SIAM J. COMPUT.
Vol. 43, No. 1, pp. 179–193

ON THE NP-HARDNESS OF MAX-NOT-2*

JOHAN HÅSTAD†

Abstract. We prove that, for any $\epsilon > 0$, given a satisfiable instance of Max-NTW (Not-2), it is NP-hard to find an assignment that satisfies a fraction $\frac{5}{8} + \epsilon$ of the constraints. This, up to the existence of ϵ, matches the approximation ratio obtained by the trivial algorithm that just picks an assignment at random, and thus the result is tight. Said equivalently, the result proves that Max-NTW is approximation resistant on satisfiable instances, and this makes complete our understanding of arity three maximum constraint satisfaction problems with regards to approximation resistance.

Key words. constraint satisfaction, approximation resistance, probabilistically checkable proofs

AMS subject classifications. 68Q17, 68W25

DOI. 10.1137/120882718

1. Introduction.

In this paper we study the approximability of maximum constraint satisfaction problems (Max-CSPs). In the generic problem we are given a large number of constraints, each affecting only a constant number of variables, and the goal is to find an assignment to the variables that satisfies the maximal number of constraints. The most common domain of these variables is given by Boolean values, and this is also the focus of this paper. Constraints can be of many forms, but the most studied case, examined here, is when each constraint is of the form of a fixed predicate, P, applied to a sequence of literals corresponding to different variables. Almost all Max-CSPs are NP-hard, and we turn to efficient approximation algorithms.

For such a maximization problem we say that an algorithm, A, is a C-approximation algorithm if it, for each instance I, outputs a number $A(I)$ such that $Opt(I) \geq A(I) \geq COpt(I)$, where $Opt(I)$ is the optimal value on instance I. Most algorithms will, at least with the help of randomness, in fact find an assignment that satisfies this $A(I)$-fraction of the constraints, but we do not set this as a formal requirement.

The simple algorithm, which picks an assignment uniformly at random, gives a lower bound for approximability. In the basic situation this is just the probability, E_P, that a random assignment satisfies the defining predicate P. It is somewhat surprising, but mounting evidence [3, 7] shows that for most predicates this is the best constant of approximability that can be guaranteed by any algorithm running in polynomial time. The predicates for which this is indeed the best constant of approximability are called *approximation resistant*.

An equivalent way to formulate approximation resistance is to say that, for any $\epsilon > 0$, it is NP-hard to distinguish instances for which the best assignment satisfies a fraction $1 - \epsilon$ of the constraints from those where this fraction is $E_P + \epsilon$. A slightly stronger property is that it is NP-hard to distinguish completely satisfiable instances from those where the best assignment satisfies a fraction $E_P + \epsilon$ of the constraints. We call such predicates *approximation resistant on satisfiable instances*.

As stated above, it follows from the work of Chan [3] and Håstad [7] that most

*Received by the editors June 27, 2012; accepted for publication (in revised form) November 18, 2013; published electronically February 4, 2014. This is the full version of the conference paper that appeared as [9]. This work was funded by ERC Advanced Investigator Grant 226203.
http://www.siam.org/journals/sicomp/43-1/88271.html
†Department of Computer Science, KTH Royal Institute of Technology, Lindstedtsvagen 3, Stockholm, S-100 44, Sweden (johanh@kth.se).

predicates are approximation resistant, but when it comes to establishing approximation resistance on satisfiable instances, much less is known. The question of whether the difference between satisfiable instances and almost satisfiable instances is real or just a technicality is, in our eyes, not fully answered. The main example separating the two cases is parity, and while there are some differences also for higher degree equations [8], also these examples have a strong smell of linear equations.

In this paper we focus on predicates of small arity and in particular on arity at most three. Using the seminal paper by Goemans and Williamson [4], it follows that all predicates of arity two are nontrivially approximable. When it comes to arity three, by combining the results of Zwick [18] and the results of Håstad [6], a predicate is approximation resistant iff it is implied by parity or its negation.

When turning to satisfiable instances, all problems of arity two can be solved perfectly, while for arity three the situation is more interesting. For parity (of any size) it is the case that if all constraints can be satisfied simultaneously, then such an assignment can be found efficiently by Gaussian elimination. At the other end, the predicates that are implied by parity (or its negation) and accept at least six of the eight inputs were proved already in [6] to be approximation resistant on satisfiable instances. In view of these results, the only approximation resistance problem of arity three that has remained open is the status of predicates accepting five inputs and being implied by parity (or its negation) when considering satisfiable instances. Such a predicate accepts the four strings accepted by parity (or its negation) and one more string. If we negate a suitable subset of inputs to make this extra string the all zero string, the predicate turns into "not two," which is true unless exactly two of the three Boolean inputs take the value true. As negating some inputs does not change the approximation resistance, we may as well study this predicate, and we call the resulting problem *Max-NTW*.

To address this problem O'Donnell and Wu [15] proved, assuming the d-to-1 conjecture of Khot [11], that Max-NTW is approximation resistant on satisfiable instances. The purpose of the current paper is to establish the same result based only on NP$\neq$P and thus make complete our knowledge with respect to approximation resistance of predicates of arity at most three. Let us briefly discuss the methods used.

We, as do previous papers establishing similar results, obtain our result by producing a probabilistically checkable proof (PCP), where the acceptance criterion is given by the target predicate (in our case "not two"). We follow the approach of [15] to a great extent, and our starting point is a projection label cover instance. Such an instance is given by two sets of variables, $u \in U$ and $v \in V$, all of which should be given labels from the sets[1] $[M]$ and $[L]$, respectively. For some pairs (u, v) we are given constraints in the form of projection operators π_{uv}, and a labeling satisfies a given constraint iff $\pi_{uv}(l_v) = l_u$.

