at the time of that paper. In particular, we prove standard NP-hardness results and do not use any sophisticated results in harmonic analysis.

As a by-product of our proofs we make some observations in the area of maximum constraint satisfaction problems. Such a problem is given by a predicate P of arity k and an instance is given by a sequence of k-tuples of literals. The task is to find an assignment such that the maximal number of the resulting k-tuples of bits satisfy P. We say that a predicate is approximation resistant if it is NP-hard to get a better approximation ratio than is obtained by simply picking a random assignment. An even stronger hardness property is to prove that is NP-hard to get a better ratio even when considering instances where all constraints can be satisfied simultaneously.

Given a predicate P of arity k we construct a predicate, P^L, of arity $3k$ by replacing each input by the exclusive-or of three bits. A straightforward extension of our techniques show that for any P, the resulting predicate P^L is approximation resistant and if P does not imply an affine condition the result also applies to satisfiable instances. As a curiosity we note that this way it is possible to construct a predicate that is approximation resistant while for satisfiable instances there is a better approximation ratio that is still strictly smaller than one but larger than the ratio given by the random assignment.

An outline of the paper is as follows. In Section 2 we give some preliminaries and the rather easy result for non-perfect completeness is given in Section 3.

The most technically interesting part of the paper is given in Section 4 where we study systems of equations where all equations can be satisfied simultaneously. Due to space limitations we postpone the proof of the hardness result to the appendix. We make some observations on constraint satisfaction problems in Section 5 and end with some final remarks in Section 6.

2 Preliminaries

We are interested in polynomials, not as formal polynomials, but rather as functions mapping $GF[2]^n$ to $GF[2]$. In particular, we freely use that $x_i^2 = x_i$ and thus any term in our polynomials can be taken to be multilinear. We start with the following standard result which we, for completeness, even prove.

Theorem 2.1. *Any multivariate polynomial P of degree d that is nonzero takes the value 1 for at least a fraction 2^{-d} of the inputs.*

Proof. The proof is by induction over n and d, with the base case of $d = 1$ which is true as each linear polynomial is unbiased.

For the induction step, suppose $P(x) = P_0(x) + x_1 P_1(x)$ and let us consider what happens for the two possible values of x_1. If both P_0 and $P_0 + P_1$ are non-zero we are done by induction. If not, as P_1 is of degree at most $d - 1$, the polynomial of the two that is non-zero is of degree at most $d - 1$. Hence this polynomial takes the value 1 for at least a fraction 2^{1-d} of *its* inputs. As the set of inputs of this polynomial constitutes half of the inputs of P, the result follows also in this case. □

It is not difficult to see that this result is tight by considering $P(x) = \prod_{i=1}^{d} x_i$, or more generally, products of d independent affine forms. It is important for us that these are the only cases of tightness. This follows from a characterization by Kasami and Tokura [7] of all polynomials that are non-zero for at most a fraction 2^{1-d} of the inputs. A consequence of their characterization is the following theorem.

Theorem 2.2. *Let P be a degree d polynomial over $GF[2]$ which factors as*

$$P(x) = Q(X) \prod_{i=1}^{r} A_i(x)$$

where A_i are linearly independent affine forms and Q does not contain any affine factor. Then the fraction of points on which $P(x) = 1$ is at least

$$2^{-r}(2^{1-(d-r)} - 2^{1-2(d-r)}),$$

if $d \neq r$ and 2^{-d} if $d = r$.

For completeness we prove Theorem 2.2 in the appendix.

As mentioned in the introduction in this paper we also obtain some results for maximum constraint satisfaction problems (Max-CSPs). Let us for completeness state some definitions. For a predicate P let $r(P)$ be the probability that a random assignment satisfies P. Note that $r(P)$ is the approximation ratio achieved by the algorithm that simply picks a random assignment independent of the instance under consideration. Let Max-P be the Max-CSP where each constraint is given by the predicate P applied to a k-tuple of literals.

Definition 2.3. *A predicate P is* approximation resistant *if, for any $\epsilon > 0$, it is NP-hard to approximate Max-P within $r(P) + \epsilon$.*

There is also a stronger notion of hardness.

Definition 2.4. *A predicate P is* approximation resistant on satisfiable instances *if, for any $\epsilon > 0$, it is NP-hard to distinguish instances of Max-P where all constraints can be satisfied simultaneously from those where only a fraction $r(P) + \epsilon$ of the constraints can be satisfied simultaneously.*

We make use of the Fourier transform and as we are dealing with polynomials over $GF[2]$ we let the inputs come from $\{0, 1\}^n$ while the output is a real number. For any $\alpha \subseteq [n]$ we have the character χ_α defined by

$$\chi_\alpha(x) = (-1)^{\sum_{i\in\alpha} x_i}$$

and the Fourier expansion of a function f is given by

$$f(x) = \sum_{\alpha\subseteq[n]} \hat{f}_\alpha \chi_\alpha(x).$$

Suppose that $R \leq L$ and we are given a projection π mapping $[L]$ to $[R]$. We define a related operator π_2 acting on sets such that $\pi_2(\beta) = \alpha$ for $\beta \subseteq [L]$ and $\alpha \subseteq [R]$ iff exactly the elements of α has an odd number of preimages that belong to β. The reason this definition is useful is that if we have an $x \in \{0, 1\}^R$ and interpret this as an element $y \in \{0, 1\}^L$ by setting $y_i = x_{\pi(i)}$ then $\chi_\beta(y) = \chi_{\pi_2(\beta)}(x)$.

