$\mathcal{NP}$-Hardness of Approximately Solving Linear Equations Over Reals [*]

Subhash Khot [†] Dana Moshkovitz [‡]

August 31, 2011

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: 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 1.5. 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

[*] This is a new and improved version of our paper [KM10] that established the same result, but under the Unique Games Conjecture.

[†] 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.

[‡] dmoshkov@mit.edu. Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

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 (w.r.t. A) is $\text{Margin}(A, eq) \doteq |r_1A(x_1) + r_2A(x_2) + r_3A(x_3)|$. 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) There exist universal constants b_0, c ($b_0 = 2$ works) such that for every $\delta > 0$, given a b_0-regular* 3Lin($\mathbb{R}$) *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 3Lin($\mathbb{R}$) 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 poly(δ) 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:

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

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:

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

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 $\forall x_i \in X, \|v_i\| \le b$, after rounding all but $\mathrm{poly}(\delta)$ 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.

Optimality of Semidefinite Programming Based Algorithms: As shown by Raghavendra [Rag08], the Unique Games Conjecture, if true, implies that for every constraint satisfaction problem[1], a certain semi-definite programming based algorithm gives the best efficient approximation for the problem (as long as $\mathcal{P} \neq \mathcal{NP}$). Similar results hold for many other types of problems, e.g., certain covering and ordering problems. In light of this, a natural question is whether one can prove for *specific* problems that an SDP-based algorithm is optimal, assuming only $\mathcal{P} \neq \mathcal{NP}$, and not relying on the Unique Games Conjecture.

Zwick [Zwi98] gave several examples of constraint satisfaction problems where each constraint depends on three variables, for which the natural semi-definite programming algorithm (with a particular rounding) yields the best possible approximation, assuming $\mathcal{P} \neq \mathcal{NP}$. These examples include the AND function and the Majority function on three variables.

Our work can be seen as continuing this line of work, showing optimality of SDP for the 3Lin($\mathbb{R}$) problem.

The Squared-ℓ_2 versus ℓ_1 Error: The SDP algorithm described above finds an assignment that minimizes the expected squared margin, i.e. $\mathbf{E}_{j\in[m]}\left[\mathrm{Margin}(A,j)^2\right]$. Thus the problem of minimizing the squared-ℓ_2 error is a computationally easy problem. However, Theorem 1 implies that minimizing the ℓ_1 error (i.e. $\mathbf{E}_{j\in[m]}\left[\mathrm{Margin}(A,j)\right]$), even approximately, is computationally hard (assuming $\mathcal{P} \neq \mathcal{NP}$). 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 $O(\delta\log(1/\delta))$.[2] 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

[1]where variables range over a constant sized alphabet, and each constraint depends on a constant number of variables.

[2]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.

$\Omega(\sqrt{\delta})$. Thus approximating the ℓ_1 error within a quadratic factor is computationally hard; 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

1.3.1 Dictatorship Test Over Reals

Similar to most 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 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*.

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:

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

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 The Reduction

The NP-hardness proof proceeds by using the dictatorship test discussed in the previous section as a gadget in a reduction. One might expect the reduction to go along the lines of Håstad's reduction for the Boolean 3Lin, however the real case confronts us with serious challenges. A key component in Håstad's reduction addresses the following problem (in the Boolean case):

The Restriction Problem. Given oracle access to a function $f : \mathbb{F}^n \to \mathbb{F}$ that is approximately a dictatorship function (for the sake of exposition, assume that for some $i_0 \in [n]$, on most points $y \in \mathbb{F}^n$ we have $f(y) = y_{i_0}$), and to a function $g : \mathbb{F}^m \to \mathbb{F}$, $m \cdot \ell = n$, test whether g is the following *restriction* of f:

$$g(x_1, \ldots, x_m) = f(x_1, \ldots, x_1, \ldots\ldots, x_m, \ldots, x_m),$$

where each x_i repeats ℓ times. The test should check a linear equation on three values of f and g (altogether).

The restriction problem can be solved in the Boolean case $\mathbb{F} = \{0, 1\}$ and for any finite field $\mathbb{F}$ via *self-correction*. The tester is as follows:

1. Pick $x \in \mathbb{F}^m$, $y \in \mathbb{F}^n$ uniformly at random.
2. Set $z = (x_1, \ldots, x_1, \ldots\ldots, x_m, \ldots, x_m) \in \mathbb{F}^n$.
3. Accept if and only if
$$g(x_1, \ldots, x_m) = f(y) + f(z - y).$$

Note that when f is a dictatorship function and g is the appropriate restriction of it, the test always accepts (in fact, linearity of f suffices). Also note that the test is linear in three values of f and g. The test works also when f is *close* to a dictatorship function $\tilde{f}$, because the points y and $z - y$ are uniformly distributed in $\mathbb{F}^n$, and with high probability, f evaluates to the correct dictatorship function $\tilde{f}$ at both the points. Note that z itself is not uniformly distributed in $\mathbb{F}^n$, but still $f(y) + f(z - y)$ yields, with high probability, the correct value $\tilde{f}(z)$.

Now consider the analogous problem for functions in Gaussian space. In this case, we can at most gurantee that with high probability over $y \sim \mathcal{N}^n$ it holds that $f(y) \approx y_{i_0}$. The tester we showed for the finite field case no longer works: even when $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$ are Gaussian distributed, the point $z - y$ may not be distributed as a Gaussian in $\mathbb{R}^n$. We instead proceed as follows. Define a subspace S of $\mathbb{R}^n$ as:

$$S := \{(x_1, \ldots, x_1, \ldots\ldots, x_m, \ldots, x_m) \mid x_1, \ldots, x_m \in \mathbb{R}\},$$

where each x_i repeats ℓ times. Let $\pi : S \mapsto \mathbb{R}^m$ denote the projection that for $1 \leq i \leq m$, picks the common coordinate from the i^{th} block. The tester is as follows:

1. Pick $y \sim \mathcal{N}^n$.
2. Write $y = y^{||} + y^{\perp}$, where $y^{||} \in S$ and $y^{\perp} \in S^{\perp}$. Set $y' = y^{||} - y^{\perp}$ and let $y^{\downarrow} \in \mathbb{R}^m$ be the vector $y^{\downarrow} := \sqrt{\ell} \cdot \pi(y^{||})$. It is easily seen that $y^{\downarrow}$ is distributed as $\mathcal{N}^m$.
3. Check that
$$g(y^{\downarrow}) = \sqrt{\ell} \cdot \frac{f(y) + f(y')}{2}.$$

It can be easily checked that if f is a dictatorship and g its appropriate restriction, then the test equation holds. Note that y, y' are both Gaussian distributed, and thus if f is close to a dictatorship $\tilde{f}$, then with high probability $f(y) \approx \tilde{f}(y)$ and $f(y') \approx \tilde{f}(y')$ and $\sqrt{\ell} \cdot \frac{f(y)+f(y')}{2} \approx \sqrt{\ell} \cdot \frac{\tilde{f}(y)+\tilde{f}(y')}{2} = g(y^{\downarrow})$ if g is the appropriate restriction of $\tilde{f}$. One caveat however is that the error involved in the approximating $\tilde{f}$ by f gets multiplied by $\sqrt{\ell}$ in this calculation and if ℓ is too large, the equation becomes rather meaningless.

How large is ℓ in hardness applications? This parameter corresponds to the "Outer PCP" (aka Label Cover) being "ℓ to 1". In standard hardness results, such as Håstad's, one uses the Parallel Repetition Theorem [Raz98] and $\ell = (1/\varepsilon)^{O(1)}$, where ε is the soundness error of the Outer PCP. Moreover, the soundness error ε usually needs to be tiny, which in turn requires ℓ to be large, and this is prohibitive in our application.

To avoid having large ℓ, we do not use parallel repetition, and work instead with the basic PCP Theorem [AS98, ALM+98]. This PCP has high soundness error (say 0.99), but is adequate for the purpose of proving Theorem 1. The reason is that Theorem 1 is also a "hige error" hardness result – we only guarantee in the NO case that a constant fraction of the equations (say 1%) fail with a good margin.

Still, working with a high error PCP seems impossible at first sight. The dictatorship test gives rise to a *list decoding* of possible dictatorship functions, rather than identifying a single dictatorship function, and this seems to call for an Outer PCP with low error. Indeed, virtually all existing hardness results rely on PCP with low error for the same reason (where one of the dictatorship coordinates in the decoded list is picked at random as a candidate label/answer for the Outer PCP). To circumvent the need for a low error PCP, we build a new Outer PCP. Suppose that the basic PCP corresponds to a set of variables $\mathcal{Z}$, a set of tests/constraints $\mathcal{C}$, and each test depends on d variables. The new Outer PCP is as follows:

1. The verifier picks independently at random k possible tests $c_1, \ldots, c_k \in \mathcal{C}$, an index $i \in [k]$, and a variable z in the test c_i.

