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

Subhash Khot*
College of Computing
Georgia Tech
khot@cc.gatech.edu

Guy Kindler†
Theory Group
Microsoft Research
gkindler@microsoft.com

Elchanan Mossel‡
Department of Statistics
U.C. Berkeley
mossel@stat.berkeley.edu

Ryan O'Donnell*
Theory Group
Microsoft Research
odonnell@microsoft.com

September 25, 2005

Abstract

In this paper we show a reduction from the Unique Games problem to the problem of approximating MAX-CUT to within a factor of $\alpha_{\text{GW}} + \epsilon$, for all $\epsilon > 0$; here $\alpha_{\text{GW}} \approx .878567$ denotes the approximation ratio achieved by the Goemans-Williamson algorithm [25]. This implies that if the Unique Games Conjecture of Khot [36] holds then the Goemans-Williamson approximation algorithm is optimal. Our result indicates that the geometric nature of the Goemans-Williamson algorithm might be intrinsic to the MAX-CUT problem.

Our reduction relies on a theorem we call Majority Is Stablest. This was introduced as a conjecture in the original version of this paper, and was subsequently confirmed in [42]. A stronger version of this conjecture called Plurality Is Stablest is still open, although [42] contains a proof of an asymptotic version of it.

Our techniques extend to several other two-variable constraint satisfaction problems. In particular, subject to the Unique Games Conjecture, we show tight or nearly tight hardness results for MAX-2SAT, MAX-q-CUT, and MAX-2LIN(q).

For MAX-2SAT we show approximation hardness up to a factor of roughly $.943$. This nearly matches the $.940$ approximation algorithm of Lewin, Livnat, and Zwick [40]. Furthermore, we show that our $.943...$ factor is actually tight for a slightly restricted version of MAX-2SAT. For MAX-q-CUT we show a hardness factor which asymptotically (for large q) matches the approximation factor achieved by Frieze and Jerrum [24], namely $1 - 1/q + 2(\ln q)/q^2$.

For MAX-2LIN(q) we show hardness of distinguishing between instances which are $(1-\epsilon)$-satisfiable and those which are not even, roughly, $(q^{-\epsilon/2})$-satisfiable. These parameters almost match those achieved by the recent algorithm of Charikar, Makarychev, and Makarychev [10]. The hardness result holds even for instances in which all equations are of the form $x_i - x_j = c$. At a more qualitative level, this result also implies that $1 - \epsilon$ vs. ϵ hardness for MAX-2LIN(q) is *equivalent* to the Unique Games Conjecture.

*Work performed while the author was at the Institute for Advanced Study. 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.

†Work performed while the author was at DIMACS and at the Institute for Advanced Study, Princeton, and was partly supported by CCR grant $\mathcal{N}$CCR-0324906 and $\mathcal{N}$DMS-0111298.

‡Supported by a Miller fellowship in Computer Science and Statistics, U.C. Berkeley

1 Introduction

The main result in this paper is a bound on the approximability of the MAX-CUT problem which matches the approximation ratio achieved by the well-known Goemans-Williamson algorithm [25]. The proof of this hardness result relies on the *Unique Games Conjecture* of Khot [36]. We also rely critically on a theorem we call *Majority Is Stablest*, which was introduced as a conjecture in the original version of this paper. For the convenience of the reader we will now briefly describe these two tools; formal statements appear in Sections 3 and 4.

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 Theorem (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 theorem 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). We also note that the Majority Is Stablest theorem has interesting applications outside of this work — to the economic theory of social choice [33] for example — and may well prove useful for other PCP-based inapproximability results. In Section 6.3 we mention interesting generalizations of the Majority Is Stablest theorem for q-ary functions, $q > 2$, which are relevant for hardness of approximation and are not resolved in full.

Despite the fact that our hardness result for MAX-CUT relies on the unproven Unique Games Conjecture, we feel it is interesting for several reasons. First, in our opinion it is remarkable that the Unique Games Conjecture should yield a *tight* 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.

Another reason we believe our result is interesting 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 [31] from two years later there has been no progress on the hardness of approximating MAX-CUT, except for the creation of a better reduction gadget [51]. As one of the most natural and simple problems to have resisted matching approximability bounds, we feel MAX-CUT 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 of other 2-variable CSPs such as MAX-2SAT; see Section 9.

Finally, instead of viewing our result as relying on the unproven Unique Games Conjecture, we can view it as being an investigation into the truth of UGC. Indeed our hardness results for both MAX-CUT and for two-variable linear equations modulo q provide explicit parameters for which the Unique Games Conjecture, if true, must hold. (Note that both problems are Unique Games themselves.) Thus our work gives a target for algorithmic attacks on the Unique Games Conjecture, which if passed will refute it.

Indeed, works subsequent to the original version of this paper have provided approximation algorithms for the Unique Games problem [50, 28, 10] improving on Khot's original algorithm [36]. In particular, in [10] Charikar, Makarychev, and Makarychev gave a semidefinite programming-based approximation algorithm for Unique Games whose approximation factor nearly matches our hardness bound for MAX-2LIN(q). The current situation is therefore that any improvement in the approximation factors for either MAX-CUT or for the more general MAX-2LIN(q) will refute the Unique Games Conjecture.

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 and discuss very recent algorithm results for the problem. The Majority Is Stablest problem is discussed in Section 4, along with its generalization to q-ary domains, $q \geq 2$. We discuss the geometric aspects of MAX-CUT and their connection with Majority Is Stablest result and the Goemans-Williamson approximation algorithm in Section 5. Our main results are stated in Section 6. Section 7 is devoted to some technical definitions, preliminaries, and Fourier analytic formulas. In Section 8 we prove our main theorem on the hardness of approximating MAX-CUT, based on the Unique Games Conjecture. In Section 9 we investigate the approximability of other binary 2-CSPs, such as MAX-2SAT. Finally, Section 11 is devoted to extending our techniques to the q-ary domain; we prove some results about noise stability in this domain and then prove our Unique Games-hardness results for MAX-q-CUT and MAX-2LIN(q) and MAX-q-CUT.

2 About 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 [35]) 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 [45] 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 [25]. 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 $\theta^* \approx 2.33 \approx 134^\circ$, 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 [44] and Bellare, Goldreich, and Sudan [2] 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 [31] via a reduction from MAX-3LIN using a gadget of Trevisan, Sorkin, Sudan, and Williamson [51]. This stands as the 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} [34, 18] and the integrality gap of their semidefinite relaxation was also proved to be α_{GW} [18], 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., [25, 17, 16, 34, 52, 15, 18]). And of course, perhaps a completely different algorithm altogether can perform better. Several papers have either explicitly ([16]) or implicitly ([18]) given the problem of improving on α_{GW} as an important research goal.

However, in this paper we show that approximating MAX-CUT to within any factor larger than α_{GW} will in fact overturn the Unique Games Conjecture.

3 About 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 the case of MAX-CUT, the vertices serve as variables and the edges as constraints. Every constraint says that two certain variables should 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 so-called 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. Usually, inapproximability results for boolean CSPs are obtained by encoding assignments to Label Cover variables 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 [31, 46, 29]). However the approach gets stuck in the case of 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 [36] 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 the '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: Given is a 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, $\gamma > 0$, there exists a constant $M = M(\eta, \gamma)$ 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 γ.*

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. Typically, though, the inner verifier will need to rely on rather deep theorems about the Fourier spectrum of boolean functions, e.g. the theorem of Bourgain [7] or of Friedgut [21].

The Unique Games Conjecture was used in [36] 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 [30] 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 [37] 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 [14] that showed 1.36 hardness for Vertex Cover. In the present paper we continue this line of research, showing an inner verifier that together with the Unique Games Conjecture yields a tight hardness result for MAX-CUT. Our inner verifier relies critically on the Majority Is Stablest theorem.

Algorithmic results for Unique Label Cover. It is natural to ask how the function $M(\eta, \gamma)$ in the Unique Games Conjecture can behave. Lower bounds on M are obtained by giving algorithms for Unique Label Cover. Several very recent results have provided such algorithms. Most relevant for this paper is the algorithm of [10], which has the following behavior for Unique Label instances with label set of size q: For any constant $\eta > 0$, on instances with optimum $1 - \eta$ it satisfies roughly a $(1/q)^{\eta/(2-3\eta)}$ fraction of edges, up to lower order powers of q. Also, for $\eta = 1/\log q$, it seems to satisfy an $\Omega(1)$ fraction of edges (at the present time the final version of [10] has not yet appeared).

4 About the Majority Is Stablest problem

To state the Majority Is Stablest problem, we need some definitions. For convenience we regard the boolean values as -1 and 1 rather than 0 and 1. Thus a boolean function is a map $f \colon \{-1, 1\}^n \to \{-1, 1\}$. We will often generalize to the case of functions $f \colon \{-1, 1\}^n \to \mathbb{R}$. In all of what follows we consider the set of strings $\{-1, 1\}^n$ to be a probability space under the uniform distribution.

First we recall the well-known notion of 'influence', introduced to computer science in [3] and studied even earlier in economics.

Definition 2. *Let $f \colon \{-1, 1\}^n \to \mathbb{R}$. Then the* influence of x_i on f *is defined by*

$$\mathrm{Inf}_i(f) = \mathop{\mathbf{E}}_{(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)} \left[\mathrm{Var}_{x_i}[f]\right].$$

(Note that for $f\colon \{-1,1\}^n \to \{-1,1\}$,

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

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 (in expectation) of the bits. This leads to the study of 'noise sensitivity,' pioneered in computer science by [32, 31, 4].

Definition 3. *Let $f\colon \{-1,1\}^n \to \mathbb{R}$ and let $-1 \leq \rho \leq 1$. The* noise stability 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 stability is defined to be*

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

(Note that for $f\colon \{-1,1\}^n \to \{-1,1\}$ we have $\mathbb{S}_\rho(f) = 2\Pr_{x,y}[f(x) = f(y)] - 1$.)

We may now state the Majority Is Stablest theorem. This result was presented as a strongly-believed conjecture in the original version of this paper. It has recently been proved in [42]. Informally, the theorem says that among all balanced boolean functions with small influences, the Majority function has the highest noise stability. Note that the assumption of small influences is necessary since the 'dictator' function $f(x) = x_i$ provably has the highest noise stability among all balanced boolean functions, for every ρ. Note that when n tends to infinity, the noise stability 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 [27] and is ultimately derived from the Central Limit theorem plus a result from an 1890's paper of Sheppard [47]). Thus we have the formal statement of the theorem:

Majority Is Stablest theorem: *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 $\mathrm{Inf}_i(f) \leq \delta$ for all $i = 1 \ldots n$, then*

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

In the remainder of this section, we shall describe why the Majority Is Stablest theorem is relevant for MAX-CUT inner verifiers.

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 an 'outer verifier' (almost always a Label Cover problem) with an 'inner verifier'. As mentioned the outer verifier for our reduction is the Unique Label Cover problem. As for the inner verifier, it is always application-specific and its acceptance predicate is tailor-made for the problem at hand, in our case MAX-CUT.

A *codeword test* is an essential submodule of an inner verifier. It is a probabilistic procedure for checking whether a given string is a codeword of an error-correcting code, most commonly the 'Long Code' (see [2]).

