Optimal Inapproximability Results for MAX-CUT and Other 2-variable CSPs?

Subhash Khot[†]
School of Mathematics
Institute for Advanced Study
Princeton, NJ
khot@ias.edu

Guy Kindler
DIMACS
Rutgers University
Piscataway, NJ
gkindler@dimacs.rutgers.edu

Elchanan Mossel[*]
Department of Statistics
U.C. Berkeley
Berkeley, CA
mossel@stat.berkeley.edu

Ryan O'Donnell[†]
School of Mathematics
Institute for Advanced Study
Princeton, NJ
odonnell@ias.edu

June 19, 2004

Preliminary version

Abstract

In this paper we give evidence that it is NP-hard to approximate MAX-CUT to within a factor of $\alpha_{\text{GW}}+\epsilon$, for all $\epsilon > 0$. Here α_{GW} denotes the approximation ratio achieved by the Goemans-Williamson algorithm [22], $\alpha_{\text{GW}} \approx .878567$. We show that the result follows from two conjectures: a) the Unique Games conjecture of Khot [33]; and, b) a very believable conjecture we call the Majority

[*]Supported by a Miller fellowship in Computer Science and Statistics, U.C. Berkeley

[†]This material is based upon work supported by the National Science Foundation under agreement Nos. DMS-0111298 and CCR-0324906 respectively. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Is Stablest conjecture. Our results suggest that the naturally hard "core" of MAX-CUT is the set of instances in which the graph is embedded on a high-dimensional Euclidean sphere and the weight of an edge is given by the squared distance between the vertices it connects.

The same two conjectures also imply that it is NP-hard to $(\beta+\epsilon)$-approximate MAX-2SAT, where $\beta \approx .943943$ is the minimum of $(2 + \frac{2}{\pi}\theta)/(3 - \cos(\theta))$ on $(\frac{\pi}{2}, \pi)$. Motivated by our proof techniques, we show that if the MAX-2CSP and MAX-2SAT problems are slightly restricted — in a way that seems to retain all their hardness — then they have $(\alpha_{\mathrm{GW}} - \epsilon)$- and $(\beta - \epsilon)$-approximation algorithms, respectively.

Though we are unable to prove the Majority Is Stablest conjecture, we give some partial results and indicate possible directions of attack. Our partial results are enough to imply that MAX-CUT is hard to $(\frac{3}{4} + \frac{1}{2\pi} + \epsilon)$-approximate ($\approx .909155$) assuming only the Unique Games conjecture. We also discuss MAX-2CSP problems over non-boolean domains and state some related results and conjectures, including a hardness results which given the Unique Games conjecture, shows that it is hard to approximate MAX-2LIN(q) to within any constant factor.

1 Introduction

The main result in this paper is a bound on the approximability of the MAX-CUT problem that matches the approximation ratio achieved by the well-known Goemans-Williamson algorithm [22]. The proof of this hardness result unfortunately relies on two unproven conjectures. These conjectures are the *Unique Games conjecture* of Khot [33] and a commonly believed conjecture we call the *Majority Is Stablest conjecture*. For the convenience of the reader, we give a brief description of the conjectures; formal statements appear in Sections 3 and 4, respectively.

Unique Games conjecture (roughly): Given a bipartite graph G, a large constant size set of labels $[M]$, and a permutation of $[M]$ written on each edge, consider the problem of trying to find a labeling of the vertices of G from $[M]$ so that each edge permutation is 'satisfied.' The conjecture is that if M is a large enough constant then it is NP-hard to distinguish instances which are 99% satisfiable from instances which are 1% satisfiable.

Majority Is Stablest conjecture (roughly): Let f be a boolean function which is equally often 0 or 1. Suppose the string x is picked uniformly at random and the string y is formed by flipping each bit of x independently with probability η; we call $\Pr[f(x) = f(y)]$ the *noise stability* of f. The conjecture states that among all f in which each coordinate has $o(1)$ 'influence,' the Majority function has the highest noise stability, up to an additive $o(1)$.

We add in passing that the name Majority Is Stablest is a bit of a misnomer in that almost all balanced boolean threshold functions are equally noise stable (see Theorem 5).

Let us discuss the reasons why we believe this result is important despite its reliance on unproven conjectures. First and foremost, we believe that it is quite remarkable these two conjectures should yield a *matching* hardness of approximation ratio for MAX-CUT, and that indeed the best factor should be the peculiar number α_{GW}. It is intriguing that the precise quantity α_{GW} should arise from a noise stability property of the Majority function, and certainly there was previously little evidence to suggest that the Goemans-Williamson algorithm might be optimal.

As regards the conjectures themselves, we strongly believe in the Majority Is Stablest conjecture. The experts in "analysis of boolean functions" whom we have consulted have agreed that the conjecture should be correct (actually, similar conjectures already appear in the literature. See e.g. Conjecture 5.1 in [30]), and every relevant piece of evidence is in concordance with the conjecture. Because of this, we believe that understanding the status of the Unique Games conjecture is the main issue.

Unlike the Majority Is Stablest conjecture, the Unique Games conjecture is far from certain to be true; in fact, there is no particularly strong evidence either for it or against it. Rather than this being a problem for our studies of MAX-CUT, we can view this paper as an investigation into the Unique Games conjecture via the lens of MAX-CUT. First, we see that the conjecture does not give an incorrectly strong hardness bound for MAX-CUT, and indeed (as it does for Vertex Cover [34]) it gives an ultimately natural bound. Second, we show that the Unique Games problem is formally easier than the problem of beating the Goemans-Williamson algorithm for MAX-CUT (modulo the Majority Is Stablest conjecture) and thus give encouragement for attacking Unique Games algorithmically.

Another reason we believe our result is important despite its reliance on conjectures is related to this last point. Since the Goemans-Williamson algorithm was published a decade ago there has been no algorithmic progress on approximating MAX-CUT. Since Håstad's classic inapproximability paper [25] five years ago there has been no progress on the hardness of approximating MAX-CUT, except for the creation of a better reduction gadget. As one of the most natural and simple problems to have resisted matching approximability bounds, we feel that it deserves further investigation and analysis. In particular, we think that regardless of the truth of the Unique Games conjecture, this paper gives interesting insight into the geometric nature of MAX-CUT. Indeed, insights we have gleaned from studying the MAX-CUT problem in this light have motivated us to give new positive approximation algorithms for variants on other 2-variable CSPs such as MAX-2SAT; see Appendix F.

Finally, we believe that the Fourier-analytic problems we consider in this work, and in particular the Majority Is Stablest conjecture itself, are of significant independent interest. First, the conjecture has interesting applications outside of this work — to the economic theory of social choice [30] for example — and will quite likely prove useful for other PCP-based inapproximability results. Second, the conjecture is an extension of, or is very similar to, several other important theorems in the analysis of boolean functions, including the KKL theorem [29] and Bourgain's theorem [8]. Third, the partial progress we make on proving this conjecture is independently interesting for the analysis of boolean functions, and it clarifies certain aspects the papers of Friedgut, Kalai, and Naor [21] (cf. Theorem 7) and Talagrand [45] (cf. Theorem 9). We note in passing that our partial progress lets us prove an inapproximability factor of .909155 for MAX-CUT assuming only the Unique Games conjecture; this is already stronger than the best known bound.

1.1 Overview of the paper

In Section 2 we describe the MAX-CUT problem and discuss its history. We then state the Unique Games conjecture in Section 3. The plausibility of the conjecture is discussed further in Appendix A. The Majority Is Stablest conjecture is presented in Section 4. Its plausibility is discussed in Appendix B, and some partial progress towards its proof is discussed in Appendix G. We discuss the geometric aspects of the Majority Is Stablest conjecture and its connection with the Goemans-Williamson approximation algorithm in Section 5. Our main results are stated in Section 6. Some conclusions and directions for further research are given in Section 7.

In Appendix C we make some technical definitions, and give some Fourier analytic formulas that are used later. We then reduce the Majority is Stablest conjecture into more useful forms in Appendix D, in Appendix E we prove the hardness of approximating MAX-CUT, based on the conjectures mentioned above. In Appendix F we investigate the approximability of other 2-CSPs, such as MAX-2SAT. In Appendix H we discuss the MAX-2LIN(q) problem (of satisfying mod q linear equations with two variables in each). We formulate some hardness results and related conjectures, and show a computation of the noise-stability of the plurality function.

2 On MAX-CUT

The MAX-CUT problem is a classic and simple combinatorial optimization problem: Given a graph G, find the size of the largest cut in G. By a cut we mean a partition of the vertices of G into two sets; the size of the cut is the number of edges with one vertex on either side of the partition. One can also consider a weighted version of the problem in which each edge is assigned a nonnegative weight and the goal is to cut as much weight as possible.

MAX-CUT is NP-complete (indeed, it is one of Karp's original NP-complete problems [32]) and so it is of interest to try to find polynomial time approximation algorithms. For maximization problems such as MAX-CUT we say an algorithm gives an α-approximation if it always returns an answer which is at least α times the optimal value; we also often relax this definition to allow randomized algorithms which in expectation give α-approximations. Crescenzi, Silvestri, and Trevisan [11] have shown that the weighted and unweighted versions of MAX-CUT have equal optimal approximation factors (up to an additive $o(1)$) and so we pass freely between the two problems in this paper.

The trivial randomized algorithm for MAX-CUT — put each vertex on either side of the partition independently with equal probability — is a 1/2-approximation, and this algorithm is easy to derandomize; Sahni and Gonzalez [41] gave the first 1/2-approximation algorithm in 1976. Following this some $(1/2 + o(1))$-approximation algorithms were given, but no real progress was made until the breakthrough 1994 paper of Goemans and Williamson [22]. This remarkable work used semidefinite programming to achieve an α_{GW}-approximation algorithm, where the constant $\alpha_{\text{GW}} \approx .878567$ is the trigonometric quantity

$$\alpha_{\text{GW}} = \min_{0<\theta<\pi} \frac{\theta/\pi}{(1-\cos\theta)/2}.$$

The optimal choice of θ is the solution of $\theta = \tan(\theta/2)$, namely $\theta^* \approx 2.33 \approx 134°$, and $\alpha_{\text{GW}} = \frac{2}{\pi \sin\theta^*}$. The geometric nature of Goemans and Williamson's algorithm might be considered surprising, but as we shall see, this geometry seems to be an inherent part of the MAX-CUT problem.

On the hardness of approximation side, MAX-CUT was proved MAX-SNP hard [40] and Bellare, Goldreich, and Sudan [3] explicitly showed that it was NP-hard to approximate MAX-CUT to any factor higher than 83/84. The hardness factor was improved to $16/17 \approx .941176$ by Håstad [27] via a reduction from MAX-3LIN using a gadget of Trevisan, Sorkin, Sudan, and Williamson [46]. This stands as current best hardness result.

Despite much effort and many improvements in the approximation guarantees of other semidefinite programming-based algorithms, no one has been able to improve on the algorithm of Goemans and Williamson. Although the true approximation ratio of Goemans-Williamson was proved to be not more than α_{GW} [31, 17] and the integrality gap of their semidefinite relaxation was also proved to be α_{GW} [17], there appears on the face of it to be plenty of possibilities for improvement. Adding triangle constraints and other valid constraints to the semidefinite program has been suggested, alternate rounding schemes have been proposed, and local modification heuristics that work for special graphs have been proven (see, e.g., [22, 15, 14, 31, 47, 16, 17]). And of course, perhaps a completely different algorithm altogether can perform better. Several papers have either explicitly ([14]) or implicitly ([17]) given the problem of improving on α_{GW} as an important research goal.

In this paper we give evidence that in fact MAX-CUT is hard to approximate within any factor larger than α_{GW}.

3 On the Unique Games conjecture

MAX-CUT belongs to the class of constraint satisfaction problems on 2 variables (2-CSPs). In a k-CSP we are given a set of variables and a set of constraints where each constraint depends on exactly k variables. The goal is to find an assignment to the variables so as to maximize the number of constraints satisfied. In case of MAX-CUT, the vertices serve as variables and the edges as constraints. Every constraint says that its two variables receive different boolean values.

Proving inapproximability results for a k-CSP is equivalent to constructing a k-query PCP with a specific acceptance predicate. Usually the Label Cover problem is a starting point for any PCP construction. Label Cover is a 2-CSP where the variables range over a large (non-boolean) domain. An inapproximability result for boolean CSPs is obtained by encoding the assignment to Label Cover via a binary code and then running PCP tests on the (supposed) encodings. This approach has been immensely successful in proving inapproximability results for k-CSPs with $k \geq 3$ (see for example [27, 42, 24]). However the approach gets stuck once we move to 2-CSPs. We seem to have no techniques for constructing (boolean) 2-query PCPs and the bottleneck seems to be the lack of an appropriate PCP outer verifier.

Khot suggested the Unique Games Conjecture in [33] as a possible direction for proving inapproximability results for some important 2-CSPs, such as Min-2SAT-Deletion, Vertex Cover, Graph-Min-Bisection and MAX-CUT. This conjecture asserts the hardness of Unique Label Cover problem:

Definition 1. *The Unique Label Cover problem $\mathcal{L}(V, W, E, [M], \{\pi^{v,w}\}_{(v,w)\in E})$ is defined as follows: We are given a regular bipartite graph with left side vertices V, right side vertices W, and a set of edges E. The goal is to assign one "label" to every vertex of the graph, where $[M]$ is the set of allowed labels. The labeling is supposed to satisfy certain constraints given by bijective maps $\sigma_{v,w} : [M] \to [M]$. There is one such map for every edge $(v, w) \in E$. A labeling "satisfies" an edge (v, w) if*

$$\sigma_{v,w}(\text{label}(w)) = \text{label}(v).$$

The optimum OPT of the unique label cover problem is defined to be the maximum fraction of edges satisfied by any labeling.

The Unique Label Cover problem is a special case of the Label Cover problem. It can also be stated in terms of 2-Prover-1-Round Games, but the Label Cover formulation is easier to work with. The Unique Games conjecture asserts that this problem is hard:

Unique Games conjecture: *For any η, $\delta > 0$, there exists a constant $M = M(\eta, \delta)$ such that it is NP-hard to distinguish whether the Unique Label Cover problem with label set of size M has optimum at least $1 - \eta$ or at most δ.*

The Unique Games conjectures asserts the existence of a powerful outer verifier that makes only 2 queries (albeit over a large alphabet) and has a very specific acceptance predicate: for every answer to the first query, there is exactly one answer to the second query for which the verifier would accept, and vice versa. Once we have such a powerful outer verifier, we can possibly construct a suitable inner verifier and prove the desired inapproximability results. However, even the inner verifier typically relies on rather deep theorems about the Fourier spectrum of boolean functions, e.g. theorems of Bourgain [8] and Friedgut [19].

The Unique Games conjecture was used in [33] to show that Min-2SAT-Deletion is NP-hard to approximate within any constant factor. The inner verifier is based on a test proposed by Håstad [26] and on Bourgain's theorem. It is also implicit in this paper that the Unique Games conjecture with an additional "expansion-like" condition on the underlying bipartite graph of the Label Cover problem would imply that Graph-Min-Bisection is NP-hard to approximate within any constant factor. Khot and Regev [34] showed that the conjecture implies that Vertex Cover is NP-hard to approximate within any factor less than 2. The inner verifier in their paper is based on Friedgut's theorem and is inspired by the work of Dinur and Safra [13] that showed 1.36 hardness for Vertex Cover. In the present paper we continue this line of research and propose a plausible direction for attacking the MAX-CUT problem. We do construct an inner verifier, but to prove its correctness, we need another powerful conjecture about the Fourier spectrum of boolean functions. This is the subject of the Majority Is Stablest conjecture.

In Appendix A we discuss the plausibility of the Unique Games conjecture. We also consider in Appendix H the analogue of the Majority function over non-boolean q-ary domains, which we call the *Plurality* function. We show that its noise stability tends to 0 as q becomes large. We believe that all balanced q-ary functions with small influences should have $o(1)$ noise sensitivity as $q \to \infty$, and perhaps even a "Plurality Is Stablest" result should hold. This could be a step towards showing equivalence between the Unique Games conjecture and hardness of two-variable linear equations mod q, or even towards proving the Unique Games conjecture itself.

4 On the Majority Is Stablest conjecture

To state the Majority Is Stablest conjecture, we need to begin with some definitions. For these definitions it is traditional and convenient to regard the boolean values TRUE and FALSE as -1 and 1 rather than 0 and 1. So let $f\colon \{-1,1\}^n \to \{-1,1\}$ be an arbitrary boolean function. In all of what follows we consider the set of strings $\{-1,1\}^n$ to be a probability space under the uniform distribution.

Definition 2. *The* influence of x_i on f *is*

$$\mathrm{Inf}_i(f) = \Pr_{x\in\{-1,1\}^n}[f(x_1,\ldots,x_n) \neq f(x_1,\ldots,x_{i-1},-x_i,x_{i+1},\ldots,x_n)].$$

Thus $\mathrm{Inf}_i(f)$, a number between 0 and 1, measures the relevance of the ith coordinate to the boolean function f.

Instead of picking x at random, flipping one bit, and seeing if this changes the value of f, we can instead flip a constant fraction of the bits.