As is standard we use the long code introduced by Bellare et al [5]. If $v \in [L]$ then the corresponding long code is a function $A : \{0, 1\}^L \to \{-1, 1\}$ where $A(x) = (-1)^{x_v}$. We want our long codes to be folded, which means that they only contain values for inputs with $x_1 = 1$. The value when $x_1 = 0$ is defined to be $-A(\bar{x})$. This ensures that the function is unbiased and that the Fourier coefficient corresponding to the empty set is 0. Let us return to studying systems of constant degree polynomials.

3 The case of non-perfect completeness

We start with the algorithm.

Theorem 3.1. *Given a system of m polynomial equations of degree d over $GF[2]$, it is possible to, in polynomial time, to find an assignment that satisfies at least $m2^{-d}$ equations.*

Proof. In fact, by Theorem 2.1, a random assignment satisfies each equation with probability 2^{-d} and thus just picking a random assignment gives a randomized algorithm fulfilling the claim of the theorem in expectation.

To get a deterministic algorithm we use the method of conditional expectations. We do not calculate the fraction of assignments that satisfies a particular equation but rather use the lower bound that at least a fraction $2^{-d'}$ of the inputs satisfies any nontrivial equation that is currently of degree d'. □

The lower bound follows rather immediately from known results.

Theorem 3.2. *For any $\epsilon > 0$ it is NP-hard to approximate Max-d-Eq within $2^{-d} + \epsilon$.*

Proof. In [6] it is proved that it is NP-hard to distinguish systems of linear equations where a fraction $1 - \epsilon$ of the equations can be satisfied from those where only a fraction $\frac{1}{2} + \epsilon$ can be satisfied. Suppose we are given an instance of this problem with m equations which, possibly by adding one to both sides of the equation, can be be assumed to be of the form

$$A_i(x) = 1.$$

Taking all d-wise products of such equations we end up with m^d equations, each of the form

$$\prod_{j=1}^{d} A_{i_j}(x) = 1,$$

which clearly is a polynomial equation of degree d. If the optimal solution to the set of linear equations satisfies δm equations then the same solution is optimal for the constructed system and satisfies $\delta^d m^d$ equations. The theorem now follows from the result of [6]. □

We remark that, by appealing to the results by Raz and Moshkovitz [8], we can even obtain results for non-constant values of ϵ.

4 Completely satisfiable systems

When studying systems where it is possible to simultaneously satisfy all equations the situation changes. Suppose we are have an equation of the form $P(x) = 1$ and suppose this equation implies the affine condition $A(x) = 1$. Then, as the system is satisfiable, we can use this equation to eliminate one variable from the system, preserving the degrees of all equations. This is done by taking any variable x_i that appears in A and replacing it by $x_i + A(x) + 1$ (note that this function does not depend on x_i as the two occurrences of this variables cancel). This substitution preserves the satisfiability of the system and the degrees of all equations and the process stops only when none of the current equations implies an affine condition.

Using Theorem 2.2 we see that when this process ends each equation is satisfied by at least a fraction $2^{1-d} - 2^{1-2d}$ of the inputs. It seems reasonable to hope that for each perfectly satisfiable system we can efficiently find an assignment that satisfies this fraction. There are two points in the outlined argument that require closer inspection. The first is the question of how to actually determine whether a polynomial equation implies an affine condition and the second is to make sure that once the process of finding implied affine conditions has ended we can indeed deterministically find a solution that satisfies the expected number of equations. Let us first address the issue of determining whether a given equation implies an affine condition.

Suppose $P(x) = 1$ implies implies $A(x) = 1$ for some unknown affine function A. Let us assume that x_1 appears in A with a nonzero coefficient. We may write

$$P(x) = P_0(x) + P_1(x)x_1$$

where neither P_0 nor P_1 depends on x_1. Consider

$$Q(x) = P(x) + A(x)P_1(x). \tag{1}$$

As x_1 appears with coefficient one in A it follows that Q does not depend on x_1 and let us assume that Q is not identically 0. Choose any values for $x_2, x_3 \ldots x_n$ to make $Q(x) = 1$ and set x_1 to make $A(x) = 0$. It follows from (1) that $P(x) = 1$ and thus we have found a counterexample to the assumed implication. We can hence conclude that we have

$$P(x) = A(x)P_1(x).$$

We claim furthermore that this procedure is entirely efficient. Namely given P and the identity of one variable occurring in A, P_1 is uniquely defined. Once P_1 is uniquely defined the rest of the coefficients of A can easily be found by solving a linear system of equations. As there are only n candidates for a variable in A and solving a linear system of equations is polynomial time we conclude that the entire process of finding possible implied affine conditions can be done in polynomial time.