It turns out that, for any $\epsilon > 0$, it is NP-hard to distinguish the situation when all constraints can be simultaneously satisfied from those where only a fraction ϵ of the constraints can be simultaneously satisfied. An interesting parameter here is the degeneracy[2] of the projections π_{uv} used in the instances constructed to prove hardness. In the known proofs of NP-hardness the degeneracy grows polynomially in ϵ^{-1}. The d-to-1 conjecture says that it is possible to obtain the same result with this

[1] Any sets of given cardinalities work equally well.

[2] This is defined to be the maximal number of elements from the large set that project onto the same element in the small set.

parameter bounded by d and just letting the sizes of the label sets go to infinity.

To prove our result we reuse several facts from [15]. Their PCP for a projection label cover instance has a parameter δ, and we use their protocol for a random choice of δ. We need one additional modification, and that is to use instances of *smooth label cover* as introduced by Khot [13]. If we think of the label cover instance as a two-prover game, these instances are constructed by sending a large set of identical questions to both provers.

Usually the key property of such instances is that unequal answers by the prover, with the longer answers, with high probability, project to unequal answers of the other prover, and this is the definition of "smooth." We use slightly more structure, and in particular we need that we have natural copies of the original game within this extended game.

As is well known, Max-CSPs are in fact in one-to-one correspondence with nonadaptive PCPs. Thus our result establishes $\frac{5}{8}$ as the tight infimum of the soundness for any nonadaptive PCP that reads three bits and has perfect completeness. This lower bound was proved by Zwick [18], while the upper bound by O'Donnell and Wu [15] was conditioned on the d-to-1 conjecture. The previously best upper bound based only on NP-completeness was $\frac{20}{27}$ by Khot and Saket [12]. For a longer discussion of these issues we refer the reader to [15].

Finally let us note that we can make a Max-NTW into a two-prover game by sending a uniformly chosen constraint to one of the provers and a random variable from that constraint to the other prover. These provers are supposed to return values to the three variables and the single variable, respectively, and the verifier accepts iff the two assignments to the common variable agree and the constraint is satisfied by the triplet. It is not difficult to see that the optimal strategy for this two-prover game is to fix an assignment returned by the variable-prover and then let the constraint-prover return this assignment to the queried variables whenever the assignment satisfies the constraint in question and otherwise flip one variable. It is not difficult to see that if the optimal assignment satisfies a fraction $\frac{5}{8}+\epsilon$ of the constraints, then this strategy succeeds with probability

$$\frac{5}{8}+\epsilon+\frac{2}{3}\left(\frac{3}{8}-\epsilon\right)=\frac{7}{8}+\frac{\epsilon}{3}.$$

This two-prover game is 3-to-1 in that for each answer from the variable-prover there are at most three answers from the constraint-prover that make the verifier accept. The completeness of this protocol is clearly one, and this is, as far as we know, the best soundness for a 3-to-1 game.

2. Preliminaries. We use mostly standard notation. We use $\{-1,1\}$-notation for Boolean variables, with -1 corresponding to "true." We have real-valued functions of Boolean variables, mapping $\{-1,1\}^n$ into the real numbers for different values of n. Many functions that we use are bounded by 1 in absolute value, but not all. We have the Fourier expansion given as

$$f(x)=\sum_{\alpha}\hat{f}_\alpha\chi_\alpha(x),$$

where χ_α is a character function defined to equal $\prod_{i\in\alpha}x_i$.

We let $[M]$ denote the set of integers $0,1,\ldots,M-1$, and we are interested in projection operators π mapping $[L]$ to $[M]$. Any such operator creates a partitioning

of $[L]$, defining the blocks to be the elements that map onto the same element. For a set $\beta \subseteq [L]$ we define $\pi(\beta)$ to be the set of projected elements; i.e.,

$$\pi(\beta) = \{i \mid \exists j \in \beta,\ \pi(j) = i\}.$$

Given $g : \{-1,1\}^L \mapsto \mathbb{R}$, we use the decomposition

$$g(y) = \sum_{\alpha} g^{\alpha}(y), \tag{1}$$

where

$$g^{\alpha}(y) = \sum_{\beta \mid \pi(\beta)=\alpha} \hat{g}_{\beta}\chi_{\beta}(y). \tag{2}$$

This decomposition in fact equals the Efron–Stein decomposition with regards to blocks in the partitioning defined above. For a longer discussion of the Efron–Stein decomposition and its properties, we refer the reader to [14].

Functions of special interest to us are the dictator functions $f(x) = x_i$, which are also known as "the long code of i." In many situations we can make sure that a function is odd (i.e., that $f(-x) = -f(x)$), which in PCP-language is called "folding over true." This is ensured by, for each pair $(x, -x)$, storing only one entry in the table which is negated before it is used if the value of $f(-x)$ is desired. Note that correct long codes are indeed odd.

We use correlated spaces as introduced by Mossel [14]. For a measure μ and real number, $p \geq 1$, we have the L^p-norm as

$$\|f\|_p = E_{\mu}[|f(x)|^p]^{1/p}.$$

The norm depends on the measure μ, but we leave this dependence implicit.

DEFINITION 2.1. *Suppose that μ is a probability measure on $(X \times Y)$. The* correlation *of X and Y under μ is*

$$\rho(X, Y, \mu) = \max E_{\mu}[f(x)g(y)],$$

where the maximum is taken over functions f and g such that $\|f\|_2 = \|g\|_2 = 1$ and $E[f] = E[g] = 0$.

It is important for us how correlated spaces behave under products and how they interact with the Efron–Stein decomposition. Given a sequence of correlated spaces $(X_i, Y_i, \mu_i)_{i=1}^n$, we define the product space $((X_i)_{i=1}^n, (Y_i)_{i=1}^n, \prod_{i=1}^n \mu_i)$, and Proposition 2.13 of [14] establishes that the correlation of this product space is bounded by the maximum correlation of any underlying space.

