Electronic Colloquium on Computational Complexity, Report No. 53 (2010)

Hardness of Approximately Solving Linear Equations Over Reals

Subhash Khot *   Dana Moshkovitz †

March 28, 2010

Abstract

In this paper, we consider the problem of approximately solving a system of homogeneous linear equations over reals, where each equation contains at most three variables.

Since the all-zero assignment always satisfies all the equations exactly, we restrict the assignments to be "non-trivial". Here is an informal statement of our result: assuming the Unique Games Conjecture, it is $\mathcal{NP}$-hard to distinguish whether there is a non-trivial assignment that satisfies $1-\delta$ fraction of the equations or every non-trivial assignment fails to satisfy a constant fraction of the equations with a "margin" of $\Omega(\sqrt{\delta})$.

We develop linearity and dictatorship testing procedures for functions $f : \mathbb{R}^n \mapsto \mathbb{R}$ over a Gaussian space, which could be of independent interest.

Our research is motivated by a possible approach to proving the Unique Games Conjecture.

1 Introduction

In this paper, we study the following natural question: given a homogeneous system of linear equations over reals, each equation containing at most three variables (call it 3LIN($\mathbb{R}$)), we seek a non-trivial approximate solution to the system. In the authors' opinion, the question is poorly understood whereas the corresponding question over a finite field, say $GF(2)$, is fairly well understood [Hås01, HK04]. Over a finite field, an equation is either satisfied or not satisfied, whereas over reals, an equation may be approximately satisfied up to a certain margin and we may be interested in the margin.

The main motivation for this research is a possible approach to proving the Unique Games Conjecture. More details appear in Section 1.4. We first describe our result and techniques and compare it with known results.

1.1 Our Result

Fix a parameter $b_0 \geq 1$. Call a 3LIN($\mathbb{R}$) system b_0-regular if every variable appears in the same number of equations, and the absolute values of the coefficients in all the equations are in the range $[\frac{1}{b_0}, b_0]$. Let X denote the set of variables so that an assignment is a map $A : X \mapsto \mathbb{R}$. For an equation $eq : r_1x_1 + r_2x_2 + r_3x_3 = 0$, and an assignment A, the margin of the equation

*khot@cims.nyu.edu. Computer Science Department, Courant Institute of Mathematical Sciences, New York University. Research supported by NSF CAREER grant CCF-0833228, NSF Expeditions grant CCF-0832795, and BSF grant 2008059.

†dana.moshkovitz@gmail.com. School of Mathematics, The Institute for Advanced Study. Research supported by NSF Grant CCF-0832797.

ISSN 1433-8092

(w.r.t. A) is $\mathrm{Margin}(A, eq) \doteq |r_1 A(x_1) + r_2 A(x_2) + r_3 A(x_3)|$. The all-zeroes assignment, $\forall x \in X, A(x) = 0$, satisfies all the equations exactly, i.e. with a zero margin. Therefore, we will be interested only in the "non-trivial" assignments. For now, think of a non-trivial assignment as one where the distribution of its values $\{A(x) | x \in X\}$ is "well-spread". Specifically, we may consider the "Gaussian distributed assignments", for which the set of values $\{A(x) | x \in X\}$ is distributed (essentially) according to a standard Gaussian. Here is an informal statement of our result:

Theorem 1. *(Informal) Assume the Unique Games Conjecture. Then there exist universal constants b_0, c ($b_0 = 2$ works) such that for every $\delta > 0$, given a b_0-regular $3\textsc{Lin}(\mathbb{R})$ system, it is $\mathcal{NP}$-hard to distinguish between:*

- *(YES Case): There is a Gaussian distributed assignment that satisfies $1 - \delta$ fraction of the equations.*
- *(NO Case): For every Gaussian distributed assignment, for at least a fraction c of the equations, the margin is at least $c\sqrt{\delta}$.*

A few remarks are in order. Since the $3\textsc{Lin}(\mathbb{R})$ instance is finite, we cannot expect the set of values $\{A(x) | x \in X\}$ to be exactly Gaussian distributed. The proof of our result proceeds by constructing a probabilistically checkable proof (PCP) over a continuous high-dimensional Gaussian space and then this "idealized" instance is discretized to obtain a finite instance. Theorem 1 holds in the idealized setting. The discretization step introduces, in the YES Case, a margin of at most γ in each equation, but γ can be made arbitrarily small relative to δ and hence this issue may be safely ignored. The distribution of values is still "close" to a standard Gaussian. We also set all variables with values larger than $O(\log(1/\delta))$ to zero. This applies to only $\mathrm{poly}(\delta)$ fraction of the variables and hence does not have any significant effect on the result. Thus our assignment, in the YES Case, satisfies in particular:

$$\forall x \in X, \ |A(x)| \le b = O(\log(1/\delta)), \qquad \mathop{\mathbf{E}}_{x \in X}\left[A(x)^2\right] = 1. \tag{1}$$

In the NO Case, our analysis extends to every assignment that satisfies (1), and the conclusion is appropriately modified (which is necessary since an assignment that satisfies (1) could still have a very skewed distribution of its values). A formal statement of the result appears as Theorem 6 in Section 2.

1.2 Optimality of Our Result, Squared-ℓ_2 versus ℓ_1 Error, and Homogeneity

Optimality: The result of Theorem 1 is qualitatively almost optimal as can be seen from a natural semi-definite programming relaxation and a rounding algorithm. Suppose there are N variables $X = \{x_1, \ldots, x_N\}$, m equations and j^{th} equation in the system is

$$r_{j1} x_{j_1} + r_{j2} x_{j_2} + r_{j3} x_{j_3} = 0.$$

Consider the following SDP relaxation where for every variable x_i, we have a vector v_i and $b = O(\log(1/\delta))$:

$$\begin{aligned} \text{Minimize} \quad & \mathbf{E}_{j \in [m]}\left[\|r_{j1} v_{j_1} + r_{j2} v_{j_2} + r_{j3} v_{j_3}\|^2\right], \\ \text{Such that} \quad & \\ & \forall x_i \in X, \ \|v_i\| \le b, \\ & \mathbf{E}_{x_i \in X}\left[\|v_i\|^2\right] = 1. \end{aligned}$$

Suppose that in the YES Case, there is an assignment A that satisfies (1) and satisfies $1-\delta$ fraction of the equations exactly. Then letting $v_i = A(x_i)v_0$ for some fixed unit vector v_0 gives a feasible solution to the SDP with the objective $O(\delta \log^2(1/\delta))$. Hence the SDP finds a feasible vector solution with the same upper bound on the objective. Suppose the SDP vectors lie in d-dimensional Euclidean space. Consider a rounding that picks a standard d-dimensional Gaussian vector r and defines an assignment $A(x_i) = \langle v_i, r\rangle$. It is easily seen that after a suitable scaling, with constant probability over the rounding scheme, we have:

$$\mathop{\mathbf{E}}_{x_i \in X}\left[A(x_i)^2\right] = 1, \quad \mathop{\mathbf{E}}_{j\in[m]}\left[|r_{j1}A(x_{j_1}) + r_{j2}A(x_{j_2}) + r_{j3}A(x_{j_3})|^2\right] \leq O(\delta \log^2(1/\delta)).$$

Thus the margin $|r_{j1}A(x_{j_1}) + r_{j2}A(x_{j_2}) + r_{j3}A(x_{j_3})|$ is at most $O(\sqrt{\delta}\log(1/\delta))$ for almost all, say 99%, of the equations. Moreover, since $\forall x_i \in X, \|v_i\| \leq b$, after rounding all but $\text{poly}(\delta)$ fraction of the variables get values bounded by $O(\log^2(1/\delta))$, and these variables can be set to zero without affecting the solution significantly.

The Squared-ℓ_2 versus ℓ_1 Error: The SDP algorithm described above finds an assignment that minimizes the expected squared margin, i.e. $\mathbf{E}_{j\in[m]}\left[\text{Margin}(A,j)^2\right]$. Thus the problem of minimizing the squared-ℓ_2 error is a computationally easy problem. However, Theorem 1 implies that modulo the UGC, minimizing the ℓ_1 error (i.e. $\mathbf{E}_{j\in[m]}\left[\text{Margin}(A,j)\right]$), even approximately, is computationally hard. In the YES Case therein, all but δ fraction of the equations are exactly satisfied, and the variables are bounded by $O(\log(1/\delta))$. Hence the ℓ_1 error is $O(\delta\log(1/\delta))$.[1] In the NO Case, for any Gaussian distributed assignment, for at least a constant fraction of the equations, the margin is at least $\Omega(\sqrt{\delta})$, and hence the ℓ_1 error is $\Omega(\sqrt{\delta})$. Thus approximating the ℓ_1 error within a quadratic factor is computationally hard (modulo UGC); this is optimal since the squared-ℓ_2 minimization implies an ℓ_1 approximation within a quadratic factor.

Homogeneity: Theorem 1 holds for a system of linear equations that is homogeneous and it is necessary therein (in the NO Case) to restrict the distribution of values of an assignment. When the system of equations is non-homogeneous, one might hope to drop the restriction on the distribution of values. However, then a simple LP can directly minimize the ℓ_1 error and hence one cannot hope for a theorem analogous to Theorem 1.

1.3 Techniques

Similar to most of the UGC-based hardness results, our result proceeds by developing an appropriate "dictatorship test". However, unlike most previous applications that use a dictatorship test over an n-dimensional boolean hypercube (or k-ary hypercube in some cases), we develop a dictatorship test over $\mathbb{R}^n$ with the standard Gaussian measure. The test is quite natural, but its analysis turns out to be rather delicate. We think that the test itself is of independent interest and provide its high level overview here.

Let $\mathcal{N}^n$ denote the n-dimensional Gaussian distribution with n independent mean 0 and variance 1 coordinates. Let $L^2(\mathbb{R}^n, \mathcal{N}^n)$ be the space of all measurable real functions $f : \mathbb{R}^n \to \mathbb{R}$ with $\|f\|_2^2 = \mathbf{E}_{x\sim\mathcal{N}^n}\left[f(x)^2\right] < \infty$. This is an inner product space with the inner product $\langle f, g\rangle \doteq \mathbf{E}_{x\sim\mathcal{N}^n}\left[f(x)g(x)\right]$.

A dictatorship is a function $f(x) = x_{i_0}$ for some fixed coordinate $i_0 \in [n]$. Given oracle access to a function $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$, we desire a probabilistic homogeneous linear test that accesses

[1] A closer examination of the proof of Theorem 1 shows that the upper bound is actually $O(\delta)$; for the equations that are not satisfied, the margin itself is distributed according to a standard Gaussian.

at most three values of f. The tests, over all choices of randomness, can be written down as a system of homogeneous linear equations over the values of f. We assume that the function f is non-trivial, i.e. $\|f\|_2^2 = 1$, and anti-symmetric, i.e. $f(-x) = -f(x)\ \forall x \in \mathbb{R}^n$. In particular, $\mathbf{E}[f] = 0$. We desire a test such that a dictatorship function is a "good" solution to the system of linear equations, whereas a function that is *far from a dictatorship*, is a "bad" solution to the system. The test we propose is a combination of a *linearity test* and a *coordinate-wise perturbation test*. A dictatorship function satisfies all the equations of the linearity test and $1-\delta$ fraction of the equations of the *coordinate-wise perturbation test*. A function that is *far from a dictatorship*, either fails "miserably" on the linearity test, or a constant fraction of the equations have a margin $\Omega(\sqrt{\delta})$ on the *coordinate-wise perturbation test*. This directly translates to the hardness gap of Theorem 1 via a UGC-based reduction (the reduction is standard, but does involve some subtle points in our context).