Once this process halts we need to implement the method of conditional expectations to find an assignment that satisfies the expected number of equations. As opposed to the case of Theorem 3.1 where we could use the lower bound of $2^{-d'}$ for the fraction of inputs that satisfy any degree-d' equation we here need to find a more accurate bound to calculate the conditional expectation. We do not know how to do this in deterministic polynomial time and hence the best we can do is to pick a random assignment and see if it satisfies the target number of equations. If it does not, we keep repicking random assignments until we are successful. We get the following theorem.

Theorem 4.1. *There is a probabilistic polynomial time algorithm that given a system of m simultaneously satisfiable equations of degree d over $GF[2]$ finds an assignment that satisfies at least $(2^{1-d} - 2^{1-2d})m$ equations.*

The proof of this theorem is essentially done and there is only one small detail to discuss. We know that we get at least $(2^{1-d}-2^{1-2d})m$ on average and we need to prove that we get this number with a some non-negligible probability. Let us argue this slightly informally. It is not hard to see that an integer-valued, positive random variable with maximum m and an integral mean attains its average with probability at least $1/m$. If the fractional part of the average is a multiple of $\frac{1}{t}$ the this probability might reduce to $1/tm$ but not further.

Let us remark that for $d = 2$ it if possible to make the algorithm deterministic. This follows from the fact that we can transform a degree 2 polynomial into a normal form from which we can read off the fraction of inputs for which it is equal to 1. We omit the details and let us turn to the lower bound.

Theorem 4.2. *For any ϵ it is NP-hard to distinguish satisfiable instances of Max-d-Eq from those where the optimal solution satisfies a fraction $2^{1-d} - 2^{1-2d} + \epsilon$ of the equations.*

Proof. Consider the predicate, P, on $6d$ variables given by

$$P(x) = \prod_{i=1}^{d} L_i(x) + \prod_{i=d+1}^{2d} L_i(x), \tag{2}$$

where $L_i(x) = x_{3i-2} + x_{3i} + x_{3i}$, i.e. each L_i is the exclusive or of three variables and no variable appears in two linear forms. Theorem 4.2 now follows from Theorem 4.3 below (which is proved in the appendix) as the probability that a random assignment satisfies P is exactly $2^{1-d} - 2^{1-2d}$. □

Theorem 4.3. *The predicate P defined by (2) is approximation resistant on satisfiable instances.*

5 Consequences for Max-CSPs

Let us draw some conclusions from the argument in the proof of Theorem 4.3. In this section, let P be an arbitrary predicate of arity k. Define P^L be the predicate of arity $3k$ obtained by replacing each input bit of P by the exclusive-or of three independent bits, similarly to constructing the predicate of the previous section. We have the following theorem.

Theorem 5.1. *For any predicate P that accepts at least one input, the predicate P^L is approximation resistant.*

Proof. (Sketch) Let $\alpha \in \{0,1\}^k$ be an input accepted by P. Define a distribution D_μ by setting $\mu_i = \alpha_i$ with probability $1-\epsilon$ and otherwise $\mu_i = \overline{\alpha_i}$, independently for each i. Otherwise follow the protocol in the proof of Theorem 4.3. The completeness of this protocol is at least $1-\epsilon$, but as ϵ is a an arbitrarily small constant and we only need almost-perfect completeness this is not a problem. The soundness analysis of this verifier is now similar to that of the analysis in the proof of Theorem 4.3 using

$$\left| E\left[\prod_{i\in S}\chi_{\beta^i}(\mu_i)\right]\right| = (1-2\epsilon)^{\sum_{i\in S}|\beta^i|},$$

resulting in an almost identical argument but with different constants. □

It is not difficult to see that for any P, P^L supports a measure that is pairwise independent. This implies that the results of Austrin and Mossel [4] would have been sufficient to give resistance assuming the unique games conjecture. In our case we get NP-hardness which is an advantage and it is also possible to get a general theorem with perfect completeness.

Theorem 5.2. *For any predicate P such that $P^{-1}(1)$ is not contained in a $(k-1)$-dimensional affine subspace of $\{0,1\}^k$, the predicate P^L is approximation resistant for satisfiable instances.*

Proof. (Sketch) We choose μ uniformly from the set of strings accepted by P. As $\sum_i \mu_i$ is not constant, the equivalent of Lemma A.2 is true with the constant $\frac{1}{2}$ replaced by some other constant strictly smaller than one. The rest of the argument is unaffected. □

It is tempting to guess that for any P that does imply an affine condition and hence Theorem 5.2 does not apply, P^L would not be approximation resistant on satisfiable instances. This does not seem to be obviously true and let us outline the problems.