LEMMA 2.2 (see [14]). *Let $(X_i \times Y_i)$ and μ_i be correlated spaces; then*

$$\rho\left((X_i)_{i=1}^n, (Y_i)_{i=1}^n, \prod_{i=1}^n \mu_i\right) \leq \max_i \rho(X_i, Y_i, \mu_i).$$

If g is a function on $(Y_i)_{i=1}^n$ and $g = \sum_S g_S$ is its Efron–Stein decomposition, then by Proposition 2.12 of [14] we have the following lemma.

LEMMA 2.3 (see [14]). *Let $(X_i \times Y_i)$ and μ_i be correlated spaces with $\rho(X_i, Y_i, \mu_i) = \rho_i$, and let $f : (X_i)_{i=1}^n \mapsto \mathbb{R}$ and $g : (Y_i)_{i=1}^n \mapsto \mathbb{R}$; then*

$$E[f(x)g_S(y)] \leq \|f\|_2 \|g_S\|_2 \prod_{i \in S} \rho_i.$$

We get the following lemma as a corollary.

LEMMA 2.4. *Let the set-up be as in Lemma* 2.3*, let* $\rho = \max_i \rho_i$*, and suppose that the Efron–Stein decomposition of* g *contains only functions of weight at least* s*; then*

$$E[f(x)g(y)] \le \rho^s \|f\|_2 \|g\|_2.$$

Proof. Taking the Efron–Stein decomposition of both functions, we get

$$E[f(x)g(y)] \le \sum_{T,S} E[f_T(x)g_S(y)] = \sum_S E[f_S(x)g_S(y)] \le \sum_S \|f_S\|_2 \rho^{|S|} \|g_S\|_2$$
$$\le \rho^s \left(\sum_S \|f_S\|_2^2\right)^{1/2} \left(\sum_S \|g_S\|_2^2\right)^{1/2} \le \rho^s \|f\|_2 \|g\|_2. \qquad \square$$

3. From label cover to a PCP. We start with a standard projection label cover instance and think of it as a two-prover game. In this game the verifier generates tuples (q_1, q_2, π) and sends question q_i to prover P_i. The prover P_2 gives an answer $a_2 \in [L]$, while P_1 answers $a_1 \in [M]$, and the verifier accepts iff $\pi(a_2) = a_1$. We here assume that π is d-to-1; in other words, for any a_1 there are exactly d different a_2 such that $\pi(a_2) = a_1$. The lemma below follows from the PCP-theorem [1] and Raz's parallel repetition theorem [16].

LEMMA 3.1. *For any* $\epsilon > 0$ *there exists a two-prover game with parameters* M*,* L*, and* d*, where the verifier uses* $O(\log n)$ *random bits such that it is NP-hard to distinguish the cases when all constraints can be simultaneously satisfied from those where the optimal strategy of the provers makes the verifier accept with probability at most* ϵ*. The sizes of* M*,* L*, and* d *can all be taken to be polynomial in* ϵ^{-1}*.*

We make the two-prover-protocol more robust by independently generating T extra copies of the question q_2. These questions are sent to both players. Thus the prover P_2 gets $T+1$ independent instances of its standard type of question, while P_1 gets T questions of the type initially sent to P_2 and one of its original type of question. Let us denote instances of these new types of questions by Q_2 and Q_1, respectively.

Both provers are supposed to answer all questions, and the extended verifier accepts if it gets the same answers from the two provers on the questions sent to both provers and if the original verifier would have accepted the answers given to the standard questions. We call this protocol the T-extended protocol.

This protocol has properties similar to those of the original, not extended, protocol. We claim that the parameters d and ϵ do not change. The former is quite obvious, as the extra questions are preserved under projection. To see that the acceptance probability in the completeness case does not decrease, note that any shared deterministic strategy for the provers in the original game can be used in the extended game to make the verifier accept with the same probability.

The fact that the soundness cannot increase follows from the fact that, for any fixed value of the T extra questions, we have a copy of the original game and the provers cannot win this subgame with probability higher than the original game.

The parameters M and L do increase in the extended game, and we reserve M and L to be used for these new values.

3.1. The PCP. We turn this extended two-prover protocol into a PCP, in a more or less standard way. For each question to one of the provers in the T-extended two-prover game we introduce a table which, in a correct proof for a correct statement, should be the long code of the answer to this question.

We have the below basic test, called NTW_δ, with a parameter δ, which is assumed to be bounded from above by $\frac{1}{2}$. We note that this test is in fact identical to the test used by O'Donnell and Wu [15].

TEST NTW_δ.

Written proof. For each question Q_1 to P_1 we have a table $f_{Q_1} : \{-1,1\}^M \mapsto \{-1,1\}$, and similarly tables $g_{Q_2} : \{-1,1\}^L \mapsto \{-1,1\}$ for questions to P_2. These tables are folded over true.

Desired property. To check that the tables form a long coding of a strategy in the T-extended game that makes the verifier of that game accept.

Verifier.

1. Choose a question (Q_1, Q_2, π) in the two-prover game.
2. Choose $x \in \{-1,1\}^M$ with the uniform probability.
3. Choose $y \in \{-1,1\}^L$ with the uniform probability.
4. Set $z_j = -y_j x_{\pi(j)}$ for all $j \in L$.
5. For each $i \in [M]$ with probability δ choose a random j such that $\pi(j) = i$, and set $z_j = y_j = x_i$.
6. Accept iff not exactly two of the bits $f_{Q_1}(x)$, $g_{Q_2}(y)$, and $g_{Q_2}(z)$ are -1.

For each δ we define two parameters, $s_\delta = c' \log(1/\delta)/\log d$ for a constant c' and $S_\delta = c'' \log(1/\delta) d^3 2^{2d} \delta^{-2}$ for a constant c''. We later find suitable values for these constants. We are now in a position to define our final test.

TEST $NTW^k_{\delta'}$.

Written proof. Same as for NTW_δ.

Desired property. Same as for NTW_δ.

Verifier.

1. Set $\delta_0 = \delta'$, and for $i = 1, \ldots, k-1$ choose δ_i such that $s_{\delta_i} = S_{\delta_{i-1}}$.
2. Pick a random $i \in [k]$ uniformly at random, and run NTW_{δ_i}.