One starts out by observing that a dictatorship function is linear. Thus, for any $\lambda, \mu \in \mathbb{R}$ such that $\lambda^2 + \mu^2 = 1$, say $\lambda = \mu = \frac{1}{\sqrt{2}}$, one can test whether

$$f(\lambda x + \mu y) = \lambda x + \mu y,$$

where $x, y \sim \mathcal{N}^n$ are picked independently. Clearly, a dictatorship function satisfies each such equation exactly. The condition $\lambda^2 + \mu^2 = 1$ ensures that the query point $\lambda x + \mu y$ is also distributed according to $\mathcal{N}^n$. Note that we assume $\|f\|_2^2 = 1$ and $\mathbf{E}[f] = 0$. Functions in $L^2(\mathbb{R}^n, \mathcal{N}^n)$ have the Hermite representation; in particular, f can be decomposed into the linear and non-linear parts:

$$f = f^{=1} + e, \quad f^{=1} = \sum_{i=1}^{n} a_i x_i, \quad \langle f^{=1}, e \rangle = 0.$$

Note that $1 = \|f\|_2^2 = \|f^{=1}\|_2^2 + \|e\|_2^2$. A simple Fourier analytic argument shows that unless $\|e\|_2^2 \le 0.01$, the linearity test fails with "large" average squared margin (and the analysis of the test is over). Therefore we may assume that $\|e\|_2^2 \le 0.01$.

Assume for now, that $e \equiv 0$ and hence the function is linear: $f = f^{=1} = \sum_{i=1}^n a_i x_i$ and $\sum_{i=1}^n a_i^2 = 1$. We introduce the *coordinate-wise perturbation test* to ensure that the coefficients $\{a_i\}_{i=1}^n$ are concentrated on a bounded set. This makes sense because for a dictatorship function, there is exactly one non-zero coefficient. The test picks a random point $x \in \mathcal{N}^n$ and for a randomly chosen δ fraction of the coordinates, each chosen coordinate is re-sampled independently from a standard Gaussian. If $\tilde{x}$ is the new point, then one tests whether

$$f(\tilde{x}) - f(x) = 0.$$

Note that for a dictatorship function, the above equation is satisfied with probability $1-\delta$, whereas with probability δ, the margin is distributed as a mean-0 variance-$\sqrt{2}$ Gaussian. On the other hand, if $f = \sum_{i=1}^n a_i x_i$ is *far from a dictatorship*, then coefficients $\{a_i\}_{i=1}^n$ are "spread-out", and with a constant probability, the margin is $\Omega(\sqrt{\delta})$. This is intuitively the idea behind the test; however the presence of the non-linear part e complicates matters considerably. Even though $\|e\|_2^2 \le 0.01$, we are dealing with margins of the order of $\sqrt{\delta}$, and the non-linear part e could potentially interfere with the above simplistic argument. We therefore need a more refined argument. We observe that since $f = f^{=1} + e$,

$$f(\tilde{x}) - f(x) = (f^{=1}(\tilde{x}) - f^{=1}(x)) + (e(\tilde{x}) - e(x)).$$

When $f^{=1} = \sum_{i=1}^{n} a_i x_i$ is "spread-out", the first term in the above equation, namely $f^{=1}(\tilde{x}) - f^{=1}(x)$, is $\Omega(\sqrt{\delta})$ with a constant probability as we observed above. The same can be concluded about the left hand side of the equation, namely $f(\tilde{x}) - f(x)$, unless the second term $e(\tilde{x}) - e(x)$ "interferes" in a very correlated manner. If this happens, then the function e must be "sensitive" to noise along a random set of δn coordinates. We add a test ensuring that e is "insensitive" to noise of comparable magnitude in a random direction. We then show that the two behaviors are contradictory, using a Fourier analytic argument that relies, in addition, on the cut-decomposition of line/ℓ_1 metrics.

1.4 Comparison with Known Results and Motivation for Studying 3LIN($\mathbb{R}$)

MinUncut: Given a graph $G(V = [N], E)$, the MinUncut problem seeks a cut in the graph that minimizes the number of edges not cut. It can be thought of as an instance of 2LIN($\mathbb{R}$) where one has variables $\{x_1, \ldots, x_N\}$, and for every edge $(i,j) \in E$, a homogeneous equation:

$$x_i + x_j = 0,$$

and the goal is to find a boolean, i.e. $\{-1,1\}$-valued assignment that minimizes the number of unsatisfied equations. Khot *et al* [KKMO07] show that assuming the UGC, for sufficiently small $\delta > 0$, given an instance that has an assignment that satisfies all but δ fraction of the equations, it is $\mathcal{NP}$-hard to find an assignment that satisfies all but $\frac{2}{\pi}\sqrt{\delta}$ fraction of the equations. This result is qualitatively similar to Theorem 1, but note that the variables are restricted to be boolean.

Balanced Partitioning: Given a graph $G(V = [N], E)$, the Balanced Partitioning problem seeks a roughly balanced cut (i.e. each side has $\Omega(N)$ vertices) in the graph that minimizes the number of edges cut. It can again be thought of as an instance of 2LIN($\mathbb{R}$) where one has variables $\{x_1, \ldots, x_N\}$, and for every edge $(i,j) \in E$, a homogeneous equation:

$$x_i - x_j = 0, \tag{2}$$

and the goal is to find a $\{-1,1\}$-valued and roughly balanced assignment that minimizes the number of unsatisfied equations. Arora *et al* [AKK+08] show that assuming a certain variant of the UGC, given an instance of Balanced Partitioning that has a balanced assignment that satisfies all but δ fraction of the equations, it is $\mathcal{NP}$-hard to find a roughly balanced assignment that satisfies all but δ^c fraction of the equations. Here $\frac{1}{2} < c < 1$ is an arbitrary constant and for every such c, the result holds for all sufficiently small $\delta > 0$. The result is again qualitatively similar to Theorem 1. In fact, the result holds even when the variables are allowed to be real valued, say in the range $[-1,1]$, as long as the set of values is "well-separated". Imagine picking a random $\lambda \in [-1,1]$ and partitioning the variables (i.e. vertices of the graph) into two sets depending on whether their value is less or greater than λ. The cut is roughly balanced if the set of values is well-separated, and the probability that an edge $(i,j) \in E$ is cut is $\frac{|x_i - x_j|}{2}$. Thus solving the 2LIN($\mathbb{R}$) instance w.r.t. ℓ_1 error is equivalent to solving the Balanced Partitioning problem.

Motivation for Studying 3LIN($\mathbb{R}$): The hardness results for the MinUncut and the Balanced Partitioning problem cited above are known only assuming the UGC. It would be a huge progress to prove these results without relying on the UGC and could possibly lead to a proof of the UGC itself. Due to the close connection of both the problems to the 2LIN($\mathbb{R}$) problem, it is natural to seek a hardness result for the 2Lin($\mathbb{R}$) problem w.r.t. the ℓ_1 error. This is the main motivation behind the work in this paper. We propose that understanding the complexity of the 3Lin($\mathbb{R}$) problem might help us make progress on the UGC: the plan would be to (1) prove Theorem 1, or perhaps a weaker form of it, without relying on the UGC and then (2) give a gap-preserving reduction from the 3Lin($\mathbb{R}$) to 2Lin($\mathbb{R}$). The first step might be doable since a hardness result for 3Lin($\mathbb{R}$) amounts to constructing a 3-query PCP, and in general 3-query PCPs are quite powerful. In particular, one is allowed to do a 3-query linearity test, which could be useful. On the negative side, PCPs over reals seem to present new difficulties, e.g. in the Gaussian space, the analogue of error correction for Hadamard Codes does not seem to work. On the positive side, the authors, in a follow-up work, are able to prove a weak hardness result for the Lin($\mathbb{R}$) problem (homogeneous system of linear equations with *unbounded* number of variables in each equation) w.r.t. the ℓ_1 error. Regarding the second step, the authors currently have a candidate reduction from 3Lin($\mathbb{R}$) to 2Lin($\mathbb{R}$) along with counterexamples showing that the reduction, as is, does not work. The authors believe that there might be a way to fix the reduction.

Guruswami and Raghavendra's Result: Our result is incomparable to that in [GR09]. Their result shows that given a system of non-homogeneous linear equations over integers (as well as over reals), with three variables in each equation, it is $\mathcal{NP}$-hard to distinguish $1-\delta$ satisfiable instances from δ satisfiable instances. The instance produced by their reduction is non-homogeneous, a good solution in the YES Case consists of large (unbounded) integer values, the result is very much about exactly satisfying equations, and in particular does not give, if any, a strong gap in terms of margins, especially relative to the magnitude of integers in a good solution.

Comparison with Results over $GF(2)$: We argue that, in order to make progress on Min-Uncut, Balanced Partitioning and UGC, studying equations over reals may be the "right" thing to do, as opposed to equations over $GF(2)$. As we discussed before, the Balanced Partitioning problem can be thought of as an instance of 2Lin($\mathbb{R}$) (as in Equation (2)) where one seeks to minimize ℓ_1 error and the set of values is a well-separated set in $[-1,1]$. Assuming a UGC variant, we know that (δ,δ^c)-gap is $\mathcal{NP}$-hard for $c>\frac{1}{2}$, whereas Theorem 1 yields a similar gap for 3Lin($\mathbb{R}$), with a stronger conclusion that a constant fraction of equations have a margin at least $\Omega(\sqrt{\delta})$. We pointed out that such a gap is also the best one may hope for. Thus the 3-variable case seems qualitatively similar to the 2-variable case in terms of hardness gap that may be expected. For equations over $GF(2)$, the two cases are qualitatively very different. Suppose one thinks of the Balanced Partitioning problem as an instance of 2Lin($GF(2)$) where a cut is a $GF(2)$ valued balanced assignment, and one introduces an equation $x_i \oplus x_j = 0$ for each edge (i,j). Its generalization to homogeneous equations with three variables, namely 3Lin($GF(2)$), turns out to be qualitatively very different. Holmerin and Khot [HK04] show a hardness gap (in terms of fraction of equations left unsatisfied by a balanced assignment) of $(\delta,\approx\frac{1}{2})$ which is qualitatively very different from the (δ,δ^c) gap that may be expected for 2Lin($GF(2)$).

1.5 Overview of the Paper

In Section 2, we formally state our main result (Theorem 6) and provide preliminaries on Hermite representation of functions in $L^2(\mathbb{R}^n,\mathcal{N}^n)$. In Section 3, we propose and analyze the *linearity test* that is used as a sub-routine in the *dictatorship test* proposed and analyzed in Section 4. The UGC-based reduction, proving our main result, is presented in Section 6. The reduction is presented first in a continuous setting and then discretized in Section 6.3.