It is true that P^L is a polynomial of degree at most k and we can use the implied affine conditions to eliminate some variables as we did in the proof of Theorem 4.1. The final stage when we have no more implied affine constraints is, however, more difficult to control. The resulting constraints are given by linear constraints jointly with the original P. By the assumption on perfect satisfiability we can conclude that the each equation is still satisfiable but not much more.

If, however, our predicate is of limited degree when viewed as a polynomial we have more information on the result. Clearly during the process of eliminating affine constraints, the degree does not increase, and in fact it decreases when we remove the known affine factor within each polynomial. We get the following conclusion.

Theorem 5.3. *Suppose predicate P of arity k is given by a polynomial of degree d that contains r linearly independent affine factors. Then if P accepts less than a fraction $2^{1-(d-r)} - 2^{1-2(d-r)}$ of the inputs, P^L is approximation resistant but not approximation resistant on satisfiable instances, unless $NP \subseteq BPP$.*

Proof. The predicate is approximation resistant by Theorem 5.1. On perfectly satisfiable instances we can run the algorithm of Theorem 4.1, and as we remove affine constraints the resulting degree is at most $d-r$. □

The simplest example of a predicate for which this theorem applies is the predicate, P, given by the equation

$$x_1(x_2x_3 + x_4x_5) = 1$$

which has $d = 3$ and is satisfied by only a fraction $\frac{3}{16}$ of the inputs. For this instantiation of P, P^L is approximation resistant but not approximation resistant for satisfiable instances. To get a hardness result for satisfiable constraints we can use Theorem 4.3 for the predicate

$$x_2x_3 + x_4x_5 = 1$$

which is approximation resistant with factor $\frac{3}{8}$ on satisfiable instances. We get a matching algorithm as the affine factor can be removed and the equations that remain are of degree 2.

Let us finally point out that all our approximation resistance results establish the stronger property of "uselessness" introduced by Austrin and Håstad [3]. This follows as we are able to bound arbitrary non-trivial characters and not only the characters appearing in the considered predicates.

6 Final words

The current paper gives optimal approximability results for satisfying the maximal number of low degree equations over $GF[2]$. The methods used in the proofs are more or less standard and thus the main contribution of this paper is to obtain tight results for a natural problem. There is a provable difference between perfectly satisfiable and almost-perfectly satisfiable systems in that we can satisfy strictly more equations in the former case. The difference is not as dramatic as in the linear case, but still striking.

For the case of Max-CSPs we obtained a few approximation resistance results for, admittedly, non-standard predicates. We feel, however, that the examples give, a not major but nonempty, contribution towards understanding the difference of approximation resistant predicates and those predicates that have this property also on satisfiable instances. Our example of an approximation resistant predicate which has another, nontrivial, approximation constant on satisfiable instances is the first of its kind. Although not surprising this result gives another piece in the puzzle to understand Max-CSPs.

Acknowledgement. I am grateful to Parikshit Gopalan for alerting me to the paper [7] and providing me with an electronic version of that paper.

References

[1] S. Arora, C. Lund, R. Motwani, M. Sudan, and M.Szegedy. Proof verification and intractability of approximation problems. *Journal of the ACM*, 45:501–555, 1998.

[2] S. Arora and S. Safra. Probabilistic checking of proofs: a new characterization of NP. *Journal of the ACM*, 45:70–122, 1998.

[3] P. Austrin and J. Håstad. On the usefullness of predicates. Unpublished manuscript, submitted to this conference, 2011.

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

[5] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs and non-approximability—towards tight results. *SIAM Journal on Computing*, 27:804–915, 1998.

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

[7] T. Kasami and N. Tokura. On the weight structure of Reed-Muller codes. *IEEE Transactions on Information Theory*, 16:752–759, 1970.

[8] D. Moshkovitz and R. Raz. Two query PCP with sub-constant error. *Journal of the ACM*, 57, 2010.

A Proof of Theorem 4.3

We reduce the standard projecting label cover instance to Max-P for this predicate P. This is the same starting point as in [6] but let us formulate it in more modern terms.

We are given a bipartite graph with vertices U and V. Each vertex $u \in U$ should be given a label $\ell(u) \in [L]$ and each vertex $v \in V$ should be given a label $\ell(v) \in [R]$. For each edge (u, v) there is a mapping $\pi_{u,v}$ and a labeling satisfies this edge iff $\pi_{u,v}(\ell(u)) = \ell(v)$.