Definition 3. *Let* $0 \le \eta \le 1$*. The* noise stability of f for noise rate η *is defined as follows: Let* x *be a uniformly random string in* $\{-1,1\}^n$ *and form* y *by flipping each bit of* x *independently with probability* η*. Then the noise stability of* f *for* η*-noise is defined to be* $\Pr_{x,y}[f(x) = f(y)]$.

With bits defined as $\{-1,1\}$, instead of looking at the probability that $f(x)$ equals $f(y)$, it will be more natural to look at the correlation between the two. If X and Y are bit-valued random variables with $\mathbf{E}[X] = \mathbf{E}[Y] = 0$, then their correlation is $\mathbf{E}[XY]$; this quantity is 1 if the variables are always equal, 0 if they are uncorrelated, and -1 if they are always unequal. Note that $\mathbf{E}[XY] = 2\Pr[X = Y] - 1$. With correlation in mind, we will rephrase the previous definition.

Definition 4. *Let* $-1 \le \rho \le 1$*. The* noise correlation of f at ρ *is defined as follows: Let* x *be a uniformly random string in* $\{-1,1\}^n$ *and let* y *be a "*ρ*-correlated" copy; i.e., pick each bit* y_i *independently so that* $\mathbf{E}[x_i y_i] = \rho$*. Then the noise correlation is defined to be*

$$\mathbb{S}_\rho(f) = \mathbf{E}_{x,y}[f(x)f(y)].$$

Finally, we need to extend all of these definitions to *real*-valued boolean functions $f\colon \{-1,1\} \to \mathbb{R}$. The definition of noise correlation does not require any change. The way to define influence is as follows:

Definition 5. *For* $f\colon \{-1,1\} \to \mathbb{R}$*, define* $\mathrm{Inf}_i(f) = \mathbf{E}_{(x_1,\ldots,x_{i-1},x_i,\ldots,x_n)}\left[\mathrm{Var}_{x_i}[f]\right]$.

We may now state the Majority Is Stablest conjecture. Informally, the conjecture says that among all balanced boolean functions with small influences, the Majority function has the highest noise correlation. Note that the assumption of small influences is necessary since the 'dictator' function $f(x) = x_i$ provably has the highest noise correlation among all balanced boolean functions, for every ρ. Note that when n tends to infinity, the noise correlation at ρ of the n-bit Majority function approaches $(1 - \frac{2}{\pi}\arccos\rho)$ (this fact was stated in a paper of Gulibaud from the 1960's [23] and is ultimately derived from the Central Limit theorem plus a result from an 1890's paper of Sheppard [43]). Thus we have the formal statement of the conjecture:

Majority Is Stablest conjecture: *Fix* $\rho \in [0,1)$*. Then for any* $\epsilon > 0$ *there is a small enough* $\delta = \delta(\epsilon,\rho) > 0$ *such that if* $f : \{-1,1\}^n \to [-1,1]$ *is any function satisfying* $\mathbf{E}[f] = 0$ *and* $\mathrm{Inf}_i(f) \le \delta$ *for all* $i = 1 \ldots n$*, then*

$$\mathbb{S}_\rho(f) \le 1 - \tfrac{2}{\pi}\arccos\rho + \epsilon.$$

The plausibility of the conjecture. We strongly believe that the Majority Is Stablest conjecture is true, as do the experts in the field whom we consulted. We note that similar, though weaker conjectures, already appear in the literature (see e.g. section 5 in [30]). Further discussion of the plausibility of the conjecture appears in Appendix B.

4.1 Why does it arise?

As described in the previous section, inapproximability results for many problems are obtained by constructing a tailor-made PCP: usually, the PCP is obtained by composing a so-called outer verifier with a so-called inner verifier. The outer verifier for our reduction came from the Unique Label Cover problem, and indeed the outer verifier is almost always a Label Cover problem. It is the inner verifier that is application-specific and its acceptance predicate is tailor-made for the problem one is interested in.

A *codeword test* is an essential submodule of the inner verifier. It is a probabilistic procedure for checking whether a given string is a codeword, most commonly a Long Code (see [3]) word.

Definition 6 (Long Code). *The Long Code over domain $[n]$ is a binary code, where each code-word is in fact the truth-table of a boolean function $f : \{-1,1\}^n \to \{-1,1\}$. The codeword encoding the 'message' $i \in [n]$ is given by the ith dictator function; i.e., the function $f(x_1, x_2, \ldots, x_n) = x_i$.*

A codeword test for the Long Code can often be extended to a full-fledged inner verifier. So in the following, we will focus only on a Long Code test. The choice of the test is determined by the problem at hand, in our case MAX-CUT. The test must read two bits from a Long Code and accept if and only if the values read are distinct. Note that a legal Long Code word is precisely the truth table of a boolean function, in which some coordinate has influence 1. Let us say that a function f is *far from being a Long Code* if all the coordinates have $o(1)$ influences.

We expect the following from a codeword test: a Long Code passes the test with probability c (called the 'completeness' parameter of the test) whereas any function far from being a Long Code passes the test with probability at most s (called the 'soundness' parameter). Once we construct a full-fledged inner verifier, the ratio s/c is exactly the inapproximability factor for MAX-CUT.

The Long Code test. As mentioned before, the test checks a given boolean function $f : \{-1,1\}^n \to \{-1,1\}$ at two random but correlated inputs x and y and checks that $f(x) \neq f(y)$. The test will be precisely a "noise stability" test for some fixed noise rate; i.e., x will be chosen uniformly at random and y will be formed by flipping each bit of x independently with probability $(1-\rho)/2$. Here ρ will be a value between -1 and 0, and therefore y is a *highly* noisy version of x, or alternatively, a moderately noisy version of $-x$. Thus (at least for legal Long Code words) we expect $f(x)$ to be quite *anticorrelated* with $f(y)$; i.e., it should pass the test with relative high probability. Specifically, the probability that the test passes is precisely $\frac{1}{2} - \frac{1}{2}\mathbb{S}_\rho(f)$.

A legal Long Code word (namely a dictator function) has noise correlation precisely ρ and thus the completeness of the Long Code test is $c = \frac{1}{2} - \frac{1}{2}\rho$. The crucial part is in analyzing the soundness parameter.

This is where the Majority Is Stablest conjecture comes in. Suppose $f\colon \{-1,1\}^n \to \{-1,1\}$ is any function that is far from being a Long Code word. By a simple trick (see Proposition 16) we can show that the Majority Is Stablest conjecture — which is stated only for $\rho \geq 0$ — implies that for $\rho < 0$ the noise correlation of f at ρ is at *least* $(1 - \frac{2}{\pi}\arccos\rho)$ (a negative number). Hence it follows that functions that are far from being a Long Code pass the test with probability *at most* $s = \frac{1}{2} - \frac{1}{2}(1 - \frac{2}{\pi}\arccos\rho) = (\arccos\rho)/\pi$.

This leads to an inapproximability ratio of

$$\frac{s}{c} = \min_{-1<\rho<0} \frac{(\arccos \rho)/\pi}{\frac{1}{2} - \frac{1}{2}\rho} = \min_{0\le\theta\le\pi} \frac{\theta/\pi}{(1-\cos\theta)/2} = \alpha_{\mathrm{GW}},$$

precisely the Goemans-Williamson constant.

5 On the geometry of MAX-CUT

We shall now try to (non-rigorously!) explain the connection between the Majority Is Stablest conjecture and the geometric picture that arises from the Goemans-Williamson algorithm. But before going further, let us first note that the approximation ratio achieved by the Goemans-Williamson algorithm arises as the solution for a trigonometric minimization problem, which in turn originates from a geometric setting. To obtain a matching inapproximability constant, it seems essential to introduce some similar geometric structure. Such a structure is present (although it is implicit in the actual proofs) in the construction of our Long Code test.

For the purposes of the following explanation, let us consider the n-dimensional discrete cube $\{-1,1\}^n$ as a subset of the n-dimensional Euclidean unit sphere (we normalize the Euclidean norm accordingly). The Majority Is Stablest conjecture essentially states that the discrete cube is a good approximation of the sphere in a certain sense.

The Goemans-Williamson algorithm. We start with a brief description of how the approximation ratio α_{GW} arises in the Goemans-Williamson algorithm. To find a large cut in a given graph $G = (V, E)$ with n vertices, the Goemans-Williamson algorithm embeds the graph in the unit sphere of $\mathbb{R}^n$, identifying each vertex $v \in V$ with a unit vector $\mathbf{x}_v$ on the sphere. The embedding is selected such that the sum

$$\sum_{(u,v)\in E} \frac{1}{2} - \frac{1}{2}\langle \mathbf{x}_u, \mathbf{x}_v \rangle, \tag{1}$$

involving the inner products of vectors associated with the endpoints of edges of G, is maximized. The maximal sum bounds from above the size of the maximum cut, since the size of every cut can be realized by associating all the vertices from one side of the cut with an arbitrary point $\mathbf{x}$ on the sphere, and associating all other vertices with $-\mathbf{x}$.

Once the embedding is set, a cut in G is obtained by choosing a random hyperplane through the origin and partitioning the vertices according to the side of the hyperplane on which their associated vectors falls. For an edge (u, v) in G, the probability that u and v lie on opposite sides of the random cut is proportional to the angle between $\mathbf{x}_u$ and $\mathbf{x}_v$. More precisely, letting $\rho = \langle \mathbf{x}_u, \mathbf{x}_v \rangle$ denote the inner product between the vectors associated with u and v, the probability that the edge (u, v) is cut is $(\arccos \rho)/\pi$.

The approximation ratio α_{GW} of the Goemans-Williamson algorithm is obtained by noting that

$$\alpha_{\text{GW}} = \min_{\rho \in [-1,1]} \left\{ \frac{(\arccos \rho)/\pi}{1/2 - 1/2\ \rho} \right\} \approx .878567 \tag{2}$$

is the smallest ratio possible between the probability of an edge to be cut and its contribution to (1). Hence the expected size of the cut obtained by the Goemans-Williamson algorithm is at least an α_{GW}-fraction of (1), and therefore it is also at least an α_{GW}-fraction of the maximum cut in G.

Cutting the sphere. In [17],, Feige and Schechtman considered the graph G_ρ whose vertices are all the vectors on the unit sphere, and two vertices are connected by an edge in G_ρ iff their inner product is roughly ρ (we do not get into the precise details). It is shown in [17] that in this graph the largest cut is obtained by any hyperplane through the origin (to state this rigorously one should define appropriate measures etc., but let us remain at a simplistic level for now). Such a hyperplane cuts an $(\arccos \rho)$-fraction of the edges in the graph.

Restricting to the cube. We would like to consider an edge-weighted graph H_ρ which is, in a non-rigorous sense, the graph induced by G_ρ on the discrete hypercube. For two vectors $\mathbf{x}, \mathbf{y}$ on the discrete cube, we define the weight of the edge $(\mathbf{x}, \mathbf{y})$ to be

$$\Pr[X = \mathbf{x} \text{ and } Y = \mathbf{y}],$$

where X and Y are ρ-correlated random elements of the discrete cube. The graph H_ρ resembles G_ρ in the sense that almost all the edge-weight in H_ρ is concentrated on edges $(\mathbf{x}, \mathbf{y})$ for which $\langle \mathbf{x}, \mathbf{y} \rangle \approx \rho$; we call such edges *typical edges.* Let us examine how good H_ρ is as an "approximation" of the graph G_ρ.

Note that the structure of H_ρ is very reminiscent of our Long Code test, mentioned above. To make the similarity even clearer, note that a cut C in H_ρ immediately defines a Boolean function f_C over the discrete cube. It is easy to observe that the size of C (namely the sum of weights of the edges that are cut) is exactly the noise stability of f — i.e., the acceptance probability of the Long Code test with parameter ρ when applied to f_C.

The size of the cut. So how large can the size of C be? If C is determined by a random hyperplane, then a typical edge is cut with probability $\approx (\arccos \rho)/\pi$. The expected size of such a cut is therefore roughly the same as the weight of the maximal cut in G_ρ (when the total weight of the edges in G_ρ is normalized to 1).

There are, however, cuts in H_ρ whose weight is larger than $(\arccos \rho)/\pi$. For example, one can partition the vertices in H_ρ according to their first coordinate, taking one side of the cut C to be the set of vectors in the discrete cube whose first coordinate is 1 and the other side of C to be the set of vectors whose first coordinate is -1. Note that this is the cut defined by the hyperplane which is perpendicular to the first coordinate. When interpreted as a function, C corresponds to the function $f_C(x) = x_1$, namely it is a correct Long Codeword. One can easily observe that the size of C is $\frac{1}{2} - \frac{1}{2}\rho$ — i.e., it is exactly the completeness of the Long Code test with parameter ρ.

The conjecture comes in. The size of one-coordinate cuts in H_ρ is larger than the best cuts achievable in G_ρ. The Majority Is Stablest conjecture implies, however, that essentially those are the only special cases, and that all other cuts in H_ρ are no larger than the maximum cut in G_ρ. That is, it implies that unless f_C depends significantly on one of the coordinates, then the size of C is at most $(\arccos \rho)/\pi + \epsilon$. Stated formally, Proposition 16 in Appendix D implies the following.

Proposition *If the Majority Is Stablest conjecture is true, then the following holds for every $\rho \in (-1, 0]$. For any $\epsilon > 0$ there is a small enough $\delta = \delta(\epsilon, \rho) > 0$ such that if C is a cut in H_ρ such that $\mathrm{Inf}_i(f_C) \le \delta$ for every i, then the size of C is at most $(\arccos \rho)/\pi + \epsilon$*

In Appendix G.3 we prove that the statement of the above Proposition holds with respect to all *hyperplane cuts* even without assuming the Majority Is Stablest conjecture.

6 Our results

6.1 Main results

Our main results regarding MAX-CUT are the following:

Theorem 1. *Assume the Unique Games conjecture and the Majority Is Stablest conjecture. Then it is NP-hard to approximate MAX-CUT to within any factor greater than the Goemans-Williamson constant,*

$$\alpha_{GW} = \min_{0 \leq \theta \leq \pi} \frac{\theta/\pi}{(1-\cos\theta)/2} = \min_{-1<\rho<0} \frac{(\arccos\rho)/\pi}{\frac{1}{2}-\frac{1}{2}\rho}.$$

Theorem 2. *Assume only the Unique Games conjecture. Then it is NP-hard to approximate MAX-CUT to within any factor greater than* $3/4 + 1/2\pi$.

We prove these results in Appendix E.

In Appendix F we discuss how our results are relevant for other 2-bit CSPs besides MAX-CUT. In particular we prove:

Theorem 3. *Assume the Unique Games conjecture and the Majority Is Stablest conjecture. Then it is NP-hard to approximate MAX-2SAT to within any factor greater than* $\beta \approx .943943$*, the number defined in the abstract of this paper and in Equation (5).*

Inspired by this, we are led to consider a slightly weaker version of MAX-2SAT called Balanced-MAX-2SAT, in which each variable appears equally often positively and negatively (see Appendix F for more details). We show that this problem *can* be approximated to within β:

Theorem 4. *Balanced-MAX-2SAT is polynomial-time approximable to within any factor less than* β.

6.2 Partial progress on the Majority Is Stablest conjecture

We now state some partial progress we have made towards proving the Majority Is Stablest conjecture.

First, as mentioned we have shown that the conjecture holds for the subclass of balanced threshold functions; this follows from the following two results:

Theorem 5. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any balanced threshold function, $f(x) = \mathrm{sgn}(a_1x_1 + \cdots a_nx_n)$*[1]*, where $\sum a_i^2 = 1$. If $\delta = \max\{|a_i|\}$, then for all $\rho \in [-1,1]$,*

$$\mathbb{S}_\rho(f) = 1 - \tfrac{2}{\pi}\arccos\rho \pm O(\delta(1-|\rho|)^{-3/2}).$$

Proposition 6. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any balanced threshold function, $f(x) = \mathrm{sgn}(a_1x_1 + \cdots a_nx_n)$, where $\sum a_i^2 = 1$. If $\delta = \max\{|a_i|\}$, then the variable achieving the maximum has influence at least $\Omega(\delta)$.*

Next, we can prove that any function with small influences has no more Fourier weight at level 1 than does Majority:

Theorem 7. *Suppose $f : \{-1,1\}^n \to [-1,1]$ satisfies $\mathrm{Inf}_i(f) \le \delta$ for all i. Then*

$$\sum_{|S|=1} \hat{f}(S)^2 \le \frac{2}{\pi} + C\delta,$$

where $C = 2(1 - \sqrt{2/\pi})$.

As a simple corollary we get a weakened version of the Majority Is Stablest conjecture:

Corollary 8. *Suppose $f\colon \{-1,1\}^n \to [-1,1]$ satisfies $\mathrm{Inf}_i(f) \le \delta$ for all i. Let $C = 2(1 - \sqrt{2/\pi})$.*

For $\rho \ge 0$ and $\mathbf{E}[f] = 0$ we have

$$\mathbb{S}_\rho(f) \le (2/\pi + C\delta)\rho + (1 - 2/\pi - C\delta)\rho^2.$$

For $\rho < 0$ we have

$$\mathbb{S}_\rho(f) \ge (2/\pi + C\delta)\rho + (1 - 2/\pi - C\delta)\rho^3.$$

Finally, we note that one of the most effective techniques in bounding $\mathbb{S}_\rho$ or Fourier weights at low levels is to first bound the Fourier weight at level 1 and then apply random restrictions; c.f. [10, 45, 8]. When performing a random restriction, the resulting function may not be balanced. Thus, in order to apply the "random restriction" technique, a bound for *non-balanced* functions with small influences is needed. We give the following generalization of Theorem 7. This theorem should be compared to the theorem Talagrand [45] stating that for every function $f : \{-1,1\}^n \to \{-1,1\}$ with $\Pr[f = 1] = p \le 1/2$ it holds that $\sum_{|S|=1} \hat{f}(S)^2 \le O(p^2 \log(1/p))$.