2 Problem Definition, Our Result, and Preliminaries

We consider the problem of approximately solving a system of homogeneous linear equations over the reals. Each equation depends on (at most) three variables. The system of equations is given by a distribution over equations, meaning different equations receive different "weights".

Definition 2 (Robust-3Lin($\mathbb{R}$) instance)**.** *Let $b_0 \geq 1$ be a parameter. A* Robust-3Lin($\mathbb{R}$) *instance is given by a set of real variables X and a distribution $\mathcal{E}$ over equations on the variables. Each equation is of the form:*

$$r_1x_1 + r_2x_2 + r_3x_3 = 0,$$

where the coefficients satisfy $|r_1|, |r_2|, |r_3| \in [\frac{1}{b_0}, b_0]$ *and* $x_1, x_2, x_3 \in X$.

Definition 3 (Assignment to Robust-3Lin($\mathbb{R}$) instance)**.** *An assignment to the variables of a* Robust-3Lin($\mathbb{R}$) *instance $(X, \mathcal{E})$ is a function $A : X \to \mathbb{R}$. An equation $r_1x_1 + r_2x_2 + r_3x_3 = 0$ is* exactly satisfied *by A if*

$$r_1A(x_1) + r_2A(x_2) + r_3A(x_3) = 0.$$

*The equation is β-*approximately satisfied *for an approximation parameter β, if*

$$|r_1A(x_1) + r_2A(x_2) + r_3A(x_3)| \leq \beta.$$

Notation. The set of variables appearing in an equation $eq : r_1x_1 + r_2x_2 + r_3x_3 = 0$ is denoted as $X_{eq} = \{x_1, x_2, x_3\}$. The assignment A will usually be clear from the context. We use the shorthand $|eq|$ to denote the *margin* $|r_1A(x_1) + r_2A(x_2) + r_3A(x_3)|$.

An assignment that assigns 0 to all variables trivially exactly satisfies all equations. Hence, we use a measure for how different the assignment is from the all-zero assignment, locally (per equation) and globally (on average over all equations):

Definition 4 (Assignment norm)**.** *Let $(X, \mathcal{E})$ be a* Robust-3Lin($\mathbb{R}$) *instance. Let $A : X \to \mathbb{R}$ be an assignment. Define the squared norm of A at equation eq to be:*

$$\|A_{eq}\|_2^2 = \mathop{\mathbf{E}}_{x \in X_{eq}} \left[A(x)^2\right].$$

Define the squared norm of A to be:

$$\|A\|_2^2 = \mathop{\mathbf{E}}_{eq \sim \mathcal{E}} \left[\|A_{eq}\|_2^2\right].$$

Remark 2.1. *We will sometimes refer to a distribution on the set of variables X induced by first picking an equation from the distribution $\mathcal{E}$ and then picking a variable at random from that equation. If $\mathcal{D}$ denotes this distribution on variables, then clearly $\|A\|_2^2 = \mathbf{E}_{x \in \mathcal{D}} \left[A(x)^2\right]$.*

Legitimate assignments A are required to be normalized $\|A\|_2^2 = 1$ and bounded $A : X \to [-b, b]$ for some parameter b. We seek to maximize:

$$val^{\beta}_{(X,\mathcal{E})}(A) \doteq \mathop{\mathbf{E}}_{eq \sim \mathcal{E}} \left[\chi_{|eq| \leq \beta} \cdot \|A_{eq}\|_2^2\right], \tag{3}$$

where $\chi_{|eq| \leq \beta}$ is indicator function of the event that $|eq| \leq \beta$. In words, we seek to maximize[2] the total squared norm of equations that are satisfied with margin of at most β.

[2] We recommend that the reader takes a pause and convinces himself/herself that this is a reasonable measure of how good an assignment is. Since an assignment may be very skewed, assigning large values to a tiny subset of variables and zero to the rest of the variables, simply maximizing the fraction of equations satisfied does not make much sense.

Definition 5 (ROBUST-3LIN($\mathbb{R}$) problem)**.** *Let $b_0 \geq 1, b \geq 0$ and $0 < \beta < 1$ be parameters. Given a* ROBUST-3LIN($\mathbb{R}$) *instance where the coefficients are in $[\frac{1}{b_0}, b_0]$ in magnitude, the problem is to find an assignment $A : X \to [-b, b]$ of norm $\|A\|_2^2 = 1$ that maximizes $val^{\beta}_{(X,\mathcal{E})}(A)$.*

We are now ready to formally state our result:

Theorem 6 (Hardness of ROBUST-3LIN($\mathbb{R}$))**.** *Assume the Unique Games Conjecture. There exist universal constants $b_0 = 2$ and $c, s > 0$, such that for any $\gamma, \delta > 0$, there is $b = O(\log(1/\delta))$, such that given an instance $(X, \mathcal{E})$ of* ROBUST-3LIN($\mathbb{R}$) *with the magnitude of the coefficients in $[\frac{1}{b_0}, b_0]$, it is $\mathcal{NP}$-hard to distinguish between the following two cases:*

- Completeness: *There is an assignment $A : X \to [-b, b]$ with $\|A\|_2^2 = 1$, such that*

$$val^{\gamma}_{(X,\mathcal{E})}(A) \geq 1 - \delta.$$

- Soundness: *For any assignment $A : X \to [-b, b]$ with $\|A\|_2^2 = 1$, it holds that*

$$val^{c\sqrt{\delta}}_{(X,\mathcal{E})}(A) \leq 1 - s.$$

We note three points: (1) The parameter γ is to be thought of as negligible compared to δ and essentially equal to 0. Our reduction is best thought of as a continuous construction on a Gaussian space, and the parameter γ arises as a negligible error involved in discretization of the construction. (2) In the YES Case, we can say more about how the "good" assignment looks like. Consider the distribution $\mathcal{D}$ induced on variables by first picking an equation $eq \in \mathcal{E}$ and then picking one of the variables in the equation. The values taken by the good assignment, w.r.t. $\mathcal{D}$, are distributed (essentially) as a standard Gaussian, and can be truncated to $b = O(\log(1/\delta))$ in magnitude without affecting the result. (3) In the NO Case, if an assignment has either values bounded in $[-1, 1]$ or values distributed, w.r.t. $\mathcal{D}$, (essentially) as a standard Gaussian, it is indeed the case that a *constant fraction* of the equations fail with a margin of at least $c\sqrt{\delta}$, proving informal Theorem 1.

2.1 Fourier Analysis Over Gaussian Space

Gaussian Space. Let $\mathcal{N}^n$ denote the n-dimensional Gaussian distribution with n independent mean-0 and variance-1 coordinates. $L^2(\mathbb{R}^n, \mathcal{N}^n)$ is the space of all real functions $f : \mathbb{R}^n \to \mathbb{R}$ with $\mathbf{E}_{x\sim\mathcal{N}^n}\left[f(x)^2\right] < \infty$. This is an inner product space with inner product

$$\langle f, g\rangle \doteq \mathop{\mathbf{E}}_{x\sim\mathcal{N}^n}[f(x)g(x)].$$

Hermite Polynomials. For a natural number j, the j'th *Hermite polynomial* $H_j : \mathbb{R} \to \mathbb{R}$ is

$$H_j(x) = \frac{1}{\sqrt{j!}} \cdot (-1)^j e^{x^2/2} \frac{d^j}{dx^j} e^{-x^2/2}.$$

The first few Hermite polynomials are $H_0 \equiv 1$, $H_1(x) = x$, $H_2(x) = \frac{1}{\sqrt{2}} \cdot (x^2 - 1)$, $H_3 = \frac{1}{\sqrt{6}} \cdot (x^3 - 3x)$, $H_4(x) = \frac{1}{2\sqrt{6}} \cdot (x^4 - 6x^2 + 3)$. The Hermite polynomials satisfy:

Claim 2.1 (Orthonormality)**.** *For every j, $\langle H_j, H_j\rangle = 1$. For every $i \neq j$, $\langle H_i, H_j\rangle = 0$. In particular, for every $j \geq 1$, $\mathbf{E}_{x\in\mathcal{N}}[H_j(x)] = 0$.*

Claim 2.2 (Addition formula)**.**

$$H_j\left(\frac{x+y}{\sqrt{2}}\right) = \frac{1}{2^{j/2}} \cdot \sum_{k=0}^{j} \sqrt{\binom{j}{k}} H_k(x) H_{j-k}(y).$$

Fourier Analysis. The multi-dimensional Hermite polynomials defined as:

$$H_{j_1,\ldots,j_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} H_{j_i}(x_i),$$

form an orthonormal basis for the space $L^2(\mathbb{R}^n, \mathcal{N}^n)$. Every function $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$ can be written as

$$f(x) = \sum_{S \in \mathbb{N}^n} \hat{f}(S)\ H_S(x),$$

where S is multi-index, i.e. an n-tuple of natural numbers, and the $\hat{f}(S) \in \mathbb{R}$ are the Fourier coefficients of f. The size of a multi-index $S = (S_1, \ldots, S_n)$ is defined as $|S| = \sum_{i=1}^{n} S_i$. The Fourier expansion of degree d is $f^{\leq d} = \sum_{|S| \leq d} \hat{f}(S) H_S(x)$, and it holds that

$$\lim_{d \to \infty} \|f - f^{\leq d}\|_2^2 = 0.$$

The *linear part* of f is $f^{=1} = f^{\leq 1} - f^{\leq 0}$. When f is anti-symmetric, i.e. $\forall x \in \mathbb{R}^n, f(-x) = -f(x)$, we have $\widehat{f}(\vec{0}) = \mathbf{E}[f] = 0$ and $f^{\leq 0} \equiv 0$.

Influence. We denote the restriction of a Gaussian variable $x \sim \mathcal{N}^n$ to a set of coordinates $D \subseteq [n]$ by $x_{|D}$. The *influence* of a set of coordinates $D \subseteq [n]$ on a function $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$ is

$$I_D(f) \doteq \mathop{\mathbf{E}}_{x_{|\overline{D}}}\left[\mathop{\mathbf{Var}}_{x_{|D}}[f(x)]\right].$$

The influence can also be expressed in terms of Fourier spectrum of f:

Proposition 2.3.

$$I_D(f) = \sum_{S \cap D \neq \phi} \hat{f}(S)^2,$$

where $S \cap D \neq \phi$ denotes that there exists $i \in D$ such that $S_i \neq 0$. Note that $S \in \mathbb{N}^n$ is a multi-index and $D \subseteq [n]$ is a subset of coordinates.