In [6] we used the fact that for any constant ϵ there are constant values for L and R such that it is NP-hard to determine whether the optimal labeling satisfies all constraints or only a fraction ϵ of the constraints, and this is all that we need also here. Using [8] one can extend this to non-constant size domains, but let us ignore this point.

As is standard, we transform the label cover instance into a PCP by long-coding a good assignment, and for each vertex u we have a table $g_u(y)$ for $y \in \{0,1\}^L$, and similarly we have a table $f_v(x)$ for $x \in \{0,1\}^R$ for each $v \in V$. As mentioned in the preliminaries we assume that these long codes are folded and hence each table is unbiased.

Before we describe how the verifier checks this PCP we define an "error" distribution, D_μ, on $2d$ bits $(\mu_i)_{i=1}^{2d}$. First pick a random bit b with uniform probability and if $b = 0$ set $\mu_i = 1$ for $1 \le i \le d$ and select values for the other d bits uniformly from the $2^d - 1$ binary strings that contains at least one 0. If $b = 1$ we do the symmetric assignment exchanging the two halves. We need two simple facts about the distribution D_μ. The first is obvious from construction.

Lemma A.1. *With probability one it is true that*

$$\prod_{i=1}^{d} \mu_i + \prod_{i=d+1}^{2d} \mu_i = 1.$$

Secondly we have.

Lemma A.2. *For any nonempty set S and $d \geq 2$, we have*

$$|E_{D_\mu}[(-1)^{\sum_{i \in S} \mu_i}]| \leq \frac{1}{2}.$$

Proof. If S is contained in one of the two halves we observe that the distribution on this half is obtained by picking a string from the uniform distribution with probability $\frac{1}{2}(1 + (2^d - 1)^{-1})$ and otherwise picking the all one string. It follows that in this case

$$|E_{D_\mu}[(-1)^{\sum_{i \in S} \mu_i}]| = \frac{1}{2}(1 - (2^d - 1)^{-1}) < \frac{1}{2}.$$

If, on the other hand, S contains inputs from both halves then by conditioning on which half gets the all one assignment it is easy to see that

$$|E_{D_\mu}[(-1)^{\sum_{i \in S} \mu_i}]| \leq (2^d - 1)^{-1} < \frac{1}{2}.$$

□

Let us return to defining our PCP by the actions of the verifier. For readability we drop the obvious subscripts on f, g and π.

1. Pick a edge (u, v) which comes with a projection constraint $\pi : [L] \mapsto [R]$.
2. Pick $x^{(i)} \in \{0,1\}^R$ and $y^{(i)} \in \{0,1\}^L$ uniformly at random, $1 \leq i \leq 2d$.
3. For each $j \in [L]$ pick an element $\mu^{(j)}$ with the distribution D_μ and construct $z^{(i)}$ by setting $z_j^{(i)} = x_{\pi(j)}^{(i)} + y_j^{(i)} + \mu_i^{(j)} \mod 2$.
4. Read the $6d$ bits[1] corresponding to $f(x^{(i)})$, $g(y^{(i)})$, and $g(z^{(i)})$. Accept if these $6d$ bits satisfy P where the three bits fed into L_i are $f(x^{(i)})$, $g(y^{(i)})$, and $g(z^{(i)})$.

We have first have the easy completeness lemma.

Lemma A.3. *The verifier accepts a correct proof of a correct statement with probability 1.*

Proof. Suppose the proof gives labels $\ell(u)$ to $\ell(v)$ to u and v, respectively. Then $g_u(y^{(i)}) = (-1)^{y_{\ell(u)}^{(i)}}$, $g_u(z^{(i)}) = (-1)^{z_{\ell(u)}^{(i)}}$, $f_v(x^{(i)}) = (-1)^{x_{\ell(v)}^{(i)}}$. As $\pi(\ell(u)) = \ell(v)$ the exclusive-or (product in the ±1 notation) of these bits equal $(-1)^{\mu_i^{\ell(u)}}$. The lemma now follows from Lemma A.1. □

[1] We interpret -1 as the bit 1 and 1 as the bit 0.

We turn to soundness.

Lemma A.4. *If the verifier accepts with probability at least $2^{1-d} - 2^{1-2d} + \epsilon$ then there is a labeling in the label cover problem that satisfies at least a fraction $c_d\epsilon^2$ of the conditions for some constant $c_d > 0$ depending only on d.*

Proof. Expand the predicate P by its multilinear expansion. Since the constant term, $\hat{P}_\emptyset$, is $2^{1-d} - 2^{1-2d}$ we conclude that given the assumption of the lemma there are non-empty sets S_1, S_2 and S_3 such that

$$|E[\prod_{i\in S_1} f(x^{(i)}) \prod_{i\in S_2} g(y^{(i)}) \prod_{i\in S_3} g(z^{(i)})]| \geq c_d\epsilon, \tag{3}$$

for some constant c_d depending only on d. We warn the reader that we abuse notation by allowing the constant c_d to change during the argument but it remains a strictly positive number depending only on d.