[1] Without loss of generality we assume the linear form is never 0.

Theorem 9. *Let ϕ be the Gaussian density function and Φ be the Gaussian distribution function. Let $U(x) = \phi(\Phi^{-1}(x)) : [0,1] \to [0, 1/\sqrt{2\pi}]$ denote the so-called "Gaussian isoperimetric function."*

Suppose $f : \{-1,1\}^n \to [-1,1]$ satisfies $\mathrm{Inf}_i(f) \le \delta$ for all i. Letting $\mu = \frac{1}{2} + \frac{1}{2}\mathbf{E}[f]$, we have

$$\sum_{|S|=1} \hat{f}(S)^2 \le 4\left(U(\mu) + \epsilon\right)^2,$$

where the error term ϵ is given by $\epsilon = \max\{1, \sqrt{|\Phi^{-1}(\mu)|}\} O(\sqrt{\delta})$.

6.3 Multi-valued functions.

It is interesting to study the extensions of the Majority is Stablest conjecture to the case of multi-valued functions of the form $f : [q]^n \to [q]$. Such a function f is *balanced*, if it obtains every element $i \in [q]$ equally often.

One natural candidate for being the generalization of Majority to the q-ary domain is the Plurality function $[q]^n \to [q]$, whose output is the most commonly appearing input. (See Appendix H for more details.) When $q \to \infty$ with n "unbounded," we are able to get an asymptotically sharp formula for the noise stability of Plurality:

Theorem 10. *The noise stability of the Plurality function for $n = \infty$ and $q \to \infty$ is $q^{-(1-\rho)/(1+\rho)+o(1)}$.*

(Note that for any constant $\rho < 1$, this approaches 0 as q tends to ∞.)

We are unable to show that the plurality function is the stablest, or even that the stability of balanced q-ary function where the influence of each coordinate is small tends to zero as q grows. However we do conjecture it. Every function $f : [q]^n \to \mathbb{R}$ can be written in a unique way as $f = \sum_{S \subset [n]} f_S$, where $f_S(x)$ depends only on $\{x_i : i \in S\}$ and $\mathbf{E}[f_S f_T] = 0$ if $S \ne T$. We define noise correlation by $\sum_S \rho^{|S|} \|f_S\|_2^2$ and the i'th influence by $\sum_{S: i \in S} \|f_S\|_2^2$.

Conjecture 24 *Let ρ, $-1 < \rho < 1$ be some fixed parameter. Then there exist positive functions $\delta_\rho, C_\rho : \mathbb{N} \to \mathbb{R}$ such that $\lim_{q \to \infty} C_\rho[q] = 0$ and such that the following holds. For every function $f : [q]^n \to [0,1]$ with $\mathbf{E}[f] = 1/q$, all of whose influences are smaller than $\delta_\rho(q)$,*

$$\sum_{S \ne \emptyset} \rho^{|S|} \|f_S\|_2^2 \le C_\rho[q]/q.$$

This conjecture leads to an inapproximability result for MAX-2LIN(q) – the problem of maximizing the number of satisfied equations, in a system of linear equations mod q with two variables appearing in each equation.

Theorem 25 *The Unique Games conjecture and Conjecture 24 together imply the following. Let $\epsilon > 0$ be any fixed parameter. Then there exists a large enough q, such that given an instance of MAX-2LIN(q) is it NP-hard to distinguish between the case where it is ϵ-satisfiable and the case where it is $(1-\epsilon)$-satisfiable.*

We have some partial results in the direction of Conjecture 24, bounding the first level weight in the generalized Fourier representation of balanced q-ary functions (details and definitions appear in Appendix H). This bound, together with the Unique Games conjecture, already gives an inapproximability result for MAX-2LIN(q).

Theorem 26 *There exists a constant K and a positive function $\delta : \mathbb{N} \to \mathbb{R}$, such that for all functions $f : [q]^n \to \{0,1\}$ with $\Pr[f = 1] = 1/q$ and which satisfy $\|f_i\|_2^2 \le \delta(q)$ for all i, it holds that*

$$\sum_i \|f_i\|_2^2 \le \frac{K \log q}{q^2}.$$

Theorem 27 *Assuming the Unique Games conjecture, the following holds. For every $\epsilon > 0$ there exists a positive δ and an integer q, such that given an instance of MAX-2LIN(q), it is NP-hard to distinguish between the case where the instance is δ-satisfiable, and the case where it is $\epsilon\delta$-satisfiable.*

7 Conclusions

We have shown that together the Unique Games conjecture and the Majority Is Stablest conjecture imply an optimal hardness of approximation constant for MAX-CUT and new hardness results for other 2-bit CSPs. Assuming that the Majority Is Stablest conjecture is true, this means that approximating the Unique Label Cover problem is formally easier than improving the approximation constant for the MAX-CUT problem. The problem of approximating the best solution (in terms of the number of satisfied equations) for a system of linear equations modulo q, over 2 variables each, is a restricted version of the Label Cover problem and is therefore even easier. This suggests that to improve the approximation constant for MAX-CUT, one should first try to tackle this MAX-2LIN(q) problem.

On the other hand, it would be interesting to prove that MAX-2LIN(q) is as hard to approximate as the Unique Label Cover problem. Theorem 27 and Theorem 25 are partial results in that direction. A reduction from the Unique Label Cover problem to MAX-2LIN(q) may increase our understanding of the Unique Games conjecture. But before such a reduction is obtained, it seems that a statement of the form "Plurality is Stablest" is needed.

As for the Majority is Stablest conjecture, proving it seems to be a formidable challenge, but a proof should have many applications other approximability problems besides MAX-CUT.

8 Acknowledgements

We thank Jean Bourgain, Steve Evans, Elad Hazan, Gil Kalai, Wenbo Li and Gideon Schechtman for stimulating discussion.

References

[1] M. Abramowitz and I. Stegun. *Handbook of mathematical functions.* Dover, 1972.

[2] D. Bakry. *L'hypercontractivité et son utilisation en théorie des semigroupes.* Springer-Verlag, 1992.

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

[4] M. Ben-Or and N. Linial. Collective coin flipping. In S. Micali, editor, *Randomness and Computation.* Academic Press, New York, 1990.

[5] I. Benjamini, G. Kalai, and O. Schramm. Noise sensitivity of boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90:5–43, 1999.

[6] R. Bhattacharya and R. Rao. *Normal approximation and asymptotic expansions.* Robert E. Krieger Publishing Company, 1986.

[7] J. Bourgain. An appendix to *Sharp thresholds of graph properties, and the k-sat problem*, by E. Friedgut. *J. American Math. Soc.*, 12(4):1017–1054, 1999.

[8] J. Bourgain. On the distribution of the Fourier spectrum of boolean functions. *Israel Journal of Mathematics*, 131:269–276, 2002.

[9] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson, and N. Linial. The influence of variables in product spaces. *Israel Journal of Mathematics*, 77:55–64, 1992.

[10] J. Bourgain and G. Kalai. Influences of variables and threshold intervals under group symmetries. *Geom. and Func. Analysis*, 7:438–461, 1997.

[11] P. Crescenzi, R. Silvestri, and L. Trevisan. On weighted vs unweighted versions of combinatorial optimization problems. *Information and Computation*, 167(1):10–26, 2001.

[12] I. Dinur, E. Friedgut, and G. Kindler. Bounded noise-insensitive functions are juntas. 2003.

[13] I. Dinur and S. Safra. The importance of being biased. In *Proc. 34th Ann. ACM Symp. on the Theory of Computing*, pages 33–42, 2002.

[14] U. Feige. *Randomized rounding of semidefinite programs — variations on the MAX-CUT example*, volume 1761 of *Lecture Notes in Computer Science*, pages 189–196. 1999.

[15] U. Feige and M. Goemans. Approximating the value of two prover proof systems, with applications to MAX-2SAT and MAX-DICUT. In *Proc. 3rd Ann. Israel Symp. on Theory of Comp. and Sys.*, pages 182–189, 1995.

[16] U. Feige, M. Karpinski, and M. Langberg. Improved approximation of MAX-CUT on graphs of bounded degree. *Electronic Colloq. on Comp. Complexity (ECCC)*, (021), 2000.

[17] U. Feige and G. Schechtman. On the optimality of the random hyperplane rounding technique for MAX-CUT. *Random Structures and Algorithms*, 20:403–440, 2002.

[18] W. Feller. *An introduction to probability theory and its applications.* John Wiley & Sons, 1968.

[19] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):474–483, 1998.

[20] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124:2993–3002, 1996.

[21] E. Friedgut, G. Kalai, and A. Naor. Boolean functions whose Fourier transform is concentrated at the first two levels. To appear, *Advances in Mathematics.*

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

[23] G. Guilbaud. *Theories of general interest, and the logical problem of aggregation*, pages 262–307. MIT PRess, 1966.

[24] V. Guruswami, J. Håstad, and M. Sudan. Hardness of approximate hypergraph coloring. In *Proc. 41st Ann. IEEE Symp. on Foundations of Comp. Sci.*, pages 149–158, 2000. To appear, SIAM Journal on Computing.

[25] J. Håstad. Some optimal inapproximability results. In *Proc. 29th Ann. ACM Symp. on the Theory of Computing*, pages 1–10, 1997.

[26] J. Håstad. On a protocol possibly useful for MIN-2SAT. Unpublished manuscript, 2001.

[27] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48:798–869, 2001.

[28] N. Johnson and S. Kotz. *Distributions in statistics: continuous multivariate distributions.* John Wiley & Sons, 1972.

[29] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proc. 29th Ann. IEEE Symp. on Foundations of Comp. Sci.*, pages 68–80, 1988.

[30] G. Kalai. A Fourier-theoretic perspective on the Concordet paradox and Arrow's theorem. *Adv. in Appl. Math.*, 29(3):412–426, 2002.

[31] H. Karloff. How good is the Goemans-Williamson MAX-CUT algorithm? *SIAM Journal on Computing*, 29(1):336–350, 1999.

[32] R. Karp. *Reducibility among combinatorial problems*, pages 85–103. Plenum Press, 1972.

[33] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th Ann. ACM Symp. on the Theory of Computing*, pages 767–775, 2002.

[34] S. Khot and O. Regev. Vertex Cover might be hard to approximate to within $2 - \epsilon$. In *Proc. 18th Ann. IEEE Confence on Computational Complexity*, 2003.

[35] D. Kleitman. Families of non-disjoint subsets. *J. Combin. Theory*, 1:153–155, 1966.

[36] H. König, C. Schütt, and N. Tomczak-Jaegermann. Projection constants of symmetric spaces and variants of Khintchine's inequality. *J. Reine Agnew. Math.*, 511:1–42, 1999.

[37] M. Lewin, D. Livnat, and U. Zwick. Improved rounding techniques for the MAX-2SAT and MAX-DICUT problems. In *Proc. 9th Ann. IPCO*, pages 67–82, 2002.

[38] E. Mossel and R. O'Donnell. On the noise sensitivity of monotone functions. In B. Chauvin, P. Flajolet, D. Gardy, and A. Mokkadem, editors, *Trends in Mathematics: Mathematics and Computer Science II*. Birkhäuser, 2002.

[39] E. Mossel, R. O'Donnell, O. Regev, B. Sudakov, and J. Steif. Non-interactive corellation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality. To appear, 2004.

[40] C. Papadimitriou and M. Yannakakis. Optimization, approximation, and complexity classes. *Journal of Computing and Sys. Sci.*, 43:425–440, 1991.

[41] S. Sahni and T. Gonzales. P-complete approximation problems. *Journal of the ACM*, 23:555–565, 1976.

[42] A. Samorodnitsky and L. Trevisan. A PCP characterization of NP with optimal amortized query complexity. In *Proc. 32nd Ann. ACM Symp. on the Theory of Computing*, pages 191–199, 2000.

[43] W. Sheppard. On the application of the theory of error to cases of normal distribution and normal correlation. *Phil. Trans. Royal Soc. London*, 192:101–168, 1899.

[44] M. Talagrand. On Russo's approximate 0-1 law. *Annals of Probability*, 22:1476–1387, 1994.

[45] M. Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996.

[46] L. Trevisan, G. Sorkin, M. Sudan, and D. Williamson. Gadgets, approximation, and linear programming. *SIAM Journal on Computing*, 29(6):2074–2097, 2000.

[47] U. Zwick. Outward rotations: a tool for rounding solutions of semidefinite programming relaxations, with applications to MAX-CUT and other problems. In *Proc. 31st Ann. ACM Symp. on the Theory of Computing*, pages 679–687, 1999.

A Is the Unique Games conjecture true?

The Unique Games conjecture seems very difficult to prove or disprove and there is not much evidence either in support of it or against it. It might even be the case that the gap-version of Unique Label Cover problem is not NP-hard, but rather lies somewhere between P and the class of NP-hard problems.

Consider the following intuition in support of the conjecture: We are given a system of linear equations modulo a prime M such that there is an assignment that satisfies 99% of the equations. Every equation has two variables (this is a natural example of a constraint satisfaction problem with a *bijective* predicate). How good an assignment can a polynomial time algorithm find? It is natural to expect that a polynomial time algorithm can find assignments that satisfy only $\phi(M)$ fraction of equations and $\phi(M) \to 0$ as $M \to \infty$. It would be quite counterintuitive if in polynomial time, one could find an assignment satisfying, say 5% of the equations *irrespective* of how large M is. Roughly speaking, the Unique Games conjecture says that this intuition is correct and it may be an argument in its support. The conjecture has very strong consequences and it would be a pity if it were false.

We also note that it is essential that the (constant) alphabet size M of a Unique Label Instance be a large enough function of η and δ, since [33] gives a polynomial time algorithm that finds a labeling satisfying $1-O(M^2\eta^{1/5}\log(1/\eta))$ fraction of edges in the case that $1-\eta$ are satisfiable.

B Is the Majority is Stablest conjecture true?

There has been a long line of work in the analysis of boolean functions studying the noise sensitivity of functions and the associated Fourier-theoretic quantities [29, 9, 44, 20, 45, 10, 19, 5, 7, 8, 21, 30, 38, 12, 39] all of the evidence therein seems to corroborate the conjecture. Unfortunately, proving the Majority Is Stablest conjecture seems like a non-trivial task; there are very few theorems in the analysis of boolean functions for which sharp constants are known. However, we can discuss the evidence in favor of the conjecture.

The Majority and weighted majority (or balanced threshold) functions play an important role in the study of noise sensitivity of boolean functions. The set of these functions is in a sense the set of all "uniformly noise-stable" functions. In [5], it is shown that a family of monotone functions is asymptotically noise sensitive if and only if it is asymptotically orthogonal to the family of balanced threshold functions;

by asymptotically noise sensitive functions it is meant those that have $\mathbb{S}_\rho(f) = o(1)$ for any constant ρ. Moreover, in Theorem 5 and Proposition 6 we prove that the Majority Is Stablest conjecture holds for all balanced threshold functions. These two facts together support the Majority Is Stablest conjecture.

Stated in terms of Fourier coefficients (see Appendix C.2), the Majority Is Stablest conjecture says that among all "non-junta-like" functions, the one which has most Fourier mass on the lower levels is the Majority function. This is because $\mathbb{S}_\rho(f)$ is a just a weighted sum of the squared Fourier coefficients of f, where coefficients at level k have weight ρ^k. Very strong support for the Majority Is Stablest conjecture is given by Bourgain [8], who showed that non-junta functions f have their Fourier tails $\sum_{|S|>k} \hat{f}(S)^2$ lower bounded by $k^{-1/2-o(1)}$. As Bourgain noted, the Majority function has precisely this tail decay and thus his theorem is "basically" optimal. In other words, Majority has the "least" Fourier weight on higher levels and therefore the "most" Fourier weight on lower levels.

It is interesting to note that the expression $\mathbb{S}_{-1/3}(f)$ plays a central role in a Fourier-theoretic approach to the Condorcet Paradox and Arrow's Theorem given by Kalai [30]. This expression determines the probability of an "irrational outcome" in a certain voting scheme. Much of [30] is devoted to the study of $\mathbb{S}_{-1/3}(f)$ and some of the arguments in [30] are similar to the argument presented here. In particular, Conjecture 5.1 of [30] states that for 'transitive' functions, which have the property that all influences are the same, the sum $\sum_{|S|\le k} \hat{f}(S)^2$ is maximized by the majority function for all k as $n \to \infty$.

Next, in [39] it is shown that Majority is essentially the maximizer for another noise stability problem, namely maximizing the kth norm of $T_\rho f$, where T_ρ is the Bonami-Beckner operator (see Appendix C) among balanced functions f for large k and $n = \infty$. This too is just one more piece of evidence in favor of the Majority Is Stablest conjecture.