Perturbation Operator. The perturbation operator (more commonly known as the Ornstein-Uhlenbeck operator) T_ρ takes a function $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$ and produces a function $T_\rho f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$ that averages the value of f over local neighborhoods:

$$T_\rho f(x) = \mathop{\mathbf{E}}_{y \in \mathcal{N}^n}\left[f(\rho x + \sqrt{1-\rho^2}\, y)\right].$$

The Fourier spectrum of $T_\rho f$ can be obtained from the Fourier spectrum of f as follows:

Proposition 2.4.

$$T_\rho f = \sum_S \rho^{|S|} \hat{f}(S)\ H_S.$$

2.2 The Unique Games Conjecture

We prove a hardness result for ROBUST-3LIN($\mathbb{R}$) under the Unique Games Conjecture of Khot [Kho02], i.e., assuming the hardness of approximating the UNIQUE-GAME problem:

Definition 7 (UNIQUE-GAME)**.** *The input to the problem is a regular graph $G = (V, E)$, a number k, and permutations on the edges $\pi_e : [k] \to [k]$. A* labeling *for the graph is a function $\varphi : V \to [k]$. An edge $e = (u, v) \in E$ is* satisfied *by labeling φ if $\pi_e(\varphi(u)) = \varphi(v)$. The task is to find a labeling $\varphi : V \to [k]$ that satisfies as many of the edges as possible.*

Conjecture 2.1 (Unique Games Conjecture [Kho02])**.** *For any constants $0 < \eta, \varepsilon < 1$ there exists $k = k(\eta, \varepsilon)$, such that given a* UNIQUE-GAME *instance on k labels, it is $\mathcal{NP}$-hard to distinguish between the case that $1 - \eta$ fraction of the edges can be satisfied and the case where only fraction ε of the edges can be satisfied.*

3 Linearity Testing

We show how to perform linearity testing for functions in $L^2(\mathbb{R}^n, \mathcal{N}^n)$ using linear equations on three variables each. Linear functions always exactly satisfy the linear equations. Functions with a large non-linear part give rise to heavy margins in the equations.

The linearity test we show resembles linearity testing in finite fields (see, e.g., [BLR93, BCH$^+$96]). We change it slightly so as to guarantee that all the queries to the function are distributed according to the Gaussian distribution.

Linearity Test:

Given oracle access to a function $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$, f anti-symmetric, i.e., $f(-x) = -f(x)$ for every $x \in \mathbb{R}^n$. Pick $x, y \sim \mathcal{N}^n$ and test:

$$f(x) + f(y) + \sqrt{2} \cdot f\left(-\frac{x+y}{\sqrt{2}}\right) = 0.$$

Note that a linear function always exactly satisfies the test's equation. The following lemma shows that if the test's equations are approximately satisfied, then the weight of f's non-linear part is small:

Lemma 3.1 (Linearity testing)**.** *Let $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$, f anti-symmetric, i.e., $f(-x) = -f(x)$ for every $x \in \mathbb{R}^n$. Then*

$$\|f - f^{=1}\|_2^2 \;\leq\; \mathop{\mathbf{E}}_{x,y\sim\mathcal{N}^n}\left[\left|f(x) + f(y) + \sqrt{2} \cdot f\left(-\frac{x+y}{\sqrt{2}}\right)\right|^2\right].$$

Proof. Since x and y are independent, the variables x, y and $-\frac{x+y}{\sqrt{2}}$ are all distributed according to $\mathcal{N}^n$. Also f is anti-symmetric. Hence,

$$\mathop{\mathbf{E}}_{x,y\sim\mathcal{N}^n}\left[\left|f(x) + f(y) + \sqrt{2} \cdot f\left(-\frac{x+y}{\sqrt{2}}\right)\right|^2\right] \;=\; 4\|f\|_2^2 - 4 \cdot \sqrt{2} \cdot \mathop{\mathbf{E}}_{x,y}\left[f(x) f\left(\frac{x+y}{\sqrt{2}}\right)\right]. \quad (4)$$

Writing in terms of the Fourier representation:

$$
\begin{aligned}
\mathop{\mathbf{E}}_{x,y}\left[f(x)f\left(\frac{x+y}{\sqrt{2}}\right)\right] &= \mathop{\mathbf{E}}_{x,y}\left[\sum_{S,T\in\mathbb{N}^n}\hat{f}(S)\hat{f}(T)H_S(x)H_T\left(\frac{x+y}{\sqrt{2}}\right)\right] \\
&= \sum_{S,T}\hat{f}(S)\hat{f}(T)\mathop{\mathbf{E}}_{x,y}\left[\prod_{i=1}^{n}H_{S_i}(x_i)H_{T_i}\left(\frac{x_i+y_i}{\sqrt{2}}\right)\right] \\
&= \sum_{S,T}\hat{f}(S)\hat{f}(T)\prod_{i=1}^{n}\mathop{\mathbf{E}}_{x,y}\left[H_{S_i}(x_i)H_{T_i}\left(\frac{x_i+y_i}{\sqrt{2}}\right)\right].
\end{aligned}
$$

By Claim 2.2,

$$
H_{T_i}\left(\frac{x_i+y_i}{\sqrt{2}}\right) = \frac{1}{2^{T_i/2}}\sum_{l=0}^{T_i}\sqrt{\binom{T_i}{l}}H_l(x_i)H_{T_i-l}(y_i).
$$

Hence,

$$
\mathop{\mathbf{E}}_{x,y}\left[f(x)f\left(\frac{x+y}{\sqrt{2}}\right)\right] = \sum_{S,T}\hat{f}(S)\hat{f}(T)\prod_{i=1}^{n}\frac{1}{2^{T_i/2}}\sum_{l=0}^{T_i}\sqrt{\binom{T_i}{l}}\mathop{\mathbf{E}}_{x}\left[H_{S_i}(x_i)H_l(x_i)\right]\mathop{\mathbf{E}}_{y}\left[H_{T_i-l}(y_i)\right].
$$

By Claim 2.1, $\mathbf{E}_y\left[H_{T_i-l}(y_i)\right]=0$, unless $l=T_i$, and $\mathbf{E}_x\left[H_{S_i}(x_i)H_l(x_i)\right]=0$, unless $l=S_i$. Thus,

$$
\begin{aligned}
\mathop{\mathbf{E}}_{x,y}\left[f(x)f\left(\frac{x+y}{\sqrt{2}}\right)\right] &= \sum_{S}\hat{f}(S)^2\cdot\left(\frac{1}{\sqrt{2}}\right)^{|S|} \\
&\leq \frac{1}{\sqrt{2}}\cdot\|f^{=1}\|_2^2+\left(\frac{1}{\sqrt{2}}\right)^2\cdot\|f-f^{=1}\|_2^2, \qquad (5)
\end{aligned}
$$

where we used $\hat{f}(\vec{0})=0$ that follows from anti-symmetry. By combining equality (4) and inequality (5),

$$
\begin{aligned}
\mathop{\mathbf{E}}_{x,y\sim\mathcal{N}^n}\left[\left|f(x)+f(y)+\sqrt{2}\cdot f\left(-\frac{x+y}{\sqrt{2}}\right)\right|^2\right] &\geq 4\|f\|_2^2-4\|f^{=1}\|_2^2-\frac{4}{\sqrt{2}}\|f-f^{=1}\|_2^2 \\
&= (4-2\sqrt{2})\|f-f^{=1}\|_2^2 \\
&\geq \|f-f^{=1}\|_2^2.
\end{aligned}
$$

□

4 Dictator Testing

In this section we devise a dictator test, i.e., a test that checks whether an anti-symmetric real function in $L^2(\mathbb{R}^n,\mathcal{N}^n)$ is a *dictator* (that is, of the form $f(x)=x_i$ for some $i\in[n]$) or *far* from a dictator. We consider a function to be *close* to a dictator if it satisfies the following definition:

Definition 8 ((J,s)-Approximate linear junta)**.** *An anti-symmetric function $f\in L^2(\mathbb{R}^n,\mathcal{N}^n)$ with linear part $f^{=1}=\sum_{i=1}^{n}a_ix_i$, is called a* (J,s)-approximate-linear-junta*, if:*

- $\|f^{=1}\|_2^2 = \sum_{i=1}^n a_i^2 \geq (1-s)\|f\|_2^2$.
- $\sum_{i:a_i^2 \leq \frac{1}{J}\|f\|_2^2} a_i^2 \leq \Gamma \cdot \|f\|_2^2$, *where* $\Gamma = 0.05$ *is an absolute constant.*

An approximate linear junta has almost all the Fourier mass on its linear part, and this linear part is concentrated on at most J coordinates: Let $I = \left\{ i \mid a_i^2 \geq \frac{1}{J}\|f\|_2^2 \right\}$. Then $|I| \leq J$, and $\|f - \sum_{i \in I} a_i x_i\|_2^2 \leq (s + \Gamma)\|f\|_2^2$.

Our test will produce equations that dictators almost always satisfy exactly. On the other hand, functions that are not even approximate linear juntas fail with large margin.

Theorem 9 (Dictator testing)**.** *There are universal constants $s, c > 0$ such that the following holds. For every sufficiently small $\delta > 0$, there is a dictator test given by a distribution $\mathcal{E}$ over equations, where each equation depends on the value of f on at most three points in $\mathbb{R}^n$. The test satisfies the following properties:*

1. Uniformity: *The distribution over $\mathbb{R}^n$ obtained from picking at random an equation and x such that $f(x)$ is queried by the equation, is Gaussian $\mathcal{N}^n$.*
2. Bound on coefficients: *All the coefficients in the equations are in $[\frac{1}{b_0}, b_0]$ in magnitude where b_0 is a universal constant ($b_0 = 2$ works).*
3. Completeness: *If $f(x) = x_i$ for some $i \in [n]$, then*
$$\mathop{\mathbf{E}}_{eq \sim \mathcal{E}} \left[\chi_{|eq|>0} \cdot \|f_{eq}\|_2^2\right] \leq \delta.$$
4. Soundness: *For any anti-symmetric function $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$, $\|f\|_2^2 = 1$, if f is not a $(\frac{10}{\Gamma\delta^2}, s)$-approximate linear junta, then*
$$\mathop{\mathbf{E}}_{eq \sim \mathcal{E}} \left[\chi_{|eq|>c\sqrt{\delta}} \cdot \|f_{eq}\|_2^2\right] \geq \frac{s}{100}.$$

Remark 4.1. *Note that it follows from the soundness guarantee that for an anti-symmetric function $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$ with arbitrary non-zero norm, if f is not a $(\frac{10}{\Gamma\delta^2}, s)$-approximate linear junta, then*
$$\mathop{\mathbf{E}}_{eq \sim \mathcal{E}} \left[\chi_{|eq|>c\sqrt{\delta}\cdot\|f\|_2} \cdot \|f_{eq}\|_2^2\right] \geq \frac{s}{100} \cdot \|f\|_2^2.$$
This is obtained by applying the theorem with the normalized version of f, i.e., $\frac{f}{\|f\|_2}$.

The test will consist of three steps: (i) Linearity test that rules out functions that are not well-approximated by their linear parts. (ii) Coordinatewise perturbation test that checks that the function does not change by re-sampling a small fraction of the coordinates. (iii) Random perturbation test that guarantees that the function does not change much if perturbing the input slightly in a random direction. We achieve the effect of this test by instead doing two correlated linearity tests, in order to keep the coefficients in the range $[\frac{1}{2}, 2]$ in magnitude.

Dictator Test:

Given oracle access to a function $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$, f anti-symmetric. With equal probability, perform one of these three tests:

1. **Linearity test** on f, as in Section 3.

2. **Coordinatewise perturbation test:**

   (a) Pick $x, y \sim \mathcal{N}^n$. Pick $\tilde{x} \sim \mathcal{N}^n$ as follows: for $i = 1, 2, \ldots, n$, independently, with probability $1-\delta$, set $\tilde{x}_i = x_i$, and with probability δ, set $\tilde{x}_i = y_i$.