First note that, as s_δ tends to infinity when δ tends to 0, we do get a well-defined sequence δ_i. We can also observe that $\log(1/\delta_{k-1})$ is a constant that depends only on d, k, and δ'. As s_δ is smaller than $\log(1/\delta)$ and S_δ larger than δ^{-2}, δ_{i+1} is smaller than $e^{-\delta_i^{-2}}$, and thus δ_{k-1} is a tower of exponentials of height at least k. Even if we are not strongly concerned with the value of our constants, we must admit that getting better constants would certainly be nice.

We study the completeness and soundness of the above test in the next section.

4. Completeness and soundness of the PCP. Let us start by showing the easy completeness.

LEMMA 4.1. *If the label-cover instance is satisfiable, then there is a proof such that the verifier in $NTW^k_{\delta'}$ always accepts.*

Proof. Consider a written proof that is given by correct long codes of a strategy that always convinces the verifier in the extended two-prover game. In this situation, the three bits read are of the form $x_{\pi(j)}, y_j, z_j$, and these either have product -1 or are all 1. In either case the verifier accepts. □

Let us turn to the more interesting problem of analyzing soundness. The key soundness lemma is the following.

LEMMA 4.2. *For any $\epsilon' > 0$ and any basic two-prover games with parameters L, M, and d there are constants $\delta' > 0$, k, and T such that if the verifier accepts in $NTW^k_{\delta'}$ with probability at least $\frac{5}{8} + \epsilon'$, then there is a strategy for the provers in the basic two-prover game that makes that verifier accept with probability ϵ'^2.*

We can conclude that for any $\epsilon' > 0$, for appropriate values of δ', k, and T, the soundness of $NTW^k_{\delta'}$ is at most $\frac{5}{8} + \epsilon'$. This follows by choosing $\epsilon < \epsilon'^2$, obtaining

parameters L, M, and d such that the soundness of the basic two-prover game is at most ϵ, and then using the values of δ', k, and T produced by Lemma 4.2.

As all involved numbers are constants, we get, by the standard translation from a PCP with a given acceptance criterion to the Max-CSP with the same predicate, our main theorem.

THEOREM 4.3. *For any $\epsilon > 0$ it is NP-hard to approximate Max-NTW within $\frac{5}{8} + \epsilon$ on satisfiable instances.*

All that remains is to prove Lemma 4.2.

Proof of Lemma 4.2. We analyze NTW_δ for a fixed value of δ. For readability we drop the subscripts on the functions f and g, as well as the parameters s and S.

Arithmetizing the predicate "not two," we see that

$$\frac{5 + f(x) + g(y) + g(z) + f(x)g(y) + f(x)g(z) + g(y)g(z) - 3f(x)g(y)g(z)}{8} \tag{3}$$

is 1 if the verifier accepts and 0 otherwise. We need to analyze the expected value of this quantity.

We claim that each of x, y, and z is uniformly random. This is obvious for x and almost obvious for y (and z) as y_j is either random or set to equal $x_{\pi(j)}$, which in its turn is random. We also note that y and z are symmetric.

We have here used the fact that for any i we, with probability δ, pick a single j with $\pi(j) = i$ and set $y_j = x_i$. A distribution where we, for each j, set $y_j = x_{\pi(j)}$ with probability δ/d would not result in a uniformly random y.

From the uniformity of x, y, and z it follows, by folding, that the first three nontrivial terms in (3) have expectation 0. For the next three terms we use the analysis of [15]; let us recall some of their arguments.

We let X be one bit, x_i in x, Y the block of bits y_j such that $\pi(j) = i$ (and which is thus a set of d bits), and define Z analogously. Then x and y have the distribution as an M-fold product of such spaces. Lemma 5.2 of [15] is as follows.

LEMMA 4.4 (after [15]). *$\rho(X, Y) \le \delta$ and $\rho(X, Z) \le \delta$.*

Now the below lemma follows from the above lemma, Lemma 2.2, and the fact that both f and g are unbiased Boolean functions and hence have L^2-norm 1.

LEMMA 4.5 (see [15]). *$|E[f(x)g(y)]| \le \delta$ and $|E[f(x)g(z)]| \le \delta$.*

The next-to-last term in (3) was also taken care of by O'Donnell and Wu, and here we only quote their Theorem 6.2.

LEMMA 4.6 (after [15]). *For any g that is odd we have $|E[g(y)g(z)]| \le \delta$.*

We turn to analyzing $E[f(x)g(y)g(z)]$, which is the most challenging term. Let us look at the Fourier expansion

$$g(y) = \sum_\alpha \hat{g}_\alpha \chi_\alpha(y)$$

and divide the terms into four parts forming functions g_i, $1 \le i \le 4$. This division is guided by our two parameters s and S. The first two functions are the large and the medium-size terms and are straightforward to define as

$$g_1(y) = \sum_{|\beta| \ge S} \hat{g}_\beta \chi_\beta(y)$$

and

$$g_2(y) = \sum_{S > |\beta| \ge s} \hat{g}_\beta \chi_\beta(y).$$

For the small sets β we define a set β to be *shattered* if for any $j_1, j_2 \in \beta$ with $j_1 \neq j_2$ we have that j_1 and j_2 give different answers to the T questions sent to both provers. Note that this in particular implies that $\pi(j_1) \neq \pi(j_2)$. We let g_3 be the small terms that are not shattered, and g_4 the small terms that are shattered; in other words,

$$g_3(y) = \sum_{|\beta|<s,\beta \text{ not shattered}} \hat{g}_\beta \chi_\beta(y)$$

and

$$g_4(y) = \sum_{|\beta|<s,\beta \text{ shattered}} \hat{g}_\beta \chi_\beta(y).$$

Obviously

$$E[f(x)g(y)g(z)] = \sum_{i=1}^{4} E[f(x)g(y)g_i(z)], \tag{4}$$

and we estimate these terms separately.