Finally we would like to point out two consequences of Majority Is Stablest conjecture that can actually be proved. The first is Theorem 7, proved in this paper (see the following subsection for more details): For all functions with small influences, the Fourier mass at level 1 is bounded by the Fourier mass at level 1 of the Majority function. In particular, this theorem already implies a weakened formulation of the Majority Is Stablest conjecture, and this weakened formulation is enough to give an improved hardness of approximation result for MAX-CUT conditional only on the Unique Games conjecture.

The second consequence of Majority Is Stablest is obtained by the well known procedure of inferring results for the Ornstein-Uhlenbeck process from results for

Beckner operators (see, e.g., [2]). Let U_ρ be the Ornstein-Uhlenbeck operator in $\mathbb{R}$; i.e., $U_\rho f(x) = \mathbf{E}[f(\rho x + \sqrt{1-\rho^2}N)]$, where N is a standard Gaussian. The Majority Is Stablest conjecture implies that for $\rho > 0$ and among all functions $f : \mathbb{R} \to \{-1, +1\}$ with Gaussian expected value 0, the f which maximizes $\mathbf{E}[f(N)U_\rho f(N)]$ is the function $f(x) = \mathrm{sgn}(x)$. Motivated by our Majority Is Stablest, Wenbo Li (private communication) recently proved that indeed $f(x) = \mathrm{sgn}(x)$ is the maximizer in the Ornstein-Uhlenbeck case. Li's result may be interpreted in the discrete world as saying that among all symmetric functions with further smoothness properties the most stable function is the Majority function.

C Some technical matters

Let us make some required definitions and technical observations.

C.1 MAX-CUT and MAX-2SAT

In this paper we deal mainly with the MAX-CUT and the MAX-2SAT problems. We give the formal definitions of these problems below.

Definition 7 (MAX-CUT). *Given an undirected graph $G = (V, E)$, the MAX-CUT problem is that of finding a partition $C = (V_1, V_2)$ which maximizes the size of the set $(V_1 \times V_2) \cap E$. Given a weight-function $w : E \to \mathbb{R}^+$, the weighted MAX-CUT problem is that of maximizing*

$$\sum_{e \in (V_1 \times V_2) \cap E} w(e).$$

Definition 8 (MAX-2SAT). *An instance of the MAX-2SAT problem is a set of boolean variables and a set of disjunctions over two literals each, where a literal is either a variable or its negation. The problem is to assign the variables so that the number of satisfied literals is maximized. Given a nonnegative weight function over the set of disjunctions, the weighted MAX-2SAT problem is that of maximizing the sum of weights of satisfied disjunctions.*

As we noted earlier [11] implies that the achievable approximation ratios for the weighted versions of the above two problems are the same, up to an additive $o(1)$, as the approximation ratios of the respective non-weighted versions. Hence in the following, we do not make a distinction between the weighted and the non-weighted versions.

C.2 Analytic notions

In this paper we treat the bit TRUE as -1 and the bit FALSE as 1; we consider functions $f\colon \{-1,1\}^n \to \mathbb{R}$ and say a function is *boolean-valued* if its range is $\{-1,1\}$. The domain $\{-1,1\}^n$ is viewed as a probability space under the uniform measure and the set of all functions $f\colon \{-1,1\}^n \to \mathbb{R}$ as an inner product space under $\langle f,g\rangle = \mathbf{E}[fg]$. The associated norm in this space is given by $\|f\|_2 = \sqrt{\mathbf{E}[f^2]}$. We also define the q norm for every q, $1 \le q < \infty$, by $\|f\|_q = (E[|f|^q])^{1/q}$. In addition, let $\|f\|_\infty = \max\{|f(x)\}$.

Fourier expansion. For $S \subseteq [n]$, let χ_S denote the parity function on S, $\chi_S(x) = \prod_{i\in S} x_i$. It is well known that the set of all such functions forms an orthonormal basis for our inner product space and thus every function $f : \{-1,1\}^n \to \mathbb{R}$ can be expressed as

$$f = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S.$$

Here the real quantities $\hat{f}(S) = \langle f, \chi_S\rangle$ are called the *Fourier coefficients* of f and the above is called the *Fourier expansion* of f. *Parseval s identity* states that $\langle f,g\rangle = \sum_S \hat{f}(S)\hat{g}(S)$ and in particular, $\|f\|_2^2 = \sum_S \hat{f}(S)^2$. Thus if f is boolean-valued then $\sum_S \hat{f}(S)^2 = 1$, and if $f : \{-1,1\}^n \to [-1,1]$ then $\sum_S \hat{f}(S)^2 \le 1$. We speak of f's squared Fourier coefficients as *weights*, and we speak of the sets S being stratified into *levels* according to $|S|$. So for example, by the *weight of f at level 1* we mean $\sum_{|S|=1} \hat{f}(S)^2$.

The Bonami-Beckner operator. For any $\rho \in [-1,1]$ we define the *Bonami-Beckner operator* T_ρ, a linear operator on the space of functions $\{-1,1\}^n \to \mathbb{R}$, by $T_\rho(f)(x) = \mathbf{E}[f(y)]$; where each coordinate y_i of y is independently chosen to be x_i with probability $\frac{1}{2} + \frac{1}{2}\rho$ and $-x_i$ with probability $\frac{1}{2} - \frac{1}{2}\rho$. It is easy to check that $T_\rho(f) = \sum_S \rho^{|S|}\hat{f}(S)\chi_S$. It is also easy to verify the following relation between T_ρ and the noise stability (see Definition 4).

Proposition 11. *Let $f\colon \{-1,1\}^n \to \mathbb{R}$ and $\rho \in [-1,1]$. Then*

$$\mathbb{S}_\rho(f) = \langle f, T_\rho f\rangle = \sum_{S\subseteq[n]} \rho^{|S|}\hat{f}(S)^2.$$

The following identity is a well-known one, giving a Fourier analytic formula for the influences of a coordinate on a function (see Definition 5).

Proposition 12. *Let* $f\colon \{-1,1\}^n \to \mathbb{R}$*. Then for every* $i \in [n]$*,*

$$\mathrm{Inf}_i(f) = \sum_{S \ni i} \hat{f}(S)^2,$$

Once we have the Fourier analytic formula for the influence, we can consider the contribution to the influence of characters of bounded size.

Definition 9. *Let* $f\colon \{-1,1\}^n \to \mathbb{R}$*, and let* $i \in [n]$*. The* k-degree influence of coordinate i on f *is defined by*

$$\mathrm{Inf}_i^{\leq k}(f) = \sum_{\substack{S \ni j \\ |S| \leq k}} \hat{f}(S)^2.$$

D Different forms of the Majority is Stablest Conjecture

Recall the Majority Is Stablest conjecture:

Majority Is Stablest conjecture: *Fix* $\rho \in [0,1)$*. Then for any* $\epsilon > 0$ *there is a small enough* $\delta = \delta(\epsilon, \rho) > 0$ *such that if* $f : \{-1,1\}^n \to [-1,1]$ *is any function satisfying*

$$\mathbf{E}[f] = 0, \textit{ and}$$

$$\mathrm{Inf}_i(f) \leq \delta \textit{ for all } i = 1 \ldots n,$$

then

$$\mathbb{S}_\rho(f) \leq 1 - \tfrac{2}{\pi} \arccos \rho + \epsilon.$$

Some simple reductions give formally weaker versions of the conjecture that are actually equivalent:

Proposition 13. *To prove the Majority Is Stablest conjecture, it suffices to prove that, for all* $\gamma > 0$*, the conjecture holds for all functions* f *satisfying* $|\hat{f}(S)| \leq (1-\gamma)^{|S|}$ *for all* $S \subseteq [n]$*.*

Proof. Fix $\rho \in [0,1)$ and $\epsilon > 0$. Let k' be the minimal k such that $\rho^k < \epsilon/4$. Choose $\gamma > 0$ to satisfy $(1-\gamma)^{-2k'} < 1 + \epsilon/4$.

Assuming the majority is stablest conjecture holds for all g satisfying $|\hat{g}(S)| \leq (1-\gamma)^{|S|}$ for all $S \subseteq [n]$ it follows that there exists a δ such that for all such g if $\mathrm{Inf}_i(g) \leq \delta$ for all i then $\mathbb{S}_\rho(g) \leq 1 - \frac{2}{\pi}\arccos\rho + \epsilon/4$.

Now let f be any function satisfying $\mathrm{Inf}_i(f) \leq \delta$ and let $g = T_{1-\gamma}f$. Clearly $\mathrm{Inf}_i(g) \leq \delta$ for all i. By the assumption above, $\mathbb{S}_\rho(g) \leq 1 - \frac{2}{\pi}\arccos\rho + \epsilon/4$. Moreover for all S with $|S| \leq k'$ we have $\hat{g}^2(S) = (1-\gamma)^{2|S|}\hat{f}^2(S)$ and therefore $\hat{f}^2(S) \leq (1+\epsilon/4)\hat{g}^2(S)$.

Now

$$\begin{aligned}
\mathbb{S}_\rho(f) &= \sum_S \hat{f}^2(S)\rho^s = \sum_{S:|S|\leq k'} \hat{f}^2(S)\rho^s + \sum_{S:|S|>k'} \hat{f}^2(S)\rho^s \\
&\leq \sum_{S:|S|\leq k'} \hat{f}^2(S)\rho^s + \epsilon/4 \leq (1+\epsilon/4)\sum_{S:|S|\leq k'} \hat{g}^2(S)\rho^s + \epsilon/4 \\
&\leq \mathbb{S}_\rho(g) + \epsilon/2 \leq 1 - \tfrac{2}{\pi}\arccos\rho + 3\epsilon/4.
\end{aligned}$$

□

Proposition 14. *To prove the Majority Is Stablest conjecture, it suffices to prove it in the case that f is monotone.*

Proof. It is known that monotone combinatorial shifting in the sense of Kleitman [35] preserves expectation, and only decreases influences [4] and noise stability [5]. □

The Majority Is Stablest conjecture also has the following simple consequences. First, we can replace influences by low-degree influences:

Proposition 15. *If the Majority Is Stablest conjecture is true, it remains true if the assumption that $\mathrm{Inf}_i(f) \leq \delta$ for all i is replaced by the assumption that $\mathrm{Inf}_i^{\leq k'}(f) \leq \delta'$, where δ' and k' are universal functions of ϵ and ρ.*

Proof. The proof is similar to the proof of Proposition 13: Fix $\rho < 1$ and $\epsilon > 0$. Choose γ such that $\rho^k(1-(1-\gamma)^{2k}) < \epsilon/4$ for all k. Let δ be chosen such that if $\mathrm{Inf}_i(g) \leq \delta$ for all i then $\mathbb{S}_\rho(g) \leq 1 - \frac{2}{\pi}\arccos\rho + \epsilon/4$. Choose $\delta' = \delta/2$ and k' such that $(1-\gamma)^{2k'} < \delta'$.

Let f be a function satisfying $\mathrm{Inf}_i^{\leq k'}(f) \leq \delta'$ and let $g = T_{1-\gamma}f$. Note that

$$\mathrm{Inf}_i(g) \leq \sum_{S:i\in S,|S|\leq k'} \hat{f}^2(S) + (1-\gamma)^{2k'} \sum_{S:i\in S,|S|\leq k'} \hat{f}^2(S) < \delta' + \delta' = \delta$$

for all i.

It now follows that $\mathbb{S}_\rho(g) \leq 1 - \frac{2}{\pi}\arccos\rho + \epsilon/4$ and therefore

$$\mathbb{S}_\rho(f) = \mathbb{S}_\rho(g) + \sum_S \hat{f}^2(S)(\rho^k(1-(1-\gamma)^k)) < 1 - \tfrac{2}{\pi}\arccos\rho + 3\epsilon/4.$$

□

Second, we get a "Reverse" Majority Is Stablest conjecture:

Proposition 16. *If the Majority Is Stablest conjecture is true, then for $\rho \in (-1, 0]$, the same conjecture holds in reverse: $\mathbb{S}_\rho(f) \geq 1 - \frac{2}{\pi}\arccos\rho - \epsilon$, and furthermore, the assumption $\mathbf{E}[f] = 0$ becomes unnecessary.*

Proof. Let $f : \{-1,1\}^n \to [-1,1]$ satisfy $\mathrm{Inf}_i(f) \leq \delta$ for all i. Let g be the odd part of f, $g(x) = (f(x) - f(-x))/2 = \sum_{|S| \text{ odd}} \hat{f}(S)x_S$. Then $\mathbf{E}[g] = 0$, $\mathrm{Inf}_i(g) \leq \mathrm{Inf}_i(f)$ for all i, and $\mathbb{S}_\rho(f) \geq \mathbb{S}_\rho(g) = -\mathbb{S}_{-\rho}(g)$, which exceeds $-(1 - \frac{2}{\pi}\arccos\rho + \epsilon)$ by the Majority Is Stablest conjecture applied to g. □

Finally, we can combine the above two consequences to get a result that will be necessary for the reduction from Unique Label Cover to 2-bit CSPs:

Proposition 17. *If the Majority Is Stablest conjecture is true, then so is the following:*

Fix $\rho \in (-1, 0]$. Then for any $\epsilon > 0$ there is a small enough $\delta = \delta(\epsilon, \rho) > 0$ and a large enough $k = k(\epsilon, \rho)$ such that if $f : \{-1,1\}^n \to [-1,1]$ is any function satisfying

$$\mathrm{Inf}_i^{\leq k}(f) \leq \delta \text{ for all } i = 1 \ldots n,$$

then

$$\mathbb{S}_\rho(f) \geq 1 - \tfrac{2}{\pi}\arccos\rho - \epsilon.$$

E Reduction from Unique Label Cover to MAX-CUT

In this section we prove Theorems 1 and 2.

E.1 The PCP

We construct a PCP that reads two bits from the proof and accepts if and only if the two bits are unequal. The completeness and soundness are c and s respectively. This implies that MAX-CUT is NP-hard to approximate within any factor greater than s/c. The reduction from the PCP to MAX-CUT is straightforward and can be considered standard. Let the bits in the proof be vertices of a graph and the tests of the verifier be the edges of the graph. The $\{-1, 1\}$ assignment to bits in the proof corresponds to a partition of the graph into two parts and the tests for which the verifier accepts correspond to the edges cut by this partition.

The completeness and soundness properties of the PCP rely on the Unique Games conjecture and the Majority Is Stablest conjecture. The Unique Label Cover instance given by the Unique Games conjecture serves as the PCP Outer Verifier. The soundness of the Long Code-based inner verifier is implied by the Majority Is Stablest conjecture.

Before we explain the PCP test, we need some notation. For $x \in \{-1, 1\}^M$ and a bijection $\sigma : [M] \to [M]$, let $x \circ \sigma$ denote the string $(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(M)})$. For x, $y \in \{-1, 1\}^M$, let xy denote the M-bit string that is the coordinatewise product of x and y.

The PCP verifier is given the Unique Label Cover instance $\mathcal{L}(V, W, E, [M], \{\sigma_{v,w}\}_{(v,w)\in E})$ given by the Unique Games conjecture. The values of η and δ will be chosen to be sufficiently small later. The verifier expects as a proof the Long Code of the label of every vertex $w \in W$. The verifier is parameterized by $\rho \in (-1, 0)$.

The PCP verifier with parameter $-1 < \rho < 0$

- Pick a vertex $v \in V$ at random and two of its neighbors $w, w' \in W$ at random. Let $\sigma = \sigma_{v,w}$ and $\sigma' = \sigma_{v,w'}$ be the respective bijections for edges (v, w) and (v, w').
- Let f and g be the supposed Long Codes of labels of w and w' respectively.
- Pick $x \in \{-1, 1\}^M$ at random.
- Pick $\mu \in \{-1, 1\}^M$ by choosing each coordinate independently to be 1 with probability $\frac{1}{2} + \frac{1}{2}\rho < \frac{1}{2}$ and -1 with probability $\frac{1}{2} - \frac{1}{2}\rho > \frac{1}{2}$.
- Accept iff
$$f(x \circ \sigma) \neq g((x \circ \sigma')\mu).$$

E.2 Completeness

It is easy to see that the completeness of the verifier is $(1-2\eta)(\frac{1}{2}-\frac{1}{2}\rho)$. Assume that the Label Cover instance has a labeling that satisfies a $1-\eta$ fraction of edges. Take this labeling and encode the labels via Long Codes. We will show that the verifier accepts with probability at least $(1-2\eta)(\frac{1}{2}-\frac{1}{2}\rho)$.

With probability $1-2\eta$, both the edges (v,w) and (v,w') are satisfied by the labeling. Let the labels of v,w,w' be $i,j,j' \in [M]$ respectively such that $\sigma(j)=i=\sigma'(j')$. The functions f,g are the Long Codes of j,j' respectively. Hence

$$f(x\circ\sigma) = x_{\sigma(j)} = x_i, \quad g((x\circ\sigma')\mu) = x_{\sigma'(j')}\mu_{j'} = x_i\mu_{j'}$$

Thus the two bits are unequal (and the test accepts) iff $\mu_{j'}=-1$ which happens with probability $\frac{1}{2}-\frac{1}{2}\rho$.