   (b) Test:
$$f(x) - f(\tilde{x}) = 0.$$

3. **Random perturbation test (in disguise):**

   (a) Pick $y, z \sim \mathcal{N}^n$. Let $x = \frac{y+z}{\sqrt{2}}$, $w = \frac{y-z}{\sqrt{2}}$, and
$$\begin{aligned}\tilde{x} &= (1-\delta)x + \sqrt{2\delta - \delta^2}w \\ &= \left(\frac{1-\delta}{\sqrt{2}} + \frac{\sqrt{2\delta-\delta^2}}{\sqrt{2}}\right) y + \left(\frac{1-\delta}{\sqrt{2}} - \frac{\sqrt{2\delta-\delta^2}}{\sqrt{2}}\right) z \\ &= \lambda_1 y + \lambda_2 z \quad (say).\end{aligned}$$

   (b) Note that λ_1, λ_2 are very close to $\frac{1}{\sqrt{2}}$. Test with equal probability:
$$f(x) - \frac{1}{\sqrt{2}} f(y) - \frac{1}{\sqrt{2}} f(z) = 0.$$
$$f(\tilde{x}) - \lambda_1 f(y) - \lambda_2 f(z) = 0.$$

Note that in the random perturbation test, $\tilde{x} = (1-\delta)x + \sqrt{2\delta-\delta^2}w$ and x is independent of w. Thus $\tilde{x}$ can indeed be thought of as a perturbation of x in a random direction. The uniformity property, as well as the bound on the coefficients, hold by the definition of the tests. Denote the distribution on all equations by $\mathcal{E}$, and the three sub-distributions by: $\mathcal{E}_l$ (linearity tests), $\mathcal{E}_c$ (coordinatewise perturbation tests), $\mathcal{E}_r$ (random perturbation tests).

Completeness: A dictator function f, being a linear function, always exactly satisfies the linearity test and the random perturbation test. As for the coordinatewise perturbation test, $\mathbf{E}_{eq \sim \mathcal{E}_c}\left[\chi_{|eq|>0} \cdot \|f_{eq}\|_2^2\right] \leq \delta \|f\|_2^2 = \delta$.

Soundness: In the following, $O(\cdot)$ and $\Omega(\cdot)$ notations will hide universal constants. We will pick s and c to be universal constants eventually, but throughout the proof, retain the dependence on them. Assume for now that $2c \leq s \leq 0.01$. The parameter δ is thought of as tending to zero.

Let $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$ be an anti-symmetric function, $\|f\|_2^2 = 1$, f is not a $(J = \frac{10}{\Gamma\delta^2}, s)$-approximate linear junta. Assume, for the sake of a contradiction, that
$$\mathop{\mathbf{E}}_{eq \sim \mathcal{E}}\left[\chi_{|eq| \leq c\sqrt{\delta}} \cdot \|f_{eq}\|_2^2\right] \geq 1 - \frac{s}{100}.$$

Denote the non-linear part of f by $e = f - f^{=1}$ (since f is anti-symmetric, $f^{\leq 0} \equiv 0$). We handle the cases that $\|e\|_2^2 \leq s$ and $\|e\|_2^2 > s$ separately.

Case $\|e\|_2^2 > s$: By Lemma 3.1, $\mathbf{E}_{eq\sim\mathcal{E}_l}\left[|eq|^2\right] \geq \|e\|_2^2 > s$. By Cauchy-Schwarz inequality, for every equation,[3] we have $|eq|^2 \leq 12\|f_{eq}\|_2^2$, so

$$s < \mathop{\mathbf{E}}_{eq\sim\mathcal{E}_l}\left[|eq|^2\right] \leq \mathop{\mathbf{E}}_{eq\sim\mathcal{E}_l}\left[\chi_{|eq|>c\sqrt{\delta}} \cdot 12\|f_{eq}\|_2^2\right] + c^2\delta \leq 12 \mathop{\mathbf{E}}_{eq\sim\mathcal{E}_l}\left[\chi_{|eq|>c\sqrt{\delta}} \cdot \|f_{eq}\|_2^2\right] + \frac{s}{3}.$$

Since the distribution $\mathcal{E}$ is average of distributions $\mathcal{E}_l, \mathcal{E}_c$, and $\mathcal{E}_r$, we get

$$\mathop{\mathbf{E}}_{eq\sim\mathcal{E}}\left[\chi_{|eq|>c\sqrt{\delta}} \cdot \|f_{eq}\|_2^2\right] \geq \frac{1}{3} \cdot \mathop{\mathbf{E}}_{eq\sim\mathcal{E}_l}\left[\chi_{|eq|>c\sqrt{\delta}} \cdot \|f_{eq}\|_2^2\right] > \frac{s}{100}.$$

This contradicts our assumption that $\mathbf{E}_{eq\sim\mathcal{E}}\left[\chi_{|eq|\leq c\sqrt{\delta}} \cdot \|f_{eq}\|_2^2\right] \geq 1 - \frac{s}{100}$.

Case $\|e\|_2^2 \leq s$: We first show that in this case, almost every equation is satisfied with margin at most $c\sqrt{\delta}$.

Lemma 4.1. *The probability that a dictator test equation chosen at random is $c\sqrt{\delta}$-approximately satisfied is at least $1 - 7\sqrt[3]{s}$.*

Proof. We begin by showing that for $x \sim \mathcal{N}^n$, $|f(x)| \geq \frac{\sqrt[3]{s}}{4}$ except with probability at most $6\sqrt[3]{s}$. When $x \sim \mathcal{N}^n$, except with probability at most $4\sqrt[3]{s}$, we have that $|e(x)|^2 \leq \frac{1}{4\sqrt[3]{s}}\|e\|_2^2 \leq \frac{s^{2/3}}{4}$. Write $f^{=1}(x) = \sum_{i=1}^n a_i x_i$. When $x \sim \mathcal{N}^n$, we have that $f^{=1}(x)$ is normal with mean 0 and variance $\sum_{i=1}^n a_i^2 = 1 - \|e\|_2^2 \geq 0.99$. Thus, except with probability at most $2\sqrt[3]{s}$, we have that $\left|f^{=1}(x)\right| \geq \sqrt{0.99}\sqrt[3]{s}$. Overall, except with probability at most $6\sqrt[3]{s}$, we have that $|f(x)| \geq \left|f^{=1}(x)\right| - |e(x)| \geq \sqrt{0.99}\sqrt[3]{s} - \frac{\sqrt[3]{s}}{2} \geq \frac{\sqrt[3]{s}}{4}$.

Assume, for the sake of a contradiction, that with probability at least $7\sqrt[3]{s}$, a dictator test equation has margin at least $c\sqrt{\delta}$. An equation has at most three variables, and each of these is distributed as $\mathcal{N}^n$. With probability at least $7\sqrt[3]{s} - 6\sqrt[3]{s} = \sqrt[3]{s}$, it also holds that the first variable, say $f(x)$, in the equation has magnitude $|f(x)| \geq \frac{\sqrt[3]{s}}{4}$. For such an equation, $\|f_{eq}\|_2^2 \geq \frac{1}{3}f(x)^2 \geq \frac{s^{2/3}}{48}$. Hence,

$$\mathop{\mathbf{E}}_{eq\sim\mathcal{E}}\left[\chi_{|eq|>c\sqrt{\delta}} \cdot \|f_{eq}\|_2^2\right] \geq \sqrt[3]{s}\ \frac{s^{2/3}}{48} > \frac{s}{100}.$$

This contradicts our assumption, and the claim follows. □

In the sequel we inspect the change in e as we perturb the input. We show that our assumptions on f (made towards a contradiction) imply that e may change somewhat as a result of a perturbation in a random direction, yet changes noticeably more as a result of a coordinatewise perturbation. We will later show that these two behaviors are contradictory.

Lemma 4.2 (e is noise-stable for random perturbation)**.** *(Under the assumptions we made towards a contradiction) Let $x, \tilde{x}$ be picked as in the random perturbation test. Then, with probability at least $1 - O(\sqrt[3]{s})$,*

$$|e(x) - e(\tilde{x})| \leq O(\sqrt[3]{s})\sqrt{\delta}.$$

[3] The linearity testing equation is of the form $f(x) + f(y) - \sqrt{2}f(z) = 0$. Here $|eq| = |f(x) + f(y) - \sqrt{2}f(z)|$ and $\|f_{eq}\|_2^2 = \frac{f(x)^2+f(y)^2+f(z)^2}{3}$.

Proof. Since the random perturbation test is performed with probability $\frac{1}{3}$, from Lemma 4.1, with probability at least $1 - O(\sqrt[3]{s})$, we have

$$\left| f(x) - \frac{1}{\sqrt{2}} f(y) - \frac{1}{\sqrt{2}} f(z) \right| \leq c\sqrt{\delta},$$
$$|f(\tilde{x}) - \lambda_1 f(y) - \lambda_2 f(z)| \leq c\sqrt{\delta}.$$

Since $f = f^{=1} + e$, and $f^{=1}$ is linear, the above inequalities are really inequalities for e:

$$\left| e(x) - \frac{1}{\sqrt{2}} e(y) - \frac{1}{\sqrt{2}} e(z) \right| \leq c\sqrt{\delta},$$
$$|e(\tilde{x}) - \lambda_1 e(y) - \lambda_2 e(z)| \leq c\sqrt{\delta}.$$

Combining the two inequalities and substituting for λ_1 and λ_2, we get:

$$|e(x) - e(\tilde{x})| \ \leq \ 2c\sqrt{\delta} + O(\sqrt{\delta})(|e(y)| + |e(z)|).$$

By Markov inequality, except with probability at most $\sqrt[3]{s}$, it holds that $|e(y)|^2 \leq \|e\|_2^2 / \sqrt[3]{s} \leq s^{2/3}$. The same applies to $e(z)$. Therefore, with probability at least $1 - O(\sqrt[3]{s})$,

$$|e(x) - e(\tilde{x})| \ \leq \ 2c\sqrt{\delta} + O(\sqrt[3]{s} \cdot \sqrt{\delta}) = O(\sqrt[3]{s})\sqrt{\delta}.$$

□

Lemma 4.3 (*e* is noise-sensitive coordinatewise)**.** *(Under the assumptions we made towards a contradiction) Let $x, \tilde{x} \sim \mathcal{N}^n$ be picked as in the coordinatewise perturbation test. Then, with probability at least $\Omega(1)$, we have*

$$|e(x) - e(\tilde{x})| \geq \Omega(\sqrt{\delta}).$$

Proof. Write $f^{=1} = \sum_{i=1}^n a_i x_i$. Since $f = f^{=1} + e$, we have

$$\begin{aligned} |e(x) - e(\tilde{x})| \ &\geq \ \left| f^{=1}(x) - f^{=1}(\tilde{x}) \right| - |f(x) - f(\tilde{x})| \\ &= \ \left| \sum_{i=1}^n a_i (x_i - \tilde{x}_i) \right| - |f(x) - f(\tilde{x})| \, . \end{aligned}$$

From Lemma 4.1, we know that except with probability $O(\sqrt[3]{s})$, the second term $|f(x) - f(\tilde{x})|$ is at most $c\sqrt{\delta}$. Thus it suffices to show that with probability $\Omega(1)$, the first term is at least $\Omega(\sqrt{\delta})$ (and to choose c and s sufficiently small).