Let us consider the first term in (4). The function g_1 consists only of terms given by Fourier coefficients of size at least S. Using the definition of the Efron–Stein decomposition given by (1), we see that it contains only terms of size at least S/d. We want to use Lemma 2.3 with the subdivision $X \times Y$ and Z, and we have the following correlation bound, which appears as Lemma 5.3 of [15].

LEMMA 4.7 (after [15]). *$\rho(X \times Y, Z) \leq (1 - \frac{\delta^2}{d^2 2^{2d+1}})$, and the same bound applies to $\rho(X \times Z, Y)$.*

By Lemma 2.4 we can conclude that

$$|E[f(x)g(y)g_1(z)]| \leq \left(1 - \frac{\delta^2}{d^2 2^{2d+1}}\right)^{S/d} \|fg\|_2 \|g_1\|_2 \leq \left(1 - \frac{\delta^2}{d^2 2^{2d+1}}\right)^{S/d} \leq e^{-\frac{\delta^2 S}{d^3 2^{2d+1}}}.$$

For the second and third terms in (4) we have

$$|E[f(x)g(y)g_i(z)]| \leq \|fg\|_2 \|g_i\|_2 \leq \|g_i\|_2,$$

and we bound these L^2-norms later in Lemmas 4.11 and 4.12, respectively. For the last term we use

$$E[f(x)g(y)g_4(z)] = \sum_{i=1}^{4} E[f(x)g_i(y)g_4(z)], \tag{5}$$

and we bound all but the last term in a very similar manner as before. To be more precise, since y and z are symmetric and the only property we used in the previous steps is that $f(x)g(y)$ has L^2-norm bounded by 1, and the same is true for $f(x)g_4(z)$, we can repeat the above argument and bound the first three terms in (5) by

$$e^{-\frac{\delta^2 S}{d^3 2^{2d+1}}} + \|g_2\|_2 + \|g_3\|_2.$$

We are left to analyze

$$E[f(x)g_4(y)g_4(z)]. \tag{6}$$

We expand the three functions by the Fourier transform, and we need to analyze

$$E\left[\sum_{\alpha,\beta,\gamma} \hat{f}_\alpha \hat{g}_\beta \hat{g}_\gamma \chi_\alpha(x)\chi_\beta(y)\chi_\gamma(z)\right].$$

Remember that all β and γ occurring in the sum are of size at most s and are shattered. In fact, any sum over β from now on contains only such terms. Moving the expectation inside the sum, we first note that for $\beta = \gamma$ and $\pi(\beta) = \alpha$ we have

$$|E[\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(z)]| = (1 - \delta/d)^{|\beta|}.$$

This follows as, unless some element $j \in \beta$ is chosen to make $z_j = y_j = x_{\pi(j)}$ under step 5 of Test NTW_δ, the expectation is $(-1)^{|\beta|}$. Furthermore, as β is shattered, the events that the different elements within β are chosen are independent, and when one element is chosen, the expectation is 0.

Most other terms are 0; let us first identify many such terms.

LEMMA 4.8. *If α, β, γ violate either*

- *$\alpha \subseteq \pi(\beta) \cup \pi(\gamma)$, or*
- *for any element in i contained in $\pi(\beta) \cup \pi(\gamma)$ but not in α we have an element $j \in \beta \cap \gamma$ such that $\pi(j) = i$,*

then

$$E[\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(z)] = 0.$$

Proof. In the first case for $i \in \alpha$ but $i \notin \pi(\beta) \cup \pi(\gamma)$ we have that x_i is uniformly random and independent of all other variables in the product.

To see the other claim take any $j \in \beta$ such that $\pi(j) = i$ and $i \notin \alpha$. If $j \notin \gamma$, then y_j is uniform and independent of all other variables in the product. The case for $j \in \gamma$ and $j \notin \beta$ follows by symmetry. ∎

For the terms not covered above we have the following lemma.

LEMMA 4.9. *For any term not covered by Lemma 4.8 and which does not satisfy $\beta = \gamma$ and $\pi(\beta) = \alpha$ we have*

$$|E[\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(z)]| \leq \delta/d.$$

Proof. Remember that we are considering only the case when β and γ are shattered. Now, take any j which is in the symmetric difference of β and γ (or which belongs to both β and γ but $\pi(j) \notin \alpha$ if $\beta = \gamma$).

In the first case assume without loss of generality that $j \in \beta$. We have two cases depending on whether we have an $j' \in \gamma$ such that $\pi(j') = \pi(j)$. (Remember that we are assuming that $j' \neq j$.) Suppose that $\pi(j) = i$, and let us fix all values except x_i, y_j, and $z_{j'}$. It is not difficult to see that $E[x_i y_j z_{j'}] = 0$, $E[x_i y_j] = \delta/d$, $E[y_j z_{j'}] = 0$, and $E[y_j] = 0$, so the lemma follows in this case.

In the second case note that, as β and γ are shattered, there is no $j' \neq j$ in these sets with $\pi(j') = \pi(j)$. As $y_j z_j = -x_{\pi(j)}$ with probability $1 - \delta/d$ and otherwise equals 1, the nonexistence of j' and the uniformity of $x_{\pi(j)}$ implies the lemma. ∎

Next we bound

$$\sum_{\alpha,\beta,\gamma} |\hat{f}_\alpha \hat{g}_\beta \hat{g}_\gamma|, \tag{7}$$

where we sum over all α, β, γ that give a nonzero value of $E[\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(z)]$.

To help the reader's intuition, let us point out that any bound, b, on the sum (7) in terms of s and d that allows s to go to infinity while making $b\delta/d$ tend to 0 is good enough for us.