E.3 Soundness

We will show that the verifier accepts with probability at most $(\arccos\rho)/\pi+\epsilon$. The analysis is standard: we use Fourier methods to show that if the test accepts with probability $(\arccos\rho)/\pi+\epsilon$, then it is possible to "decode" the Long Codes and "extract" a labeling for the Label Cover instance that satisfies a significant fraction of the edges. This is a contradiction provided we choose the soundness of Label Cover to be small enough.

The probability of acceptance can be arithmetized as

$$\Pr[\text{acc}] = \mathop{\mathbf{E}}_{v,w,w',x,\mu}\left[\frac{1}{2}-\frac{1}{2}f(x\circ\sigma)g((x\circ\sigma')\mu)\right].$$

Fix v,w,w' for the time being and analyze the expectation over x,μ. The functions f, g can be expanded using their Fourier representations. We get

$$\frac{1}{2}-\frac{1}{2}\mathop{\mathbf{E}}_{x,\mu}\left[\sum_{S,S'\subseteq[M]}\hat{f}(S)\hat{g}(S')\chi_S(x\circ\sigma)\chi_{S'}(x\circ\sigma')\chi_{S'}(\mu)\right].$$

Noting that

$$\chi_S(x\circ\sigma)=\chi_{\sigma(S)}(x), \quad \chi_{S'}(x\circ\sigma')=\chi_{\sigma'(S')}(x),$$

we have

$$\frac{1}{2}-\frac{1}{2}\mathop{\mathbf{E}}_{x,\mu}\left[\sum_{S,S'}\hat{f}(S)\hat{g}(S')\chi_{\sigma(S)}(x)\chi_{\sigma'(S')}(x)\chi_{S'}(\mu)\right]$$

The expectation over x vanishes unless $\sigma(S) = \sigma'(S')$; note that in this case S and S' have the same size. Also, $\mathbf{E}_\mu[\chi_{S'}(\mu)] = \rho^{|S'|} = \rho^{|S|}$. Hence we get

$$\frac{1}{2} - \frac{1}{2} \sum_{S,S' \,:\, \sigma(S)=\sigma'(S')} \rho^{|S|} \hat{f}(S)\hat{g}(S').$$

If $\Pr[\mathrm{acc}] \geq (\arccos\rho)/\pi + \epsilon$, then for at least an $\epsilon/2$ fraction of $v \in V$, the expectation (over the choice of w, w') of the above expression is at least $(\arccos\rho)/\pi + \epsilon/2$. Fix any such "good" $v \in V$. Note that after fixing v, the vertices w, w' are identically distributed. Taking expectation over w, w' and rewriting,

$$\frac{1}{2} - \frac{1}{2}\sum_S \rho^{|S|} \mathop{\mathbf{E}}_w[\hat{f}(\sigma^{-1}(S))] \mathop{\mathbf{E}}_{w'}\left[\hat{g}(\sigma'^{-1}(S'))\right] \geq (\arccos\rho)/\pi + \epsilon/2$$

This is same as saying

$$\frac{1}{2} - \frac{1}{2}\sum_S \rho^{|S|} \left(\mathop{\mathbf{E}}_w[\hat{f}(\sigma^{-1}(S))]\right)^2 \geq (\arccos\rho)/\pi + \epsilon/2$$

which implies that

$$\sum_S \rho^{|S|} \left(\mathop{\mathbf{E}}_w[\hat{f}(\sigma^{-1}(S))]\right)^2 \leq 1 - \tfrac{2}{\pi}\arccos\rho - \epsilon. \tag{3}$$

Now define a function $h : \{-1,1\}^M \mapsto [-1,1]$ as follows:

$$h(x) = \mathop{\mathbf{E}}_w[f(x \circ \sigma)].$$

Clearly, $\hat{h}(S) = \mathbf{E}_w[\hat{f}(\sigma^{-1}(S))]$ and therefore (3) can be written as

$$\sum_S \rho^{|S|}\hat{h}(S)^2 \leq 1 - \tfrac{2}{\pi}\arccos\rho - \epsilon.$$

We now assume the Majority Is Stablest conjecture and apply Proposition 17 to conclude that h has at least one coordinate, say j, with k-degree influence at least δ. We shall give the label j to v. In this way, all "good" $v \in V$ are labeled. Now since $\mathrm{Inf}_j^{\leq k}(h) \geq \delta$, we have

$$\delta \leq \sum_{\substack{S \ni j \\ |S| \leq k}} \hat{h}(S)^2 = \sum_{\substack{S \ni j \\ |S| \leq k}} \mathop{\mathbf{E}}_w[\hat{f}(\sigma^{-1}(S))]^2 \leq \sum_{\substack{S \ni j \\ |S| \leq k}} \mathop{\mathbf{E}}_w[\hat{f}(\sigma^{-1}(S))^2] = \mathop{\mathbf{E}}_w\left[\mathrm{Inf}^{\leq k}_{\sigma^{-1}(j)}(f)\right]. \tag{4}$$

For any $w \in W$, define the set of candidate labels for w to be

$$\mathrm{Cand}[w] = \{i \in [M] \colon \mathrm{Inf}_i^{\leq k}(f) \geq \delta/2\}.$$

Since $\sum_i \mathrm{Inf}_i^{\leq k}(f) \leq k$, we conclude that $|\mathrm{Cand}[w]| \leq 2k/\delta$. Equation (4) implies that for at least a $\delta/2$ fraction of neighbors w of v we have $\mathrm{Inf}_{\sigma^{-1}(j)}^{\leq k}(f) \geq \delta/2$ and therefore $\sigma^{-1}(j) \in \mathrm{Cand}[w]$. Now we label each vertex $w \in W$ by choosing a random element of $\mathrm{Cand}[w]$ (or any label if this set is empty). It follows that among the set of edges adjacent to "good" vertices v, at least a $(\delta/2)(\delta/2k)$ are satisfied in expectation. Thus it follows that there is labeling for all vertices which satisfies an $(\epsilon/2)(\delta/2)(\delta/2k)$ fraction of all edges. Choosing the soundness of the Label Cover instance to be small enough gives a contradiction.

E.4 Hardness of MAX-CUT

Theorem 1 follows easily from the completeness and soundness properties of the PCP. The completeness is $(1-2\eta)(\frac{1}{2} - \frac{1}{2}\rho)$ and the soundness is $(\arccos \rho)/\pi - \epsilon$. Here η and ϵ can be made arbitrarily small. Thus MAX-CUT is NP-hard to approximate within factor

$$\frac{(\arccos \rho)/\pi}{\frac{1}{2} - \frac{1}{2}\rho} - \epsilon'$$

where we can choose any $\rho \in (-1, 0)$ and ϵ' can be made arbitrarily small. The minimum of the above quantity is precisely the Goemans-Williamson constant α_{GW}.

Using the second result in Corollary 8 — which is unconditional — we can get a weaker hardness of approximation result for MAX-CUT that doesn't depend on the Majority Is Stablest conjecture and only uses the Unique Games conjecture. Repeating the analysis from Appendix E.3 it is easy to see that the resulting hardness factor is

$$\min_{-1<\rho<0} \frac{\frac{1}{2} - \frac{1}{2}[(2/\pi)\rho + (1-2/\pi)\rho^3]}{\frac{1}{2} - \frac{1}{2}\rho} = 3/4 + 1/2\pi \approx .909155.$$

We thus obtain Theorem 2.

F Other 2-bit CSPs

The same method used to prove hardness of approximation for MAX-CUT can be used to give improved hardness of approximation for another important 2-bit CSP, namely MAX-2SAT. Recall that the input to a MAX-2SAT problem is a collection of clauses, i.e. disjunctions, of at most 2 variables; the goal is to find an assignment that satisfies as many clauses as possible.

Assume the Unique Games and Majority Is Stablest conjectures. The natural inner verifier test for MAX-2SAT is this: With probability 1/2 test $f(x \circ \sigma) \vee g((x \circ \sigma')\mu)$; with probability 1/2 test $-f(x \circ \sigma) \vee -g((x \circ \sigma')\mu)$. It is easy to check that this leads to an acceptance probability of $\frac{3}{4} - \frac{1}{4}\mathbb{S}_\rho(h)$. The dictator passes this test with probability $\frac{3}{4} - \frac{1}{4}\rho$; the Majority Is Stablest conjecture implies that no function with small low-degree influences can pass this test with probability exceeding $\frac{3}{4} - \frac{1}{4}(1 - \frac{2}{\pi}\arccos\rho) + \epsilon$. This leads to a hardness of approximation ration of

$$\beta = \min_{-1<\rho<0} \frac{\frac{3}{4} - \frac{1}{4}(1 - \frac{2}{\pi}\arccos\rho)}{\frac{3}{4} - \frac{1}{4}\rho} \approx .943943. \tag{5}$$

This is our Theorem 3.

Note that β is smaller than the best unconditional hardness factor known for MAX-2SAT, $21/22 \approx .954545$, due to Håstad [27] (using the gadget of Bellare, Goldreich, and Sudan [3]); as well, the best algorithm known for MAX-2SAT, due to Lewin, Livnat, and Zwick [37], achieves an approximation ratio of .9401 which is close to and smaller than β.

Our conjectures and methodology do not seem to improve the hardness factors for other 2-bit CSPs beyond α_{GW}. Consider the MAX-2ConjSAT problem, in which the input is a collection of *conjunctions* of (at most) 2 variables and the goal is to satisfy as many conjunctions as possible. The natural inner verifier test is this: With probability 1/2 test $f(x \circ \sigma) \wedge g((x \circ \sigma')\mu)$; with probability 1/2 test $-f(x \circ \sigma) \wedge -g((x \circ \sigma')\mu)$. This leads to an acceptance probability of $\frac{1}{4} - \frac{1}{4}\mathbb{S}_\rho(h)$. Assuming the Majority Is Stablest conjecture, we get the same hardness of approximation for MAX-DICUT as we do for MAX-CUT, α_{GW}, since $(\frac{1}{4} - \frac{1}{4}(1 - \frac{2}{\pi}\arccos\rho))/(\frac{1}{4} - \frac{1}{4}\rho) = ((\arccos\rho)/\pi)/(\frac{1}{2} - \frac{1}{2}\rho)$. In some sense this may not be surprising since the best algorithm known for this problem ([37] again) already achieves an approximation ratio of .8740, which is nearly α_{GW}. In fact, the same paper achieves .8740 even for the most general problem, MAX-2CSP in which arbitrary 2-bit constraints are allowed.

Motivated by these results we are led to conjecture that MAX-2SAT *is* polynomial-time approximable to within any factor less than β and that MAX-2CSP, MAX-DICUT, MAX-2ConjSAT, MAX-2LIN, etc. are all polynomial-time approximable to within any factor less than α_{GW}. We will now show that these bounds *are* achievable for a slight weakening of the problems.

Definition 10. *Given a 2-bit CSP, by its* balanced *version we mean the problem with the restriction that every input instance* $\{C_1, \ldots, C_m\}$ *has the following property: for each* $i = 1 \ldots n$*, the expected number of constraints satisfied when* x_i *is set to* 1 *and the other variables are set uniformly at random is equal to the expected number of constraints satisfied when* x_i *is set to* -1 *and the other variables are set uniformly at random.*

As an example, Balanced-MAX-2SAT is the MAX-2SAT problem with the additional constraint that each variable appears positively and negatively in equally many clauses (in the weighted case, with equal total weight).

We contend that the balanced versions of 2-bit CSPs ought to be equally hard as their general versions; the intuition is that if more constraints are expected to be satisfied if x_i is set to, say, 1 rather than -1, it is a "free hint" that the x_i should be set to TRUE. Note that the reductions we suggest from Unique Label Cover to MAX-2SAT, MAX-2ConjSAT, etc. produce balanced instances, and thus we get the same hardness of approximation bounds, β and α_{GW}, for the balanced problems (conditional on the two conjectures).

We can prove unconditionally that Balanced-MAX-2SAT is polynomial-time approximable to within any factor less than β, and that MAX-2CSP, MAX-DICUT, MAX-2ConjSAT, MAX-2LIN, etc. are all polynomial-time approximable to within any factor less than α_{GW}. By way of illustration, we prove Theorem 4:

Proof. The algorithm is essentially the same as that used by Goemans-Williamson. The input is a collection of clauses C of the form $(y \vee z)$, where $y = \sigma_i x_i$ and $z = \sigma_j x_j$ for some variables x_i and x_j and signs σ_i and σ_j. Arithmetizing each clause with $-1 \vee -1 = 1$, $-1 \vee 1 = 1$, $1 \vee -1 = 1$, $1 \vee 1 = 0$, we get $\frac{3}{4} - \frac{1}{4}y - \frac{1}{4}z - \frac{1}{4}y \cdot z$. Thus we have the objective function

$$\text{OBJ} = \sum_{C=(y \vee z)} \tfrac{3}{4} - \tfrac{1}{4}y - \tfrac{1}{4}z - \tfrac{1}{4}y \cdot z.$$

The condition that the instance is balanced is precisely equivalent to the condition that the linear terms cancel out. (This holds true by definition for all balanced 2-bit CSP problems.) Thus in fact

$$\text{OBJ} = \sum_{C=(y \vee z)} \tfrac{3}{4} - \tfrac{1}{4}y \cdot z.$$

Hence the optimum value of the Balanced-MAX-2SAT instance is

$$\mathrm{OPT} = \max\ \mathrm{OBJ} \qquad \text{subject to } x_i \in \{-1, 1\} \text{ for all } i.$$

Following Goemans-Williamson we directly relax this to a semidefinite program by replacing x_i with a high-dimensional vector v_i, subject to $v_i \cdot v_i = 1$, and solving; in polynomial time we can find a solution $\{v_i\}$ which achieves $\mathrm{SDP} - \epsilon$, where SDP denotes the optimal value of the semidefinite program. We now round by picking r to be a random Gaussian vector and setting $x_i = \mathrm{sgn}(r \cdot v_i)$. Recalling from [22] that this gives $\mathbf{E}[x_i \cdot x_j] = 1 - \frac{2}{\pi}\arccos(v_i \cdot v_j)$, we have for any clause $(y \vee z) = (\sigma_i x_i \vee \sigma_j x_j)$,

$$\mathbf{E}[\tfrac{3}{4} - \tfrac{1}{4}(\sigma_i x_i)\cdot(\sigma_j x_j)] = \tfrac{3}{4} - \tfrac{1}{4}(1 - \tfrac{2}{\pi}\arccos(\sigma_i v_i \cdot \sigma_j v_j)) \geq \beta(\tfrac{3}{4} - \tfrac{1}{4}(\sigma_i v_i \cdot \sigma_j v_j)),$$

where we have used the definition of β and the fact that it is unchanged if we let ρ range over $[-1, 1]$. It follows that $\mathbf{E}[\mathrm{OBJ}] \geq \beta\mathrm{SDP} \geq \beta\mathrm{OPT}$ and the proof is complete. □

G Partial progress on the Majority Is Stablest conjecture

G.1 A bound on the weight at level 1

If the Majority Is Stablest conjecture is true, then it is true for extremely small ρ, and thus we must be able to prove that functions with small influences have no more weight at level 1 than Majority has, viz., $\frac{2}{\pi}$ (up to $o(1)$). We now prove Theorem 7.

Proof. Let ℓ denote the linear part of f, $\ell(x) = \sum_{i=1}^n \hat{f}(\{i\})x_i$. We have that $|\hat{f}(\{i\})| \leq \mathrm{Inf}_i(f) \leq \delta$ for all i. Now $\sum_{|S|=1} \hat{f}(S)^2 = \|\ell\|_2^2$ and

$$\begin{aligned} \|\ell\|_2^2 &= \langle f, \ell \rangle \\ &\leq \|f\|_\infty \|\ell\|_1 \\ &\leq \|\ell\|_1. \end{aligned}$$

Since all of ℓ's coefficients are small, smaller than δ, we expect ℓ to behave like a Gaussian with mean zero and standard deviation $\|\ell\|_2$; such a Gaussian has L^1-norm equal to $\sqrt{2/\pi}\|\ell\|_2$. Several error bounds on the Central Limit Theorem exist to this effect; the sharpest is a result of König, Schütt, and Tomczak-Jaegermann [36] which implies that $\|\ell\|_1 \leq \sqrt{2/\pi}\|\ell\|_2 + (C/2)\delta$. Thus

$$\|\ell\|_2^2 \leq \sqrt{2/\pi}\|\ell\|_2 + (C/2)\delta,$$

hence $\|\ell\|_2 \leq \sqrt{1/2\pi} + \sqrt{1/2\pi + C\delta/2}$ and therefore $\|\ell\|_2^2 \leq 2/\pi + C\delta$. □

The results of König, Schütt, and Tomczak-Jaegermann were used in [21] in order to prove that Boolean functions whose Fourier transform is concentrated at the first two levels are "close" to the dictator function.

Given Theorem 7, it is easy to prove Corollary 8.

Proof. Recall that $\mathbb{S}_\rho(f) = \sum_S \rho^{|S|}\hat{f}(S)^2$ and $\sum_S \hat{f}(S)^2 \leq 1$. For $\rho \geq 0$ and $\mathbf{E}[f] = 0$ we have that $\hat{f}(\emptyset)^2 = 0$ and that $\mathbb{S}_\rho(f)$ is maximized when as much weight as possible is put at level 1 and the rest is at level 2. By Theorem 7, at most $2/\pi + C\delta$ weight can be on level 1. The first result follows.