Recall that the test picks the pair $(x, \tilde{x})$ as follows: First pick a set $D \subseteq [n]$ by including in it every $i \in [n]$ independently with probability δ. Pick $x, y \sim \mathcal{N}^n$ independently. For every $i \notin D$, set $\tilde{x}_i = x_i$, and for every $i \in D$, set $\tilde{x}_i = y_i$. Thus for a fixed D,

$$\sum_{i=1}^n a_i (x_i - \tilde{x}_i) = \sum_{i \in D} a_i (x_i - y_i),$$

which is a normal variable with mean 0 and variance $2 \sum_{i \in D} a_i^2$. We will show that the variance is at least $\Gamma\delta$ with probability 0.9 over the choice of D. Whenever this happens, the random variable exceeds $\Omega(\sqrt{\delta})$ in magnitude with probability $\Omega(1)$ and we are done.

Let $I = \left\{i \in [n] \mid a_i^2 \leq \frac{1}{J}\right\}$ be the "non-influential" coordinates. Since f is not a (J,s)-approximate linear junta, and $\|e\|_2^2 \leq s$, we must have $\sum_{i\in I} a_i^2 \geq \Gamma$. A standard Hoeffding bound now shows that for a random choice of set D, the sum $\sum_{i\in I\cap D} a_i^2$ is at least half its expected value with probability at least 0.9 and the expected value is $\delta \sum_{i\in I} a_i^2$ which is at least $\Gamma\delta$.

$$\Pr_D\left[\left|\sum_{i\in I\cap D} a_i^2 - \delta\sum_{i\in I} a_i^2\right| \geq \frac{\delta}{2}\sum_{i\in I} a_i^2\right] \leq 2\cdot\exp\left(-\frac{2(\frac{\delta}{2}\sum_{i\in I} a_i^2)^2}{\sum_{i\in I} a_i^4}\right) \leq 2\cdot\exp\left(-\frac{J}{2}\cdot\Gamma\delta^2\right) \leq 0.1,$$

where we noted that $\sum_{i\in I} a_i^4 \leq \frac{1}{J}\sum_{i\in I} a_i^2$ and $J = \frac{10}{\Gamma\delta^2}$.

□

The rest of the proof is devoted to showing that Lemma 4.2 and Lemma 4.3 cannot both hold, i.e., a function cannot be noise stable for random perturbation, yet noise sensitive for coordinatewise perturbation. Towards this end, we will construct from e a new function e' (that happens to be $\{0,1\}$-valued) for which the expected squared change as a result of coordinatewise perturbation is much larger than the expected squared change as a result of random perturbation:

Lemma 4.4. *(Under the assumptions we made towards a contradiction, and in particular, assuming Lemma 4.2 and Lemma 4.3) There is a function e' such that:*

1. $\mathbf{E}_{(x,\tilde{x})\sim R}\left[|e'(x) - e'(\tilde{x})|^2\right] \leq O(\sqrt[3]{s})$,
2. $\mathbf{E}_{(x,\tilde{x})\sim C}\left[|e'(x) - e'(\tilde{x})|^2\right] \geq \Omega(1)$,

where R is the distribution over pairs in the random perturbation test, and C is the distribution over pairs in the coordinatewise perturbation test.

The proof of Lemma 4.4 appears in Section 5. For sufficiently small s, Lemma 4.4 leads to a contradiction by the following claim:

Claim 4.5. *For any function $h \in L^2(\mathbb{R}^n, \mathcal{N}^n)$,*

$$\mathop{\mathbf{E}}_{(x,\tilde{x})\sim R}\left[|h(x) - h(\tilde{x})|^2\right] \geq \mathop{\mathbf{E}}_{(x,\tilde{x})\sim C}\left[|h(x) - h(\tilde{x})|^2\right],$$

where R is the distribution over pairs in the random perturbation test, and C is the distribution over pairs in the coordinatewise perturbation test.

Proof. The expectation $\mathbf{E}_{(x,\tilde{x})\sim C}\left[|h(x) - h(\tilde{x})|^2\right]$ is given by the following expression:

$$\mathop{\mathbf{E}}_D\left[\mathop{\mathbf{E}}_{x_{|\overline{D}}}\left[\mathop{\mathbf{E}}_{x_{|D},\tilde{x}_{|D}}\left[|h(x) - h(\tilde{x})|^2\right]\right]\right],$$

where the set of coordinates $D \subseteq [n]$ is chosen by including each $i \in [n]$ in D independently with probability δ. Using $\mathbf{Var}_x[F(x)] = \frac{1}{2}\mathbf{E}_{x,x'}\left[(F(x) - F(x'))^2\right]$ and the notion of influence as discussed in the preliminaries, the above expression can be re-written as:

$$\mathop{\mathbf{E}}_D\left[\mathop{\mathbf{E}}_{x_{|\overline{D}}}\left[2\mathop{\mathbf{Var}}_{x_{|D}}[h(x)]\right]\right] = 2\mathop{\mathbf{E}}_D[I_D(h)] = 2\mathop{\mathbf{E}}_D\left[\sum_{S\cap D\neq\phi}\hat{h}(S)^2\right] = 2\sum_S \hat{h}(S)^2 \Pr_D[S\cap D\neq\phi].$$

For every multi-index $S \in \mathbb{N}^n$, we have: $\Pr_D[S \cap D \neq \phi] = 1-(1-\delta)^{\#S} \leq 1-(1-\delta)^{|S|}$. Here $|S| = \sum_{i=1}^n S_i$ and $\#S$ denotes the number of S_i that are non-zero, and hence we have $\#S \leq |S|$. Therefore, the expectation is at most

$$2\sum_S \hat{h}(S)^2 \cdot (1-(1-\delta)^{|S|}).$$

On the other hand, the expectation $\mathbf{E}_{(x,\tilde{x})\sim R}\left[|h(x)-h(\tilde{x})|^2\right]$ is given by the following expression, for $\rho = 1-\delta$:

$$2\mathop{\mathbf{E}}_x\left[h(x)^2\right] - 2\mathop{\mathbf{E}}_{x,w}\left[h(x)h(\rho x+\sqrt{1-\rho^2}w)\right].$$

We have $\mathbf{E}_{x,w}\left[h(x)h(\rho x+\sqrt{1-\rho^2}w)\right] = \langle h, T_\rho h\rangle = \sum_S \hat{h}(S)^2\rho^{|S|}$ and $\mathbf{E}_x\left[h(x)^2\right] = \sum_S \hat{h}(S)^2$, so the expectation is

$$2\sum_S \hat{h}(S)^2(1-(1-\delta)^{|S|}).$$

□

This concludes the proof of Theorem 9 assuming Lemma 4.4.

5 Proof of Lemma 4.4

In this section we prove Lemma 4.4. Assume that a function $e \in L^2(\mathbb{R}^n, \mathcal{N}^n)$ with $\|e\|_2^2 \leq s$ satisfies:

- With probability at least $1 - O(\sqrt[3]{s})$ over $(x,\tilde{x}) \sim R$, it holds that

$$|e(x)-e(\tilde{x})| \leq d_R = O(\sqrt[3]{s})\sqrt{\delta}. \tag{6}$$

- With probability at least $\Omega(1)$ over $(x,\tilde{x}) \sim C$, it holds that

$$|e(x)-e(\tilde{x})| \geq d_C = \Omega(\sqrt{\delta}). \tag{7}$$

We show how to obtain a function $e' \in L^2(\mathbb{R}^n, \mathcal{N}^n)$ (in fact $\{0,1\}$-valued) that satisfies:

- $\mathbf{E}_{(x,\tilde{x})\sim R}\left[|e'(x)-e'(\tilde{x})|^2\right] \leq O(\sqrt[3]{s})$.
- $\mathbf{E}_{(x,\tilde{x})\sim C}\left[|e'(x)-e'(\tilde{x})|^2\right] \geq \Omega(1)$.

To this end, we construct two graphs on $\mathbb{R}^n$, $G_R = (\mathbb{R}^n, E_R)$ and $G_C = (\mathbb{R}^n, E_C)$, representing the function e under random perturbation and under coordinatewise perturbation, respectively. The graphs are infinite, and we will be abusing notation in the following, but all the arguments can be made precise by replacing sums by integrals wherever appropriate.

Perturbation Graphs. The graphs G_R and G_C have labels on their vertices and weights on their edges. The label of a vertex $x \in \mathbb{R}^n$ is $e(x)$.

The graph G_R has edges between pairs $(x, \tilde{x})$ such that: (i) The labels on the endpoints are bounded, $|e(x)|, |e(\tilde{x})| \le 1$. (ii) $|e(x) - e(\tilde{x})| \le d_R$. The weight of the edge $w_R(x, \tilde{x})$ is the probability that $(x, \tilde{x})$ is chosen in the random perturbation test. The total edge weight is $w_R(E_R) \ge 1 - O(\sqrt[3]{s})$ from Hypothesis (6) and the observation that $\|e\|_2^2 \le s$ and thus for $x \in \mathcal{N}^n$, $|e(x)| \le 1$ except with probability $\sqrt{s}$.

The graph G_C has edges between pairs $(x, \tilde{x})$ such that: (i) The labels on the endpoints are bounded, $|e(x)|, |e(\tilde{x})| \le 1$. (ii) $|e(x) - e(\tilde{x})| \ge d_C$. The weight of the edge $w_C(x, \tilde{x})$ is the probability that $(x, \tilde{x})$ is chosen in the coordinate-wise perturbation test. The total edge weight is $w_C(E_C) \ge \Omega(1)$ from Hypothesis (7) and since $\|e\|_2^2 \le s$.

Cuts in Perturbation Graphs. We will construct a cut $\mathcal{C} : \mathbb{R}^n \to \{0, 1\}$, and this will be our function $e' \equiv \mathcal{C}$. Denote by $w_R(\mathcal{C})$ and $w_C(\mathcal{C})$, the weight of the edges in the graphs G_R and G_C respectively that are cut by $\mathcal{C}$. The cut $\mathcal{C}$ will satisfy:

1. (Small E_R weight is cut:) $\quad w_R(\mathcal{C}) \le O(\sqrt[3]{s})$.
2. (Large E_C weight is cut) $\quad w_C(\mathcal{C}) \ge \Omega(1)$.

Let us first check that this proves Lemma 4.4: When choosing $(x, \tilde{x})$ as in the random perturbation test, the probability that the pair $(x, \tilde{x})$ is separated is at most $w_R(\mathcal{C}) + (1 - w_R(E_R)) \le O(\sqrt[3]{s})$. When choosing $(x, \tilde{x})$ as in the coordinatewise perturbation test, the probability the pair $(x, \tilde{x})$ is separated is at least $w_C(\mathcal{C}) \ge \Omega(1)$.