Definition 4. *The Long Code over domain $[n]$ is a binary code in which the message space is in fact the set of truth tables of boolean functions $f : \{-1,1\}^n \to \{-1,1\}$. 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, i.e. a dictator, is the truth table of a boolean function in which one 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 (note that this is not a standard notion of being far from a codeword, but rather a notion tailored for our proof technique).

We expect the following from a codeword test: a correct Long Code word 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 will be the inapproximability factor for MAX-CUT.

The Long Code test. As mentioned before, our Long Code test will need to take a boolean function $f : \{-1,1\}^n \to \{-1,1\}$, pick two inputs x and y, and check that $f(x) \neq f(y)$. In fact our test will be precisely a 'noise stability' test for some fixed noise rate ρ; i.e., x will be chosen uniformly at random and y will be formed by flipping each bit of x independently with probability $\frac{1}{2} - \frac{1}{2}\rho$. 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 relatively high probability. Recalling Definition 3, we see that the probability a given function f passes our test is precisely $\frac{1}{2} - \frac{1}{2}\mathbb{S}_\rho(f)$.

A legal Long Code word, i.e. a dictator function, has noise stability precisely ρ and thus the completeness of the Long Code test is $c = \frac{1}{2} - \frac{1}{2}\rho$. The crucial aspect of our test is the analysis of the soundness parameter.

This is where the Majority Is Stablest theorem 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 7.4) we can show that the Majority Is Stablest theorem (which is stated only for $\rho \geq 0$) implies that for $\rho < 0$ the noise stability 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* $s = \frac{1}{2} - \frac{1}{2}(1 - \frac{2}{\pi}\arccos\rho) = (\arccos\rho)/\pi$.

Choosing $\rho < 0$ as we please, this leads to an inapproximability ratio of

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

precisely the Goemans-Williamson constant.

4.1 History of the Majority Is Stablest problem

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 (some examples, roughly in chronological order: [32, 8, 48, 22, 49, 9, 21, 4, 6, 7, 23, 33, 41, 43, 13]). Building on the intuition gathered from this past work, we were motivated to make the Majority Is Stablest conjecture in the originial version of the paper. We discuss these relevant previous results below.

The Majority and weighted majority (or balanced threshold) functions have always played an important role in the study of noise sensitivity of boolean functions. This family of functions is, in a sense, the set of all "uniformly noise-stable" functions. In [4], 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.

Stated in terms of Fourier coefficients (see Section 7.2), the Majority Is Stablest theorem 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. Some strong evidence in favor of the Majority Is Stablest theorem was given by Bourgain [7], 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.

The expression $\mathbb{S}_{-1/3}(f)$ played a central role in a Fourier-theoretic approach to the Condorcet Paradox and Arrow's Theorem given by Kalai [33]. This expression determines the probability of an "irrational outcome" in a certain voting scheme. Much of [33] is devoted to the study of $\mathbb{S}_{-1/3}(f)$ and in particular, it is conjectured there (Conjecture 5.1) 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. Although this conjecture turns out to be false, the corollaries of the conjecture in [33] are implied by the fact that Majority is the stablest transitive function, and this is a consequence of the Majority Is Stablest theorem.

Finally, in [43] it was 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 Section 7) among balanced functions f for large k and $n = \infty$.

For some interesting special cases of the Majority Is Stablest theorem we can give short and simple proofs. These are shown in Section 10.

4.2 Generalizations to the q-ary domain

Our methods can also be used to obtain hardness results for constraint satisfaction problems over variables ranging over larger domains $[q]$. In the q-ary regime we need a multi-valued analogue of the Majority Is Stablest theorem. Before we can formulate the appropriate analogue, we need to specify what we mean by 'q-ary functions' and also to define the notions of noise stability and influences for them.

The obvious generalization of a boolean function to the q-ary regime would be a function of the form $f : [q]^n \to [q]$. However, as we did for boolean functions, we will consider a continuous relaxation of the range. Specifically, define

$$\Delta_q = \{(x_1, \dots, x_q) \in [0,1]^q : \textstyle\sum x_i = 1\},$$

which can be thought of as the space of probability distributions over $[q]$. We will consider functions $f : [q]^n \to \Delta_q$; this generalizes functions $f : [q]^n \to [q]$ if we identify the elements $a \in [q]$ in f's range with the points $(0, \dots, 0, 1, 0, \dots, 0) \in \Delta_q$.

Definition 5. *Let $-\frac{1}{q-1} \le \rho \le 1$ and let x and y be $[q]^n$-valued random variables. We say that x and y are a ρ-*correlated pair *if x is uniformly distributed on $[q]^n$, and y is formed from x by choosing each y_i so that $\Pr[y_i = a] = \delta_{\{x_i=a\}}\rho + \frac{1-\rho}{q}$ for each a, independently for each i. Note that for $0 \le \rho \le 1$, it is equivalent to say that each coordinate y_i is independently chosen to be x_i with probability ρ and is a uniformly random element of $[q]$ otherwise.*

Definition 6. *Let $f : [q]^n \to \Delta_q$ and let $-\frac{1}{q-1} \le \rho \le 1$. The* noise stability of f at ρ *is defined to be*

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

where x and y are a ρ-correlated pair. Equivalently, we may define the noise stability of functions $g : [q]^n \to \mathbb{R}$ via

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

and then denoting by f^i the ith coordinate projection of f, we have $\mathbb{S}_\rho(f) = \sum_{i=1}^n \mathbb{S}_\rho(f^i)$.

We remark that when f's range is simply $[q]$ (as embedded in Δ_q), the quantity $\mathbb{S}_\rho(f)$ is simply the probability that $f(x) = f(y)$ when x and y are a ρ-correlated pair.

The definition of influences is very similar to that in the boolean case:

Definition 7. *Let* $f : [q]^n \to \Delta_q$. *For* $1 \le i \le n$, *the* influence of the ith coordinate on f *is defined to be*

$$\mathrm{Inf}_i(f) = \mathop{\mathbf{E}}_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n}[\mathrm{Var}_{x_i}[f^i(x_1,\ldots,x_n)]].$$

We say that $f : [q]^n \to \Delta_q$ is 'balanced' if $\mathbf{E}[f^i] = 1/q$ for each i. The most obvious generalization of the Majority function to the q-ary domain is the Plurality function, which on input $x \in [q]^n$ outputs the most common value for x_i (tie-breaking is unimportant). It is natural to ask whether a "Plurality Is Stablest" theorem holds. This question is still open, and we present it as a conjecture. For this purpose, define

$$\mathrm{PlurStab}(q,\rho) = \lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{Plurality}_{n,q}).$$

The limit in the formula above indeed exists, and there appears to be no closed formula for it; however we provide an exact description of it in Theorem 8 in Section 6.

Plurality Is Stablest Conjecture. *Fix* $q \ge 2$ *and* $-\frac{1}{q-1} \le \rho \le 1$. *Then for any* $\epsilon > 0$ *there is a small enough* $\delta = \delta(\epsilon,\rho,q)$ *such that if* $f : [q]^n \to [q]$ *is any balanced* q*-ary function with* $\mathrm{Inf}_i(f) \le \delta$ *for all* $i = 1 \ldots n$, *then*

$$\mathbb{S}_\rho(f) \le \mathrm{PlurStab}(q,\rho) + \epsilon$$

Note that in the case $q = 2$, Sheppard's formula gives $\mathrm{PlurStab}(2,\rho) = 1 - \frac{2}{\pi}\arccos\rho$, which is the noise stability of Majority; there is also a closed formula for $q = 3$ ([26, 12]). For large values of q we give asymptotics which hold up to a $1 + o_q(1)$ factor in Section 6. For the reader's convenience, we remark here that

$$\mathrm{PlurStab}(q,\rho) = \tilde{\Theta}\left((1/q)^{(1-\rho)/(1+\rho)}\right).$$

Although we don't have Plurality Is Stablest, a result of [42] generalizing Majority Is Stablest serves us almost equally well. This result bounds the stability of a function in terms of the behavior of correlated Gaussians. To state it, we need one more definition:

Definition 8. *Let* $\mu \in [0,1]$ *and* $\rho \in [0,1]$. *Let* X *and* Y *denote normal random variables with mean* 0 *and covariance matrix* $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. *We define*

$$\Lambda_\rho(\mu) = \Pr[X \ge t \text{ and } Y \ge t],$$

where t *is chosen so that* $\Pr[X \ge t] = \mu$.

MOO theorem: *Fix* $q \ge 2$ *and* $\rho \in [0,1)$. *Then for any* $\epsilon > 0$ *there is a small enough* $\delta = \delta(\epsilon,\rho,q) > 0$ *such that if* $f : [q]^n \to [0,1]$ *is any function satisfying* $\mathbf{E}[f] = \mu$ *and* $\mathrm{Inf}_i(f) \le \delta$ *for all* $i = 1 \ldots n$, *then*

$$\mathbb{S}_\rho(f) \le \Lambda_\rho(\mu) + \epsilon.$$

As a result we have that the noise stability of any balanced $f : [q]^n \to \Delta_q$ is essentially at most $q\Lambda_\rho(1/q)$. We give the asymptotics of this quantity in Section 6 and they are extremely close to those of $\mathrm{PlurStab}_\rho(q)$; in particular, they are the same up to a constant multiplicative factor.

5 On the geometry of MAX-CUT

We shall now try to explain (non-rigorously) the connection between the Majority Is Stablest theorem 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 Goemans-Williamson arises as the solution of 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 in the construction of our Long Code test, although it is only implicit in the actual proofs.

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 theorem 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 maximum 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 fall. 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

$$\alpha_{\text{GW}} = \min_{-1\le\rho\le 1} \frac{(\arccos \rho)/\pi}{\frac{1}{2} - \frac{1}{2}\rho} \approx .878567 \tag{2}$$

is the smallest ratio possible between the probability of an edge being 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 [18], Feige and Schechtman considered the graph G_ρ whose vertices are all the vectors on the unit sphere and in which 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 [18] 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 this discussion.) Such a hyperplane cuts an $(\arccos \rho)/\pi$-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 about $(\arccos \rho)/\pi$. 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 $f_C(x) = x_1$; i.e., it is a correct Long Code word. 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 Majority Is Stablest theorem 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 theorem 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 7.4 in Section 7.3 says the following.

Proposition *For any $\rho \in (-1, 0]$ and 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 $(\arccos \rho)/\pi + \epsilon$*

6 Our results

In this section we formally state our main results.

6.1 Hardness for MAX-CUT and 2-bit CSPs

Our main result regarding MAX-CUT is the following:

Theorem 1. *Assume the Unique Games Conjecture. Then for every constant $-1 < \rho < 0$ and $\epsilon > 0$, it is NP-hard to distinguish instances of MAX-CUT that are at least $(\frac{1}{2} - \frac{1}{2}\rho)$-satisfiable from instances that are at most $((\arccos \rho)/\pi + \epsilon)$-satisfiable. In particular, choosing $\rho = \rho^*$, where*

$$\rho^* = \underset{-1<\rho<0}{\operatorname{argmin}} \frac{(\arccos \rho)/\pi}{\frac{1}{2} - \frac{1}{2}\rho} \approx -.689,$$

implies that it is NP-hard to approximate MAX-CUT to within any factor greater than the Goemans-Williamson constant $\alpha_{GW} \approx .878567$.