For $\rho < 0$ (and $\mathbf{E}[f]$ arbitrary), $\mathbb{S}_\rho(f)$ is minimized when as much weight as possible is put at level 1 and the rest is at level 3. The second result again follows from Theorem 7. □

G.2 Improved weight bounds for level 1

In this subsection we prove optimal bounds on the weight of the first level for (not necessarily balanced) functions with low sensitivity. The bound will generalize Theorem 7. It should be compared to a theorem of Talagrand [45]:

Theorem 18. *(Talagrand) Suppose $f : \{-1,1\}^n \to \{-1,1\}$ satisfies $\Pr[f = 1] = p \leq 1/2$. Then*

$$\sum_{|S|=1} \hat{f}(S)^2 \leq O(p^2 \log(1/p)).$$

Our bound in Theorem 9 improves on Theorem 18 for functions with small influences.

One technical tool we will use is the Berry-Eséen theorem, which gives error bounds for the Central Limit Theorem. We use the following version in [18]:

Theorem 19. *(Berry-Eséen) Let $X_1, \ldots, X_n$ be a sequence of independent random variables satisfying $\mathbf{E}[X_j] = 0$ for all j, $(\sum_{j=1}^n \mathbf{E}[X_j^2])^{1/2} = \sigma$, and $\sum_{j=1}^n \mathbf{E}[|X_j|^3] =$*

ρ_3. *Let* $Q = \sigma^{-1}(X_1 + \cdots + X_n)$, *let* F *denote the cumulative distribution function of* Q, $F(x) = \Pr[Q \le x]$, *and let* Φ *denote the cumulative distribution function of a standard normal random variable. Then*

$$\sup_x (1 + |x|^3)|F(x) - \Phi(x)| \le O(\rho_3/\sigma^3).$$

In particular, if A *is any interval in* $\mathbb{R}$, $|\Pr[Q \in A] - \Pr[N(0,1) \in A]| \le O(\rho_3/\sigma^3)$.

We now prove Theorem 9.

Proof. It will be more convenient to work with the $[0,1]$ valued-function $g = \frac{1}{2} + \frac{1}{2}f$ and prove that $\sum_{|S|=1} \hat{g}(S)^2 \le \left(U(\mu) + \max\{1, \sqrt{|\Phi^{-1}(\mu)|}\}O(\sqrt{\delta})\right)^2$. Note that $\mu = \mathbf{E}[g]$. We will assume without loss of generality that $\mu \ge 1/2$ (otherwise look at $\frac{1}{2} - \frac{1}{2}f$).

Let τ denote $(\sum_{|S|=1} \hat{g}(S)^2)^{1/2}$. As in the proof of Theorem 7, we let ℓ be the linear part of g and we know that all of ℓ's coefficients are at most $\delta/2$. The function $L = \ell/\tau = \sum_S \hat{g}(S)\chi_S/\tau$ is a sum of independent random variables $X_S = \hat{g}(S)\chi_S/\tau$. Clearly $\mathbf{E}[X_S] = 0$ for all S. Moreover, $\sum_S \mathbf{E}[X_S^2] = 1$ and $\sum_S \mathbf{E}[X_S^3] \le \max_S |X_S| \le \delta/(2\tau)$.

Now $\tau^2 = \langle g, \ell \rangle$ and therefore $\tau =< g, L >$. We will show below that

$$\tau = \langle g, L \rangle \le U(\mu) + \max\{1, |\Phi^{-1}(\mu)|\}O(\delta/\tau). \tag{6}$$

Multiplying by τ implies that

$$(\tau - \frac{U(\mu)}{2})^2 \le U^2(\mu)/4 + \max\{1, |\Phi^{-1}(\mu)|\}O(\delta).$$

which in turn implies that

$$\tau \le U(\mu) + \max\{1, \sqrt{|\Phi^{-1}(\mu)|}\}O(\sqrt{\delta})$$

Finally, we will conclude that

$$\tau^2 \le \left(U(\mu) + \max\{1, \sqrt{|\Phi^{-1}(\mu)|}\}O(\sqrt{\delta})\right)^2.$$

We now prove (6). Let t be a number such that $\Pr[L > t] = \mu$. Since g is a $[0,1]$ valued-function it follows that $\langle g, L \rangle \le \mathbf{E}[\mathbf{1}_{L>t} L]$.

Letting F denote the cumulative distribution function of L the Berry-Eséen theorem implies that $\sup_x (1 + |x|^3)|F(x) - \Phi(x)| \leq O(\delta/\tau)$. In particular, $|\Pr[L > t] - \Pr[N(0,1) > t]| \leq O(\delta/(\tau(1+t^3))$ and hence

$$|\mu - \Phi(-t)| \leq O(\frac{\delta}{\tau(1+t^3)}) \tag{7}$$

Note that the function U satisfies

$$U'(x) = \phi'(\Phi^{-1}(x)) \times ((\Phi^{-1}(x))') = -\Phi^{-1}(x)\phi(\Phi^{-1}(x))\frac{1}{\phi(\Phi^{-1}(x))} = -\Phi^{-1}(x).$$

Therefore $U''(x) = -1/\phi(\Phi^{-1}(x)) = -1/U(x)$. It follows that U is concave.

We now estimate $U(\mu) - \phi(t)$. Since U' is a monotone function, it follows that

$$\begin{aligned} |U(\mu) - \phi(t)| &= |U(\Phi(-t)) - U(\mu)| \leq |\Phi(-t) - \mu| \max\{|U'(\Phi(-t))|, |U'(\mu)|\} \quad (8) \\ &\leq \max\{|t|, \Phi^{-1}(\mu)\} O(\delta/(\tau(1+t^3))) \leq \max\{1, |\Phi^{-1}(\mu)|\} O(\delta/\tau). \end{aligned}$$

Further,

$$\begin{aligned} < g, L > &\leq \mathbf{E}[\mathbf{1}_{L>t} L] = t\Pr[L > t] + \int_t^\infty \Pr[L > x]\, dx \\ &= t\Pr[L > t] + \int_t^\infty \Pr[N(0,1) > x]\, dx + \int_t^\infty (F(x) - \Phi(x))\, dx \\ &= t\mu - t\Phi(-t) + \phi(t) + \int_t^\infty (F(x) - \Phi(x))\, dx \\ &\leq \phi(t) + |t| \cdot |\mu - \Phi(-t)| + \int_t^\infty |F(x) - \Phi(x)|\, dx \\ &\leq \phi(t) + \frac{|t|}{1+|t|^3} O(\delta/\tau) + O(\delta/\tau) \int_t^\infty 1/(1+|x|^3)\, dx \qquad \text{((7) and Berry-Eséen)} \\ &= \phi(t) + O(\frac{\delta}{\tau(1+t^2)}) \qquad \text{((8))} \\ &\leq U(\mu) + \max\{1, |\Phi^{-1}(\mu)|\} O(\delta/\tau). \end{aligned}$$

which proves (6) as needed. □

G.3 Weighted majorities

In this subsection we show that the Majority Is Stablest conjecture holds for weighted majority functions. We consider function of the form $\text{sgn}(a_1 x_1 + \cdots a_n x_n)$ where $\sum a_i^2 = 1$, it is further assumed that for all $x \in \{-1,1\}^n$ it holds that $\sum a_i x_i \neq 0$.

In order to prove Theorem 5 we will need another version of the Central Limit Theorem with error bounds, this one for multidimensional random variables. The following theorem is from [6, Corollary 16.3]:

Theorem 20. *Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be independent random variables taking values in $\mathbb{R}^k$ satisfying:*

- $\mathbf{E}[\mathbf{X}_j] = 0, j = 1 \dots n;$
- $n^{-1} \sum_{j=1}^{n} \mathrm{Cov}(\mathbf{X}_j) = V$, *where* Cov *denotes the variance-covariance matrix;*
- λ *is the smallest eigenvalue of* V, Λ *is the largest eigenvalue of* V*;*
- $\rho_3 = n^{-1} \sum_{j=1}^{n} \mathbf{E}[|\mathbf{X}_j|^3] < \infty$.

Let Q_n denote the distribution of $n^{-1/2}(\mathbf{X}_1 + \cdots + \mathbf{X}_n)$, let $\Phi_{0,V}$ denote the distribution of the k-dimensional Gaussian with mean 0 and variance-covariance matrix V, and let $\eta = C\lambda^{-3/2}\rho_3 n^{-1/2}$, where C is a certain universal constant.

Then for any Borel set A,

$$|Q_n(A) - \Phi_{0,V}(A)| \le \eta + B(A),$$

where $B(A)$ is the following measure of the boundary of A: $B(A) = 2\sup_{y \in \mathbb{R}^k} \Phi_{0,V}((\partial A)^{\eta'} + y)$, where $\eta' = \Lambda^{1/2}\eta$ and $(\partial A)^{\eta'}$ denotes the set of points within distance η' of the topological boundary of A.

We now prove Theorem 5.

Proof. Since f is antisymmetric, we only need to prove the result for $\rho \in [0, 1]$. Let x and y be ρ-correlated uniformly random strings, let $X_j = a_j x_j$, $Y_j = a_j y_j$, and $\mathbf{X}_j = (X_j, Y_j) \in \mathbb{R}^2$. Let Q_n denote the distribution of $\mathbf{X}_1 + \cdots + \mathbf{X}_n = n^{-1/2}(\sqrt{n}\,\mathbf{X}_1 + \cdots + \sqrt{n}\,\mathbf{X}_n)$. Since $\mathbb{S}_\rho(f) = 2\Pr[f(x) = f(y)] - 1$, we are interested in computing $2Q_n(A_{++} \cup A_{--}) - 1$, where A_{++} denotes the positive quadrant of $\mathbb{R}^2$ and A_{--} denotes the opposite quadrant.

We shall apply Theorem 20. We have $\mathbf{E}[\mathbf{X}_j] = 0$ for all j. We have $\mathrm{Cov}(\sqrt{n}\,\mathbf{X}_j) = na_i^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, and thus $V = n^{-1} \sum \mathrm{Cov}(\sqrt{n}\,\mathbf{X}_j) = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. The eigenvalues of V are $\lambda = 1 - \rho$ and $\Lambda = 1 + \rho$. Since $|\sqrt{n}\,\mathbf{X}_j|$ is $\sqrt{2n}\,|a_i|$ with probability 1, $\rho_3 = n^{-1} \sum \mathbf{E}[|\sqrt{n}\,\mathbf{X}_j|^3] = 2^{3/2} n^{1/2} \sum |a_i|^3 \le 2^{3/2} n^{1/2} \delta$. Thus $\eta = O(1)\delta(1 - \rho)^{-3/2}$ and $\eta' = (1 + \rho)^{1/2}\eta = O(\eta)$.

It is well known (see, e.g., [1, 26.3.19]) that $\Phi_{0,V}(A_{++}) = \Phi_{0,V}(A_{--}) = 1/2 - (1/2\pi)\arccos(\rho)$, and it is easy to check that $B(A_{++} \cup A_{--}) = O(\eta')$. Thus by Theorem 20 we get $Q_n(A_{++} \cup A_{--}) = 1 - (\arccos\rho)/\pi \pm O(\eta)$ and the theorem follows. □

We now prove Proposition 6

Proof. We may assume without loss of generality that $\delta = a_1 \geq a_2 \geq \cdots \geq a_n \geq 0$. Letting X_i denote the random variable $a_i x_i$, we will prove that $\mathrm{Inf}_1(f) \geq \Omega(\delta)$ by proving that

$$\Pr[|X_2 + \cdots + X_n| \leq \delta] \geq \Omega(\delta). \tag{9}$$

Let C be the constant hidden in the $O(\cdot)$ in the final part of the Berry-Eséen theorem, Theorem 19. In proving (9) we may also assume without loss of generality that

$$1 - 100C^2\delta^2 \geq 1/4; \tag{10}$$

i.e., δ is not too large. We may also assume C is an integer.

Let $m = 100C^2{+}2$. We will split into two cases, depending on the magnitude of a_m. In either case, we shall apply the Berry-Eséen theorem to the sequence $X_m, \ldots, X_n$. We have

$$\sigma = (\sum_{j=m}^{n} \mathbf{E}[X_j]^2)^{1/2} = (\sum_{j=m}^{n} a_j^2)^{1/2} \geq (1 - (m-2)\delta^2)^{1/2} \geq (1 - 100C^2\delta^2)^{1/2} \geq 1/2,$$

where we have used (10). We also have $\rho_3 = \sum_{j=m}^{n} \mathbf{E}[|X_j|^3] \leq \sum_{j=m}^{n} a_m \mathbf{E}[X_j^2] = a_m\sigma^2$, so the error term in the conclusion of the theorem, $O(\rho_3/\sigma^3)$, is at most $Ca_m/\sigma \leq 2Ca_m$.

Case 1: $a_m \leq \frac{1}{10C}\delta$. In this case, by the Berry-Eséen theorem we have that

$$\Pr[X_m + \cdots + X_n \in [0,\delta]] \geq \Phi([0,\delta]) - 2Ca_m \geq \delta\phi(\delta) - \delta/5 \geq .04\delta,$$

where we have used the fact that $\phi(\delta) \geq .24$ for $\delta \leq 1$. On the other hand, since $a_2, \ldots, a_{m-1}$ are all at most δ, it is easy to fix particular signs $y_i \in \{-1, 1\}$ such that $\sum_{i=2}^{m-1} a_i y_i \in [-\delta, 0]$. These signs occur with probability 2^{-m+2}, which is at least 2^{-100C^2}. Thus with probability at least $.04 \cdot 2^{-100C^2}\delta = \Omega(\delta)$ both events occur, and $|X_2 + \cdots + X_n| \leq \delta$ as desired.

Case 2: $a_m \geq \frac{1}{10C}\delta$. In this case, we apply the Berry-Eséen theorem to the interval $[-10C\delta, 10C\delta]$ and merely use the fact that $a_m \leq \delta$. We conclude that

$$\begin{aligned}\Pr[X_m + \cdots + X_n \in [-10C\delta, 10C\delta]] &\geq \Phi([-10C\delta, 10C\delta]) - 2C\delta \\ &\geq 20C\delta \cdot \phi(10C\delta) - 2C\delta \geq 20C\delta \cdot \tfrac{1}{\sqrt{2\pi}}(1 - (10C\delta)^2/2) - 2C\delta \geq 4C\delta,\end{aligned}$$

where we have used (10) in the last step to infer $1 - (10C\delta)^2/2 \geq 5/8$. Given $X_m + \cdots + X_n = t \in [-10C\delta, 10C\delta]$, it is easy to choose particular signs $y_2, \ldots, y_{m-1}$ such that $t + \sum_{i=2}^{m-1} a_i y_i \in [-\delta, \delta]$. This uses the fact that each a_i is at least $\frac{1}{10C}\delta$ and hence $\sum_{i=2}^{m-1} a_i \geq 100C^2 \frac{1}{10C}\delta \geq 10C\delta$; it also uses the fact that each a_i is at most δ. Once again, these signs occur for $x_2, \ldots, x_{m-1}$ with probability at least 2^{-100C^2}. Thus $|X_2 + \cdots + X_n| \leq \delta$ happens with probability at least $4C2^{-100C^2}\delta = \Omega(\delta)$, as desired. □

H Mod q linear equations over two variables.

The MAX-2LIN(q) problem is that of maximizing the number of satisfied equations in a given system of linear equations over $\mathbb{Z}_q$, where in each equation there are exactly two variables. The MAX-CUT problem can be viewed as a special case of the MAX-2LIN(2) problem, by taking the equation $v + u = 1$ for every edge $\{u, v\}$ in the graph. It would be interesting to apply the ideas we used in the reduction to MAX-CUT, to get conditional hardness-of-approximation results for MAX-2LIN(q), or for other 2CSP problems over non-binary domains.

In particular, it would be nice if we could show that for every positive ϵ there exists a large-enough q, such that (assuming the Unique Games conjecture) it is hard to distinguish between the case where a given MAX-2LIN(q) instance is $(1-\epsilon)$-satisfiable, and the case where it is not even ϵ-satisfiable. Note that we could use the hardness of the above gap problem as a starting point of the reduction to MAX-CUT, instead of the Unique Label Cover problem. In order to show hardness for that problem, it seems that one must use a q-ary version of the Majority is Stablest conjecture.

In this section we state a conjecture regarding the stability of q-ary valued functions (functions of the form $f : [q]^n \to [q]$), that together with the Unique Games conjecture implies that the $(\epsilon, 1-\epsilon)$-gap version of MAX-2LIN(q) is NP-hard. By proving a weaker statement, we show that the Unique Games conjecture implies that MAX-2LIN(q) is hard to approximate up to any arbitrary constant, where the gap lies between some positive $\epsilon\delta$ and δ, rather than between ϵ and $1-\epsilon$ as above.

We also state some conjectures regarding the stability of q-ary balanced functions, where by "balanced" we mean that each value $i \in [q]$ is obtained with the same probability over a random input. In addition, we present the plurality function as a candidate for the stablest q-ary function among balanced q-ary functions where the influences of every coordinate is small enough, and compute its stability.