2. The verifier sends the tuple $u = (c_1, \ldots, c_k)$ to the first prover and the tuple $(c_1, \ldots, c_{i-1}, z, c_{i+1}, \ldots, c_k)$ to the second prover.

3. Both provers are supposed to answer with the values of all the variables in the tuple they were given.

4. The verifier checks that provers' answers are consistent and satisfy the tests.

Note that this outer PCP is as sound as the basic PCP. Moreover, it is "$\ell = d$ to 1" where each constraint depends on d variables for a fixed constant d. The crux of the analysis is that a short list of each prover's answers in this PCP translate (with high probability) into just one answer for a random coordinate $i \in [k]$ on which the basic PCP test is actually performed. Thus, via this Outer PCP, we convert the list decoding setting into a unique decoding setting, and allow the reduction to go through. We make the argument formal by using the technique of correlated sampling [KT02, Hol09] to choose a consistent element from two lists, one for each prover.

Due to the specific Outer PCP construction, our reduction maps instances of SAT of size N to instances of 3LIN($\mathbb{R}$) of size $N^{O(k)}$, $k = (1/\delta)^{O(1)}$, where δ is the parameter of Theorem 1. Hence, the reduction incurs a blow-up of $N^{(1/\delta)^{O(1)}}$ in the size. This blow-up matches the blow-up predicted by the recent work of Arora, Barak and Steurer [ABS10] for unique games.

We remark that the actual analysis is much more complex than hinted here. The reason is that the 3LIN($\mathbb{R}$) instance constructed by the reduction consists of several functions $f : \mathbb{R}^n \mapsto \mathbb{R}$ that could have widely varying norms, whereas list decoding via dictatorship testing can be extracted only from functions with non-negligible norms, and the eventual prover strategies have to be weighted delicately according to these norms.

1.5 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 (which we do) and then (2) give a gap-preserving reduction from the 3LIN($\mathbb{R}$) to 2LIN($\mathbb{R}$). 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 MinUncut, 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($GF(2)$).

1.6 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 reduction, proving our main result, is presented in Section 5. The soundness analysis is first presented in a simplified setting and then in the general setting. The entire reduction is presented in a continuous setting and then discretized in Section 5.7.

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:*

$$r_1 x_1 + r_2 x_2 + r_3 x_3 = 0,$$

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*

$$r_1 A(x_1) + r_2 A(x_2) + r_3 A(x_3) = 0.$$

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

$$|r_1 A(x_1) + r_2 A(x_2) + r_3 A(x_3)| \leq \beta.$$

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* $|r_1 A(x_1) + r_2 A(x_2) + r_3 A(x_3)|$.

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[3] the total squared norm of equations that are satisfied with margin of at most β.

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($\mathbb{R}$))**.** *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*

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

[3] 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.

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 Distributions: Entropy and Distance

The *entropy* of a probability distribution D over a discrete probability space Ω is

$$H(D) \doteq \sum_{a \in \Omega} D(a) \log \frac{1}{D(a)}.$$

Entropy satisfies the following properties:

Proposition 2.5. *[CT91] For distributions $D, D_1, \ldots, D_k$ over Ω,*

- *Range: $0 \leq H(D) \leq \log|\Omega|$; the lower bound is attained by constant distributions; the upper bound is attained by the uniform distribution.*
- *Concavity: $H(\frac{1}{k}\sum_{i=1}^k D_i) \geq \frac{1}{k}\sum_{i=1}^k H(D_i)$.*
- *Sub-additivity: $H(D_1 \ldots D_k) \leq \sum_{i=1}^k H(D_i)$.*

The *statistical distance* between distributions D_1 and D_2 over a discrete probability space Ω is

$$\Delta(D_1, D_2) \doteq \frac{1}{2}\sum_{a \in \Omega} |D_1(a) - D_2(a)|.$$

A distribution with nearly maximal entropy is close to uniform:

Proposition 2.6. *[CT91]*

$$\log |\Omega| - H(D) \geq \frac{1}{2 \ln 2} \|D - \mathsf{Uniform}\|_1^2.$$

The squared *Hellinger distance* between D_1 and D_2 is

$$\Delta_H^2(D_1, D_2) \doteq \frac{1}{2} \sum_{a \in \Omega} \left(\sqrt{D_1(a)} - \sqrt{D_2(a)} \right)^2 = 1 - \sum_{a \in \Omega} \sqrt{D_1(a) D_2(a)}.$$

We have the following connection between the Hellinger distance and the statistical distance:

Proposition 2.7. *[Pol02]*

$$\Delta_H^2(D_1, D_2) \leq \Delta(D_1, D_2) \leq \sqrt{2} \Delta_H(D_1, D_2).$$

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

- $\|f^{=1}\|_2^2 = \sum_{i=1}^n a_i^2 \geq (1-s)\|f\|_2^2$.
- $\sum_{i: a_i^2 \leq \frac{1}{J}\|f\|_2^2} a_i^2 \leq \Gamma \cdot \|f\|_2^2$.

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 8 (Dictator testing)**.** *For every constant $0 < \Gamma \leq 0.01$, there are 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 $(\frac{10}{\Gamma\delta^2}, s, \Gamma)$-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 $(\frac{10}{\Gamma\delta^2}, s, \Gamma)$-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, $\mathbf{E}_{eq \sim \mathcal{E}_c}\left[\chi_{|eq|>0} \cdot \|f_{eq}\|_2^2\right] \le \delta \|f\|_2^2 = \delta$.

Soundness: In the following, $O(\cdot)$ and $\Omega(\cdot)$ notations will hide universal constants. Let Γ be the given constant in Theorem 8. We will pick s and c to be constants eventually, but throughout the proof, retain the dependence on the parameters. Assume for now that $2c \le s \le 0.01$ and $\sqrt[3]{s} \ll \sqrt{\Gamma}$. 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 $(J = \frac{10}{\Gamma\delta^2}, s, \Gamma)$-approximate linear junta. Assume, for the sake of a contradiction, that

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

Denote the non-linear part of f by $e = f - f^{=1}$ (since f is anti-symmetric, $f^{\le 0} \equiv 0$). We handle the cases that $\|e\|_2^2 \le s$ 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,[4] 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 $1 - O(\sqrt[3]{s})$,*

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

[4] 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

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

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

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

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

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

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

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

□

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*

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

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

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

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 $\Omega(\sqrt{\Gamma\delta})$ (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 $\Omega(\sqrt{\Gamma\delta})$ in magnitude with probability $\Omega(1)$ and we are done.

Let $I = \left\{ i \in [n] \mid a_i^2 \le \frac{1}{J} \right\}$ be the "non-influential" coordinates. Since f is not a (J, s, Γ)-approximate linear junta, and $\|e\|_2^2 \le s$, we must have $\sum_{i \in I} a_i^2 \ge \Gamma$. 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 $\Gamma\delta$.

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

where we noted that $\sum_{i\in I} a_i^4 \le \frac{1}{J}\sum_{i\in I} a_i^2$ 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. $\mathbf{E}_{(x,\tilde{x})\sim R}\left[|e'(x) - e'(\tilde{x})|^2\right] \le O\left(\frac{\sqrt[3]{s}}{\sqrt{\Gamma}}\right)$,
2. $\mathbf{E}_{(x,\tilde{x})\sim C}\left[|e'(x) - e'(\tilde{x})|^2\right] \ge \Omega(1)$,

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 4.1. For sufficiently small s, Lemma 4.4 leads to a contradiction by the following claim:

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

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

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 $\rho = 1 - \delta$:

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

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 8 assuming Lemma 4.4.

4.1 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
$$|e(x) - e(\tilde{x})| \geq d_C = \Omega(\sqrt{\Gamma\delta}). \tag{7}$$

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

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

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, $|e(x)|, |e(\tilde{x})| \leq 1$. (ii) $|e(x) - e(\tilde{x})| \leq d_R$. 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 $w_R(E_R) \geq 1 - O(\sqrt[3]{s})$ from Hypothesis (6) and the observation that $\|e\|_2^2 \leq s$ and thus for $x \in \mathcal{N}^n$, $|e(x)| \leq 1$ 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, $|e(x)|, |e(\tilde{x})| \leq 1$. (ii) $|e(x) - e(\tilde{x})| \geq d_C$. 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 $w_C(E_C) \geq \Omega(1)$ from Hypothesis (7) and since $\|e\|_2^2 \leq s$.