Recall that the main result of Goemans and Williamosn [25] is an algorithm which, given instances of MAX-CUT with fractional optimum at least $\frac{1}{2} - \frac{1}{2}\rho$ (where $\rho \le \rho^*$), outputs a solution with value at least $(\arccos \rho)/\pi - \epsilon$ (where $\epsilon > 0$ can be an arbitrarily small constant). Thus our Unique Games-hardness theorem precisely matches the algorithmic guarantee of Goemans and Williamson for *all* $-1 < \rho \le \rho^*$. For ρ very close to -1, by considering the Taylor expansion of $\arccos \rho$, we have the following corollary:

Corollary 2. *Assume the Unique Games Conjecture. Then for all sufficiently small $\eta > 0$, it is NP-hard to distinguish instances of MAX-CUT that are at least $(1-\eta)$-satisfiable from instances that are at most $(1-(2/\pi)\sqrt{\eta})$-satisfiable.*

We prove Theorem 1 in Section 8.

In Section 9 we apply our techniques for other 2-bit CSPs besides MAX-CUT. In particular we prove:

Theorem 3. *Assume the Unique Games Conjecture. Then it is NP-hard to approximate MAX-2SAT to within any factor greater than β, where*

$$\beta = \min_{\frac{\pi}{2} \le \theta \le \pi} \frac{2 + (2/\pi)\theta}{3 - \cos\theta} \approx .943.$$

The proof of Theorem 3 actually implies that MAX-2SAT is hard to approximate to within any factor greater than β, even if restricted to instances where each variable appears equally often positively and negatively (see Section 9 for more details). We show that for this restricted problem, called Balanced-MAX-2SAT, the approximation bound β is tight; i.e., it *can* be approximated to within any factor smaller than β:

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

6.2 Special cases of the Majority Is Stablest theorem

Some special cases of the Majority Is Stablest theorem are of independent interest.

First, it should be noted that the Majority function is not a 'unique' optimizer, in the sense that every weighted threshold that does not depend largely on any one coordinate is equally noise-stable:

Theorem 5. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any balanced threshold function, namely of the form $f(x) = \mathrm{sgn}(a_1x_1 + \cdots a_nx_n)$[1]. Let $\delta = \max_i \{\mathrm{Inf}_i(f)\}$. Then for all $\rho \in [-1,1]$,*

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

It is also of interest to consider the case where ρ tends to zero. It is easy to see that in this case the Majority Is Stablest theorem implies that the weight of a Boolean function on the first level of its Fourier transform is essentially bounded by $2/\pi$. We give an easy and direct proof of this fact:

Theorem 6. *Suppose $f : \{-1,1\}^n \to [-1,1]$ satisfies $\mathrm{Inf}_i(f) \le \delta$ for all i. Then*

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

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

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

We also can give a direct proof of an improved version of Theorem 6 which depends on the mean of f; as the mean becomes small enough, this result approaches a result of Talagrand [49] (which states that for every function $f : \{-1,1\}^n \to \{-1,1\}$ with $\Pr[f = 1] = p \le 1/2$ it holds that $\sum_{|S|=1} \hat{f}(S)^2 \le O(p^2 \log(1/p))$):

Theorem 7. *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

$$\epsilon = \max\{1, \sqrt{|\Phi^{-1}(\mu)|}\} \cdot O(\sqrt{\delta}).$$

This theorem is sharp up to the error term, as can be observed by considering restrictions of Majority functions — i.e., symmetric threshold functions. Note that for x small, $U(x) \sim x\sqrt{2\ln(1/x)}$; this is why our result is comparable with Talagrand's.

6.3 Larger domains: q-ary functions

In this section we state our results for q-ary functions and for q-ary constraint satisfaction problems. We will be concerned with two such 2-CSPs. The first is MAX-q-CUT, the problem of partitioning a graph into q parts so as to maximize the number of edges between parts. The second is is MAX-2LIN(q): Given an integer $q \ge 2$, the MAX-2LIN(q) problem is to maximize the number of satisfied equations in a given system of linear equations modulo q, where exactly two variables appear in each equation. See Section 11.1 for formal definitions.

Stability estimates. Our hardness results are based in part on the following analysis of the noise stability of q-ary functions, as discussed in Section 4. We first obtain an exact analytic expression for the noise stability of the plurality function.

Theorem 8. *Fix q and $-\frac{1}{q-1} \le \rho \le 1$. Then*

$$\lim_{n\to\infty} \mathbb{S}_\rho(\mathrm{Plurality}_n) = qI(q,\rho),$$

where $I(q,\rho)$ is defined as follows: Let $(U_1, V_1) \ldots, (U_q, V_q)$ be a set of q i.i.d. normal vectors with mean 0 and covariance matrix $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$; then

$$I(q,\rho) = \Pr[U_1 = \max_{1\le i\le q} U_i, \quad V_1 = \max_{1\le i\le q} V_i].$$

Further, the quantity $I(q,\rho)$ is precisely equal the key quantity called $I(\rho)$ (with $q = k$) in Frieze and Jerrum's paper on MAX-q-CUT [24] (see also [12]).

As a corollary of Theorem 8, and a result of de Klerk et al. [12] (see also [24]) which gives the asymptotics of $I(q,\rho)$, we obtain the following:

Corollary 9. *For every fixed $0 \le \rho < 1$, we have*

$$\mathrm{PlurStab}(q,\rho) \sim \left(\frac{1}{q-1}\right)^{(1-\rho)/(1+\rho)} (4\pi \ln(q-1))^{-\rho/(1+\rho)} \frac{\Gamma(1/(1+\rho))^2}{(1-\rho^2)^{1/2}}, \tag{3}$$

where the $\sim$ indicates that the ratio of the two sides is 1 as $q \to \infty$, and Γ is the gamma function.

Since we do not have the Plurality Is Stablest conjecture we cannot actually use Corollary 9 in our hardness results. Instead we use the MOO theorem which is stated in terms of the function $\Lambda_\rho(\mu)$; therefore we need bounds on its asymptotics. Slightly improving the estimate from Lemma 11.1 of [12], we have:

Proposition 6.1. *Denote by ϕ the Gaussian density function $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, and let $N(x) = \int_x^\infty \phi$ denote the Gaussian tail probability function. For any $0 \le \mu < 1/2$, let $t > 0$ be the number such that $N(t) = \mu$. Then for all $0 \le \rho \le 1$,*

$$\Lambda_\rho(\mu) \le (1+\rho) \cdot \frac{\phi(t)}{t} \cdot N\Big(t\sqrt{\tfrac{1-\rho}{1+\rho}}\Big). \tag{4}$$

Note that in the case $\mu = 1/2$, $\Lambda_\rho(1/2) = \frac{1}{2}(1 - \frac{2}{\pi}\arccos\rho)$, and the case $\mu > 1/2$ can be easily reduced to the case $\mu < 1/2$. Also, it is relatively easy to see that the right hand side of (4) becomes a lower bound on $\Lambda_\rho(\mu)$ if the $(1+\rho)$ factor is removed.

In fact, we are mainly interested in the case where $\mu \to 0$ ($t \to \infty$). In this case, it turns out that (4) holds as an equality up to a $1 + o_\mu(1)$ factor (even if ρ is a function of μ). This yields the following:

Corollary 10. *Let $\mu \to 0$ and let $t = t(\mu)$ be defined as in Proposition 6.1. Then the following holds:*

1. *For any $\rho = \rho(\mu)$, $0 < \rho < 1$,*

$$\Lambda_\rho(\mu) \overset{\mu\to 0}{\sim} (1+\rho)\cdot\mu\cdot N\Big(t\sqrt{\tfrac{1-\rho}{1+\rho}}\Big)$$

   (where by '$\sim$' we mean that the ratio between the two sides tends to one).

2. *If $0 < \rho < 1$ is fixed, then*

$$\Lambda_\rho(\mu) \overset{\mu\to 0}{\sim} \mu^{2/(1+\rho)}(4\pi\ln(1/\mu))^{-\rho/(1+\rho)}\frac{(1+\rho)^{3/2}}{(1-\rho)^{1/2}}.$$

3. *For any fixed $0 < \eta < 1$,*

$$q\Lambda_{1-\eta}(1/q) \le (1/q)^{\eta/(2-\eta)}.$$

4. *For any $\lambda = \lambda(q) \in (0,1)$, let $\rho = 1 - \frac{\lambda}{\ln q}$. Then*

$$q\Lambda_\rho(1/q) \le 1 - \sqrt{2/\pi}\cdot\sqrt{\lambda} + o_{\lambda\to 0}(1) + o_{q\to\infty}(1).$$

Part 2 of Corollary 10 is due de Klerk et al. [12]. It implies that $q\Lambda_\rho(1/q)$ and $\mathrm{PlurStab}_\rho(q)$ have the same asymptotics as q tends to infinity, up to a small multiplicative constant. The other statements of Corollary 10 are proven in Section 11.3.

Hardness Results. We now move to stating our hardness results for q-ary domains. For MAX-q-CUT we show that assuming the Unique Games Conjecture, it is impossible to essentially improve on the approximation ratios for MAX-q-CUT achieved by Frieze and Jerrum [24] by more than an additive ϵ.

Theorem 11. *Assume the Unique Games Conjecture. Then for every $\epsilon > 0$ it is NP-hard to distinguish $(1-\epsilon)$-satisfiable instances of MAX-q-CUT from instances that are at most $(1 - 1/q + (2\ln q)/q^2 + O(\ln\ln q)/q^2)$-satisfiable.*

Our hardness result for MAX-2LIN(q) is formulated in terms of $\Lambda_q(\mu)$, discussed above.

Theorem 12. *Assume the Unique Games Conjecture. Then for every $q \geq 2$, $\rho \in [0,1]$ and $\epsilon > 0$, given an instance of MAX-2LIN(q), it is NP-hard to distinguish between the case where it is at least $(\rho + \frac{1}{q}(1-\rho) - \epsilon)$-satisfiable and the case where it is at most $(q\Lambda_\rho(\frac{1}{q}) + \epsilon)$-satisfiable. Furthermore, this holds even for instances in which all equations are of the form $x_i - x_j = c$.*

Using the asymptotics of $\Lambda_\rho(\mu)$ given above in Corollary 10, we have:

Corollary 13. *Assume the Unique Games Conjecture. Then for every fixed $\eta > 0$ there exists $q_0 = q_0(\eta)$ such that for every fixed $q > q_0$ the following holds. Given an instance of MAX-2LIN(q), it is NP-hard to distinguish between the case where the instance is at least $(1 - \eta)$-satisfiable and the case where it is at most $(1/q)^{\eta/(2-\eta)}$-satisfiable.*

Corollary 14. *Assume the Unique Games Conjecture, and let $\lambda = \lambda(q) \in (0,1)$. Given an instance of MAX-2LIN(q), it is NP-hard to distinguish between the case where the instance is at least $(1 - \frac{\lambda}{\ln q})$-satisfiable and the case where it is at most s-satisfiable, where*

$$s = 1 - \sqrt{2/\pi} \cdot \sqrt{\lambda} + o_{\lambda \to 0}(1) + o_{q \to \infty}(1).$$

Note that MAX-2LIN(q) is itself essentially an instance of Unique Label Cover, except for the fact that the variable/equation structure need not be bipartite. But in fact, it is easy to observe that the "non-bipartite" version of the Unique Games Conjecture is equivalent to the usual Unique Games Conjecture [38] (up to a factor of 2 in the soundness). Hence Theorem 12 and its corollaries may be viewed as concerning the allowable parameter tradeoffs in the Unique Games Conjecture. In particular, Corollary 13 implies:

Corollary 15. *The Unique Games Conjecture holds if and only if it holds as follows: For every $\eta > 0$ and label set size q (sufficiently large as a function of η), it is NP-hard to distinguish whether the Unique Label Cover problem with label set size q has optimum at least $1 - \eta$ or at most $(1/q)^{\eta/(2-\eta)}$.*

(The factor of 2 lost in soundness from passing to a bipartite version can be absorbed since the soundness obtained in the proof of Corollary 13 is actually stronger by a factor of $(\log q)^{\Omega(1)}$.)