Not all terms of the form (3) appear in the expansion of P but as we can bound any such term and we make some use of this fact in Section 5 we treat an arbitrary term.

First note if $S_2 \neq S_3$ the expectation in (3) is zero as for any i in the symmetric difference we get a factor $g(y^{(i)})$ or $g(z^{(i)})$ that is independent of the other factors and as g is folded the expectation of such a term is 0. To get a non-zero value we also need $S_1 = S_3$ as otherwise negating $x^{(i)}$ in the symmetric difference we get cancelling terms. Thus we need to study

$$E\left[\prod_{i\in S} f(x^{(i)})g(y^{(i)})g(z^{(i)})\right]. \tag{4}$$

Expanding each function by the Fourier transform we get the expectation

$$E\left[\prod_{i\in S}\left(\sum_{\alpha^i,\beta^i\gamma^i} \hat{f}_{\alpha^i}\hat{g}_{\beta^i}\hat{g}_{\gamma^i}\chi_{\alpha^i}(x^{(i)})\chi_{\beta^i}(y^{(i)})\chi_{\gamma^i}(z^{(i)})\right)\right]. \tag{5}$$

If we mentally expand this product of sums and look at the expectation of each term we see, as $y_j^{(i)}$ is independent of all other variables, that terms with $\gamma^i \neq \beta^i$ give contribution 0. The same is true if $\pi_2(\beta^i) \neq \alpha^i$. Let μ_i denote the vector $(\mu_i^{(j)})_{j=1}^L$ then

$$\chi_{\pi_2(\beta^i)}(x^{(i)})\chi_{\beta^i}(y^{(i)})\chi_{\beta^i}(z^{(i)}) = \chi_{\pi_2(\beta^i)}(x^{(i)})\chi_{\beta^i}(y^{(i)})\chi_{\beta^i}(x_\pi^{(i)}+y^{(i)}+\mu_i) = \chi_{\beta^i}(\mu_i),$$

and thus (5) reduces to

$$E\left[\prod_{i\in S}\left(\sum_{\beta^i} \hat{f}_{\pi_2(\beta^i)}\hat{g}_{\beta^i}^2\chi_{\beta^i}(\mu_i)\right)\right]. \tag{6}$$

We have

$$\prod_{i\in S}\chi_{\beta^i}(\mu_i)=\prod_{j\in\cup_i\beta^i}(-1)^{\sum\mu_i^{(j)}} \tag{7}$$

where the sum in the exponent is over the set of i such that $j\in\beta^i$. By Lemma A.2 it follows that the absolute value of the expectation of (7) is bounded by

$$2^{-|\cup_i\beta^i|}\leq 2^{-\sum_{i\in S}|\beta^i|/2d},$$

and hence we can conclude that it follows that from the assumption of the lemma that

$$E_{u,v}\left[\prod_{i\in S}\left(\sum_{\beta^i}|\hat{f}_{\pi_2(\beta^i)}|\hat{g}_{\beta^i}^2 2^{-|\beta^i|/2d}\right)\right]\geq c_d\epsilon. \tag{8}$$

As S is nonempty and any factor is bounded from above by one we conclude that

$$E_{u,v}\left[\sum_{\beta}|\hat{f}_{\pi_2(\beta)}|\hat{g}_{\beta}^2 2^{-|\beta|/2d}\right]\geq c_d\epsilon. \tag{9}$$

Cauchy-Schwarz inequality implies that

$$\begin{aligned}\sum_{\beta}|\hat{f}_{\pi_2(\beta)}|\hat{g}_{\beta}^2 2^{-|\beta|/2d} &\leq \left(\sum_{\beta}\hat{g}_{\beta}^2\right)^{1/2}\left(\sum_{\beta}\hat{f}_{\pi_2(\beta)}^2\hat{g}_{\beta}^2 2^{-|\beta|/d}\right)^{1/2} &(10)\\ &\leq \left(\sum_{\beta}\hat{f}_{\pi_2(\beta)}^2\hat{g}_{\beta}^2 2^{-|\beta|/d}\right)^{1/2}. &(11)\end{aligned}$$

And thus from (9), and $E[X^2]\geq E[X]^2$ we can conclude that

$$E_{u,v}\left[\sum_{\beta}\hat{f}_{\pi_2(\beta)}^2\hat{g}_{\beta}^2 2^{-|\beta|/d}\right]\geq c_d^2\epsilon^2. \tag{12}$$

We can now extract a probabilistic labeling using the standard procedure. For each u we choose a set β with probability $\hat{g}_{\beta}^2$ and return a random element in β. Similarly for each v we choose a set α with probability $\hat{f}_{\alpha}^2$ and return a random element in α. The expected fraction of satisfied constraints is at least

$$E_{u,v}\left[\sum_{\beta}\hat{f}_{\pi_2(\beta)}^2\hat{g}_{\beta}^2\frac{1}{|\beta|}\right] \tag{13}$$

and as

$$\frac{1}{x} \geq \frac{1}{d} 2^{-x/d}$$

for any $x \geq 1$ we have that (13) is at least $\frac{c_d^2}{d}\epsilon^2$ and, adjusting the value of c_d this completes the proof of Lemma A.4. □

Theorem 4.3 now follows from Lemma A.4 and Lemma A.3 by the standard way of transforming a PCP with an acceptance criteria given by a predicate P to a hardness result for the corresponding constraint satisfaction problem Max-P.