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:) $\quad w_R(\mathcal{C}) \leq O\left(\frac{\sqrt[3]{s}}{\sqrt{\Gamma}}\right)$.
2. (Large E_C weight is cut) $\quad w_C(\mathcal{C}) \geq \Omega(1)$.

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 $w_R(\mathcal{C}) + (1 - w_R(E_R)) \leq O\left(\frac{\sqrt[3]{s}}{\sqrt{\Gamma}}\right)$. When choosing $(x, \tilde{x})$ as in the coordinatewise perturbation test, the probability the pair $(x, \tilde{x})$ is separated is at least $w_C(\mathcal{C}) \geq \Omega(1)$.

Lemma 4.6. *There is a distribution over cuts such that:*

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

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 $e(x) \geq \lambda$, 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. □

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 4.6.
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 4.7. *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 $p_R \leq O\left(\frac{\sqrt[3]{s}}{\sqrt{\Gamma}}\right)$.*
- *For every edge $(x, \tilde{x}) \in E_C$, the probability that $(x, \tilde{x})$ is cut by $\mathcal{C}$, is at least $p_C \geq \Omega(1)$.*

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 4.6, 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 $M \cdot p_{R,0} \leq O\left(\frac{\sqrt[3]{s}}{\sqrt{\Gamma}}\right)$.

If $(x, \tilde{x}) \in E_C$, then by Lemma 4.6 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 $\mathcal{C}$. $\square$

The above Lemma 4.7 shows that

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

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

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

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

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.

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

We now show the NP-hardness of Robust-3Lin($\mathbb{R}$) and prove Theorem 6. The reduction is from hardness of "GapCSP" that follows from the PCP Theorem.

5.1 Constraint Satisfaction Problems.

A constraint satisfaction problem (CSP) is given by a set of variables Z and a set of constraints $\mathcal{C}$. Each constraint depends on d variables, where d is a parameter. Each variable takes values from some finite alphabet Σ. A constraint $c \in \mathcal{C}$ restricts the set of assignments its variables may assume, thus defining some subset of Σ^d. A CSP can be represented as a bipartite graph $G = (\mathcal{C}, Z, E)$, where there is an edge between a constraint $c \in \mathcal{C}$ and a variable $z \in Z$ if z appears in the constraint c. We call an edge (c, z) a *test*. The degrees of the $\mathcal{C}$ vertices are d. We will refer to *regular* CSPs, in which the degrees of all the Z vertices are the same as well. I.e., every variable appears in the same number of constraints. An assignment to a constraint is an assignment to its variables that satisfies the constraint (i.e., is in the subset of Σ^d defined by the constraint). We say that an assignment to the constraint is *consistent* with an assignment to a variable z in it, if restricted to the variable, the two assignments are the same. We will be interested in the value of the CSP, i.e., under the best (maximizing) assignments to $\mathcal{C}$ and Z, what is the fraction of edges $e = (c, z)$ that give rise to consistent assignments? The following hardness result is well-known:

Theorem 9 (PCP Theorem, [BFLS91, AS98, ALM+98])**.** *There are a finite alphabet Σ, a dependency d, and a constant $\eta < 1$, such that given a CSP instance $(Z, \mathcal{C})$, it is NP-hard to distinguish between the case that its value is 1 and the case that its value is at most η. One may take $\Sigma = \{0, 1\}$ and $d = 3$.*

Remark 5.1. *The standard starting point for most hardness reductions is the so-called Label Cover problem with low soundness. The problem is known to be hard by combining the PCP Theorem with Raz's Parallel Repetition Theorem [Raz98]. However, our reduction uses only the basic PCP Theorem and the soundness η could be close to* 1.

5.2 The Reduction (PCP Construction)

We reduce a CSP instance $(Z, \mathcal{C})$ as in Theorem 9, to a Robust-3Lin($\mathbb{R}$) instance $(X, \mathcal{E})$. To simplify the presentation, we show a non-discretized construction, having variables for all points in real-space. We later explain how to discretize the construction.

Let $k = \frac{10^4}{\Gamma^2 \delta^4}$ be a parameter (Γ is the global constant from the definition of an approximate linear junta; δ is from the statement of Theorem 6). Denote the number of bits representing an assignment to the variables in $(k+1)$ constraints of the CSP by $N \doteq (k+1) d \log |\Sigma|$. Denote the difference between the number of bits required to represent an assignment to all the variables in a constraint and the number of bits required to represent an assignment to just one variable, by $\Delta \doteq (d-1) \log |\Sigma|$. Note that with $\Sigma = \{0,1\}$ and $d = 3$, we may take $\Delta = 2$. The construction of the Robust-3Lin($\mathbb{R}$) instance is as follows (note that it incurs a blow-up of exponent k in the size, compared to the CSP instance):

Variables: There are two types of variables:

U **variables:** For every choice of $(k+1)$ constraints of the CSP, $u = (c_1, \ldots, c_{k+1})$, and every $x \in \mathbb{R}^{2^N}$, there is a variable. We denote the assignment to those $\mathbb{R}^{2^N}$ variables associated with u by $A_u : \mathbb{R}^{2^N} \to \mathbb{R}$. Supposedly, $A_u(x) = x_a$ where a is an assignment to the variables of u (in bit representation). We assume, by *folding*[5], that:

- A_u is anti-symmetric, i.e. $\forall x \in \mathbb{R}^{2^N},\ A_u(-x) = -A_u(x)$.
- A_u corresponds to a legal assignment a, in the following sense: Let $\mathcal{H}^u \subseteq \mathbb{R}^{2^N}$ be the subspace spanned by all standard basis vectors $e_a \in \mathbb{R}^{2^N}$ corresponding to assignments a to the variables of u (in bit representation) satisfying all the constraints $c_1, \ldots, c_{k+1}$.

  Then, for all $x \in \mathbb{R}^{2^N}$, for all $\nu \in (\mathcal{H}^u)^\perp \subseteq \mathbb{R}^{2^N}$,

$$A_u(x + \nu) = A_u(x).$$

V **variables:** For every choice of k constraints, a coordinate $i \in [k+1]$, and a variable z of the CSP, $v = (c_1, \ldots, c_{i-1}, z, c_{i+1}, \ldots, c_{k+1})$, and every $x \in \mathbb{R}^{2^{N-\Delta}}$, there is a variable. We denote the assignment to those $\mathbb{R}^{2^{N-\Delta}}$ variables associated with v by $A_v : \mathbb{R}^{2^{N-\Delta}} \to \mathbb{R}$. Supposedly, $A_v(x) = x_{a'}$ where a' is an assignment to the variables of v (in bit representation). We again use folding to ensure:

- A_v is anti-symmetric, i.e. $\forall x \in \mathbb{R}^{2^{N-\Delta}},\ A_v(-x) = -A_v(x)$.

[5]Folding means we can ensure $A_u(-x) = -A_u(x)$ by letting a single variable and its negation represent the two values, instead of having two separate variables. Similarly, we can ensure $A_u(x + \nu) = A_u(x)$ by letting the same variable represent both values.

- A_v corresponds to a legal assignment a', in the following sense: Let $\mathcal{H}^v \subseteq \mathbb{R}^{2^{N-\Delta}}$ be the subspace spanned by all standard basis vectors $e_{a'} \in \mathbb{R}^{2^{N-\Delta}}$ corresponding to assignments a' to the variables of v (in bit representation) satisfying all the cosntraints $c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1}$.

  Then, for all $x \in \mathbb{R}^{2^{N-\Delta}}$, for all $\nu \in (\mathcal{H}^v)^\perp \subseteq \mathbb{R}^{2^{N-\Delta}}$,

$$A_v(x+\nu) = A_v(x).$$

Equations: The distribution over equations: Pick independently at random CSP constraints, $c_1, \ldots, c_{k+1} \in \mathcal{C}$, a distinguished constraint, $i \in [k+1]$, and a variable z appearing in the constraint c_i. Let $u = (c_1, \ldots, c_{k+1})$, $v = (c_1, \ldots, c_{i-1}, z, c_{i+1}, \ldots, c_{k+1})$, $e = (u, v)$. 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 8 with parameter δ.
- $\mathcal{E}_v$: Perform dictator testing on A_v as in Theorem 8 with parameter δ.
- $\mathcal{E}_e$: A_v is supposed to encode an assignment to all the variables in u, except for the $(d-1)$ variables missing in v. Let $I \subseteq \mathbb{R}^{2^N}$ be the subspace consisting of all points x where $x_a = x_b$ if a and b agree on the assignment to the variables in v (this is the subspace that corresponds to A_v).