Recently, a result of Charikar, Makarychev, and Makarychev [10] showed that the parameters in Corollary 15 are almost optimal. They give an algorithm for Unique Label Cover with label set size q that, given an instance with optimum $(1 - \eta)$, outputs an assignment which satisfies at least a $(1/q)^{\eta/(2-3\eta)}$-fraction of the constraints.

7 Definitions and technical preliminaries

In this section we give some definitions and make some technical observations concerning the Majority Is Stablest theorem, reducing it to a form which is useful for our MAX-CUT reduction.

7.1 MAX-CUT and MAX-2SAT

For the majority of this paper we will be concerned with the MAX-CUT problem; we will also later consider the MAX-2SAT problem. We give the formal definitions of these problems below.

Definition 9 (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 10 (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 this paper we do not make a distinction between the weighted and the non-weighted versions.

7.2 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 $\|f\|_2 = \sqrt{\mathbf{E}[f^2]}$.

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. *Plancherel'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 3).

Proposition 7.1. *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 2).

Proposition 7.2. *Let $f\colon \{-1,1\}^n \to \mathbb{R}$. Then for every $i \in [n]$,*

$$\mathrm{Inf}_i(f) = \sum_{S\ni i} \hat{f}(S)^2. \tag{5}$$

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

Definition 11. *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*

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

7.3 Different forms of the Majority Is Stablest theorem

Recall the Majority Is Stablest theorem:

Majority Is Stablest theorem: *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*

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

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

then

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

In the MAX-CUT reduction we need a slightly altered version of the Majority Is Stablest theorem. First, we can replace influences by low-degree influences:

Proposition 7.3. *The Majority Is Stablest theorem 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. 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

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

for all i.

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

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

□

Second, we need to treat the case of negative ρ:

Proposition 7.4. *The Majority Is Stablest theorem is true 'in reverse' for* $\rho \in (-1, 0]$*. That is,* $\mathbb{S}_\rho(f) \geq 1 - \frac{2}{\pi}\arccos\rho - \epsilon$*, 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 theorem applied to g. □

Combining the above two propositions we get the result that will be used in our reduction from Unique Label Cover to 2-bit CSPs:

Proposition 7.5. *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*

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

then

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

8 Reduction from Unique Label Cover to MAX-CUT

In this section we prove Theorem 1.

8.1 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 theorem. 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 theorem.

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 $(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(M)})$. For x, $\mu \in \{-1, 1\}^M$, let $x\mu$ denote the M-bit string that is the coordinatewise product of x and μ.

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. Using a result from [37] we may assume the bipartite graph is regular on the V side, so that choosing a uniformly random vertex $v \in V$ and a random neighbor w of v yields a uniformly random edge (u, w). We assume that the Unique Label Cover instance is either $(1-\eta)$-satisfiable or at most γ-satisfiable, where we will choose the values of η and γ 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 for MAX-CUT 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_w and $f_{w'}$ be the supposed Long Codes of the 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
$$f_w(x \circ \sigma) \neq f_{w'}((x \circ \sigma')\mu).$$

8.2 Completeness

It is easy to see that the completeness of the verifier is at least $(1-2\eta)(\frac{1}{2}-\frac{1}{2}\rho)$. 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 at least $1-2\eta$, 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 $f_w, f_{w'}$ are the Long Codes of j,j' respectively. Hence

$$f_w(x\circ\sigma) = x_{\sigma(j)} = x_i, \quad f_{w'}((x\circ\sigma')\mu) = x_{\sigma'(j')}\mu_{j'} = x_i\mu_{j'}$$

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

8.3 Soundness

We prove soundness in the contrapositive direction, as is usual in PCP proofs: Assume that some supposed Long Codes f_w cause the PCP verifier to accept with probability at least $(\arccos\rho)/\pi+\epsilon$. We use Fourier methods to "list-decode" the Long Codes and extract a labeling for the Unique Label Cover instance that satisfies some $\gamma'=\gamma'(\epsilon,\rho)$ fraction of its edges. Since this constant does not depend on the Label Cover label set size M, we can take M large enough in the Unique Games Conjecture to get soundness $\gamma<\gamma'$, as required.

We first analyze the probability of acceptance for the PCP verifier by arithmetizing it as follows:

$$\begin{aligned}
\Pr[\text{acc}] &= \mathop{\mathbf{E}}_{v,w,w',x,\mu}\left[\frac{1}{2}-\frac{1}{2}f_w(x\circ\sigma)f_{w'}((x\mu)\circ\sigma')\right]\\
&= \frac{1}{2}-\frac{1}{2}\cdot\mathop{\mathbf{E}}_{v,x,\mu}\left[\mathop{\mathbf{E}}_{w,w'}[f_w(x\circ\sigma)f_{w'}((x\mu)\circ\sigma')]\right]\\
&= \frac{1}{2}-\frac{1}{2}\cdot\mathop{\mathbf{E}}_{v,x,\mu}\left[\mathop{\mathbf{E}}_{w}[f_w(x\circ\sigma)]\cdot\mathop{\mathbf{E}}_{w'}[f_{w'}((x\mu)\circ\sigma')]\right] && \text{(using independence of } w \text{ and } w')\\
&= \frac{1}{2}-\frac{1}{2}\cdot\mathop{\mathbf{E}}_{v,x,\mu}[g_v(x)g_v(x\mu)] && \text{(where we define } g_v(z)=\mathop{\mathbf{E}}_{w\sim v}[f_w(z\circ\sigma_{v,w})])\\
&= \frac{1}{2}-\frac{1}{2}\cdot\mathop{\mathbf{E}}_{v}[\mathbb{S}_\rho(g_v)].
\end{aligned}$$

Now if $\Pr[\text{acc}]\ge(\arccos\rho)/\pi+\epsilon$, then for at least an $\epsilon/2$ fraction of $v\in V$,

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

We say that such a vertex v is "good". For every good v, we apply the Majority Is Stablest theorem in the guise of Proposition 7.5 to conclude that g_v 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. For a good v, since $\mathrm{Inf}_j^{\le k}(g_v)\ge\delta$, we have

$$\delta \le \sum_{\substack{S\ni j\\|S|\le k}}\widehat{g_v}(S)^2 = \sum_{\substack{S\ni j\\|S|\le k}}\mathop{\mathbf{E}}_{w}[\widehat{f_w}(\sigma^{-1}(S))]^2 \le \sum_{\substack{S\ni j\\|S|\le k}}\mathop{\mathbf{E}}_{w}[\widehat{f_w}(\sigma^{-1}(S))^2] = \mathop{\mathbf{E}}_{w}\left[\mathrm{Inf}^{\le k}_{\sigma^{-1}(j)}(f_w)\right]. \tag{6}$$

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

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

Since $\sum_i \mathrm{Inf}_i^{\leq k}(f_w) \leq k$, we conclude that $|\mathrm{Cand}[w]| \leq 2k/\delta$. Inequality (6) implies that for every good v, at least a $\delta/2$ fraction of neighbors w of v have $\mathrm{Inf}^{\leq k}_{\sigma^{-1}_{v,w}(j)}(f) \geq \delta/2$, 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 $(\delta/2)(\delta/2k)$-fraction are satisfied in expectation. Thus it follows that there is labeling for all vertices which satisfies a $\gamma' = (\epsilon/2)(\delta/2)(\delta/2k)$ fraction of all edges. This completes the proof of soundness.

8.4 Completion of the proof of Theorem 1

We have just shown how to reduce Unique Label Cover instances to MAX-CUT instances with completeness $(\frac{1}{2} - \frac{1}{2}\rho)(1 - 2\eta)$ and soundness $(\arccos \rho)/\pi + \epsilon$, where η and ϵ can be made arbitrarily small. The main statement of Theorem 1 follows by slightly modifying ρ to move the completeness correction $(1 - 2\eta)$ into the soundness correction ϵ. This result implies a hardness of approximation factor of

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

for any constant $-1 < \rho < 0$ and $\epsilon > 0$; choosing $\rho = \rho^*$ as stated in the theorem yields the desired hardness factor $\alpha_{\text{GW}} + \epsilon$.

9 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.

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 theorem 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

$$\beta = \min_{-1<\rho<0} \frac{\frac{3}{4} - \frac{1}{4}(1 - \frac{2}{\pi}\arccos \rho)}{\frac{3}{4} - \frac{1}{4}\rho} \approx .943943. \tag{7}$$

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 [31] (using the gadget of Bellare, Goldreich, and Sudan [2]); as well, the best algorithm known for MAX-2SAT, due to Lewin, Livnat, and Zwick [40], achieves an approximation ratio of .9401 which is close to and smaller than β.

Our methodology does 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 $\frac{1}{4} - \frac{1}{4}\mathbb{S}_\rho(h)$. By the Majority Is Stablest theorem, we get the same hardness of approximation for MAX-2ConjSAT 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 ([40] 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 12. *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

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

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 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 $x_i = \operatorname{sgn}(r \cdot v_i)$. Recalling from [25] 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. $\square$

10 Special cases of the Majority Is Stablest theorem

In this section we describe special cases of the Majority Is Stablest theorem which are of particular interest, and have elementary proofs.

10.1 Weighted majorities

In this subsection we prove Theorem 5, which makes the point that the majority function is not unique as a noise stability maximizer, in the sense that all weighted majority functions with small influences have the same noise stability, i.e., $1 - \frac{2}{\pi}\arccos\rho$. The proof requires a multidimensional version of the Central Limit Theorem with error bounds. We have the following version from [5, Corollary 16.3]:

Theorem 16. *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent random variables taking values in $\mathbb{R}^k$ satisfying:*

- $\mathbf{E}[\mathbf{X}_j] = 0, j = 1 \ldots n$;
- $n^{-1}\sum_{j=1}^{n} \mathrm{Cov}(\mathbf{X}_j) = V$, *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,

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

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.

Theorem 5 follows from the following two propositions.

Proposition 10.1. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any balanced threshold function, $f(x) = \mathrm{sgn}(a_1x_1 + \cdots a_nx_n)$[2], where $\sum a_i^2 = 1$. Let $\delta = \max\{|a_i|\}$. 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 10.2. *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 $\sum a_i^2 = 1$. Let $\delta = \max\{|a_i|\}$. Then $\max_i\{\mathrm{Inf}_i(f)\} \geq \Omega(\delta)$.*

We prove the two propositions below.