H.1 Some notation

This section deals mostly with functions over a q-ary domain, namely functions that are defined over the set $[q]^n$, endowed with the uniform measure. Functions that map $[q]^n$ into $[q]$ are called q-ary valued, or simply q-ary.

Relaxed q-ary functions. Let $f : [q]^n \to [q]$, be a q-ary function. For every $a \in [q]$ we let $f^a : [q]^n \to \{0, 1\}$ be the indicator of the event $f = a$, namely $f^a(x) = \mathbf{1}_{\{f(x)=a\}}$. It is easy to see that the vector $(f^a)_{a\in[q]}$ completely describes f. This vector representation allows us to consider *relaxed q-ary functions*, whose values are distributions over $[q]$.

A relaxed q-ary function is a vector $(f^a)_{a\in[q]}$ of functions of the form $f^a : [q]^n \to [0, 1]$, which satisfies

$$\forall\ x \in [q]^n\ , \qquad \sum_{a\in[q]} f^a(x) = 1\ .$$

For simplicity, we will still denote $f : [q]^n \to [q]$ for relaxed q-ary functions f. If every function f^a in f obtains only $\{0, 1\}$ values, we will identify the vector with a q-ary function, and sometimes call it a *proper* q-ary function.

A relaxed q-ary function f is called *balanced*, if

$$\forall\ a \in [q]\ , \qquad \mathop{\mathbf{E}}_{x\in[q]^n}[f^a(x)] = 1/q\ .$$

Noise correlation. For a point $x \in [q]^n$, we define an *η-correlated* x to be the random variable y , obtained by setting each coordinate y_i independently to be x_i with probability η, and a uniformly random element in $[q]$ otherwise. We denote the distribution of y by $N_\eta(x)$.

The *noise correlation* of a real valued function $f : [q]^n \to \mathbb{R}$ for noise rate η, is defined by

$$\mathbb{S}_\eta(f) = \mathbf{E}[f(x)f(y)]\ ,$$

where x is uniformly distributed in $[q]^n$, and $y \sim N_\eta(x)$ is an η-correlated x. For a relaxed q-ary function f, we define the noise correlation of f η by

$$\mathbb{S}_\eta(f) = \sum_{a \in [q]} \mathbb{S}_\eta(f^a) \ .$$

Note that for the case of a proper q-ary function f, its noise correlation is given by

$$\mathbb{S}_\eta(f) = \Pr_{\substack{x \in [q]^n \\ y \sim N_\eta(x)}} [f(x) = f(y)] \ .$$

For example, the η-noise stability of the dictator q-ary function $f(x) = x_i$ is $\eta + (1 - \eta)(1 - 1/q)$.

Fourier expansion. One may encode the elements of $[q]$ as some finite abelian group and write Fourier expansions of real-valued functions in terms of that basis. We prefer a more abstract approach. For $x \in [q]^n$ we write x_S for $\{x_i : i \in S\}$. The space of functions $X = \{f : [q]^n \to \mathbb{R}\}$ is an orthogonal sum of spaces $X = \oplus_{S \subset [n]} X_S$ where X_S denote the space of all functions $f : [q]^n \to \mathbb{R}$ such that

- $f(x)$ depends only on x_S for all x,
- f is orthogonal to all functions in the spaces $X_{S'}$ for $S' \subsetneq S$.

We will write $f : [q]^n \to \mathbb{R}$ as

$$f(x) = \sum_{S \subset [n]} f_S(x) \tag{11}$$

(we will sometime replace $f_{\{i\}}$ by f_i) where $f_S(x)$ is the projection of f to the space X_S. We will refer to(11) as the *Fourier expansion* of f.

Inner products. We use the natural inner-product for real valued functions over q-ary domains, namely $\langle f, g \rangle = \mathbf{E}_x[f(x)g(x)]$. Note that the second property above easily implies that for every two functions $f, g : [q]^n \to \mathbb{R}$ and every two *different* subsets $S, T \subseteq [n]$, f_S and g_T are orthogonal, namely $\langle f_S, g_T \rangle = 0$. The following analogues of Plancharel's and Parsavel's identity follow:

$$\langle f, g \rangle = \sum_{S \subset [n]} \langle f_S, g_S \rangle \ , \quad \text{and} \qquad \|f\|_2^2 = \sum_{S \subset [n]} \|f_S\|_2^2 \ .$$

It is easy to verify the following claim.

Claim 21. *Let f be a relaxed q-ary function over $[q]^n$. Then*

$$\sum_{a\in[q]\,,\,S\subseteq[n]} \|f_S^a\|_2^2 \le 1\ , \tag{12}$$

and equality holds if and only if f is proper.

The Fourier expansion leads to a convenient formula for the noise stability of a real-valued function. It is easy to see that if $f_S \in X_S$ then $\mathbb{S}_\eta(f_S) = (1-\eta)^{|S|}\|f_S\|_2^2$. This implies that

Claim 22. *For every real-valued function $f : [q]^n \to \mathbb{R}$ and every $\eta \in [0,1]$,*

$$\mathbb{S}_\eta(f) = \sum_S (1-\eta)^{|S|}\|f_S\|_2^2 \tag{13}$$

Influences. Using the Fourier expansion of f, we can define the influence of the i'th variable on a real-valued function f by $\mathrm{Inf}_i(f) = \sum_{S\ni i}\|f_S\|_2^2$. We also define the low degree and high degree influence of i by

$$\mathrm{Inf}_i^{\le k}(f) = \sum_{\substack{S:\ i\in S\\ |S|\le k}} \|f_S\|_2^2\ , \quad \text{and} \qquad \mathrm{Inf}_i^{>k}(f) = \sum_{\substack{S:\ i\in S\\ |S|>k}} \|f_S\|_2^2\ .$$

For a relaxed q-ary function f, we define the influence of i by $\mathrm{Inf}_i(f) = \sum_{a\in[q]} \mathrm{Inf}_i(f^a)$, and similarly for low degree and high degree influence.

The Beckner operator. We define the analogue of the Beckner operator for functions over q-ary domains. For a real valued $f : [q]^n \to [q]$ and a parameter $\rho \in [0,1]$, let

$$T_\rho(f)(x) = \mathbf{E}_{y\sim N_\rho(x)}[f(y)]\ .$$

For relaxed q-ary functions f, we define $T_\rho(f)$ by letting $T_\rho(f)^a = T_\rho(f^a)$. Note that $T_\rho(f)$ is indeed a relaxed q-ary function, and that if f is balanced then so is $T_\rho(f)$. It is also easy to observe, using the basic properties of the Fourier expansion, that for real-valued functions f, $T_\rho(f) = \sum_S \rho^{|S|} f_S$. This implies that for real-valued functions $\mathbb{S}_\rho(f) = \sum_S \rho^{|S|}\|f_S\|_2^2$. It is thus natural to define for a real valued f,

$$T_\rho(f) = \textstyle\sum_S \rho^{|S|} f_S, \quad \mathbb{S}_\rho(f) = \sum_S \rho^{|S|}\|f_S\|_2^2, \tag{14}$$

for all $\rho \in [-1,1]$.

H.2 Stability of balanced q-ary functions

We are interested in the noise stability of *balanced* q-ary functions. We conjecture that the noise stability of such functions, where the influence of each coordinate is very small, tends to zero as q tends to infinity.

Conjecture 23. *Let η, $0 < \eta < 1$, be some fixed parameter. Then there exist positive functions $\delta_\eta, S_\eta : \mathbb{N} \to \mathbb{R}$ such that $\lim_{q\to\infty} S_\eta[q] = 0$ and such that for every balanced relaxed q-ary function $f : [q]^n \to [q]$, all of whose influences are bounded by $\delta_\eta(q)$, the η-noise correlation of f is at most $S_\eta(q)$.*

To prove Conjecture 23 it suffices to show that if all influences of a $[0, 1]$-valued function f as above are small enough, then $\mathbb{S}_\eta(f) = o_{q\to\infty}(1/q)$. The following conjecture therefore implies Conjecture 23.

Conjecture 24. *Let ρ, $-1 < \rho < 1$ be some fixed parameter. Then there exist positive functions $\delta_\rho, C_\rho : \mathbb{N} \to \mathbb{R}$ such that $\lim_{q\to\infty} C_\rho[q] = 0$ and such that the following holds. For every function $f : [q]^n \to [0, 1]$ with $\mathbf{E}[f] = 1/q$, all of whose influences are smaller than $\delta_\rho(q)$,*

$$\sum_{S\neq\emptyset} \rho^{|S|} \|f_S\|_2^2 \leq C_\rho(q)/q$$

Conjecture 23 and the Unique Games conjecture together, imply that the $(\epsilon, 1-\epsilon)$-gap version of MAX-2LIN(q) is NP-hard.

Theorem 25. *The Unique Games conjecture and Conjecture 23 together imply the following. Let $\epsilon > 0$ be any fixed parameter. Then there exists a large enough q, such that given an instance of MAX-2LIN(q) is it NP-hard to distinguish between the case where it is ϵ-satisfiable and the case where it is $(1 - \epsilon)$-satisfiable.*

Theorem 25 is proven in Subsection H.4.

We cannot prove Conjecture 24, but we can show the following "first-level" version of it. This first-level bound suffices to prove Theorem 27, which shows arbitrary hardness for MAX-2LIN(q) assuming only the Unique Games conjecture.

Theorem 26. *There exists a constant K and a positive function $\delta : \mathbb{N} \to \mathbb{R}$, such that for all functions $f : [q]^n \to [0, 1]$ with $\mathbf{E}[f] \leq 1/q$ and which satisfy $\|f_i\|_2^2 \leq \delta(q)$ for all i, it holds that*

$$\sum_i \|f_i\|_2^2 \leq \frac{K \log q}{q^2}.$$

The proof of Theorem 26 appears in the next subsection. The following theorem, showing that for large enough q MAX-2LIN(q) is hard to approximate to within any constant factor, follows from Theorem 26.

Theorem 27. *Assuming the Unique Games conjecture, the following holds. For every $\epsilon > 0$ there exists a positive δ and an integer q, such that given an instance of MAX-2LIN(q), it is NP-hard to distinguish between the case where the instance is δ-satisfiable, and the case where it is $\epsilon\delta$-satisfiable.*

Theorem 27 is proven in Subsection H.4.

H.3 Proof of Theorem 26

The proof of Theorem 26 follows that of Proposition 2.2 in [45]. We will need the following sub-Gaussian estimate.

Lemma 28. *Let $\epsilon > 0$. Then there exists a $\delta = \delta(q, \epsilon)$ such that if $Z_1, \ldots, Z_n : [q] \to \mathbb{C}$ where $[q]$ is equipped with the uniform measure and*

- $\mathbf{E}[Z_i] = 0$ *for all i,*
- $\mathbf{E}[|Z_i|^2] = \sigma_i^2$ *where $\sum_{i=1}^n \sigma_i^2 \le \sigma^2$ and $\sigma_i^2 \le \sigma^2\delta^2$ for all i.*

Then for all $t \ge 2$:

$$\Pr[|\sum_{i=1}^{n} Z_i| \ge \sigma t] \le 4\left(\exp(-\frac{t^2}{8}) + \frac{\epsilon}{1+t^3}\right).$$

Proof. Clearly it suffices to prove that if Z_i are real-valued random variables, satisfying the conditions above, $\epsilon > 0$ and $t \ge 1$ then

$$\Pr[|\sum_{i=1}^{n} Z_i| \ge \sigma t] \le 2\left(\exp(-\frac{t^2}{2}) + \frac{\epsilon/8}{1+t^3}\right).$$

We apply the Berry-Eséen Theorem 19. We first note that since the variables Z_i obtain at most q values, it follows that $\|Z_i\|_\infty^2 \le q\|Z_i\|_2^2$ or $\|Z_i\|_\infty \le \sqrt{q}\|Z_i\|_2 \le \sqrt{q}\sigma\delta$.

It now follows that $\sum_i \mathbf{E}[|Z_i|^3] \le \sqrt{q}\sigma\delta \sum \mathbf{E}[|Z_i|^2] \le \sqrt{q}\sigma^3\delta$. Thus Theorem 19 implies that there exists a constant C such that for $t \ge 1$

$$\Pr[|\sum_{i=1}^{n} Z_i| \ge \sigma t] \le \frac{C\sqrt{q}\delta}{1+t^3} + 2\Pr[N(0,1) > t] \le 2\left(\exp(-\frac{t^2}{2}) + \frac{\epsilon/8}{1+t^3}\right).$$

provided that $t \ge 1$ and $\delta = \delta(q, \epsilon)$ is sufficiency small. □

We now prove Theorem 26.

Proof. Let h denote the linear part of f, $h(x) = \sum_{i=1}^n f_i(x_i)$. We may assume that $\sigma^2 := \sum_{i=1}^n \|f_i\|_2^2 \geq \frac{1}{q^2}$.

Let $t_0 = \sqrt{8 \log q}$ and $\epsilon = 1/(100q^4)$. Let δ be chosen to satisfy Lemma 28. Let $\delta' = \delta/q^2$. If $\|f_i\|_2^2 \leq \delta'$ for all i then by Lemma 28 we obtain that

$$\begin{aligned}
\sigma^2 = \sum_i \|f_i\|_2^2 &= \int h(x)f(x)dx \leq \int \Big(\sigma t_0 f(x) + (h(x) - \sigma t_0)\mathbf{1}_{\{h(x) > \sigma t_0\}}\Big)dx \\
&\leq \frac{\sigma t_0}{q} + \int_{t=0}^{\infty} \mathrm{Pr}_x[(h(x) - \sigma t_0)\mathbf{1}_{\{h(x)>\sigma t_0\}} > t]dt \\
&= \frac{\sigma t_0}{q} + \int_{t=\sigma t_0}^{\infty} \mathrm{Pr}_x[(h(x) > t]dt \\
&\leq \frac{\sigma t_0}{q} + \sigma \int_{t=t_0}^{\infty} \mathrm{Pr}_x[|h(x)| > \sigma t])dt \\
&\leq \frac{\sigma t_0}{q} + \sigma \int_{t=t_0}^{\infty} 4e^{-t^2/8}dt + \sigma\epsilon \int_{t=t_0}^{\infty} \frac{1}{1+t^3}dt \\
&\leq \frac{\sigma t_0}{q} + 4\sigma e^{-t_0^2/8} + \sigma\epsilon \leq K\sigma\frac{\sqrt{\log q}}{q}
\end{aligned}$$

for some constant K. We thus conclude that $\sigma^2 \leq K^2 \frac{\log q}{q^2}$ as needed. □

H.4 Hardness for MAX-2LIN(q)

In this subsection we prove Theorem 27 and Theorem 25. But first we need some technical facts.

Claim 29. *Let $f : [q]^n \to [q]$ be a generalized q-ary function. Then $\sum_i \mathrm{Inf}_i^{\leq k}(f) \leq k$.*

Proof. By definition,

$$\sum_i \mathrm{Inf}_i^{\leq k}(f) = \sum_a \sum_i \mathrm{Inf}_i^{\leq k}(f^a) \leq k \sum_a \sum_S \|f_S^a\|_2^2 \leq k ,$$

where the final inequality follows from (12). □

Claim 30. *Let $f : [q]^n \to [q]$ be a relaxed q-ary function, and let $\rho \in [0, 1]$ be some parameter. Then for every $i \in [n]$,*

$$\mathrm{Inf}_i^{>k}(T_\rho(f)) \leq \rho^{2k}$$

Proof. By definition,

$$\begin{aligned}\mathrm{Inf}_i^{>k}(T_\rho(f)) &= \sum_{a\in[q]} \mathrm{Inf}_i^{>k}(T_\rho(f^a)) = \sum_{a\in[q]} \sum_{\substack{S\ :\ i\in S\\ |S|>k}} \|T_\rho(f^a)_S\|_2^2 \\ &= \sum_{a\in[q]} \sum_{\substack{s\ :\ i\in S\\ |S|>k}} \rho^{2|S|}\|f_S^a\|_2^2 \le \rho^{2k} \sum_{a\in[q]} \sum_{\substack{s\ :\ i\in S\\ |S|>k}} \|f_S^a\|_2^2 \le \rho^{2k} ,\end{aligned}$$

where the last inequality follows from (12). □

The above claim yields the following version of Conjecture 23.

Corollary 31. *Let η, $0 < \eta < 1$, be some fixed parameter. Then assuming Conjecture 23 holds implies that there exist positive functions $\delta'_\eta, S'_\eta : \mathbb{N} \to \mathbb{R}$, and an integer valued function $k = k_\eta : \mathbb{N} \to \mathbb{N}$, such that $\lim_{q\to\infty} S'_\eta[q] = 0$ and such that the following holds.*

For every balanced relaxed q-ary function $f : [q]^n \to [q]$, all of whose low-degree influences $\mathrm{Inf}_i^{\le k} f$ are bounded by $\delta'_\eta(q)$, the η-noise correlation of f is at most $S'_\eta(q)$.

Proof. Let δ_η and S_η be as in Conjecture 23, and let $\delta'_\eta = \delta_\eta/2$ and $S'_\eta = 2S_\eta$. Then obviously $\lim_{q\to\infty} S'_\eta[q] = 0$.