  Pick $x \sim \mathcal{N}^{2^N}$. Write

$$x = x^{||} + x^{\perp},$$

  for $x^{||} \in I$, $x^\perp \in I^\perp$. Let $x' = x^{||} - x^\perp$. Let $x^\downarrow \in \mathbb{R}^{2^{N-\Delta}}$ be such that $x^\downarrow_{a'} = 2^{\Delta/2} \cdot x^{||}_a$ for every assignment $a' \in \{0,1\}^{N-\Delta}$ and assignment $a \in \{0,1\}^N$ where a and a' are consistent on the variables in v (note that by the definition of $x^{||}$ it does not matter which a one picks). Produce the equation:

$$2^{\Delta/2} \cdot \frac{A_u(x) + A_u(x')}{2} = A_v(x^\downarrow).$$

Note that the normalization factors are introduced appropriately so that $x^\downarrow \sim \mathcal{N}^{2^{N-\Delta}}$. Note also that a random query lands in U with probability $\frac{5}{9}$, then uniform over $u \in U$, and then for a fixed $u \in U$, Gaussian distributed over $\mathbb{R}^{2^N}$. A random query lands in V with probability $\frac{4}{9}$, then uniform over $v \in V$ (by regularity of the CSP instance), and then for a fixed $v \in V$, Gaussian distributed over $\mathbb{R}^{2^{N-\Delta}}$.

Let s_0 (slightly redefined) and c_0 be the constants for the dictator testing theorem, Theorem 8, so for any n, for any anti-symmetric function $f \in L^2(\mathbb{R}^n, \mathcal{N}^n)$,

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

Note that Theorem 8 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 constant Γ itself will be chosen to be small enough. The constants s and c for the Robust-3Lin($\mathbb{R}$) hardness theorem, Theorem 6, depend appropriately on s_0 and c_0.

5.3 Properties of Folded Assignments

In this section we prove some properties of the assignments that follow from folding. The first property is that the linear parts of the assignments are themselves folded:

Claim 5.1 (linear part folded)**.** *For every $u = (c_1, \ldots, c_{k+1}) \in \mathcal{C}^{k+1}$, for every $x \in \mathbb{R}^{2^N}$, for every $\nu \in (\mathcal{H}^u)^\perp$,*

$$A_u^{=1}(x + \nu) = A_u^{=1}(x).$$

Proof. Note that by linearity it suffices to prove that $A_u^{=1}(\nu) = 0$.

$$A_u^{=1}(\nu) = \sum_{i=1}^{2^N} \mathop{\mathbf{E}}_{x \sim \mathcal{N}^n} [A_u(x) x_i] \nu_i = \mathop{\mathbf{E}}_{x \sim \mathcal{N}^n} [A_u(x) \langle x, \nu \rangle].$$

We can write every $x \in \mathbb{R}^{2^N}$ as $x = x^{||} + x^\perp$ where $x^{||} \in \mathcal{H}^u$ and $x^\perp \in (\mathcal{H}^u)^\perp$, and get:

$$\mathop{\mathbf{E}}_{x \sim \mathcal{N}^n} [A_u(x) \langle x, \nu \rangle] = \mathop{\mathbf{E}}_{x \sim \mathcal{N}^n} \left[A_u(x^{||} + x^\perp) \langle x^\perp, \nu \rangle \right].$$

Since A_u is folded:

$$\mathop{\mathbf{E}}_{x \sim \mathcal{N}^n} \left[A_u(x^{||} + x^\perp) \langle x^\perp, \nu \rangle \right] = \mathop{\mathbf{E}}_{x \sim \mathcal{N}^n} \left[A_u(x^{||}) \langle x^\perp, \nu \rangle \right].$$

But $x^{||}$ and $x^\perp$ are independent, and since $-x^\perp \in (\mathcal{H}^u)^\perp$, we have

$$\mathop{\mathbf{E}}_{x^\perp} \left[\langle x^\perp, \nu \rangle \right] = \mathop{\mathbf{E}}_{x^\perp} \left[\frac{\langle x^\perp, \nu \rangle + \langle -x^\perp, \nu \rangle}{2} \right] = 0.$$

Thus, $A_u^{=1}(\nu) = 0$. □

Similarly, one can prove:

Claim 5.2 (linear part folded)**.** *For every $v = (c_1, \ldots, c_{i-1}, z, c_{i+1}, \ldots, c_{k+1})$, for every $x \in \mathbb{R}^{2^{N-\Delta}}$, for every $\nu \in (\mathcal{H}^v)^\perp$,*

$$A_v^{=1}(x + \nu) = A_v^{=1}(x).$$

The second property we observe is a decomposition of the linear part into summands corresponding to satisfying assignments for the CSP instance:

Claim 5.3 (linear part decomposed)**.** *Let $u = (c_1, \ldots, c_{k+1})$. Write the coefficient vector of $A_u^{=1}$ as $f_u = \sum_{a \in \{0,1\}^N} \hat{f}_u(a) e_a$. Then $\hat{f}_u(a) \neq 0$ implies that $a \in \{0,1\}^N$ is a satisfying assignment to the variables in u (in bit representation).*

Proof. By Claim 5.1, $A_u^{=1} \in \mathcal{H}^u$. The claim follows from the definition of $\mathcal{H}^u$. □

Claim 5.4 (linear part decomposed)**.** *Let $v = (c_1, \ldots, c_{i-1}, z, c_{i+1}, \ldots, c_{k+1})$. Write the coefficient vector of $A_v^{=1}$ as $f_v = \sum_{a' \in \{0,1\}^{N-\Delta}} \hat{f}_v(a') e_{a'}$. Then $\hat{f}_v(a') \neq 0$ implies that $a' \in \{0,1\}^{N-\Delta}$ is a satisfying assignment to the variables in v (in bit representation).*

5.4 Completeness

Assume that there is an assignment $A_0 : Z \to \Sigma$ to the CSP variables that satisfies all the constraints in $\mathcal{C}$. We construct from it an assignment $A : X \to \mathbb{R}$ for the Robust-3Lin($\mathbb{R}$) instance $(X, \mathcal{E})$: For every vertex $u \in U$, let $A_u = x_a$ where a is the assignment of A_0 to the variables of u (in bit representation). Note that the folding constraints hold and $\|A_u\|_2^2 = 1$.

For every vertex $v \in V$, let $A_v = x_{a'}$ where a' is the assignment A_0 to the variables of v (in bit representation). Note that the folding constraints hold and $\|A_v\|_2^2 = 1$.

Hence, $\|A\|_2^2 = 1$. Consider CSP constraints $c_1, \ldots, c_{k+1} \in \mathcal{C}$, a distinguished constraint, $i \in [k+1]$, and a variable z appearing in the constraint c_i. Let $u = (c_1, \ldots, c_{k+1})$, $v = (c_1, \ldots, c_{i-1}, z, c_{i+1}, \ldots, c_{k+1})$, $e = (u, v)$. By Theorem 8,

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

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

The equations from $\mathcal{E}_e$ are exactly satisfied. This is because $x = x^{||} + x^{\perp}$ and

$$2^{\Delta/2} \cdot \frac{A_u(x) + A_u(x')}{2} = 2^{\Delta/2} \cdot A_u(x^{||}) = 2^{\Delta/2} \cdot x_a^{||} = x_{a'}^{\downarrow} = A_v(x^{\downarrow}).$$

Thus,

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

Overall, we have $val^0_{(X,\mathcal{E})} \geq 1 - \delta$. 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.

5.5 Soundness: Simplified Setting

Assume that for any assignment to the CSP instance, at most η fraction of the (constraint,variable) pairs are consistent. Fix an assignment $A : X \to [-b, b]$, $\|A\|_2^2 = 1$. We first consider a simplified setting in which for every u and v, $\|A_u\|_2^2 = 1$, $\|A_v\|_2^2 = 1$. This setting will allow us to demonstrate the main idea of the proof without getting into many of the technicalities that the general case involves. We will show that

$$\mathop{\mathbf{E}}_{eq\sim\mathcal{E}}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right] \geq (1 - \sqrt[3]{\eta})s.$$

Note that this is enough by slightly redefining s and since $\eta < 1$ is an absolute constant. Rewrite the above inequality as:

$$\mathop{\mathbf{E}}_{e}\left[\mathop{\mathbf{E}}_{eq\sim\mathcal{E}^e}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right]\right] \geq (1 - \sqrt[3]{\eta})s. \tag{8}$$

In the sequel, we will partition the e's into two sets $E_1 \cup E_2$ where the fraction of E_2 is at most $\frac{J^2}{k+1} + \sqrt{\eta}$ and $J := \frac{10}{\Gamma\delta^2}$. The latter expression can be made smaller than $\sqrt[3]{\eta}$ for sufficiently large k. Thus, it suffices to show that the contribution of every edge $e \in E_1$ towards (8) is lower bounded as:

$$\mathop{\mathbf{E}}_{eq\sim\mathcal{E}^e}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right] \;\geq\; 2s. \tag{9}$$

Pick independently at random CSP constraints $c_1, \ldots, c_{k+1} \in \mathcal{C}$, a distinguished constraint, $i \in [k+1]$, and a variable z appearing in the constraint c_i. Let $u = (c_1, \ldots, c_{k+1})$, $v = (c_1, \ldots, c_{i-1}, z, c_{i+1}, \ldots, c_{k+1})$, $e = (u, v)$.