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

Proof of Proposition 10.1. 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 16. 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 16 we get $Q_n(A_{++} \cup A_{--}) = 1 - (\arccos\rho)/\pi \pm O(\eta)$ and the theorem follows. $\square$

Proof of Proposition 10.2. We may assume without loss of generality that $\delta = a_1 \ge a_2 \ge \cdots \ge a_n \ge 0$. Letting X_i denote the random variable $a_i x_i$, we will prove that $\mathrm{Inf}_1(f) \ge \Omega(\delta)$ by proving that

$$\Pr[|X_2 + \cdots + X_n| \le \delta] \ge \Omega(\delta). \tag{8}$$

Let C be the constant hidden in the $O(\cdot)$ in the final part of the Berry-Esseen theorem, Theorem 18. In proving (8) we may also assume without loss of generality that

$$1 - 100C^2\delta^2 \ge 1/4; \tag{9}$$

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-Esseen theorem to the sequence $X_m, \ldots, X_n$. We have

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

where we have used (9). We also have $\rho_3 = \sum_{j=m}^{n} \mathbf{E}[|X_j|^3] \le \sum_{j=m}^{n} a_m \mathbf{E}[X_j^2] = a_m\sigma^2$, so the error term in the conclusion of the theorem, $O(\rho_3/\sigma^3)$, is at most $Ca_m/\sigma \le 2Ca_m$.

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

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

where we have used the fact that $\phi(\delta) \ge .24$ for $\delta \le 1$. 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 $|X_2 + \cdots + X_n| \le \delta$ as desired.

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

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

where we have used (9) 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. $\square$

10.2 Bounds for the weight on the first level

Applying the Majority Is Stablest theorem for extremely small ρ, it follows that functions with small influences have no more weight at level 1 than Majority has, viz., $\frac{2}{\pi}$ (up to $o(1)$). This fact, stated in Theorem 6, has a very elementary proof which also provides a better bound on the additive term corresponding to the maximal influence:

Proof of Theorem 6. 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 [39] 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 $\|\ell\|_2^2 \leq 2/\pi + C\delta$. $\square$

In the following we improve the bound on the weight of the first level for not necessarily balanced functions with low influences. This result should be compared to the following theorem of Talagrand [49]:

Theorem 17. *(Talagrand) Suppose $f : \{-1,1\}^n \to \{-1,1\}$ satisfies $\Pr[f=1] = p \leq 1/2$. Then*

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

We will need the following Berry-Esseen theorem, which gives error bounds for the Central Limit Theorem [19]:

Theorem 18. *(Berry-Esseen) 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 $\sum_{j=1}^n \mathbf{E}[|X_j|^3] = \rho_3$. Let $Q = \sigma^{-1}(X_1+\cdots+X_n)$, let F denote the cumulative distribution function of Q, $F(x) = \Pr[Q \leq x]$, and let Φ denote the cumulative distribution function of a standard normal random variable. Then*

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

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

We now prove Theorem 7.

Proof of Theorem 7. It will be more convenient to work with the $[0,1]$ valued-function $g = \frac{1}{2} + \frac{1}{2}f$ and prove that $\sum_{|S|=1} \hat{g}(S)^2 \leq \left(U(\mu) + \max\{1, \sqrt{|\Phi^{-1}(\mu)|}\}O(\sqrt{\delta})\right)^2$. Note that $\mu = \mathbf{E}[g]$. We will assume without loss of generality that $\mu \geq 1/2$ (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 6, 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 $\sum_S \mathbf{E}[X_S^3] \leq \max_S |X_S| \leq \delta/(2\tau)$.

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

$$\tau = \langle g, L \rangle \leq U(\mu) + \max\{1, |\Phi^{-1}(\mu)|\}O(\delta/\tau). \tag{10}$$

Multiplying by τ implies that

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

which in turn implies that

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

Finally, we will conclude that

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

We now prove (10). Let t be a number such that $\Pr[L > t] = \mu$. Since g is a $[0,1]$ valued-function it follows that $\langle g, L \rangle \leq \mathbf{E}[\mathbf{1}_{L>t} L]$.

Letting F denote the cumulative distribution function of L, the Berry-Esseen 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

$$|\mu - \Phi(-t)| \leq O\left(\frac{\delta}{\tau(1+t^3)}\right) \tag{11}$$

Note that the function U satisfies

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

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

$$\begin{aligned} |U(\mu) - \phi(t)| &= |U(\Phi(-t)) - U(\mu)| \leq |\Phi(-t) - \mu| \max\{|U'(\Phi(-t))|, |U'(\mu)|\} \\ &\leq \max\{|t|, \Phi^{-1}(\mu)\}O(\delta/(\tau(1+t^3))) \leq \max\{1, |\Phi^{-1}(\mu)|\}O(\delta/\tau). \end{aligned} \tag{12}$$

Further,

$$\begin{aligned}
\langle g, L\rangle &\le \mathbf{E}[\mathbf{1}_{L>t}L] = t\Pr[L>t] + \int_t^\infty \Pr[L>x]\,dx \\
&= t\Pr[L>t] + \int_t^\infty \Pr[N(0,1)>x]\,dx + \int_t^\infty (F(x)-\Phi(x))\,dx \\
&= t\mu - t\Phi(-t) + \phi(t) + \int_t^\infty (F(x)-\Phi(x))\,dx \\
&\le \phi(t) + |t|\cdot|\mu - \Phi(-t)| + \int_t^\infty |F(x)-\Phi(x)|\,dx \\
&\le \phi(t) + \frac{|t|}{1+|t|^3}O(\delta/\tau) + O(\delta/\tau)\int_t^\infty 1/(1+|x|^3)\,dx \qquad \text{((11) and Berry-Esseen)} \\
&= \phi(t) + O\left(\frac{\delta}{\tau(1+t^2)}\right) \qquad \text{(by (12))} \\
&\le U(\mu) + \max\{1, |\Phi^{-1}(\mu)|\}O(\delta/\tau).
\end{aligned}$$

which proves (10) as needed. □

11 Constraint satisfaction problems over $[q]$

So far in this paper we have mostly focused on 2-CSPs in which the variables are binary — i.e., take values in the alphabet $\{-1, 1\}$. The exception is the Unique Label Cover problem, which can be thought of as a 2-CSP where the set of values a variable can take is very large. In this section we develop our techniques for 2-CSPs over such large alphabets; specifically, the alphabet $[q] = \{1, 2, \ldots, q\}$ for $q \ge 2$. We will be concerned in particular with the MAX-2LIN(q) and MAX-q-CUT (i.e. Approximate Graph q-Coloring) problems, and we will mostly be interested in the asymptotics when $q \to \infty$.

11.1 Γ-MAX-2LIN(q) and MAX-q-CUT

The MAX-q-CUT problem is a natural generalization of MAX-CUT; its formal definition is as follows:

Definition 13 (MAX-q-CUT). *Given a weighted graph $G = (V, W)$ where $W : V \times V \to \mathbb{R}_+$, the* MAX-$q$-CUT *problem is that of finding a partition of V into q sets $V_1, \ldots, V_q$ in such a way as to maximize the weight of edges between the different parts, $\sum_{i \ne j} \sum_{v \in V_i, w \in V_j} W(v, w)$.*

The MAX-2LIN(q) problem is defined as follows:

Definition 14 (MAX-2LIN(q)). *Given a system of m linear equations mod q each having at most 2 variables, along with nonnegative weights $w_1, \ldots, w_m$ for these equations, the* MAX-2LIN(q) *problem is to find an assignment to the variables maximizing the total weight of satisfied equations.*

Observe that the MAX-2-CUT problem can be viewed as a special case of the MAX-2LIN(2) problem, by associating the vertices with variables and the edges with equations $x_u - x_v = 1$. However for larger q, MAX-q-CUT is more naturally viewed as a special case of the problem of finding assignments for 2-variable linear *inequations* mod q.

The best approximation algorithm for MAX-q-CUT was obtained by Frieze and Jerrum in [24]; this paper gives a $(1 - 1/q + 2(\ln q)/q^2)$-approximation. The best NP-hardness result known is due to Kann, Khanna, Lagergren, and Panconesi [] who proved that $(1 - 1/(34q))$-approximating is hard. As for the approximability of MAX-2LIN(q), the best algorithm known was given very recently by Charikar, Makarychev, and Makarychev in [], as discussed in Subsection 6.3 (their algorithm is actually for the more general problem of not-necessarily-bipartite Unique Label Cover with label size q). When the optimal fraction of satisfiable constraints is $1 - \eta$, their semidefinite programming algorithm produces a solution satisfying about a fraction $(1/q)^{\eta/(2-3\eta)}$. The best known NP-hardness results come from a recent work of Feige and Reichman [20]. They show that it is NP-hard to approximate MAX-2LIN(q) to within a factor of $1/q^{\beta}$ for some universal constant $\beta > 0$; however this hardness is located at a gap of ϵ vs. ϵ/q^{β}. In particular, given an instance with optimum fraction $1 - \eta$, Feige and Reichman can only show that it is NP-hard to find a solution with value $1 - C\eta$ for some relatively small constant $C > 1$. Thus with current knowledge, given a $(1 - \eta)$-satisfiable instance, we don't know whether one can satisfy almost all the constraints in polynomial time, or whether it is impossible to go beyond a very small fraction of them.

A special case of the MAX-2LIN(q) problem which seems somewhat easier algorithmically ([],[]) occurs when all the equations in the instance are of the form $x_i - x_j = c_{ij}$.

Definition 15 (Γ-MAX-2LIN(q)). Γ-MAX-2LIN(q) *is the special case of MAX-2LIN(q) in which each equation is of the form $x_i - x_j = c_{ij}$.*

Our hardness results will hold even for Γ-MAX-2LIN(q). The Γ notation is essentially from Håstad [31]; we use it because our results actually hold equally well for the problem of satisfying equations of the form $x_i x_j^{-1} = c_{ij}$ over any fixed abelian group Γ of order q, not just $\mathbb{Z}_q$.

11.2 Analytic notions

We would like to generalize our notions of noise stability, influences, and Fourier expansions to q-ary functions, $f : [q]^n \to [q]$. Some of the definitions below were already given in Subsection 4.2, but we repeat them here for clarity and convenience.

The way we treat the finite set $[q]$ in the domain and in the range of q-ary functions will be different. In the domain, $[q]$ and $[q]^n$ will be treated simply as finite probability spaces under the uniform measure, with no extra structure. In the range, we would like to embed $[q]$ into a larger space. Recall that for boolean function we identified the range with the two points $-1, 1 \in \mathbb{R}$ and then considered relaxed functions taking values in their convex hull. In the q-ary case we identify the elements of $[q]$ with the standard basis vectors in $\mathbb{R}^q$. A relaxed q-ary function will thus map $[q]^n$ into the simplex which is the convex hull of these vectors.

Definition 16. *Let Δ_q denote the $(q-1)$-dimensional simplex naturally embedded in $\mathbb{R}^q$; i.e., the convex hull of the q standard basis vectors. We call functions $f : [q]^n \to \Delta_q$* relaxed q-ary functions.

We will also define the notion of a balanced function:

Definition 17. *A function $f : [q]^n \to [q]$ is called* balanced *if it obtains each value $i \in [q]$ in its range equally often. A relaxed function $f : [q]^n \to \Delta_q$ is called* balanced *if $\mathbf{E}[f(x)] = (1/q, \ldots, 1/q)$.*

Since a relaxed q-ary function f maps $[q]^n$ into $\mathbb{R}^q$, it can be viewed as a vector $f = (f_1, \ldots, f_q)$ of real-valued functions over $[q]^n$. We define the noise stability both for real-valued and Δ_q-valued functions.