For every q, let $\rho = \rho_\eta(q)$ be selected so that $\rho < 1$, and yet

$$\max_{s\in\mathbb{N}}\{(1-\eta)^s(1-\rho^{2s})\} < S_\eta(q) .$$

We let $k = k_\eta(q)$ be selected so that $\sum_{s>k} \rho^{2k} < \delta_\eta(q)/2$.

Now let f be a balanced relaxed q-ary function where for every i, $\mathrm{Inf}_i^{\le k}(f) \le \delta'_\eta(q)$. Let $g = T_\rho(f)$ (g is also balanced). Using Claim 22, we have

$$\begin{aligned}\mathbb{S}_\eta(f) - \mathbb{S}_\eta(g) &= \sum_{a\in[q]} \sum_{S} (1-\eta)^{|S|}(1-\rho^{2|S|})\|f_S^a\|_2^2 \\ &\le \max_{s\in\mathbb{N}}\{(1-\eta)^s(1-\rho^{2s})\} < S_\eta(q) ,\end{aligned}$$

where the first inequality follows from (12). We thus have

$$\mathbb{S}_\eta(f) \le \mathbb{S}_\eta(g) + S_\eta(q) \tag{15}$$

Note that from Claim 30 and (12), it follows that for every i,

$$\begin{aligned}\mathrm{Inf}_i(g) &= \mathrm{Inf}_i^{\le k}(g) + \mathrm{Inf}_i^{>k}(g) \le \mathrm{Inf}_i^{\le k}(f) + \rho^{2k} \\ &\le \delta'_\eta(q) + \delta_\eta(q)/2 \le \delta_\eta(q) ,\end{aligned}$$

where the second inequality follows from the choice of ρ and k. Since g is also balanced, it follows that Conjecture 23 applies to it, and therefore

$$\mathbb{S}_\eta(g) \le S_\eta(q) \ .$$

Together with (15) and the choice of S'_η, the corollary follows. □

We are now ready to prove Theorem 25.

Proof of Theorem 25. Let ϵ' be a small constant to be determined later. Let V be an instance of Unique Label Cover where it is hard to distinguish between ϵ' and $(1-\epsilon')$ satisfiability. Let M be the number of labels required (M is constant that depends only on ϵ').

We would like to transform our instance to an instance J of MAX-2LIN(q) (where q is constant), and maintain an $(\epsilon, 1-\epsilon)$-gap. Our construction requires a q-ary version of the folded long-code. Let us start by defining what this means.

Definition 11 (folded functions). *Let $f : [q]^M \to [q]$ be some relaxed q-ary function, and assume that q is prime. It is said that f is folded if*

- *For every $a, b \in \mathbb{Z}_q^*$ and every $x \in [q]^n$,* $\qquad f^a(x) = f^{ab}(bx)$.
- *For every $a, b \in \mathbb{Z}_q^*$ and every $x \in [q]^n$,* $\qquad f^a(x) = f^{(a+b)}(x + (b, b, \ldots, b))$.

Definition 12 (q-ary long-code). *Let q be a prime. The q-ary long-code of an element $i \in [M]$ is the q-ary function $f : [q]^M \to [q]$ defined by $f(x) = x_i$.*

Folded long-codes. We construct a system J of linear equations over the "supposed q-ary long-codes" of the assignments of all the nodes in the given instance V. We may not assume that we get correct long-code words, but by a standard PCP trick we may assume that all words are folded (this preserves the linearity of the equations). Note that a folded relaxed q-ary function must also be balanced!

Constructing J. Fix $\eta \ll \epsilon/2$ to be some positive constant, and let q be a prime large enough so that $S'_\eta(q) < \epsilon/8$. Instead of a system of linear equations, we generate a distribution of linear equations (which is roughly the same thing), by the following process:

- Pick a vertex $v \in V$ at random and two of its neighbors $w, w' \in W$ at random. Let $\sigma = \sigma_{v,w}$ and $\sigma' = \sigma_{v,w'}$ be the respective bijections for edges (v, w) and (v, w').

- Let f and g be the supposed q-ary long codes of labels of w and w' respectively.
- Pick $x \in [q]^M$ at random.
- Pick $y \sim N_\eta(x)$.
- The selected equation is $\qquad f(x \circ \sigma) = g(y \circ \sigma')$.

Completeness. It is easy to see that the completeness of the verifier is $(1-2\epsilon')(\eta + (1-\eta)(1-1/q))$ (this is very similar to Subsection E.1). This is more than the needed $(1-\epsilon)$, if ϵ' is selected to be small enough.

Soundness. Suppose that there exists an assignment which satisfies at least an ϵ fraction of the equations. We will show how this leads to an assignment for V satisfying some constant fraction of its constraints. This constant will be independent of M, and therefore by increasing M we will be able to get ϵ' to be smaller than that constant. This will suffice to prove soundness.

We follow a path analogous to that of Subsection E.3, of arithmetizing the probability of reaching a satisfied equation via the random selection process:

$$\epsilon \leq \Pr[\mathrm{acc}] = \mathop{\mathbf{E}}_{v,w,w'}\Big[\mathop{\mathbf{E}}_{x,y}\big[\sum_{a\in[q]} f_w^a(x) = f_{w'}^a(y)\big]\Big] = \mathop{\mathbf{E}}_{v}\left[\sum_{a\in[q]} \mathop{\mathbf{E}}_{x,y}\Big[\mathop{\mathbf{E}}_{w,w'}[f_w^a(x) f_{w'}^a(y)]\Big]\right]$$

$$= \mathop{\mathbf{E}}_{v}\left[\sum_{a\in[q]} \mathop{\mathbf{E}}_{x,y}\Big[\mathop{\mathbf{E}}_{w}[f_w^a(x)]\mathop{\mathbf{E}}_{w'}[f_{w'}^a(y)]\Big]\right] \qquad (w \text{ and } w' \text{ are independent given } v)$$

$$= \mathop{\mathbf{E}}_{v}\left[\sum_{a\in[q]} \mathop{\mathbf{E}}_{x,y}\Big[g_v^a(x) g_v^a(y)\Big]\right] \qquad (\text{where } g_v = \mathbf{E}_w[f_w])$$

$$= \mathop{\mathbf{E}}_{v}\Big[\mathbb{S}_\eta(g_v)\Big]$$

From the above inequality we have that for at least an $\epsilon/4$ fraction of $v \in V$, $\mathbb{S}_\eta(g_v) \geq \epsilon/4$. Now since the functions f_w are balanced, g_v is also balanced, and we may apply Conjecture 23 to it, to conclude that g_v has at least one coordinate j with $\mathrm{Inf}_j^{\leq k}(g_v) \geq \delta'_\eta(q)$. It follows, as in (4), that

$$\delta'_\eta(q) \leq \sum_a \mathrm{Inf}_j^{\leq k}(g_v) \leq \mathop{\mathbf{E}}_{w}\left[\sum_a \mathrm{Inf}_{\sigma^{-1}(j)}^{\leq k}(f_w^a)\right] = \mathop{\mathbf{E}}_{w}\Big[\mathrm{Inf}_{\sigma^{-1}(j)}^{\leq k}(f_w)\Big]$$

From here, the proof of soundness follows exactly as in Subsection E.3 after Equation 4. □

The proof of Theorem 27 is very similar to that of Theorem 27, and we will not repeat it completely. Instead, we will just point out the differences between the two proofs.

Proof of Theorem 27. First, let us note the following technical fact which follows from Theorem 26, that we will use instead of Conjecture 23: Let f be a balanced relaxed q-ary function, which satisfies $\|f_i^a\|_2^2 \leq \delta(q)$ for every i and every $a \in [q]$, where δ is as in Theorem 26. Then

$$\mathbb{S}_\eta(f) \leq \frac{K\eta \log q}{q} + \eta^2$$

. This follows immediately from (13) and (12). Also recall that the η-noise stability of the dictatorship function is $\eta + (1-\eta)(1-1/q)$.

Now our construction of the equation system will be the same as in the proof of Theorem 25, but with different parameters: We will pick η and q so that

$$\frac{\frac{K\eta \log q}{q} + \eta^2}{\eta + (1-\eta)(1-1/q)} \leq \delta$$

The rest of the proof is almost identical. □

H.5 Plurality

While we cannot prove Conjecture 23, it is still interesting to speculate about the optimal parameters for which it is true. The plurality function is a possible candidate for being an extremal example with respect to this conjecture. The Plurality function $h : [q]^n \to [q]$ is a generalization of Majority to non-binary domains: $h(x) = a$ if a maximizes $|\{i : x_i = a\}|$ (when there is more than one maximizer, choose arbitrarily). In Theorem 10 below we show that the stability of Plurality is $q^{-(1-\rho)/(1+\rho)+o(1)}$.

We now compute the noise stability of the Plurality function. For $i \in [q]$ we let u_i denote the number of i labelled coordinate in x and let v_i denote the number of i labelled coordinates of y.

Letting $p = 1/q$, u_i and v_i are both $Bin(n, p)$. It is easy to see that $\mathbf{E}[u_i] = \mathbf{E}[v_i] = pn$, $\mathrm{Var}[u_i] = \mathrm{Var}[v_i] = np(1-p)$ and $\mathrm{Cov}[v_i, v_j] = \mathrm{Cov}[u_i, u_j] = -np^2$. Moreover, $\mathrm{Cov}[v_i, u_i] = n\rho p(1-p)$ and $\mathrm{Cov}[v_i, u_j] = -n\rho p^2$.

Let $\tilde{u}_i$ be u_i normalized to have expected value 0 and variance 1. Define $\tilde{v}_i$ similarly. Then $\mathrm{Cov}[\tilde{v}_i, \tilde{v}_j] = -np^2/np(1-p) = -p/(1-p)$, $\mathrm{Cov}[\tilde{v}_i, \tilde{u}_i] = \rho$ and $\mathrm{Cov}[\tilde{v}_i, \tilde{u}_j] = -\rho p/(1-p)$.

Let $U_1, \ldots, U_n$ be i.i.d. standard normals. Let $V_1, \ldots, V_n$ be i.i.d standard normals. Consider the variables

$$W_i = \sqrt{\tfrac{q}{q-1}} \left(U_i - \tfrac{1}{q} \textstyle\sum_{j=1}^q U_j \right), \quad X_i = \sqrt{\tfrac{q}{q-1}} \left(V_i - \tfrac{1}{q} \textstyle\sum_{j=1}^q V_j \right).$$

It is easy to see that $Var[X_i] = Var[W_i] = 1$ and that $\mathrm{Cov}[X_i, X_j] = \mathrm{Cov}[W_i, W_j] = -\frac{p}{1-p}$. Let's assume that $\mathbf{E}[U_i V_j] = \rho \delta_{i,j}$. Then repeating the calculations above, we see that $\mathrm{Cov}[W_i, X_i] = \rho$ and $\mathrm{Cov}[W_i, X_j] = -\rho p/(1-p)$.

It now follows that the probability that color i is chosen as the plural in both u and v is asymptotically given by

$$\Pr[X_i = \max_j X_j, W_i = \max_j W_j] = \Pr[U_i = \max_j U_j, V_i = \max_j V_j].$$

The probability that the plural is the same in x and y is therefore given by $q\Pr[U_1 = \max_j U_j, V_1 = \max_j V_j]$. Finally, we will introduce independent normal Gaussians $Y_1, \ldots, Y_p$ and $Z_1, \ldots, Z_p$ and define

$$U_i = \frac{1}{\sqrt{1+\eta^2}} \left(Y_i + \eta Z_i \right), V_i = \frac{1}{\sqrt{1+\eta^2}} \left(Y_i - \eta Z_i \right), \tag{16}$$

where $(1-\eta^2)/(1+\eta^2) = \rho$. The event $\{U_1 = \max_j U_j\}$ is a.s. the same as the event $\{\forall i : Y_1 + \eta Z_1 \geq Y_i + \eta Z_i\}$. Similarly the event $\{V_1 = \max_j V_j\}$ is a.s. the same as the event $\{\forall i : Y_1 - \eta Z_1 \geq Y_i - \eta Z_i\}$.

Let's condition on the value a of Y_1 and the value b of Z_1. We then get the equations $Y_i + \eta Z_i \leq a + \eta b$ and $Y_i - \eta Z_i \leq a - \eta b$ for all $i > 1$. These equations are equivalent to

$$Y_i \leq a, \quad \tfrac{Y_i - a}{\eta} + b \leq Z_i \leq \tfrac{a - Y_i}{\eta} + b. \tag{17}$$

There is no explicit formula for the probability of (17) for general a and b (see [28, Chapter 6]).

Instead, we will write a closed form expression using the Gaussian distribution function Φ and density function ϕ. Clearly the probability of the expression written above is given by:

$$\Psi(a,b) = \int_{-\infty}^{a} \phi(y) \left(\Phi(\frac{a-y}{\eta} + b) - \Phi(\frac{y-a}{\eta} + b) \right) dy.$$

We thus get that the probability that 1 is the plural in both x and y is given by

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Psi^{q-1}(a,b)\phi(a)\phi(b)dadb.$$

The stability of plurality is now easily computed to be:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}q\Psi^{q-1}(a,b)\phi(a)\phi(b)dadb.$$

H.6 Asymptotic behavior of plurality

In this section we establish the asymptotic behavior as $q \to \infty$ of Plurality, proving Theorem 10.

We wish to estimate the probability that $U_1 = \max_j U_j$ and that $V_1 = \max_j V_j$. Write $\sigma = \sqrt{1+\eta^2}$. Write $\tilde{U}_i$ for σU_i and $\tilde{V}_i$ for σV_i. Write Y for Y_1 and Z for Z_1. Clearly,

$$\begin{aligned}\Pr[U_1 = \max_j U_j,\, V_1 = \max_j V_j] = \; & \tfrac{1}{2}\Pr[U_1 = \max_j U_j,\, V_1 = \max_j V_j \mid Z \le 0] \\ + \; & \tfrac{1}{2}\Pr[U_1 = \max_j U_j,\, V_1 = \max_j V_j \mid Z \ge 0] \\ \le \; & \Pr[\forall 2 \le i \le q : \tilde{U}_i \le Y], \qquad (18)\end{aligned}$$

where U, V, Y and Z are defined as in (16). We are to bound the probability (18).

We will use the well known estimate that for $a \ge 2\sigma$,

$$\tfrac{1}{2a}\phi(a) \le \Pr[Y \ge a] \le \tfrac{1}{a}\phi(a), \quad \tfrac{\sigma}{2a}\phi(\tfrac{a}{\sigma}) \le \Pr[\tilde{U}_i \ge a] \le \tfrac{\sigma}{a}\phi(\tfrac{a}{\sigma}).$$

Taking $a = \sqrt{2\sigma^2 g(q)\log q}$, we thus obtain $\Pr[Y \ge a] \le \phi(a) \le q^{-\sigma^2 g(q)}$ and

$$\Pr[\forall i : \tilde{U}_i \le a] = \Big(1 - \Pr[\tilde{U}_1 \ge a]\Big)^{q-1} \le \Big(1 - \frac{\sigma}{2a}\phi(\frac{a}{\sigma})\Big)^{q-1} \le \Big(1 - \frac{\sigma}{2a}q^{-g(q)}\Big)^{q-1}.$$

The expression on the right hand size is smaller than $q^{-\sigma^2}$ when $g(q) \le 1 - 2\log(a\sigma)/\log q - \log\log q/\log q$. Note that the probability in (18) is bounded by the sum of the probability that $Y \ge a$ and the probability that $\tilde{U}_i \le a$ for all i. It follows that the probability that 1 is the plurality both in x and y is bounded by $q^{-\sigma^2} + q^{-\sigma^2 g(q)} = q^{-\sigma^2+o(1)}$. Therefore, the probability that the plurality in x and y agree is bounded above by $q^{-\sigma^2+1+o(1)} = q^{-\eta^2+o(1)} = q^{-(1-\rho)/(1+\rho)+o(1)}$. This proves the upper bound.

To prove the lower bound, note that either by a direct calculation or by the FKG inequality it follows that

$$P[\tilde{U}_i \le a,\, \tilde{V}_i \le a] \ge (1 - P[\tilde{U}_i \ge a])^2.$$

Therefore by independence,

$$\Pr[\forall i:\ \tilde{U}_i \le a, \tilde{V}_i \le a] \ge \Big(1 - \Pr[\tilde{U}_1 \ge a]\Big)^{2q-2} \ge \Big(1 - \frac{\sigma}{a}\phi(\frac{a}{\sigma})\Big)^{2q-2} \ge \big(1 - q^{-g(q)}\big)^{2q-2}.$$

Thus taking $g(q) = 1$, the probability that $\tilde{U}_i, \tilde{V}_i$ for $2 \le i \le q$ are all smaller than $a = \sqrt{2\sigma^2 \log q}$ is $\Omega(1)$.

On the other hand it is easy to see again that $\Pr[Y \ge a+1] \ge q^{-\sigma^2 - o(1)}$ and since $\Pr[|Z| \le 1] = \Omega(1)$ it follows that

$$\Pr[\min\{Y+Z, Y-Z\} \ge \max_{2 \le i \le q}\{\tilde{U}_i, \tilde{V}_i\}] \ge q^{-\sigma^2 - o(1)}.$$

The probability that the plurality in x and y agree is therefore lower bounded by $q^{-\sigma^2+1-o(1)} = q^{-(1-\rho)/(1+\rho)+o(1)}$. This proves the lower bound.