Case A_u is not a $(\frac{10}{\Gamma\delta^2}, 100s_0, \Gamma)$-approximate linear junta. Since $\|A_u\|_2^2 = 1$, in this case we are done, since by the analysis of the dictatorship test,

$$\mathop{\mathbf{E}}_{eq\sim\mathcal{E}_u}\left[\chi_{|eq|>c_0\sqrt{\delta}}\|A_{eq}\|_2^2\right] \geq s_0 \geq 6s.$$

Therefore,

$$\mathop{\mathbf{E}}_{eq\sim\mathcal{E}^e}\left[\chi_{|eq|>c_0\sqrt{\delta}}\|A_{eq}\|_2^2\right] \geq 2s.$$

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

Thus we are left with the case where both A_u and A_v are $(\frac{10}{\Gamma\delta^2}, 100s_0, \Gamma)$-approximate linear juntas. Let $J \doteq \frac{10}{\Gamma\delta^2}$.

Write the coefficients vector of the linear part $A_u^{=1}$ as $f_u = \sum_{a\in\{0,1\}^N} \hat{f}_u(a)e_a$. Let $L_u \doteq \left\{a \in \{0,1\}^N \,\middle|\, \hat{f}_u(a)^2 \geq \frac{1}{J}\right\}$. Let A_u's approximating junta $G_u : \mathbb{R}^{2^N} \to \mathbb{R}$ be

$$G_u(x) \doteq \sum_{a\in L_u} \hat{f}_u(a)x_a.$$

Write the coefficients vector of the linear part $A_v^{=1}$ as $f_v = \sum_{a\in\{0,1\}^{N-\Delta}} \hat{f}_v(a)e_a$. Let $L_v \doteq \left\{a \in \{0,1\}^{N-\Delta} \,\middle|\, \hat{f}_v(a)^2 \geq \frac{1}{J}\right\}$. Let A_v's approximating linear junta $G_v : \mathbb{R}^{2^{N-\Delta}} \to \mathbb{R}$ be

$$G_v(x) \doteq \sum_{a\in L_v} \hat{f}_v(a)x_a.$$

Then,

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

Note that G_u and G_v contain at most J summands. For fixed u, over the choice of v, the probability that there exist assignments $a \neq b$ in G_u whose restrictions to v are identical is at most $\frac{J^2}{k+1}$. Let the edges e where this happens be in E_2, and let us assume from now on that this does not happen.

By folding (Claim 5.3), all a with non-zero coefficients in f_u (and hence in G_u) correspond to satisfying assignments to the variables of u, and (Claim 5.4) all a' with non-zero coefficients in f_v (and hence in G_v) correspond to satisfying assignments to the variables of v.

For $x \in \mathbb{R}^{2^N}$, let $x^{||}$, $x^{\perp}$, x', be as in the definition of the equations. Define $A_{uv}(x) \doteq \frac{A_u(x)+A_u(x')}{2}$, and

$$G_{uv}(x) \doteq \frac{G_u(x) + G_u(x')}{2} = \sum_{a\in L_u} \hat{f}_u(a)x_a^{||}.$$

Claim 5.5.

$$\|G_{uv} - A_{uv}\|_2^2 \leq (\Gamma + 100 s_0).$$

Proof. By Cauchy-Schwartz inequality,

$$\begin{aligned}
\|G_{uv} - A_{uv}\|_2^2 &= \mathop{\mathbf{E}}_{x}\left[\left(\frac{G_u(x)+G_u(x')}{2} - \frac{A_u(x)+A_u(x')}{2}\right)^2\right] \\
&= \mathop{\mathbf{E}}_{x}\left[\left(\frac{G_u(x)-A_u(x)}{2} + \frac{G_u(x')-A_u(x')}{2}\right)^2\right] \\
&\leq \frac{1}{2}\|G_u - A_u\|_2^2 + \frac{1}{2}\|G_u - A_u\|_2^2 \\
&\leq (\Gamma + 100 s_0).
\end{aligned}$$

□

Claim 5.6.

$$\|G_{uv}\|_2^2 = 2^{-\Delta}\|G_u\|_2^2.$$

Proof.

$$\begin{aligned}
\|G_{uv}\|_2^2 &= \mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2N}}\left[\left(\sum_{a\in L_u}\hat{f}_u(a)x_a^{||}\right)^2\right] \\
&= \mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2N}}\left[\sum_{a,b\in L_u}\hat{f}_u(a)\hat{f}_u(b)x_a^{||}x_b^{||}\right] \\
&= \sum_{a,b\in L_u}\hat{f}_u(a)\hat{f}_u(b)\mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2N}}\left[x_a^{||}\cdot x_b^{||}\right].
\end{aligned}$$

By our assumption, for every $a \neq b \in L_u$, $\mathbf{E}_{x\sim\mathcal{N}^{2N}}\left[x_a^{||}\cdot x_b^{||}\right] = 0$. Hence, we are left with:

$$\begin{aligned}
\|G_{uv}\|_2^2 &= \sum_{a\in L_u}\hat{f}_u(a)^2\cdot\mathop{\mathbf{E}}_{x}\left[(x_a^{||})^2\right] \\
&= 2^{-\Delta}\|G_u\|_2^2.
\end{aligned}$$

□

Case $\mathbf{E}_{x\sim\mathcal{N}^{2N}}\left[\left(2^{\Delta/2}G_{uv}(x) - G_v(x^{\downarrow})\right)^2\right] \geq 2^{\Delta/2+3}\cdot\sqrt{\Gamma + 100 s_0}$. First note that by Cauchy-Schwarz inequality,

$$\mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2N}}\left[\left(2^{\Delta/2}G_{uv}(x) - G_v(x^{\downarrow})\right)^2\right] \leq 2^{\Delta}\|G_{uv}\|_2^2 + \|G_v\|_2^2 + 2\cdot 2^{\Delta/2}\|G_{uv}\|_2\|G_v\|_2 \leq 4. \quad (10)$$

Applying inequality (10), we have (again using Cauchy-Schwarz):

$$\begin{aligned}
\mathop{\mathbf{E}}_{e\sim\mathcal{E}_e}\left[|eq|^2\right] &= \mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2N}}\left[\left(2^{\Delta/2}A_{uv}(x) - A_v(x^{\downarrow})\right)^2\right] \\
&= \mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2N}}\left[\left(2^{\Delta/2}(A_{uv}(x) - G_{uv}(x)) + (G_v(x^{\downarrow}) - A_v(x^{\downarrow})) + (2^{\Delta/2}G_{uv}(x) - G_v(x^{\downarrow}))\right)^2\right] \\
&\geq \mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2N}}\left[\left(2^{\Delta/2}G_{uv}(x) - G_v(x^{\downarrow})\right)^2\right] - 2\cdot 2^{\Delta/2}\sqrt{\Gamma+100s_0}\cdot\sqrt{4} - 2\cdot\sqrt{\Gamma+100s_0}\cdot\sqrt{4} \\
&\geq 2\sqrt{\Gamma+100s_0}. \qquad (11)
\end{aligned}$$

For an equation $2^{\Delta/2}\cdot\frac{A_u(x)+A_u(x')}{2} - A_v(x^{\downarrow}) = 0$, by Cauchy-Schwarz,

$$|eq|^2 \leq (2^{\Delta}+2^{\Delta}+1)(A_u(x)^2 + A_u(x')^2 + A_v(x^{\downarrow})^2).$$