B Proof of Theorem 2.2

The goal of the current section is the prove Theorem 2.2. We remind the reader that this result follows from the characterization by Kasami and Tokura [7] of all codewords of the Reed-Muller code that have weight at most twice the minimal weight. We do not need their full characterization and as their proof is not very easy to follow we include a proof here.

The bound of Theorem 2.2 is sharp as it is obtained by

$$x^\alpha(x^\beta + x^\gamma)$$

where α, β, and γ are disjoint multi-indices where the size of α is r and the sizes of β and γ both are $d-r$.

Proof. We prove the statement by induction and we establish $d=2$ as the base case below to avoid degenerate cases. The following easy observations are useful for us.

- As $d-r$ cannot equal 1, it follows that for any d-degree polynomial that is not a product of affine factors, the number of ones is at least $\frac{3}{2}2^{-d}$.
- The bound never exceeds 2^{1-d}, and thus this bound is always sufficient (but not always possible).

The statement for general d and r follows from the case of $d-r$ and 0 and hence we may focus on the case of $r=0$ (but of course use arbitrary r in the inductive statements).

A fact that is important for us is that factorization is not unique. In particular if A and A' are two affine factors of a polynomial P then so is $1+A+A'$ as

$$(1+A+A')A = A'A.$$

Thus if we have several affine factors we can construct new affine factors by taking the sum of of an even number of such factors added with the constant 1 or the sum of an odd number of such factors.

Another useful fact is that $A(x)$ is a factor of a polynomial P iff $A(x) = 0$ implies $P(x) = 0$ or equivalently if $P(y) = 1$ implies $A(y) = 1$. The non-obvious direction of this statement follows from the discussion in the proof of Theorem 4.1 of identifying implied affine constraints.

Let us address the case of $d = 2$, which can be established by the normal form of degree two polynomials, but let us follow a different path to prepare for the general proof. As stated above, the interesting case is when P has no affine factor and let us write $P(x) = P_0(x) + x_1 P_1(x)$.

Consider setting x_1 to its two values and as P does not contain an affine factor neither of the induced polynomials can be identically 0. Furthermore if neither of these settings result in a polynomial with an affine factor we are done by induction. Let us finally assume that $x_1 = 0$ results in an affine factor, which we by an affine change of coordinates can assume is x_2. Thus we can assume that

$$P(x) = x_2 A_2(x) + x_1 A_1(x),$$

for two affine functions A_1 and A_2 where we can assume that x_i does not appear in $A_i(x)$ as it can be replaced by 1 giving the same result. The main case is that the collection of $x_1, x_2, A_1(x)$, and $A_2(x)$ form independent affine functions and in this case the fraction of inputs for which P is one is exactly 3/8 which is the claimed bound. We have a number of cases to consider when the four functions are not independent.

1. $A_1(x) \equiv 1$. If $A_2(x) = x_1$ then x_1 is a factor of P while if $A_2(x) = 1 + x_1$ then P is one with probability 3/4. Finally if $A_2(x)$ is independent of x_1 this probability is 1/2.

2. $A_1(x) = x_2$ makes x_2 a factor of P.

3. $A_1(x) = 1+x_2$ makes $Pr[P(x) = 1] = 1/2$ unless we have $A_2(x) \equiv 1$ when this probability is 3/4.

4. $A_1(x)$ is independent of x_2 in which case, by an affine change of variables we can assume that $A_1(x) = x_3$. Now since $A_2(x)$ does not contain x_2, is linearly dependent of x_1 and x_3 and is not a factor of $x_1 x_3$ (ruling out also $(1 + x_1 + x_3)$), $A_2(x)$ must equal one of the functions 1, $1 + x_1$ or $1 + x_3$. In these cases it is easy to check that $Pr[P(x) = 1]$ takes the values 3/4, 1/2 and 1/2, respectively.

This finishes the case $d = 2$ end we turn to the general case. Not surprisingly, also here we end up analyzing a number of cases.

As in the case $d = 2$ we can assume that P has no affine factors but the polynomial resulting when substituting $x_1 = 0$ gives a polynomial with at least one affine factor. Picking a full set of linearly independent factors and making an affine transformation we can assume that

$$P(x) = x_2 x^{\beta} P_2(x) + x_1 P_1(x)$$

where β is a possible empty multi-index and P_2 has no affine factors. First let us consider affine factors in P_1.