Lemma 5.1. *There is a distribution over cuts such that:*

- *Every edge $(x, \tilde{x}) \in E_R$ is cut with probability at most $p_{R,0} \le O(\sqrt[3]{s})\sqrt{\delta}$.*
- *Every edge $(x, \tilde{x}) \in E_C$ is cut with probability at least $p_{C,0} \ge \sqrt{\delta}$.*

Proof. The distribution over cuts is defined by picking at random $\lambda \in [-1, 1]$. For every $x \in \mathbb{R}^n$ we define $\mathcal{C}'(x) = 1$ if $e(x) \ge \lambda$, and $\mathcal{C}'(x) = 0$ otherwise. A pair $(x, \tilde{x})$ is cut if and only if λ falls between $e(x)$ and $e(\tilde{x})$. If $e(x), e(\tilde{x}) \in [-1, 1]$, this happens with probability $\frac{|e(x) - e(\tilde{x})|}{2}$. The lemma follows from the construction of the graph. $\square$

We construct the cut $\mathcal{C}$ in a randomized way as follows: Let $M = \lceil 1/p_{C,0} \rceil$.

1. For $i = 1, \ldots, M$, draw a cut $\mathcal{C}_i$ from the distribution in Lemma 5.1.
2. Let $I \subseteq [M]$ be chosen by including every $i \in [M]$ in I independently with probability $\frac{1}{2}$.
3. Let $\mathcal{C}(x) = \bigoplus_{i \in I} \mathcal{C}_i(x)$.

Lemma 5.2. *The following hold:*

- *For every edge $(x, \tilde{x}) \in E_R$, the probability that $(x, \tilde{x})$ is cut by $\mathcal{C}$, is at most $p_R \le O(\sqrt[3]{s})$.*
- *For every edge $(x, \tilde{x}) \in E_C$, the probability that $(x, \tilde{x})$ is cut by $\mathcal{C}$, is at least $p_C \ge \Omega(1)$.*

Proof. Note that an edge is cut by $\mathcal{C}$ if and only if it is cut by an odd number of cuts $\mathcal{C}_i, i \in I$.

If $(x, \tilde{x}) \in E_R$, then by Lemma 5.1, it is cut by any specific $\mathcal{C}_i$ with probability at most $p_{R,0}$. Hence the probability that it is cut by $\mathcal{C}$ is at most $M \cdot p_{R,0} \le O(\sqrt[3]{s})$.

If $(x, \tilde{x}) \in E_C$, then by Lemma 5.1 and the choice of M, with constant probability, the edge is cut by at least one $\mathcal{C}_i, i \in [M]$. Since I is a random subset of $[M]$ of half the size, with constant probability, the edge is cut by an odd number of $\mathcal{C}_i, i \in I$, and hence by $\mathcal{C}$. ☐

The above Lemma 5.2 shows that

$$\mathbf{E}\left[w_R(\mathcal{C})\right] \le p_R \cdot w_R(E_R) \le p_R, \quad \text{and}$$

$$\mathbf{E}\left[w_C(\mathcal{C})\right] \ge p_C \cdot w_C(E_C) \ge \Omega(1) \cdot \Omega(1) = \Omega(1) = p^*.$$

It follows that there must exist a cut $\mathcal{C}$ such that both these hold simultaneously:

$$w_R(\mathcal{C}) \le \frac{4 \cdot p_R}{p^*} = O(\sqrt[3]{s}) \qquad \text{and} \qquad w_C(\mathcal{C}) \ge \frac{p^*}{2} = \Omega(1).$$

Indeed, by an averaging argument, the first condition holds with probability at least $1 - \frac{p^*}{4}$ and the second condition holds with probability at least $\frac{p^*}{2}$, and hence both conditions hold simultaneously with probability at least $\frac{p^*}{4}$. This completes the proof of Lemma 4.4.

6 Hardness of Robust-3LIN($\mathbb{R}$)

Armed with the dictator test from Section 4, we are ready to show the UGC-hardness of Robust-3Lin($\mathbb{R}$) and prove Theorem 6. For that, we show a reduction from Unique-Game to Robust-3Lin($\mathbb{R}$). Let s_0 (slightly redefined) and c_0 be the constants for the dictator testing theorem, Theorem 9, so for any anti-symmetric function $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$,

$$\mathop{\mathbf{E}}_{eq}\left[\chi_{|eq| \le c_0\sqrt{\delta}\|f\|_2} \|f_{eq}\|_2^2\right] \ge (1 - s_0)\|f\|_2^2 \;\; \Rightarrow \;\; f \text{ is } \left(\frac{10}{\Gamma\delta^2}, 100 s_0\right)\text{-approximate linear junta.}$$

Note that Theorem 9 remains correct if the parameters s_0 and c_0 are made smaller, so w.l.o.g. we can assume that these parameters can be made sufficiently small if needed. The constants s and c for the Robust-3Lin($\mathbb{R}$) hardness theorem, Theorem 6, depend appropriately on s_0 and c_0. In fact, setting $s \doteq \frac{s_0}{50}$ and $c \doteq \frac{c_0}{10}$ works. Let δ be the completeness parameter from the statement of the Robust-3Lin($\mathbb{R}$) hardness theorem, Theorem 6, and $b = O(\log(1/\delta))$. We use the Unique Games Conjecture with completeness $1 - \eta$ and soundness ε for sufficiently small $\eta, \varepsilon > 0$ and and let $k = k(\eta, \varepsilon)$ be the corresponding number of labels. Given a Unique-Game instance $(G = (V, E), k, \{\pi_e\})$, we reduce to a Robust-3Lin($\mathbb{R}$) instance $(X, \mathcal{E})$. For convenience, we first describe a non-discretized construction (having variables for every real point in $\mathbb{R}^k$), and then explain how to discretize the construction and obtain an efficient reduction. The non-discretized construction is as follows:

- **Variables:** There is a variable for every vertex $v \in V$ and every $x \in \mathbb{R}^k$. We denote the assignment to the variables associated with v by $A_v : \mathbb{R}^k \to \mathbb{R}$. We assume, by *folding*[4], that A_v is anti-symmetric, i.e. $\forall x \in \mathbb{R}^k,\ A_v(-x) = -A_v(x)$. Supposedly, $A_v(x) = x_i$ where $i \in [k]$ is a label to v.

[4]Folding means that we have just one variable for every pair $x, -x \in \mathbb{R}^k$, and we define $A_v(-x) = -A_v(x)$.

- **Equations:** The distribution over equations: pick a random edge $e = (u, v) \in E$. Sample an equation according to the following distribution $\mathcal{E}^e$: With equal probability,
  - $\mathcal{E}_u$: Perform dictator testing on A_u as in Theorem 9 with parameter δ.
  - $\mathcal{E}_v$: Perform dictator testing on A_v as in Theorem 9 with parameter δ.
  - $\mathcal{E}_e$: pick $x \sim \mathcal{N}^k$ and produce the equation:

$$A_u(\pi_e(x)) - A_v(x) = 0,$$

where $\pi_e(x) \doteq (x_{\pi_e(1)}, \ldots, x_{\pi_e(k)})$.

Note that the coefficients of the equations are in $[\frac{1}{2}, 2]$ in magnitude. Note also that every query is distributed uniformly over $v \in V$, and then for a fixed v, Gaussian distributed over $\mathbb{R}^k$.

6.1 Completeness

Assume that there is a labeling $\varphi : V \to [k]$ that satisfies $1 - \eta$ fraction of the edges in the Unique-Game instance $(G = (V, E), k, \{\pi_e\})$. We construct from it an assignment $A : X \to \mathbb{R}$ for the Robust-3Lin($\mathbb{R}$) instance $(X, \mathcal{E})$. For every vertex $v \in V$, we let $A_v(x) = x_{\phi(v)}$.

Note that G is a regular graph and the values of the variables are distributed according to a standard Gaussian, and hence $\|A\|_2^2 = 1$. Pick a random edge $e = (u, v) \in E$. The probability that e is not satisfied by the labeling is at most η and the equations incident on unsatisfied edges contribute at most $O(\eta)$ towards the overall norm on the equations. Let us therefore concentrate on the case that e is satisfied by φ. By Theorem 9,

$$\mathop{\mathbf{E}}_{eq \sim \mathcal{E}_u} \left[\chi_{|eq|>0} \cdot \|A_{eq}\|_2^2\right] \le \delta.$$

$$\mathop{\mathbf{E}}_{eq \sim \mathcal{E}_v} \left[\chi_{|eq|>0} \cdot \|A_{eq}\|_2^2\right] \le \delta.$$

For every $x \in \mathbb{R}^k$, we have $A_u(\pi_e(x)) = x_{\pi_e(\varphi(u))} = x_{\varphi(v)} = A_v(x)$. Thus

$$\mathop{\mathbf{E}}_{eq \sim \mathcal{E}_e} \left[\chi_{|eq|>0} \cdot \|A_{eq}\|_2^2\right] = 0.$$

Overall, we have $val^0_{(X,\mathcal{E})} \ge 1 - \delta - O(\eta) \ge 1 - 2\delta$ by choosing η small enough. Finally, we can truncate all the variables whose magnitude exceeds $b = O(\log(1/\delta))$ to zero. The norm on equations involving these variables is at most, say δ^4, and this does not affect the result.

6.2 Soundness

Assume that any assignment to the Unique Game instance $(G, k, \{\pi_e\})$ satisfies at most ε fraction of the edges. Fix an assignment $A : X \to [-b, b]$, $\|A\|_2^2 = 1$. We will show that

$$\mathop{\mathbf{E}}_{eq \sim \mathcal{E}} \left[\chi_{|eq|>c\sqrt{\delta}} \|A_{eq}\|_2^2\right] \ge s \doteq \frac{s_0}{50}.$$

Rewrite the above inequality as:

$$\mathop{\mathbf{E}}_{e \in E} \left[\mathop{\mathbf{E}}_{eq \sim \mathcal{E}^e} \left[\chi_{|eq|>c\sqrt{\delta}} \|A_{eq}\|_2^2\right]\right] \ge \frac{s_0}{50}. \tag{8}$$

Define $\|A_e\|_2^2 \doteq \frac{1}{2}\left(\|A_u\|_2^2 + \|A_v\|_2^2\right)$. By uniformity, $\mathbf{E}_{e\in E}\left[\|A_e\|_2^2\right] = 1$. Also, since the assignment is bounded by $[-b,b]$, $\|A_e\|_2^2 \le b^2$. We will partition the set of edges into three sets $E = E_1 \cup E_2 \cup E_3$ where:

- $E_1 \doteq \{e \mid \|A_e\|_2 \le 2c/c_0\}$.
- E_3 has at most $O(\varepsilon/\delta^4)$ fraction of the edges.

It suffices to show that the contribution of every edge towards (8) is lower bounded as:

$$\mathop{\mathbf{E}}_{eq\sim\mathcal{E}^e}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right] \ \ge\ \frac{s_0}{24}\ \|A_e\|_2^2 - \chi_{e\in E_1}\cdot\frac{s_0}{24}(2c/c_0)^2 - \chi_{e\in E_3}\cdot b^2, \tag{9}$$

where $\chi_{e\in E_i}$ is a $\{0,1\}$-valued indicator variable for $e \in E_i$. Indeed, taking expectation of both the sides over all edges $e \in E$, we see that

$$\mathop{\mathbf{E}}_{e\in E}\left[\mathop{\mathbf{E}}_{eq\sim\mathcal{E}^e}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right]\right] \ge \frac{s_0}{24} - \frac{s_0}{96} - b^2\cdot\Pr\left[e\in E_3\right] \ge \frac{s_0}{48} - O(\varepsilon/\delta^4)b^2 \ge \frac{s_0}{50},$$

since ε can be chosen to be sufficiently small, and $2c/c_0 \le \frac{1}{2}$.