Thus, we have $|eq|^2 \leq 3\cdot(2^{\Delta+1}+1)\|A_{eq}\|_2^2 \leq 2^{\Delta+3}\|A_{eq}\|_2^2$, and so:

$$\mathop{\mathbf{E}}_{eq\in\mathcal{E}_e}\left[|eq|^2\right] \leq \mathop{\mathbf{E}}_{eq\in\mathcal{E}_e}\left[\chi_{|eq|>c\sqrt{\delta}}\ \ 2^{\Delta+3}\|A_{eq}\|_2^2\right] + c^2\delta.$$

Therefore,

$$\mathop{\mathbf{E}}_{eq\in\mathcal{E}_e}\left[\chi_{|eq|>c\sqrt{\delta}}\ \|A_{eq}\|_2^2\right] \geq \frac{2\sqrt{\Gamma+100s_0}}{2^{\Delta+3}} - \frac{c^2\delta}{2^{\Delta+3}} \geq \frac{\sqrt{\Gamma+100s_0}}{2^{\Delta+3}} \geq 6s.$$

And we are done, since:

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

Thus we are left with the case that:

$$\mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2N}}\left[\left(2^{\Delta/2}G_{uv}(x) - G_v(x^{\downarrow})\right)^2\right] \leq 2^{\Delta/2+3}\cdot\sqrt{\Gamma+100s_0}.$$

In other words (using the fact that $\|G_u\|_2^2, \|G_v\|_2^2 \geq 1-(\Gamma+100s_0)$),

$$\mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2N}}\left[2^{\Delta/2}G_{uv}(x)G_v(x^{\downarrow})\right] \quad\geq\quad 1 - 2^{\Delta/2+4}\sqrt{\Gamma+100s_0}. \qquad (12)$$

We define two probability distributions over possible assignments to the variables in v (in bit representation): D_{uv} and D_v. For every $a\in L_u$, the distribution D_{uv} assigns probability $2^{-\Delta}\frac{\hat{f}_u(a)^2}{\|G_{uv}\|_2^2}$ to the restriction of a to v, which we denote $a_{|v}$ (recall that there are no two a's in G_{uv} with the same restriction to v). Every other assignment gets probability 0. For every $b\in L_v$, the distribution D_v assigns probability $2^{-\Delta}\frac{\hat{f}_v(b)^2}{\|G_v\|_2^2}$ to b. Every other assignment gets probability 0. Also define a distribution D_u over the possible assignments to the variables in u (in bit representation). D_u assigns probability $\frac{\hat{f}_u(a)^2}{\|G_u\|_2^2}$ to every $a\in L_u$, and assigns 0 to all other a's. Note that the probability assigned by D_u to $a\in L_u$ is same as the probability assigned by D_{uv} to $a|_v$. First, we argue that the Hellinger distance between D_{uv} and D_v is small:

Claim 5.7.

$$\Delta_H^2(D_{uv}, D_v) \leq 2^{\Delta/2+5}\sqrt{\Gamma+100s_0}.$$

Proof. We expand $\mathbf{E}_{x\sim\mathcal{N}^{2^N}}\left[2^{\Delta/2}G_{uv}(x)G_v(x^{\downarrow})\right]$:

$$\begin{aligned}
&= 2^{\Delta/2}\cdot \mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2^N}}\left[\left(\sum_{a\in L_u}\hat{f}_u(a)x^{||}_a\right)\left(\sum_{b\in L_v}\hat{f}_v(b)x^{\downarrow}_b\right)\right]\\
&= 2^{\Delta/2}\cdot \sum_{a\in L_u, b\in L_v}\hat{f}_u(a)\hat{f}_v(b)\mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2^N}}\left[x^{||}_a x^{\downarrow}_b\right]\\
&= \sum_{a\in L_u, a_{|v}\in L_v}\hat{f}_u(a)\hat{f}_v(a_{|v})\mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2^N}}\left[(x^{\downarrow}_b)^2\right]\\
&\le \sum_{a\in L_u, a_{|v}\in L_v}\sqrt{\hat{f}_u(a)^2\hat{f}_v(a_{|v})^2}\\
&\le \sum_{b\in\{0,1\}^{N-\Delta}}\sqrt{D_{uv}(b)D_v(b)} + 4(\Gamma+100s_0),
\end{aligned}$$

where the last inequality holds for $(\Gamma+100s_0)\le\frac{1}{4}$. The claim now follows from inequality (12). □

From Proposition 2.7, we get a bound on the statistical distance between D_{uv} and D_v:

$$\Delta(D_{uv}, D_v) \le 2^{\Delta/4+3}\cdot\sqrt[4]{\Gamma+100s_0}.$$

5.5.1 A Strategy for the CSP

Using the bound on the statistical distance between the distributions, we describe a probabilistic strategy for the CSP instance. This implies a deterministic strategy that achieves at least the same value. The probabilistic strategy is as follows:

Give a constraint c and a variable z appearing in c:

1. Use shared randomness to choose a random index $i\in[k+1]$ and a (multi-set of) random constraints $w=\{c_1,\ldots,c_{i-1},c_{i+1},\ldots,c_{k+1}\}$. Let $u=(c_1,\ldots,c_{i-1},c=c_i,c_{i+1},\ldots,c_{k+1})$ and $v=(c_1,\ldots,c_{i-1},z,c_{i+1},\ldots,c_{k+1})$.

2. Use *correlated sampling* [KT02, Hol09] to decide on an assignment to w in the following manner: Pick an infinite sequence of random pairs (a',p), where a' is an assignment to w and p is a probability, i.e., a number between 0 and 1. Let $D^{\downarrow}_u$ be the restriction of D_u to w. Let $D^{\downarrow}_v$ be the restriction of D_v to w.

   - For u, the assignment a'_u to w is the first pair (a'_u,p) in the sequence such that $D^{\downarrow}_u(a'_u)\le p$.
   - For v, the assignment a'_v to w is the first pair (a'_v,p) in the sequence such that $D^{\downarrow}_v(a'_v)\le p$.

3. Obtain an assignment to the distinguished constraint $c=c_i$ by picking an assignment a^*_u to u (i.e. the $(k+1)$ constraints) from D_u, conditioned on its restriction to w being a'_u. Restrict a^*_u to the distinguished constraint to get its assignment.

   Obtain an assignment to the variable z by picking an assignment a^*_v to v (i.e. the k constraints and the variable z) from D_v, conditioned on its restriction to w being a'_v. Restrict a^*_v to z to get its assignment.

Since D_{uv} and D_v are close in statistical distance, so are $D^{\downarrow}_{uv}$ and $D^{\downarrow}_{v}$. In particular, we have that (i) a'_u is distributed as $D^{\downarrow}_{u}$. (ii) a'_v is distributed as $D^{\downarrow}_{v}$. (iii) except with probability at most $2\Delta(D_{uv}, D_v)$, we have $a'_u = a'_v$. Let us concentrate on this case. a^*_u is distributed as D_u, and a^*_v is distributed as D_v. In fact, a'_u defines uniquely a^*_u. The probability that a^*_u does not agree with a^*_v on z is at most $\Delta(D_{uv}, D_v)$.

Overall, we get consistent assignments to c and z with probability at least

$$1 - 2^{\Delta/4+5} \cdot \sqrt[4]{\Gamma + 100 s_0}.$$

For sufficiently small Γ and s_0 this is at least $\sqrt{\eta}$. By the soundness of the CSP, the fraction of e's for which this can happen is at most $\sqrt{\eta}$. These edges are added to E_2.

5.6 Soundness: The General Setting

In general, it does not necessarily hold for every u, v that $\|A_u\|_2^2 = 1$, $\|A_v\|_2^2 = 1$. Instead, the prover may put very low norm on some of the A_u, A_v. This gives the prover the freedom not to decide on assignments to certain u, v. Fortunately, (i) the prover must put significant norm on significant number of the u, v (as the total norm is 1 and the assignment is bounded); (ii) equations involving a table A_u with *high* norm and a table A_v with *low* norm (or vice versa) are likely to fail with large margin. Let us begin by proving the second point:

Lemma 5.8 (Norm gap $\Rightarrow$ dissatisfaction)**.** *For $e = (u, v)$, define $N_e^2 \doteq \frac{5}{9}\|A_u\|_2^2 + \frac{4}{9}\|A_v\|_2^2$. Assume that $N_e \geq 2^{\Delta/2} \cdot c/c_0$ and $(\|A_u\|_2 - \|A_v\|_2)^2 \geq 2^{2\Delta+4}(\Gamma + 100 s_0) N_e^2$. Then,*

$$\mathop{\mathbf{E}}_{eq \sim \mathcal{E}^e}\left[\chi_{|eq|>c\sqrt{\delta}} \|A_{eq}\|_2^2\right] \geq s_0 2^{-\Delta-2} N_e^2.$$

Proof. By Cauchy-Schwarz inequality (we use the definition of A_{uv} from the previous section, $A_{uv}(x) = \frac{A_u(x)+A_u(x')}{2}$),

$$\begin{aligned}\mathop{\mathbf{E}}_{eq \in \mathcal{E}_e}\left[|eq|^2\right] &= \mathop{\mathbf{E}}_{x \sim \mathcal{N}^{2N}}\left[\left|2^{\Delta/2} \cdot A_{uv}(x) - A_v(x^{\downarrow})\right|^2\right] \\ &\geq \left(2^{\Delta/2} \cdot \|A_{uv}\|_2 - \|A_v\|_2\right)^2. \qquad (13)\end{aligned}$$

Note that (again by Cauchy-Schwarz), $\|A_{uv}\|_2^2 \leq \|A_u\|_2^2$. Thus, if $\|A_v\|_2 \geq 2 \cdot 2^{\Delta/2}\|A_u\|_2$, we are done by inequality (13), since $\left(2^{\Delta/2} \cdot \|A_{uv}\|_2 - \|A_v\|_2\right)^2 \geq \|A_v\|_2^2/4 \geq N_e^2/16$.