Definition 18. *Let $-\frac{1}{q-1} \le \rho \le 1$ and let x and y be $[q]^n$-valued random variables. We say that x and y are a ρ-*correlated pair *if x is uniformly distributed on $[q]^n$, and y is formed from x by choosing each y_i so that $\Pr[y_i = a] = \delta_{\{x_i = a\}}\rho + \frac{1-\rho}{q}$ for each a, independently for each i. Note that for $0 \le \rho \le 1$, it is equivalent to say that each coordinate y_i is independently chosen to be x_i with probability ρ and is a uniformly random element of $[q]$ otherwise.*

Definition 19. *Let* $f : [q]^n \to \Delta_q$ *and let* $-\frac{1}{q-1} \le \rho \le 1$*. The* noise stability of f at ρ *is defined to be*

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

where x *and* y *are a* ρ*-correlated pair. Equivalently, we may define the noise stability of functions* $g : [q]^n \to \mathbb{R}$ *via*

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

and then denoting by f^i *the* i*th coordinate projection of* f*, we have* $\mathbb{S}_\rho(f) = \sum_{i=1}^n \mathbb{S}_\rho(f^i)$*.*

We remark that when f's range is simply $[q]$ (as embedded in Δ_q), the quantity $\mathbb{S}_\rho(f)$ is simply the probability that $f(x) = f(y)$ when x and y are a ρ-correlated pair. For example, the noise stability at ρ of a dictator function $f : [q]^n \to [q]$ is equal to $\rho + \frac{1}{q}(1-\rho)$.

The definition of influences is very similar to that in the boolean case:

Definition 20. *Let* $f : [q]^n \to \Delta_q$*. For* $1 \le i \le n$*, the* influence of the ith coordinate on f *is defined to be*

$$\mathrm{Inf}_i(f) = \mathop{\mathbf{E}}_{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n}[\mathrm{Var}_{x_i}[f^i(x_1,\dots,x_n)]].$$

The space X of all functions $f : [q]^n \to \mathbb{R}^d$ (we use either $d = q$ or $d = 1$) is an inner product space with inner product

$$\langle f, g\rangle = \mathbf{E}_x[\langle f(x), g(x)\rangle]$$

and associated norm denoted $\|\cdot\|$. Given $x \in [q]^n$, write x_S for $\{x_i : i \in S\}$. It is well known that X can be written as an orthogonal sum of spaces $X = \oplus_{S\subset[n]} X_S$, where X_S denotes the space of all functions $f : [q]^n \to \mathbb{R}^d$ such that

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

Thus we can write any $f : [q]^n \to \mathbb{R}^q$ as

$$f(x) = \sum_{S\subset[n]} f_S(x), \tag{13}$$

where $f_S(x)$ is the projection of f onto the space X_S. Parseval's identity holds for this expansion:

$$\|f\|_2^2 = \sum_{S\subseteq[n]} \|f_S\|_2^2.$$

For $-\frac{1}{q-1} \le \rho \le 1$ we can define the Bonami-Beckner on X in the obvious way, $T_\rho(f)(x) = \mathbf{E}_y[f(y)]$, where y is ρ-correlated to x. We have that if $f : [q]^n \to \Delta_q$ then $T_\rho f$ also has range Δ_q, and that if f is balanced then so too is $T_\rho f$. We also have that $\mathbb{S}_\rho(f) = \langle f, T_\rho f\rangle$. The formula for noise stability from Proposition 7.1 holds in this setting:

$$\mathbb{S}_\rho(f) = \sum_{S\subseteq[n]} \rho^{|S|}\|f_S\|_2^2; \tag{14}$$

this follows from the following easy proposition, familiar from the boolean case:

Proposition 11.1.

$$T_\rho(f) = \sum_{S \subseteq [n]} \rho^{|S|} f_S.$$

Proof. It is easy to see that for each x, $(T_\rho f)(x)$ is a polynomial in ρ. Therefore it suffices to prove the claim for $0 \le \rho \le 1$. Clearly, T_ρ is linear and therefore it suffices to show that if $f \in X_S$ then $T_\rho f = \rho^{|S|} f$. From the definition of the space X_S it follows that for every subset $S' \subsetneq S$ and for every vector of values z of size $|S'|$ it holds that $\mathbf{E}[f(y) \mid y_{S'} = z] = 0$. Now for $0 \le \rho \le 1$,

$$T_\rho f(x) = \sum_{S' \subset S} \rho^{|S'|}(1-\rho)^{|S|-|S'|} \cdot \mathbf{E}[f(y) \mid y_{S'} = x_{S'}] = \rho^{|S|} f(x),$$

as needed. □

In a similar fashion, it is easy to verify that a formula similar to (5) holds in the q-ary case:

$$\mathrm{Inf}_i(f) = \sum_{S \ni i} \|f_S\|_2^2.$$

Finally, we define low-degree influences as in Definition 11:

Proposition 11.2. *Let $f : [q]^n \to \mathbb{R}^d$ and $k \ge 1$. Then we define*

$$\mathrm{Inf}_i^{\le k}(f) = \sum_{\substack{S \ni i \\ |S| \le k}} \|f_S\|_2^2.$$

11.3 Stability estimates

In this subsection we analyze the plurality function and give estimates on $\Lambda_\rho(\mu)$, proving Theorem 8, Proposition 6.1, and Corollary 10. We also prove some other estimates on $\Lambda_\rho(\mu)$ for very small ρ parameters that are required for our MAX-q-CUT reduction. We begin with the proof of Theorem 8.

Proof of Theorem 8. Suppose we choose $x \in [q]^n$ at random and let y be a ρ-correlated copy of x. For each $i \in [q]$, let u_i denote the number of coordinates in x taking the value i, and let v_i denote the number of coordinates of y taking the value i. We wish the to apply the Central Limit Theorem (specifically, Theorem 16) to calculate the stability of plurality, which is given by

$$q\Pr\Big[u_1 \ge \max_{1\le i\le q} u_i \text{ and } v_1 \ge \max_{1\le i\le q} v_i\Big] = q\Pr\Big[n - \textstyle\sum_{i\ge 2} u_i \ge \displaystyle\max_{2\le i\le q} u_i \text{ and } n - \textstyle\sum_{i\ge 2} v_i \ge \displaystyle\max_{2\le i\le q} v_i\Big].$$

Let us define the following vectors in $\mathbb{R}^{q-1}$: let $e_2, \ldots, e_q$ denote the $q-1$ unit vectors, let r denote the vector $(1/q, \ldots, 1/q)$, let r_1 denote the 0 vector, and let $r_i = e_i - r$ for $i = 2 \ldots q$. Consider random variables $(\mathbf{X}_i, \mathbf{Y}_i)$ taking values in $\mathbb{R}^{2q-2}$ where

$$\Pr[(\mathbf{X}_i, \mathbf{Y}_i) = (r_a, r_b)] = \frac{\rho}{q}\delta_{\{a=b\}} + \frac{1-\rho}{q^2}. \tag{15}$$

Note that $(u_2, \ldots, u_q, v_2, \ldots, v_q) = nr + \sum_{i=1}^n (\mathbf{X}_i, \mathbf{Y}_i)$ where $(\mathbf{X}_i, \mathbf{Y}_i)$ are the i.i.d. random variables given in (15) and that the vectors (X_i, Y_i) have mean 0. Writing A for the $(q-1) \times (q-1)$ matrix given by $A_{i,j} = \delta_{\{i=j\}}/q - 1/q^2$, the covariance matrix V of $(\mathbf{X}_i, \mathbf{Y}_i)$ is given by

$$V = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \otimes A = \begin{pmatrix} A & \rho A \\ \rho A & A \end{pmatrix}.$$

Thus the eigenvalues of V are $1/q, 1/q^2, \rho/q$ and ρ/q^2. Finally, the third norm of $(\mathbf{X}_i, \mathbf{Y}_i)$ is at most 4. We may thus apply Theorem 16 to obtain:

$$\lim_{n\to\infty} \Pr\Big[u_1 \geq \max_{1\leq i\leq q} u_i \text{ and } \quad v_1 \geq \max_{1\leq i\leq q} v_i\Big] = \Pr\Big[-\textstyle\sum_{i=2}^q N_i \geq \max_{2\leq i\leq q} N_i \text{ and } -\sum_{i=2}^q M_i \geq \max_{2\leq i\leq q} M_i\Big]$$

where $(N_i, M_i)_{i=2}^q$ is a normal vector with covariance matrix V. Letting $N_1 = -\sum_{i=2}^q N_i$ and $M_1 = -\sum_{i=2}^q M_i$ we see that $(N_1, \ldots, N_q, M_1, \ldots, M_q)$ is a zero mean normal vector with covariance matrix $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right) \otimes B$, where B is the $q \times q$ matrix given by $B_{i,j} = \delta_{\{i=j\}}/q - 1/q^2$. Finally, let $(U_1, V_1), \ldots, (U_q, V_q)$ be a collection of i.i.d. mean zero normal vectors in $\mathbb{R}^2$, where (U_i, V_i) has the covariance matrix $\left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. It is then easy to see that $(N_q, \ldots, N_q, M_q, \ldots, M_q)$ has the same covariance matrix as the normal vector

$$\sqrt{\frac{q}{q-1}} \left(U_1 - \frac{1}{q}\sum_{j=1}^q U_j, \ldots, U_q - \frac{1}{q}\sum_{j=1}^q U_j, V_1 - \frac{1}{q}\sum_{j=1}^q V_j, \ldots, U_q - \frac{1}{q}\sum_{j=1}^q V_j \right).$$

Since both vectors are normal, they have the same distribution. Hence the stability of Plurality is given by

$$q\Pr\Big[M_1 = \max_{1\leq j\leq q} M_j \text{ and } N_1 = \max_j N_j\Big] = q\Pr\Big[U_1 = \max_{1\leq j\leq q} U_j \text{ and } V_1 = \max_j V_j\Big],$$

and this completes the proof of Theorem 8. □

Let us now move to discussing estimates of $\Lambda_\rho(\mu)$, proving Proposition 6.1.