We apply the Cauchy–Schwarz inequality to (7) to get the bound

$$\left(\sum_{\alpha} f_\alpha^2\right)^{1/2} \left(\sum_{\alpha} \left(\sum_{\beta,\gamma} |\hat{g}_\beta \hat{g}_\gamma|\right)^2\right)^{1/2}, \tag{8}$$

where the inner sum is over pairs β and γ that could appear together with a given α. In particular, for any $i \in \alpha$, at least one of β and γ contains an element that projects onto i, and for $i \notin \alpha$, if β contains an element j such that $\pi(j) = i$, then j belongs also to γ .

The first factor of (8) is bounded by 1; let us look at the second factor. Expanding the square, we get a sum of the form

$$\sum_{\beta,\gamma,\beta'\gamma'} |\hat{g}_\beta \hat{g}_\gamma \hat{g}_{\beta'} \hat{g}_{\gamma'}|, \tag{9}$$

and we claim that each term that appears is a *projective double cover*. This is defined to mean for any i that appears in $\pi(\beta \cup \gamma \cup \beta' \cup \gamma')$ there are at least two elements in $\beta \cup \gamma \cup \beta' \cup \gamma'$ that project onto this element. That each term that appears in (9) is a projective double cover follows essentially from the above discussion. Namely, for $i \in \alpha$ this i must have a preimage in $\beta \cup \gamma$ as well as in $\beta' \cup \gamma'$. On the other hand, for any $i \notin \alpha$, it has either no preimage or two preimages in each of $\beta \cup \gamma$ and $\beta' \cup \gamma'$.

Furthermore, we claim that any term appears for at most 2^s different α's. This follows, as α contains any element in the symmetric difference of $\pi(\beta)$ and $\pi(\gamma)$ but may or may not contain any element in the intersection of these two sets, making the number of possible α's at most 2^s.

The idea of the following lemma is from [5]. The bound of the lemma is slightly different from the bound claimed in the conference version of this work [9], due to a slight glitch in the earlier proof. The exact value of the bound is, however, not important, as it affects only implicit constants.

LEMMA 4.10. *Suppose* $\sum_\beta \hat{g}_\beta^2 = 1$ *and that each set* β *occurring is of size at most* s *and is shattered. Then the sum* (9) *taken over distinct projective double covers is at most* $(81d/2)^{2s}$.

Proof. As a first attempt, define a function $G : \{-1, 1\}^M \mapsto \mathbb{R}$ through its Fourier coefficients by setting

$$\hat{G}_\alpha = \sum_{\pi(\beta)=\alpha} |\hat{g}_\beta|.$$

Applying the Cauchy–Schwarz inequality, we get

$$\hat{G}_\alpha^2 \le d^s \sum_{\pi(\beta)=\alpha} \hat{g}_\beta^2,$$

and hence

$$\|G\|_2 \le d^{s/2} \|g\|_2 \le d^{s/2}. \tag{10}$$

We also note that G is of degree at most s. Let us consider $E[G(x)^4]$ by expanding

$$\left(\sum_{\alpha} \hat{G}_\alpha \chi_\alpha(x)\right)^4$$

and taking the expectation of each term. It is not difficult to see that the result is a sum of the type (9) but where the sum is over *projective even covers*. A quadruple is a projective even cover iff, for any i, the number of elements in $\beta \cup \gamma \cup \beta' \cup \gamma'$ that project onto i is even. By definition, G is a function of degree at most s, and thus, by the standard hypercontractive estimate $\|G\|_q \leq \sqrt{q-1}^s \|G\|_2$, we have

$$\|G\|_4 \leq 3^{s/2}\|G\|_2 \leq (3d)^{s/2}. \tag{11}$$

We conclude that the sum (9) over projective even covers is bounded by this number raised to the fourth power.

To instead study the sum over projective double covers we make four probabilistic variants $G^{(j)}$, $1 \leq j \leq 4$, of G and study $E[G^{(1)}G^{(2)}G^{(3)}G^{(4)}]$. Each of these functions has the same distribution, and the functions are generated independently and let us describe the construction of $G^{(1)} : \{-1,1\}^{2M} \mapsto \mathbb{R}$.

For each term $\hat{g}_\beta$ we randomly create a set $F(\beta) \subseteq 2M$ by, for each $j \in \beta$ with $\pi(j) = i$, randomly including i, $i+M$, or both elements, where we use each alternative with probability $1/3$. We now set each Fourier coefficient of $G^{(1)}$ as

$$\hat{G}^{(1)}_\alpha = \sum_{F(\beta)=\alpha} |\hat{g}_\beta|.$$

Now consider

$$E[G^{(1)}G^{(2)}G^{(3)}G^{(4)}],$$

by replacing each function by its Fourier expansion and taking the expectation of each term separately. This expectation equals a sum of the type (9), where the sum is over $\beta, \gamma, \beta', \gamma'$, which, under the above probabilistic procedure, produce an even cover of the new variables. Let us analyze the probability that any such candidate appears in the sum.

As each set is of size at most s and we have a double cover, we have $|\pi(\beta \cup \gamma \cup \beta' \cup \gamma')| \leq 2s$. Now take any i in this set, and let us analyze the probability that both x_i and x_{i+M} appear an even number of times. Suppose that there are elements $j \in \beta$ and $j' \in \beta'$ which project onto i (the other cases being symmetric). Fixing the choice of the number of occurrences of x_i and x_{i+M} in $F(\gamma)$ and $F(\gamma')$, it is not difficult to see that the probability that both variables appear an even number of times is at least $2/9$.