Let $\prod_{i=1}^r A_i(x)$ be the affine factors that appear in P_1. We have two cases depending whether each A_i that might appear in the factorization, together with affine forms that appear in the first product (i.e. the coordinate functions given by x_2 and the elements of β) are independent.

Suppose these functions are not independent and hence that we can choose the factorization such that $A_1(x)$ only depends on x_2 and the variables in β. Let us look at the point x^0 where x_2 and all elements of β equals 1. If $A_1(x^0) = 1$ then $A_1(x)$ is a factor of P and this is a contradiction of the assumptions. If, on the other hand $A_1(x_0) = 0$ then the sets of points where $x_2 x^\beta P_2(x) = 1$ and $x_1 P_1(x) = 1$ are disjoint and as these sets are each of (relative) size at least 2^{-d} we get $Pr[P(x) = 1] \geq 2^{1-d}$ and the lemma follows also in this case.

Thus we can assume that the affine forms in P_1 are independent of x_2 and the variables in β and by an affine transformation we can assume that

$$P(x) = x_2 x^\beta P_2(x) + x_1 x^\gamma P_1'(x), \tag{14}$$

where β and γ are disjoint multi-indices and no more affine factors can be pulled out of P_2 or P_1'.

Let us analyze what happens for the four possible simultaneous assignments of values of x_1 and x_2. When both are 0 we get a function that is identically 0 which is not good for us but in the other cases we get polynomials of degree $d-1$ and we now analyze the structure of these polynomials.

When $x_1 = 0$ and $x_2 = 1$ we get $x^\beta P_2(x)$, when $x_1 = 1$ and $x_2 = 0$ we get $x^\gamma P_1'(x)$, and in the final case we have

$$W(x) = x^\beta P_2(x) + x^\gamma P_1'(x).$$

Note that W is not identically 0 as $(x_1 + x_2)$ then would have been an affine factor of P.

Suppose first that neither P_2 nor P_1' is identically one (and thus we have some non-affine factors in both these cases). Then $Pr[P(x) = 1]$ is at least

$$\frac{1}{4}(\frac{3}{2}2^{1-d} + \frac{3}{2}2^{1-d} + 2^{1-d}) = 2^{1-d}$$

proving the bound in this case. By symmetry we may thus assume that $P_2 \equiv 1$. If $\gamma = \emptyset$ or W does not contain any affine factor then $Pr[P(x) = 1]$ is at least

$$\frac{1}{4}(2^{1-d} + 2^{1-d} + 2^{2-d} - 2^{3-2d}),$$

which exactly equals the claimed bound. Thus we can assume that γ is non-empty and W contains an affine factor $A(x)$ and remember that

$$W(x) = x^\beta + x^\gamma P_1'(x). \tag{15}$$

Suppose A only depends on variables in β. If it is fixed to 1 by setting all variables in β to one, then $A(x)$ is a factor of x^β and hence also of $W(x) + x^\beta =$

$x^\gamma P_1'(x)$ and hence also of $P(x)$, contradicting assumptions. If $A(x)$ is forced to 0 by this assignment then setting any variable in γ (remember it is non-empty and disjoint from β) to 0 and we get $W(x) = 0$ while $x^\beta = 1$ and $x^\gamma P_1'(x) = 0$ contradicting (15). Thus we can assume that $A(x)$ depends on some variable outside β.

If we can fix some variable in γ to 0, the variables of β to one and A to zero we get a contradiction to (15) as $x^\beta = 1$ while $W(x) = 0$ and $x^\gamma = 0$. If the size of γ is at least 2 or W contains at least two different affine factors we claim that this must be possible. Suppose first that the size of γ is at least 2

As $A(x)$ does not only depend on variables in β we can fix these variables to one, and then pick a suitable variable in γ to fix to 0 without fixing the value of $A(x)$. We can then fix additional variables to make $A(x) = 0$ obtaining the desired contradiction.

If we have one more affine factor A' of W then one of A, A' and $1 + A + A'$ is a factor of W and does not depend on the first variable of γ (and some variable outside β). It follows again that we can make $x^\beta = 1$, $x^\gamma = 0$ and $W(x) = 0$.

The only remaining case is when γ is of size one and W has one affine factor. In this case, by induction $Pr[P(x) = 1]$ is at least

$$\frac{1}{4}(2^{1-d} + 2 \cdot \frac{1}{2}(2^{3-d} - 2^{5-2d})).$$

For d at least 4 this is at least 2^{1-d} and we are done. Finally note that in the case $d = 3$, P_1' as well as the co-factor of A in W are of degree at most one and hence if they do not contain an additional factor they must be the constant 1 and the lemma follows also in this case. □