Proof of Proposition 6.1. The proof of Lemma 11.1 in [12] gives

$$\Lambda_\rho(\mu) = \frac{1}{2\pi\sqrt{1-\rho^2}\cdot t^2} \exp\left(-\frac{t^2}{1+\rho}\right) \int_0^\infty\int_0^\infty \exp(-g(u,v))\, du\, dv, \tag{16}$$

where

$$g(u,v) = \frac{u+v}{1+\rho} + \frac{(u-v)^2 + 2(1-\rho)uv}{2(1-\rho^2)t^2}.$$

Since the range of integration is $u, v \geq 0$ we have $g(u,v) \geq h(u,v)$, where

$$h(u,v) = \frac{u+v}{1+\rho} + \frac{(u-v)^2}{2(1-\rho^2)t^2}.$$

We can therefore replace g by h in the integral and get

$$\int_0^\infty\int_0^\infty \exp(-g(u,v))\, du\, dv \leq \int_0^\infty\int_0^\infty \exp(-h(u,v))\, du\, dv \tag{17}$$

$$= \sqrt{2\pi}(1+\rho)\sqrt{1-\rho^2}\cdot t \cdot \exp\left(\frac{1-\rho}{1+\rho}\cdot\frac{t^2}{2}\right)\cdot N\left(t\sqrt{\tfrac{1-\rho}{1+\rho}}\right), \tag{18}$$

where the integral computation follows straightforwardly after the change of variables $r = u+v$, $s = u-v$. Combining (16) and (17) completes the proof. □

Proposition 6.1 leads to the asymptotic estimates stated in Corollary 10:

Proof of Corollary 10. Part 1 follows simply from the well-known fact that $N(t) \sim \phi(t)/t$ as $t \to \infty$. As a side note, it is simple to see from its definition that $\mu \cdot N(t\sqrt{(1-\rho)/(1+\rho)})$ is a lower bound on $\Lambda_\rho(\mu)$. Part 2 of the Corollary is Lemma 11.1 of de Klerk et al. [12]. Parts 3 and 4 follow straightforwardly from Part 1; the bound written in Part 3 actually neglects an additional negative power of $\ln q$ for simplicity. $\square$

Finally, our hardness result for MAX-q-CUT requires the following more careful analysis of $\Lambda_\rho(\mu)$ in the case that ρ is very small:

Proposition 11.3. *Let $\mu > 0$ be small and let $0 < \rho \le \frac{1}{\ln^3(1/\mu)}$. Then*

$$\Lambda_\rho(\mu) \le \mu \left(\mu + \rho \cdot 2\mu \ln(1/\mu) \cdot (1 + O(\tfrac{\ln\ln(1/\mu)}{\ln(1/\mu)} + \tfrac{\ln\ln(1/\rho)}{\ln(1/\mu)}))\right).$$

Proof. Recall that $\Lambda_\rho(\mu) = \Pr[X \ge t, X' \ge t]$, where X is a standard Gaussian, $X' = \rho X + \sqrt{1-\rho^2}\, Y$ with Y an independent standard Gaussian, and $t = N^{-1}(\mu)$. (We use the functions ϕ and N from Proposition 6.1.) The probability that $X \ge t$ is μ, so we need to show that

$$\Pr[X' \ge t \mid X \ge t] \le \mu + \rho \cdot 2\mu \ln(1/\mu) \cdot (1 + O(\tfrac{\ln\ln(1/\mu)}{\ln(1/\mu)} + \tfrac{\ln\ln(1/\rho)}{\ln(1/\mu)})). \tag{19}$$

Let us first estimate

$$\Pr[X' \ge t \mid X = t(1+\alpha)],$$

for $1/\ln(1/\mu) \le \alpha \le \ln(1/\rho)$. We have

$$\begin{aligned} \Pr[X' \ge t \mid X = t(1+\alpha)] &= \Pr[\rho X + \sqrt{1-\rho^2}\, Y \ge t \mid X = t(1+\alpha)] \\ &\le \Pr[Y \ge (t - \rho t(1+\alpha))/\sqrt{1-\rho^2}] \\ &\le \Pr[Y \ge t - \rho t(1+2\alpha)], \end{aligned}$$

where we have used that $\rho \le 1/\ln(1/\mu) \le \alpha$. Now $\Pr[Y \ge t-\beta] \le \Pr[Y \ge t] + \beta\phi(t-\beta) = \mu + \beta\phi(t-\beta)$. We can upper-bound $\phi(t-\beta)$ by expanding its definition and using the well-known fact $\phi(t) \le tN(t) + O(1/t^2)$ along with $N(t) = \mu$. With our particular $\beta = \rho t(1+2\alpha)$ we get that

$$\phi(t-\beta) \le t\mu(1 + O(1/\ln(1/\mu))),$$

where this also uses $\alpha \le \log(1/\rho)$ and $\rho \le 1/\ln^3(1/\mu)$. We thus conclude that

$$\Pr[X' \ge t \mid X = t(1+\alpha)] \le \mu + \rho \cdot 2\mu\ln(1/\mu) \cdot (1 + O(\alpha)), \tag{20}$$

where we have also used $t \le \sqrt{2\ln(1/\mu)}$ and $\alpha \ge 1/\ln(1/\mu)$.

Our next task is to estimate $\Pr[X \ge t(1+\alpha) \mid X \ge t]$. This is quite straightforward using $N(x) \sim \phi(x)/x$ and $N(t) = \mu$; the result is that

$$\Pr[X \ge t(1+\alpha) \mid X \ge t] \le (C\mu^2 \ln(1/\mu))^\alpha \tag{21}$$

for some universal constant $C < \infty$ (where we used $\alpha \ge 1/\ln(1/\mu)$).

Set $\delta = A(\frac{\ln\ln(1/\mu)}{\ln(1/\mu)} + \frac{\ln\ln(1/\rho)}{\ln(1/\mu)})$ and $K = B\frac{\ln(1/\rho)}{\ln(1/\mu)} > 1$, where A and B are large universal constants to be chosen later. They will be chosen so that $\delta < K$ (this is possible because $\rho \le 1/\ln^3(1/\mu)$). Write $\gamma = \Pr[(1+\delta)t \le X \le (1+K)t \mid X \ge t]$. We estimate

$$\begin{aligned} \Pr[X' \ge t \mid X \ge t] \le{}& (1-\gamma)\Pr[X' \ge t \mid t \le X \le t(1+\delta)] \\ &+ \gamma \Pr[X' \ge t \mid (1+\delta)t \le X \le (1+K)t] \\ &+ \Pr[X > (1+K)t \mid X \ge t]. \end{aligned}$$

Since it's clear that $\Pr[X' \geq t \mid X \in [a,b]] \leq \Pr[X' \geq t \mid X = b]$, we can use (20) to bound the sum of the first two terms by

$$\mu + \rho \cdot 2\mu \ln(1/\mu)(1 + O(\delta)) + \gamma \cdot O(K\rho \cdot \mu \ln(1/\mu)).$$

The third term, and also γ, are bounded using (21). This gives an overall bound of

$$\Pr[X' \geq t \mid X \geq t] \leq \mu + \rho \cdot 2\mu \ln(1/\mu)(1+O(\delta)) + (C\mu^2 \ln(1/\mu))^{\delta} \cdot O(K\rho \cdot \mu \ln(1/\mu)) + (C\mu^2 \ln(1/\mu))^K.$$

It is now relatively easy to check that we can take A and B large enough so that the above quantity is bounded as in (19), completing the proof. (One can take A so that the third term above is smaller than $B \cdot \rho\mu \ln\ln \frac{1}{\rho}$, and then take B large enough so that both $K > \delta$ and the last term is smaller than $\rho \cdot \mu \ln(1/\mu) \cdot \delta$.) $\square$

Proposition 11.3 leads to a *lower bound* on the stability of a q-ary function with noise $\rho = -\frac{1}{q-1}$.

Proposition 11.4. *For any $q \geq 2$ there is a small enough $\delta = \delta(q) > 0$, such that any function $f : [q]^n \to \Delta_q$ with $\mathrm{Inf}_i(f) \leq \delta$ for all $i = 1, \ldots, n$ satisfies*

$$\mathbb{S}_{-\frac{1}{q-1}}(f) \geq 1/q - (2\ln q)/q^2 - C \cdot (\ln\ln q)/q^2,$$

where $C < \infty$ is a universal constant.

Proof. Let $f^i : [q]^n \to [0,1]$ denote the ith coordinate function of f, and let $\mu_i = \mathbf{E}[f^i]$. Then

$$\begin{aligned}
\mathbb{S}_{-\frac{1}{q-1}}(f^i) &= \|f^i_\emptyset\|_2^2 - \tfrac{1}{q-1}\sum_{|S|=1}\|f^i_S\|_2^2 + \tfrac{1}{q-1}\sum_{|S|=2}\|f^i_S\|_2^2 - \cdots \\
&\geq \|f^i_\emptyset\|_2^2 - \tfrac{1}{q-1}\sum_{|S|=1}\|f^i_S\|_2^2 - \tfrac{1}{q-1}\sum_{|S|=2}\|f^i_S\|_2^2 - \cdots \\
&= 2\mu_i^2 - \mathbb{S}_{\frac{1}{q-1}}(f^i).
\end{aligned}$$

Choosing δ to be small enough as function of q, we obtain from the MOO theorem that $\mathbb{S}_{\frac{1}{q-1}}(f^i) \leq \Lambda_{\frac{1}{q-1}}(\mu_i) + \epsilon$, where ϵ is, say, $1/q^3$. It thus suffices to prove that

$$\sum_{i=1}^{q}\left[2\mu_i^2 - \Lambda_{\frac{1}{q-1}}(\mu_i)\right]^+ \geq 1/q - (2\ln q)/q^2 - O(\ln\ln q)/q^2. \tag{22}$$

We prove this using Proposition 11.3 and the fact that $\sum \mu_i = 1$. We will first carry out the estimates assuming that all μ_i's satisfy $\frac{1}{q-1} \leq \frac{1}{\ln^3(1/\mu_i)}$.

Suppose that $\mu_i \geq (1/q)^{1/10}$ for some i. Proposition 11.3 implies in this case that $\Lambda_{\frac{1}{q-1}}(\mu_i) \leq \mu_i^2 + O(\mu^{1/10})$ and so the ith summand already contributes at least $.5\mu_i^2 \geq .5(1/q)^{1/5}$ to the sum in (22), and so the inequality there holds. We may therefore assume that all μ_i's are at most $(1/q)^{1/10}$. With this in hand, Proposition 11.3 tells us that $\Lambda_{\frac{1}{q-1}}(\mu_i) \leq F(\mu_i)$, where

$$F(\mu_i) = \mu_i^2 + \frac{1}{q-1} \cdot 2\mu_i^2 \ln(1/\mu_i) \cdot (1 + C\tfrac{\ln\ln q}{\ln q})$$

and $C < \infty$ is some universal constant. Thus

$$\sum_{i=1}^{q}\left[2\mu_i^2 - \Lambda_{\frac{1}{q-1}}(\mu_i)\right]^+ \geq \sum_{i=1}^{q}(2\mu_i^2 - F(\mu_i)). \tag{23}$$

It's not hard to check that $2\mu_i^2 - F(\mu_i)$ is a convex function of μ_i so long as μ_i is at most a certain universal constant smaller than 1 (which it is when q is sufficiently large, since all μ_i's are at most $(1/q)^{1/10}$). Using $\sum_{i=1}^{q} \mu_i = 1$ we conclude that the right side of (23) is minimized when all μ_i's are equal to $1/q$, in which case it equals $1/q - (2\ln q)/q^2 - O(\ln\ln q)/q^2$; thus (22) is verified.

Finally, we consider the possibility that not all μ_i's satisfy $\frac{1}{q-1} \leq \frac{1}{\ln^3(1/\mu_i)}$. In this case, some of the term-by-term inequalities going into (23) may no longer hold. For such inequalities, though, the left-hand side term is always nonnegative and the right-hand side term is exponentially small in a power of q. Hence (23) still holds up to an additive term exponentially small in q, which is negligible; thus the argument above is unaffected. □

11.4 Hardness results for MAX-q-CUT and Γ-MAX-2LIN(q)

This section is devoted to the proofs of Theorems 11 and 12. The proofs are similar to that of Theorem 1 in Section 8, so we omit some details.

As a preliminary technical step, we need the analogue of Proposition 7.3 for the MOO theorem and for Proposition 11.4.