As the random construction is independent for different values of i, we see that the probability that any individual term, which is a double cover, appears in the sum corresponding to $E[G^{(1)}G^{(2)}G^{(3)}G^{(4)}]$ is at least $(2/9)^{2s}$. By Hölder's inequality we have

$$E[G^{(1)}G^{(2)}G^{(3)}G^{(4)}] \leq \|G^{(1)}\|_4 \|G^{(2)}\|_4 \|G^{(3)}\|_4 \|G^{(4)}\|_4.$$

Furthermore, the L^2-norm of $G^{(j)}$ is bounded in the same way as that of G, and as $G^{(j)}$ is of degree at most $2s$, we conclude that

$$\|G^{(j)}\|_4 \leq 3^s \|G^{(j)}\|_2 \leq (9d)^{s/2}. \tag{12}$$

Summing up, we get total estimate

$$(9/2)^{2s}(9d)^{2s} \leq (81d/2)^{2s} \tag{13}$$

for the sum over all double covers, and the proof is complete. □

Summing up, we get that $|E[f(x)g(y)g(z)]|$ is bounded by (remember that each term in the sum (9) can appear at most 2^s times)

$$\sum_{\beta} |\hat{f}_{\pi(\beta)}\hat{g}_{\beta}^2|(1-\delta/d)^{|\beta|} + (81d)^s\delta/d + 2e^{-\frac{\delta^2 S}{d^3 2^{2d+1}}} + 2\|g_2\|_2 + 2\|g_3\|_2. \tag{14}$$

We first take care of the last term using the smoothness property.

LEMMA 4.11. *$E[\|g_3\|_2] \leq (s^2T^{-1})^{1/2}$. The expectation is taken over a random question Q_1 that can be asked jointly with Q_2.*

Proof. We have

$$\|g_3\|_2^2 = \sum \hat{g}_{\beta}^2,$$

where the sum is taken over β of that are of size at most s and which are not shattered by π. For any $j_1, j_2 \in \beta$ such that $j_1 \neq j_2$ we have that the probability that they give the same values to the T questions sent to both provers is at most $\frac{1}{T}$, and as we have less than s^2 such pairs, the probability that any individual β is not shattered is bounded by s^2T^{-1}. Since

$$\sum \hat{g}_{\beta}^2 \leq 1,$$

the lemma follows. □

We proceed to bound $\|g_2\|_2$, and this time the random choice of δ is the key mechanism.

LEMMA 4.12. *$E[\|g_2\|_2] \leq k^{-1/2}$. The expectation is taken over a random value of i in $NTW_{\delta'}^k$.*

Proof. If i is chosen in the protocol, then

$$\|g_2\|_2^2 = \sum_{s_{\delta_i} \leq |\beta| \leq S_{\delta_i}} \hat{g}_{\beta}^2.$$

These summation intervals are disjoint, and the sum over all β is bounded by 1. The lemma follows. □

Now by setting the constant c' in the definition of s sufficiently small, we can ensure that

$$(81d)^s\delta/d \leq \sqrt{\delta},$$

and similarly, by setting the constant c'' in the definition of S large enough, we get

$$2e^{-\frac{\delta^2 S}{d^3 2^{2d+1}}} \leq \sqrt{\delta}.$$

Adding all our terms together and using that $\delta_0 = \delta'$ is the largest of the δ-values in $NTW_{\delta'}^k$ while δ_{k-1} is the smallest such value, we can conclude that when taking the expectation for the full test $NTW_{\delta'}^k$

$$|E[f(x)g(y)g(z)]| \leq E\left[\sum_{\beta} |\hat{f}_{\pi(\beta)}\hat{g}_{\beta}^2|(1-\delta_{k-1}/d)^{|\beta|}\right] + 2\delta'^{1/2} + 2k^{-1/2} + 2\left(\frac{s_{\delta_{k-1}}^2}{T}\right)^{1/2}.$$

Let us choose $\delta' \leq (\epsilon'/10)^2$ and $k \geq (10/\epsilon')^2$ and finally set $T \geq 100 s^2_{\delta_{k-1}}/\epsilon'^2$. Then by collecting all terms in (3), we see that the verifier accepts with probability at most

$$\frac{5 + 3\epsilon'/10 + 3\sum_\beta |\hat{f}_{\pi(\beta)}\hat{g}^2_\beta|(1-\delta_{k-1}/d)^{|\beta|} + 18\epsilon'/10}{8}.$$

This implies that if the verifier in $NTW^k_{\delta'}$ accepts with probability at least $\frac{5}{8} + \epsilon'$, then

$$E\left[\sum_\beta |\hat{f}_{\pi(\beta)}\hat{g}^2_\beta|(1-\delta_{k-1}/d)^{|\beta|}\right] \geq \epsilon',$$

where the sum is over β which are of size at most $s_{\delta_{k-1}}$ and shattered by π. By the Cauchy–Schwarz equation we have

$$\begin{aligned}\sum_\beta |\hat{f}_{\pi(\beta)}\hat{g}^2_\beta|(1-\delta_{k-1}/d)^{|\beta|} &\leq \left(\sum_\beta \hat{f}^2_{\pi(\beta)}\hat{g}^2_\beta(1-\delta_{k-1}/d)^{2|\beta|}\right)^{1/2}\left(\sum_\beta \hat{g}^2_\beta\right)^{1/2} \\ &\leq \left(\sum_\beta \hat{f}^2_{\pi(\beta)}\hat{g}^2_\beta(1-\delta_{k-1}/d)^{2|\beta|}\right)^{1/2},\end{aligned}$$

and using $E[X^2] \geq E[X]^2$, we can conclude that

$$E\left[\sum_\beta \hat{f}^2_{\pi(\beta)}\hat{g}^2_\beta(1-\delta_{k-1}/d)^{2|\beta|}\right] \geq \epsilon'^2. \tag{15}$$

Now consider the following probabilistic strategy for P_1 and P_2 in the basic two-prover game.