Assume therefore that $\|A_v\|_2 \leq 2 \cdot 2^{\Delta/2}\|A_u\|_2$. Then, $\|A_u\|_2^2 \geq 2^{-\Delta} N_e^2$. If there is no $(\frac{10}{\Gamma\delta^2}, 100 s_0, \Gamma)$-linear approximating junta G_u for A_u, then we are also done, since by the dictator testing:

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

And as $\|A_u\|_2 \geq 2^{-\Delta/2} N_e \geq c/c_0$,

$$\mathop{\mathbf{E}}_{eq \sim \mathcal{E}_u}\left[\chi_{|eq|>c\sqrt{\delta}} \|A_{eq}\|_2^2\right] \geq s_0 2^{-\Delta} N_e^2.$$

Hence, assume that there is a linear approximating junta G_u for A_u, $\|G_u - A_u\|_2^2 \leq (\Gamma + 100 s_0)\|A_u\|_2^2$. Let $G_{uv} = \frac{G_u(x)+G_u(x')}{2}$. We have (using the triangle inequality):

$$\|A_{uv}\|_2^2 \leq (\|G_{uv}\|_2 + \|A_{uv} - G_{uv}\|_2)^2.$$

By Claim 5.5 (adapted to the case that $\|A_u\|_2^2$ is not necessarily 1), and Claim 5.6, we have:

$$2^\Delta \cdot \|A_{uv}\|_2^2 \le 2^\Delta \cdot (2^{-\Delta/2}\|G_u\|_2 + \sqrt{\Gamma+100s_0}\|A_u\|_2)^2 \le \|A_u\|_2^2 + 2^{\Delta+1}\sqrt{\Gamma+100s_0}\|A_u\|_2^2.$$

By Claim 5.6, and since G_u is orthogonal to $(A_u - G_u)$, we have:

$$\begin{aligned}
2^\Delta \cdot \|A_{uv}\|_2^2 &\ge 2^\Delta \cdot (\|G_{uv}\|_2^2 - \|G_u\|_2\|A_u - G_u\|_2) \\
&\ge \|G_u\|_2^2 - 2^\Delta\|A_u\|_2\|A_u - G_u\|_2 \\
&\ge \|A_u\|_2^2 - (\Gamma+100s_0)\|A_u\|_2^2 - 2^\Delta\sqrt{\Gamma+100s_0}\|A_u\|_2^2
\end{aligned}$$

Overall,

$$\|A_u\|_2 \cdot \left(1 - (2^\Delta+1)\sqrt{\Gamma+100s_0}\right) \le 2^{\Delta/2} \cdot \|A_{uv}\|_2 \le \|A_u\|_2 \cdot \left(1 + 2^\Delta\sqrt{\Gamma+100s_0}\right).$$

Since $(\|A_u\|_2 - \|A_v\|_2)^2 \ge 2^{2\Delta+4}(\Gamma+100s_0)N_e^2$, we have

$$\begin{aligned}
\left|2^{\Delta/2} \cdot \|A_{uv}\|_2 - \|A_v\|_2\right| &\ge |\|A_u\|_2 - \|A_v\|_2| - (2^\Delta+1)\sqrt{\Gamma+100s_0}\|A_u\|_2 \\
&\ge |\|A_u\|_2 - \|A_v\|_2| - 2(2^\Delta+1)\sqrt{\Gamma+100s_0}N_e \\
&\ge (2^{\Delta+1}-2)\sqrt{\Gamma+100s_0}N_e.
\end{aligned}$$

Substituting in inequality (13) yields the lemma. □

Note that the total contribution to the norm of equations from $\mathcal{E}^e$ where $N_e \le 2^{\Delta/2} \cdot c/c_0$ (let us denote the set of such e's by E_0) is at most $\mathbf{E}_{e\in E_0}\left[N_e^2\right] \le 2^\Delta(c/c_0)^2$. Choosing c sufficiently small, we may ignore these equations. We therefore assume henceforth that $N_e \ge 2^{\Delta/2} \cdot c/c_0$. Further we assume that $(\|A_u\|_2 - \|A_v\|_2)^2 \le 2^{2\Delta+4}(\Gamma+100s_0)N_e^2$. From what we argued in Lemma 5.8, it follows that the expectation $\mathbf{E}_{eq\sim\mathcal{E}^e}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right]$ is large for e's for which this does not hold.

From our assumptions we get, in particular, $\|A_u\|_2^2, \|A_v\|_2^2 \ge \frac{1}{10}N_e^2$. Hence, there must be $(\frac{10}{\Gamma\delta^2}, 100s_0, \Gamma)$-linear approximating juntas G_u for A_u and G_v for A_v; otherwise, the equations fail with significant margin, as in the proof of Lemma 5.8. Moreover, we have:

Claim 5.9.

$$\min\left\{\frac{\|G_u\|_2}{\|G_v\|_2}, \frac{\|G_v\|_2}{\|G_u\|_2}\right\} \ge 1 - 2^{\Delta+6}\sqrt{\Gamma+100s_0}.$$

Proof.

$$\begin{aligned}
|\|G_u\|_2 - \|G_v\|_2| &= |(\|G_u\|_2 - \|A_u\|_2) + (\|A_v\|_2 - \|G_v\|_2) + (\|A_u\|_2 - \|A_v\|_2)| \\
&\le (\|A_u\|_2 - \|G_u\|_2) + (\|A_v\|_2 - \|G_v\|_2) + |\|A_u\|_2 - \|A_v\|_2| \qquad (14)
\end{aligned}$$

We have $\|A_u\|_2^2 - \|G_u\|_2^2 \le (\Gamma+100s_0)\|A_u\|_2^2$. Since

$$\|A_u\|_2^2 - \|G_u\|_2^2 = (\|A_u\|_2 + \|G_u\|_2)(\|A_u\|_2 - \|G_u\|_2) \ge (\|A_u\|_2 - \|G_u\|_2)\|A_u\|_2,$$

we get that

$$\|A_u\|_2 - \|G_u\|_2 \le (\Gamma+100s_0)\|A_u\|_2.$$

Applying a similar reasoning to A_v and substituting in (14),

$$\begin{aligned}
|\|G_u\|_2 - \|G_v\|_2| &\le (\Gamma+100s_0)\|A_u\|_2 + (\Gamma+100s_0)\|A_v\|_2 + 2^{\Delta+2}\sqrt{\Gamma+100s_0}N_e \\
&\le 2^{\Delta+3}\sqrt{\Gamma+100s_0}N_e
\end{aligned}$$

The claim follows, noticing that for sufficiently small Γ, s_0 it holds that $\frac{N_e^2}{\|G_u\|_2^2}, \frac{N_e^2}{\|G_v\|_2^2} \le 20$. □

Consider the case that

$$\mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2^N}}\left[\left(2^{\Delta/2}G_{uv}(x)-G_v(x^{\downarrow})\right)^2\right]\geq 2^{\Delta/2+4}\sqrt{\Gamma+100s_0}N_e^2.$$