Proposition 11.5. *Both the MOO theorem and Proposition 11.4 remain 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. (Sketch.) The proof is essentially the same as that of Proposition 7.3; one requires the following facts:

$$\sum_i \mathrm{Inf}_i^{\leq k}(f) \leq k,$$

$$\sum_{|S|>k} \|(T_{1-\gamma} f)_S\|_2^2 \leq (1-\gamma)^{2k},$$

which indeed hold for q-ary functions $f : [q]^n \to [0,1]$, as can easily be seen from the facts in Subsection 7.2. □

We will also need to define the q-ary analogue of the Long Code:

Definition 21 (q-ary Long Code). *The q-ary Long Code of an element* $i \in [M]$ *is the q-ary function* $f : [q]^M \to [q]$ *defined by* $f(x) = x_i$.

Hardness of MAX-q-CUT. We now begin the reduction from Unique Label Cover to MAX-q-CUT. As mentioned, the reduction is similar to the one in Section 8; the main difference is that we use q-ary Long Codes and we fix ρ to be $-\frac{1}{q-1}$. As in Section 8, we start with a given instance $\mathcal{L}(V, W, E, [M], \{\sigma_{v,w}\})$ of Unique Label Cover, and construct an instance of MAX-q-CUT, presented here as a PCP verifier:

The PCP verifier for MAX-q-CUT

- 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_w and $f_{w'}$ be the supposed q-ary Long Codes of the labels of w and w' respectively.
- Pick $(x, y) \in [q]^M$ to be a $(-\frac{1}{q-1})$-correlated pair; in other words, pick $x \in [q]^M$ uniformly at random and form y by choosing each y_i independently to be a uniformly random element of $[q] \setminus \{x_i\}$.

- Accept iff

$$f_w(x \circ \sigma) \neq f_{w'}(y \circ \sigma').$$

Completeness. If the Unique Label Cover instance has a $(1-\eta)$-satisfying assignment, then the PCP verifier accepts the Long Code encoding of this assignment with probability at least $1-2\eta$. This is because whenever the PCP verifier chooses edges (v,w) and (v,w') that are properly labeled, the verifier accepts with probability 1. The Unique Games Conjecture allows us to take η to be an arbitrarily small positive constant.

Soundness. Our goal is to show that if the PCP verifier accepts with probability exceeding $1-1/q+(2\ln q)/q^2 + O(\ln\ln q)/q^2$, then we can derive an assignment for the Unique Label Cover instance that satisfies at least some $\gamma' = \gamma'(q)$ fraction of its edges, independent of the label set size M.

As in Section 8, we analyze the soundness by writing the success probability of the PCP in terms of the noise stability of certain averages of the f_w's. (We view these supposed Long Codes $f_w : [q]^n \to [q]$ as having the relaxed range Δ_q.) If the noise stability is large, the MOO theorem implies the existence of influential coordinates, which in turn are used to derive an assignment for the Unique Label Cover instance.

The probability that the PCP verifier accepts is given by

$$\begin{aligned}
\Pr[\text{acc}] &= \mathop{\mathbf{E}}_{v,w,w',x,y}\big[1 - \langle f_w(x\circ\sigma), f_{w'}(y\circ\sigma')\rangle\big] \\
&= 1 - \mathop{\mathbf{E}}_{v,x,y}\left[\mathop{\mathbf{E}}_{w,w'}[\langle f_w(x\circ\sigma), f_{w'}(y\circ\sigma')\rangle]\right] \\
&= 1 - \mathop{\mathbf{E}}_{v,x,y}\left[\langle \mathop{\mathbf{E}}_{w}[f_w(x\circ\sigma)], \mathop{\mathbf{E}}_{w'}[f_{w'}(y\circ\sigma')]\rangle\right] && \text{(using independence of } w \text{ and } w'\text{)} \\
&= 1 - \mathop{\mathbf{E}}_{v,x,y}\big[\langle g_v(x), g_v(y)\rangle\big] && \text{(where we define } g_v(z) = \mathop{\mathbf{E}}_{w\sim v}[f_w(z\circ\sigma_{v,w})]) \\
&= 1 - \mathop{\mathbf{E}}_{v}\big[\mathbb{S}_{-\frac{1}{q-1}}(g_v)\big].
\end{aligned}$$

We now proceed as in the proof of Theorem 1, using Proposition 11.4 in place of the Majority Is Stablest theorem. In particular, writing $\epsilon = (\ln\ln q)/q^2$, we have that if $\Pr[\text{acc}] \geq 1 - 1/q + (2\ln q)/q^2 + (C+1)\cdot(\ln\ln q)/q^2$ then there is some $\epsilon/2$ fraction of "good" v's with $\mathbb{S}_{-\frac{1}{q-1}}(g_v) \leq 1/q - (2\ln q)/q^2 - (C+1/2)\cdot(\ln\ln q)/q^2$. By Proposition 11.4 such g_v's must have large low-degree influential coordinates, which we can use as Label Cover labels for their v's.

The remainder of the soundness proof is just as it is in the proof of Theorem 1 in Section 8. The only difference arises in the analogue of (6) and is essentially notational; we replace this line with:

$$\delta \leq \sum_{\substack{S\ni j\\ |S|\leq k}} \|(g_v)_S\|_2^2 = \sum_{\substack{S\ni j\\ |S|\leq k}} \|\mathop{\mathbf{E}}_{w}[(f_w)_{\sigma^{-1}(S)}]\|_2^2 \leq \sum_{\substack{S\ni j\\ |S|\leq k}} \mathop{\mathbf{E}}_{w}[\|(f_w)_{\sigma^{-1}(S)}\|_2^2] = \mathop{\mathbf{E}}_{w}\left[\mathrm{Inf}^{\leq k}_{\sigma^{-1}(j)}(f_w)\right].$$

With this proof of soundness in hand, the proof of Theorem 11 is now complete.

Hardness of MAX-2LIN(q). We move on to our hardness result for Γ-MAX-2LIN(q) and the proof of Theorem 12. The proof is very similar to the one we gave for MAX-q-CUT; the only new technique is the use of the old PCP trick of *folding*.

Definition 22 (Additive folding). *Let $f : [q]^M \to [q]$, where the $[q]$ in the domain is viewed as $\mathbb{Z}_q$, the integers mod q. We say that f is* folded *if for every $c \in \mathbb{Z}_q$ and $x \in (\mathbb{Z}_q)^M$ it holds that $f(x+(c,c,\ldots,c)) = f(x)+c$.*

Our PCP verifier for Γ-MAX-2LIN(q) will be able to assume that all the supposed q-ary Long Codes f_w with which it works are folded. This can be done by only making queries $f_w(x)$ when $x_1 = 0$, and simulating other queries using the assumption that the function is folded. In other words, to query $f_w(x_1, \ldots, x_n)$ the verifier instead queries $f(0, x_2 - x_1, \ldots, x_n - x_1)$ and computes the value $f(0, x_2 - x_1, \ldots, x_n - x_1) - x_1$. Note that Long Code functions (dictators) are folded, and that a folded q-ary functions must be balanced.

We now give our verifier for Γ-MAX-2LIN(q), parameterized by $0 < \rho < 1$. Given an instance of Unique Label Cover, it proceeds as follows:

The PCP verifier for Γ-MAX-2LIN(q) with parameter $0 < \rho < 1$

- 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_w and $f_{w'}$ be the *folded* supposed q-ary Long Codes of the labels of w and w' respectively.
- Pick $(x, y) \in [q]^M$ to be a ρ-correlated pair.
- Accept iff $f_w(x \circ \sigma) = f_{w'}(y \circ \sigma')$; i.e., iff $f_w(x \circ \sigma) - f_{w'}(y \circ \sigma') = 0$.

This verifier indeed yields a distribution over 2-variable linear equations mod q of the form "$x_i - x_j = c$"; note that since the verifier ensures the functions f_w are folded, the acceptance predicates $f_w(x \circ \sigma) - f_{w'}(y \circ \sigma') = 0$ will really be of the form $(f_w(x') - x'_1) - (f_{w'}(y') - y'_1) = 0$.

Analysis of this PCP verifier's completeness and soundness proceeds very much as it did in the previous proofs. The completeness is at least $(1 - 2\eta)$ times the noise stability at ρ of a q-ary Long Code function; i.e., $(1 - 2\eta)(\rho + \frac{1}{q}(1 - \rho))$. Soundness is again analyzed by arithmetizing the PCP verifier's acceptance probability, which in this case yields

$$\Pr[\text{acc}] = \mathop{\mathbf{E}}_{v}[\mathbb{S}_\rho(g_v)].$$

The functions $g_v : [q]^M \to \Delta_q$ are balanced, being the averages of folded and thus balanced f_w's. Hence their q projection functions $(g_v)^i : [q]^M \to [0, 1]$ all have mean equal to $\frac{1}{q}$. We may thus use the MOO theorem directly (instead of Proposition 11.4) and bound soundness by $q\Lambda_\rho(\frac{1}{q}) + \epsilon$. This completes the proof of Theorem 12.

12 Acknowledgements

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

References

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

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

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

[4] 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.

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

[6] 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.

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

[8] 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.

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

[10] M. Charikar, K. Makarychev, and Y. Makarychev. Manuscript, 2005.

[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] E. de Klerk, D. Pasechnik, and J. Warners. Approximate graph colouring and MAX-k-CUT algorithms based on the theta function. To appear in *Journal of Combinatorial Optimization*.

[13] I. Dinur, E. Friedgut, G. Kindler, and R. O'Donnell. On the Fourier tails of bounded functions over the discrete cube. Manuscript.

[14] 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.

[15] Feige, Karpinski, and Langberg. Improved approximation of MAX-CUT on graphs of bounded degree. *ALGORITHMS: Journal of Algorithms*, 43, 2002.

[16] 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.

[17] 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.

[18] 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.

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

[20] U. Fiege and D. Reichman. On systems of linear equations with two variables per equation. In *APPROX-RANDOM*, pages 117–127, 2004.

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

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

[23] E. Friedgut, G. Kalai, and A. Naor. Boolean functions whose Fourier transform is concentrated at the first two levels. *Adv. in Appl. Math.*, 29:427–437, 2002.

[24] A. Frieze and M. Jerrum. Improved approximation algorithms for MAX k-CUT and MAX BISECTION. In E. Balas and J. Clausen, editors, *Integer Programming and Combinatorial Optimization*, volume 920, pages 1–13. Springer, 1995.

[25] 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.

[26] M. Goemans and D. Williamson. Approximation algorithms for MAX-3-CUT and other problems via complex semidefinite programming. *Journal of Computing and Sys. Sci.*, 68:442–470, 2004.

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

[28] A. Gupta and K. Talwar. Approximating Unique Games. In *Proc. 17th Ann. ACM-SIAM Symposium on Discrete Algorithms*, 2006.

[29] 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.

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

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

[32] 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.

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

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

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

[36] 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.

[37] 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.

[38] S. Khot and N. Vishnoi. The Unique Games Conjecture, integrality gap for cut problems and embeddability of negative type metrics into l1. In *Proc. 46rd Ann. IEEE Symp. on Foundations of Comp. Sci.*, 2005.

[39] 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.

[40] 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.

[41] 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.

[42] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Proc. 46th Ann. IEEE Symp. on Foundations of Comp. Sci.*, 2005.

[43] 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. *Israel Journal of Mathematics*. To appear.

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

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

[46] 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.

[47] 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.

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

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

[50] L. Trevisan. Approximation algorithms for Unique Games. In *Proc. 46th Ann. IEEE Symp. on Foundations of Comp. Sci.*, 2005.

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

[52] 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.