Add the same independent random T copies of q_2 to each of the two questions, and look at the corresponding tables f and g in the PCP. Pick sets α and β with probabilities $\hat{f}^2_\alpha$ and $\hat{g}^2_\beta$, respectively. Look at the elements of these sets and what answers they give to the added questions. Take the element that defines the lexicographically first value on these added questions, and return the answer of this element to the real question. We claim that if $\pi(\beta) = \alpha$ and β is shattered, then this strategy succeeds. This follows as each element of β and α gives different values to the answers to the added questions, and thus choosing the element that gives the lexicographically first value (or any other uniquely defined value) as these values gives coordinated strategies such that the answers also respect π in the essential coordinate. It follows that the success probability of this strategy is at least

$$E\left[\sum_\beta \hat{f}^2_{\pi(\beta)}\hat{g}^2_\beta\right], \tag{16}$$

where the sum is over shattered β. Comparing this to the expression (15), we see that this success probability is at least ϵ'^2. This completes the proof of Lemma 4.2. ∎

5. Conclusions. As the ideas contained in [6] did not seem sufficient for proving our main theorem, it is instructive to see what additional ideas were used. Note that the idea of choosing a random bias δ was used in [6] to prove approximation resistance of Max-3Sat on satisfiable instances. Khot later realized (in an, as far as we know, unpublished note) that this complication was not needed by starting the reduction with a smooth instance of label cover.

An important part of the current paper is combining these two ingredients. While the interaction of these two ideas is not technically difficult, the resulting constants are very poor. We are not aware of any other approximation resistance result where the size blow-up as a function of the parameter ϵ is equally dramatic. While smoothness is very convenient, it is not clear whether the choice of a random noise parameter δ is needed or just an artifact of the analysis. In our eyes it is quite possible that a fixed small value of δ could be sufficient to prove our main theorem.

Another important part of the current paper is using the correlated spaces of Mossel and the seemingly simple, but very powerful, results on what happens to these spaces under products.

We also note that another key ingredient is the final step, where we manage to coordinate the strategies of P_1 and P_2 in the basic two-prover game. In previous similar arguments it has been enough to choose a random element in the picked sets α and β. This is not sufficient, in the current situation, as the resulting acceptance probability would be much smaller than the soundness in the basic two-prover game. This seems inherent in that the small noise parameter (which is essentially δ/d) results in the need to analyze sets of cardinality larger than d, and thus picking a random element from such a set would not violate the soundness of the label cover game (which is larger than $1/d$ in all constructions). To work with a higher noise parameter than $1/d$ seems incompatible with making y (and z) have the uniform distribution. Of course there could be other ways to take care of this problem.

This idea of coordinating the strategies of P_1 and P_2 could probably be used in some previous arguments in other papers, but it is not clear to us that it would result in a significant strengthening of any previous result.

It is an interesting open question to what extent the current methods can be used to eliminate the need for the d-to-1 conjecture in other situations. Huang [10] extended the results of O'Donnell and Wu (also assuming the d-to-1 conjecture) to prove that for any arity $k \geq 4$ the predicate which accepts all odd strings and one even string is approximation resistant on satisfiable instances. It is likely that our proof could be adapted to this situation, but, on the other hand, Wenner [17] has already established this result using other methods (which, however, cannot handle our case, $k = 3$), giving much better constants.

We finally note that we get that not-two is useless on satisfiable instances in the sense of [2].

Acknowledgments. I thank Sangxia Huang and Cenny Wenner for discussions relating to this paper, and I am also grateful to Oded Goldreich for some comments on the presentation, and to John Wright for pointing out the consequence for the 3-to-1 conjecture. The paper has also benefited from a careful reading by the referees.

REFERENCES

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

[2] P. Austrin and J. Håstad, *On the usefulness of predicates*, ACM Trans. Comput. Theory, 5 (2013), pp. 1–24.
[3] S. O. Chan, *Approximation resistance from pairwise independent subgroups*, in Proceedings of the 45th Annual ACM Symposium on Theory of Computing, 2013, pp. 447–456.
[4] M. Goemans and D. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. ACM, 42 (1995), pp. 1115–1145.
[5] J. Håstad, *Clique is hard to approximate within* $n^{1-\epsilon}$, Acta Math., 182 (1999), pp. 105–142.
[6] J. Håstad, *Some optimal inapproximability results*, J. ACM, 48 (2001), pp. 798–859.
[7] J. Håstad, *On the approximation resistance of a random predicate*, Comput. Complex., 18 (2009), pp. 413–434.
[8] J. Håstad, *Satisfying degree-d equations of* $GF[2]^n$, in Proceedings of APPROX 2011, Lecture Notes in Comput. Sci. 6845, Springer, New York, 2011, pp. 242–253.
[9] J. Håstad, *On the NP-hardness of Max-Not-*2, in Proceedings of APPROX 2012, Lecture Notes in Comput. Sci. 7408, Springer, New York, 2012, pp. 170–181.
[10] S. Huang, *Approximation resistance on satisfiable instances for predicates strictly dominating parity*, ECCC Report 12-040, 2012.
[11] S. Khot, *On the power of unique* 2*-prover* 1*-round games*, in Proceedings of the 34th ACM Symposium on Theory of Computing, 2002, pp. 767–775.
[12] S. Khot and R. Saket, *A* 3*-query non-adaptive PCP with perfect completeness*, in Proceedings of the 21st Annual Conference on Computation, IEEE Computer Society, Piscataway, NJ, 2006, pp. 159–169.
[13] S. Khot, *Hardness results for coloring* 3*-colorable* 3*-uniform hypergraphs*, in Proceedings of the 43rd Annual IEEE Symposium of Foundations of Computer Science, 2002, pp. 23–32.
[14] E. Mossel, *Gaussian bounds for noise correlation of functions*, Geomet. Funct. Anal., 19 (2010), pp. 1713–1756.
[15] R. O'Donnell and Y. Wu, *Conditional hardness for satisfiable* 3*-CSPs*, in Proceedings of the 41st ACM Symposium on Theory of Computing, 2009, pp. 493–502.
[16] R. Raz, *A parallel repetition theorem*, SIAM J. Comput., 27 (1998), pp. 763–803.
[17] C. Wenner, *Circumventing d-to-*1 *for approximation resistance of satisfiable predicates strictly containing parity of width four*, in Proceedings of APPROX 2012, Lecture Notes in Comput. Sci. 7408, Springer, New York, 2012, pp. 325–337.
[18] U. Zwick, *Approximation algorithms for constraint satisfaction problems involving at most three variables per constraint*, in Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, 1998, pp. 201–210.