Now we show that (9) holds for every edge. For $e \in E_1 \cup E_3$, this holds trivially as the right hand side is non-positive. It remains to show that (9) holds for every $e \in E_2 \subseteq \overline{E_1}$ and to define the appropriate partition $\overline{E_1} = E_2 \cup E_3$. Towards this end, fix an edge $e \in E_2$ so that $\|A_e\|_2 \ge 2c/c_0$. We need to show (9), that is:

$$\mathop{\mathbf{E}}_{eq\sim\mathcal{E}^e}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right] \ \ge\ \frac{s_0}{24}\|A_e\|_2^2.$$

Case $(\|A_u\|_2 - \|A_v\|_2)^2 \ge s_0\|A_e\|_2^2$: By Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathop{\mathbf{E}}_{eq\in\mathcal{E}_e}\left[|eq|^2\right] &= \mathop{\mathbf{E}}_{x\in\mathcal{N}^k}\left[\left(A_u(\pi_e(x)) - A_v(x)\right)^2\right] \\
&\ge \|A_u\|_2^2 - 2\|A_u\|_2\|A_v\|_2 + \|A_v\|_2^2 \\
&= \left(\|A_u\|_2 - \|A_v\|_2\right)^2 \\
&\ge s_0\|A_e\|_2^2.
\end{aligned} \tag{10}$$

We have[5] $|eq|^2 \le 2\|A_{eq}\|_2^2$ and thus:

$$\mathop{\mathbf{E}}_{eq\in\mathcal{E}_e}\left[|eq|^2\right] \le \mathop{\mathbf{E}}_{eq\in\mathcal{E}_e}\left[\chi_{|eq|>c\sqrt{\delta}}\cdot 2\|A_{eq}\|_2^2\right] + c^2\delta.$$

So,

$$\mathop{\mathbf{E}}_{eq\in\mathcal{E}_e}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right] \ge \frac{s_0}{2}\cdot\|A_e\|_2^2 - \frac{1}{2}c^2\delta \ge \frac{s_0}{8}\cdot\|A_e\|_2^2.$$

Since the distribution over equations $\mathcal{E}^e$ is an average of $\mathcal{E}_e$, $\mathcal{E}_u$ and $\mathcal{E}_v$, we get the following lower bound as desired:

$$\mathop{\mathbf{E}}_{eq\in\mathcal{E}^e}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right] \ge \frac{s_0}{24}\cdot\|A_e\|_2^2. \tag{11}$$

[5] For an equation of the form $eq : A(y) - A(x) = 0$, we have $|eq|^2 = |A(y)-A(x)|^2 \le (A(y)^2 + A(x)^2) = 2\|A_{eq}\|_2^2$.

Case $(\|A_u\|_2 - \|A_v\|_2)^2 \leq s_0\|A_e\|_2^2$: Since $\|A_e\|_2^2 = \frac{1}{2}\left(\|A_u\|_2^2 + \|A_v\|_2^2\right)$, for s_0 small enough, it holds that $\|A_u\|_2^2, \|A_v\|_2^2 \geq \frac{1}{4}\|A_e\|_2^2$.

Sub-case A_u is not a $(\frac{10}{\Gamma\delta^2}, 100s_0))$-approximate linear junta. In this sub-case, by the analysis of the dictatorship test,

$$\mathop{\mathbf{E}}_{eq\in\mathcal{E}_u}\left[\chi_{|eq|>c_0\sqrt{\delta}\|A_u\|_2}\|A_{eq}\|_2^2\right] \geq s_0\|A_u\|_2^2.$$

Since $\|A_u\|_2 \geq \frac{1}{2}\|A_e\|_2 \geq \frac{c}{c_0}$, the above inequality implies:

$$\mathop{\mathbf{E}}_{eq\in\mathcal{E}_u}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right] \geq \frac{s_0}{4}\|A_e\|_2^2.$$

Since the distribution over equations $\mathcal{E}^e$ is an average of $\mathcal{E}_e$, $\mathcal{E}_u$ and $\mathcal{E}_v$, we get the following lower bound as desired:

$$\mathop{\mathbf{E}}_{eq\in\mathcal{E}^e}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right] \geq \frac{s_0}{12}\cdot\|A_e\|_2^2.$$

Sub-case A_v is not a $(\frac{10}{\Gamma\delta^2}, 100s_0)$-approximate linear junta. This sub-case is handled similarly as above.

Thus we are left with the case where both A_u and A_v are $(\frac{10}{\Gamma\delta^2}, 100s_0)$-approximate linear juntas and moreover $\|A_u\|_2, \|A_v\|_2 \geq \frac{1}{2}\|A_e\|_2 \geq \frac{c}{c_0}$. In particular, there exist linear forms $l_u(x) = \sum_{i\in I_u} a_i x_i$ and $l_v(x) = \sum_{i\in I_v} b_i x_i$ such that $|I_u|, |I_v| \leq J$ and,

- $\|l_u - A_u\|_2^2 \leq (\Gamma + 100s_0)\|A_u\|_2^2$.
- $\|l_v - A_v\|_2^2 \leq (\Gamma + 100s_0)\|A_v\|_2^2$.

Sub-case $\pi_e(I_u)\cap I_v = \phi$. In this sub-case,

$$\begin{aligned}
\mathop{\mathbf{E}}_{eq\in\mathcal{E}_e}\left[|eq|^2\right] &= \mathop{\mathbf{E}}_{x\in\mathcal{N}^k}\left[(A_u(\pi_e(x)) - A_v(x))^2\right] \\
&= \|A_u\|_2^2 + \|A_v\|_2^2 - 2\mathop{\mathbf{E}}_{x\in\mathcal{N}^k}\left[A_u(\pi_e(x))A_v(x)\right] \\
&= 2\|A_e\|_2^2 - 2\mathop{\mathbf{E}}_{x\in\mathcal{N}^k}\left[A_u(\pi_e(x))A_v(x)\right] \\
&\geq \|A_e\|_2^2,
\end{aligned}$$

provided we bound the second term by $\|A_e\|_2^2$. This inequality is similar to (in fact stronger than) Equation (10) and thus enough to get the desired inequality as in Equation (11). It remains to upper bound:

$$\mathop{\mathbf{E}}_{x\in\mathcal{N}^k}\left[A_u(\pi_e(x))A_v(x)\right] = \mathop{\mathbf{E}}_{x\in\mathcal{N}^k}\left[(l_u(\pi_e(x)) + (A_u(\pi_e(x)) - l_u(\pi_e(x))))\cdot(l_v(x) + (A_v(x) - l_v(x)))\right]. \tag{12}$$

Note that since $\pi_e(I_u)\cap I_v = \phi$, we have $\mathbf{E}_{x\in\mathcal{N}^k}\left[l_u(\pi_e(x))l_v(x)\right] = 0$. Using Cauchy-Schwarz inequality, we bound

$$\mathop{\mathbf{E}}_{x\in\mathcal{N}^k}\left[l_u(\pi_e(x))(A_v(x) - l_v(x))\right] \leq \sqrt{\Gamma + 100s_0}\cdot\|A_u\|_2\|A_v\|_2 \leq \sqrt{\Gamma + 100s_0}\|A_e\|_2^2,$$

and the remaining term is bounded similarly. Thus we get an overall bound of $2\sqrt{\Gamma + 100s_0}\|A_e\|_2^2$ on (12) which is good enough since $\Gamma = 0.05$ and s_0 can be chosen to be sufficiently small.

Sub-case remaining: It remains to consider edges $e = (u, v) \in E$ where both A_u and A_v are $(\frac{10}{\Gamma\delta^2}, 100s_0)$-approximate linear juntas with $\pi_e(I_u) \cap I_v \neq \phi$. The fraction of such edges is at most $O(\varepsilon/\delta^4)$, since assigning a label $i \in I_u, j \in I_v$ at random satisfies the edge with probability $\Omega(\delta^4)$ and the soundness of the Unique Game instance is at most ε. These edges are classified as being in E_3, completing the proof.

6.3 Discretization

Let us briefly explain how the construction can be discretized. Define $L \doteq kb$, $\alpha = \gamma\delta/3b$. To obtain a discrete construction, for every vertex $v \in V$, replace $\mathbb{R}^k$ with a tiling of $[-L, L)^k$ by the cube $[0, \alpha)^k$. The new variables correspond to representatives of the shifted cube $[0, \alpha)^k$. In every equation, replace each occurrence of a variable with the appropriate representative. Replace each equation that depends on a variable outside of $[-L, L)^k$ by an equation $0 = 0$. Note that the probability that a Gaussian $x \sim \mathcal{N}^k$ falls outside of the cube $[-L, L)^k$ is at most $\frac{2}{\sqrt{2\pi}b}e^{-k^2b^2/2} \leq \delta/4b^2$.

Since k, b, γ and δ are constants, the construction is of polynomial size. Completeness and soundness follow from the completeness and soundness of the non-discrete construction: In the completeness case, by assigning the representatives their dictator values, the values effectively substituted to the other variables may shift by α compared to their original dictator values. This may cause equations that were exactly satisfied to become only 3α-approximately satisfied. It may also change the squared norm (on each equation, and on average over all equations), by an additive $O(\alpha b) \leq O(\gamma\delta)$. Additionally, we may lose the norm on the equations that were replaced with $0 = 0$, but this norm is at most $O(\delta)$. Using appropriate normalization of the dictators, we attain $val^{\gamma}_{(X,\mathcal{E})} \geq 1 - O(\delta)$.

In the soundness case, an assignment to the discretized construction induces an assignment to the non-discretized construction, and one can apply the soundness analysis we have. One needs to account for the norm on equations that were replaced by $0 = 0$, but again this norm is at most $O(\delta)$. This concludes the proof of Theorem 6.

References

[AKK+08] S. Arora, S. A. Khot, A. Kolla, D. Steurer, M. Tulsiani, and N. Vishnoi. Unique games on expanding constraint graphs are easy: extended abstract. In *Proc. 40th ACM Symp. on Theory of Computing*, pages 21–28, 2008.

[BCH+96] M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi, and M. Sudan. Linearity testing in characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996.

[BLR93] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993.

[GR09] V. Guruswami and P. Raghavendra. Hardness of solving sparse overdetermined linear systems: A 3-query PCP over integers. *ACM Trans. Comput. Theory*, 1(2):1–20, 2009.

[Hås01] J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.

[HK04] J. Holmerin and S. Khot. A new PCP outer verifier with applications to homogeneous linear equations and max-bisection. In *Proc. 36th ACM Symp. on Theory of Computing*, pages 11–20, 2004.

[Kho02] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symp. on Theory of Computing*, pages 767–775, 2002.

[KKMO07] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other two-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de