We follow the argument in the simplified setting and see what needs to be changed when A_u, A_v are not necessarily of norm 1. In inequality (10) the upper bound of 4 should be replaced by $(\|A_u\|_2+\|A_v\|_2)^2$. This change implies subsequent changes in inequality (11): the first $\sqrt{4}$ should be replaced by $\|A_u\|_2(\|A_u\|_2+\|A_v\|_2)$ and the second $\sqrt{4}$ should be replaced by $\|A_v\|_2(\|A_u\|_2+\|A_v\|_2)$. The sum of two error terms in inequality (11) is thus bounded by $2^{\Delta/2+1}\sqrt{\Gamma+100s_0}(\|A_u\|_2+\|A_v\|_2)^2\leq 2^{\Delta/2+3}\sqrt{\Gamma+100s_0}N_e^2$, giving a lower bound of $2^{\Delta/2+3}\sqrt{\Gamma+100s_0}N_e^2$ in inequality (11). This lower bound suffices for the subsequent argument to go through and derive the conclusion that an appropriate measure of equations fail, i.e.

$$\mathop{\mathbf{E}}_{eq\in\mathcal{E}_e}\left[\chi_{|eq|>c\sqrt{\delta}}\,\|A_{eq}\|_2^2\right]\geq\frac{2^{\Delta/2+3}\sqrt{\Gamma+100s_0}N_e^2}{2^{\Delta+3}}-\frac{c^2\delta}{2^{\Delta+3}}\geq 2^{-\Delta/2-1}\sqrt{\Gamma+100s_0}N_e^2.$$

So we are left with the case that:

$$\mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2^N}}\left[\left(2^{\Delta/2}G_{uv}(x)-G_v(x^{\downarrow})\right)^2\right]\leq 2^{\Delta/2+4}\sqrt{\Gamma+100s_0}N_e^2.$$

We can deduce an inequality similar to inequality (12), using Claim 5.9:

$$\mathop{\mathbf{E}}_{x\sim\mathcal{N}^{2^N}}\left[\frac{2^{\Delta/2}G_{uv}(x)}{\|G_u\|_2}\cdot\frac{G_v(x^{\downarrow})}{\|G_v\|_2}\right]\quad\geq\quad 1-2^{\Delta+15}\sqrt{\Gamma+100s_0}. \tag{15}$$

The bound on the squared Hellinger distance (Claim 5.7) goes through (in fact, since we start with an inequality that is already normalized by $\|G_u\|_2$, $\|G_v\|_2$, the last inequality in Claim 5.7, introducing a normalization error, is unnecessary). We end up with a bound on the statistical distance:

$$\Delta(D_{uv},D_v)\leq 2^{\Delta/2+8}\sqrt[4]{\Gamma+100s_0}.$$

5.6.1 Deriving a Strategy for the CSP

Assume on the contrary that

$$\mathop{\mathbf{E}}_{eq\sim\mathcal{E}}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right]<s. \tag{16}$$

We will derive a (randomized) assignment to the constraints and variables of the original CSP, such that the probability that a random constraint and a variable in it are consistent is more than η, reaching a contradiction.

Let E_1 be the set of all $e=(u,v)$ with

1. $\mathbf{E}_{eq\sim\mathcal{E}^e}\left[\chi_{|eq|>c\sqrt{\delta}}\|A_{eq}\|_2^2\right]<s_0 2^{-\Delta-2}N_e^2$.

2. $N_e\geq 2^{\Delta/2}\cdot c/c_0$.

Note that by Lemma 5.8,

$$\forall e \in E_1, \quad (\|A_u\|_2 - \|A_v\|_2)^2 \le 2^{2\Delta+4}(\Gamma + 100 s_0) N_e^2. \tag{17}$$

Let $E_{2,1}$ be the set of all e with $N_e < 2^{\Delta/2} \cdot c/c_0$. Let $E_{2,2}$ be all the $e \notin E_1 \cup E_{2,1}$. Thus,

$$s > \mathop{\mathbf{E}}_{eq \sim \mathcal{E}} \left[\chi_{|eq| > c\sqrt{\delta}} \|A_{eq}\|_2^2 \right] \ge \frac{1}{|E|} \sum_{e \in E_{2,2}} s_0 2^{-\Delta-2} N_e^2.$$

So, there is little norm outside of E_1:

$$\frac{1}{|E|} \sum_{e \notin E_1} N_e^2 = \frac{1}{|E|} \sum_{e \in E_{2,1}} N_e^2 + \frac{1}{|E|} \sum_{e \in E_{2,2}} N_e^2 \le 2^{\Delta} \cdot (c/c_0)^2 + 2^{\Delta+2}(s/s_0) \le \theta,$$

where θ can be made sufficiently small by choosing s and c appropriately. Assume also that θ satisfies, from Equation (17) and appropriate choice of Γ, s_0, that $\forall e \in E_1, \max\{\|A_u\|_2^2, \|A_v\|_2^2\} \le (1+\theta)\|A_u\|_2^2$, and $\theta \le \frac{1}{100}$.

Association Scheme

Given a constraint c^* and a variable z^* in it, we design a (randomized) scheme that associates: (i) to the constraint c^*, a tuple u containing it (the tuple u does not depend on the variable z^*, given c^*) (ii) to the variable z^*, a tuple v containing it (the tuple v does not depend on the constraint c^*, given z^*) (iii) w.h.p., $e = (u, v)$ is an edge.

For the sake of analysis, it is convenient to also design a scheme that associates, to the pair (c^*, z^*), a pair (u', v') where u' contains c^*, v' contains z^*, $e' = (u', v')$ is an edge; note however that this scheme depends on both c^* and z^*.

These schemes work as follows:

- Pick an infinite sequence of random tuples $(i, c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1}, w)$, where $i \in [k+1]$ is an index, $c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1} \in \mathcal{C}$ are CSP constraints, and w is a number between 0 and b^2.
- With a CSP constraint c^* associate $u = (c_1, \ldots, c_{i-1}, c^*, c_{i+1}, \ldots, c_{k+1})$, where $(i, c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1}, w)$ is the first tuple with $w \le \|A_u\|_2^2$.
- With a CSP variable z^* associate $v = (c_1, \ldots, c_{i-1}, z^*, c_{i+1}, \ldots, c_{k+1})$, where $(i, c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1}, w)$ is the first tuple with $w \le \|A_v\|_2^2$.
- With a CSP constraint-variable pair (c^*, z^*) associate $e' = (u', v')$ with $u' = (c_1, \ldots, c_{i-1}, c^*, c_{i+1}, \ldots, c_{k+1})$, $v' = (c_1, \ldots, c_{i-1}, z^*, c_{i+1}, \ldots, c_{k+1})$ where $(i, c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1}, w)$ is the first tuple with $w \le \max\left\{\|A_u\|_2^2, \|A_v\|_2^2\right\}$.

A Strategy for the CSP. The association scheme we just described gives rise to a strategy for the CSP instance:

1. Using the scheme, associate a tuple u to constraint c^* and a tuple v to variable z^*.
2. Using the strategy described in the simplified setting, given u decide on an assignment to c^*, and given v decide on an assignment to z^*.

The strategy succeeds if c^*, z^* are given consistent values.

Claim 5.10. *Fix (c^*, z^*). With probability at least $1 - 2^{\Delta+10}\sqrt{\Gamma + 100s_0}$ over the randomness in the strategy, conditioned on the event that $e' = (u', v')$ associated with the pair (c^*, z^*) is in E_1, we have that u' is associated with c^*, v' is associated with z^*, and the strategy succeeds for (c^*, z^*).*

Proof. Assume that the pair $e' = (u', v')$ associated with the pair (c^*, z^*) is in E_1. Let u and v be the tuples associated with c^* and z^* respectively. The probability that $u \neq u'$ or $v \neq v'$ is at most

$$\begin{aligned}
\frac{\max\left\{\|A_{u'}\|_2^2, \|A_{v'}\|_2^2\right\} - \min\left\{\|A_{u'}\|_2^2, \|A_{v'}\|_2^2\right\}}{\max\left\{\|A_{u'}\|_2^2, \|A_{v'}\|_2^2\right\}} &\leq \frac{\left|\|A_{u'}\|_2^2 - \|A_{v'}\|_2^2\right|}{N_{e'}^2} \\
&\leq \frac{1}{N_{e'}^2} \cdot \left|\|A_{u'}\|_2 - \|A_{v'}\|_2\right| \cdot (\|A_{u'}\|_2 + \|A_{v'}\|_2) \\
&\leq \frac{1}{N_{e'}^2} \cdot \left|\|A_{u'}\|_2 - \|A_{v'}\|_2\right| \cdot \frac{9}{4} \cdot \left(\frac{5}{9}\|A_{u'}\|_2 + \frac{4}{9}\|A_{v'}\|_2\right) \\
&\leq \frac{1}{N_{e'}^2} \cdot \left|\|A_{u'}\|_2 - \|A_{v'}\|_2\right| \cdot \frac{9}{4} \cdot \sqrt{\frac{5}{9}\|A_{u'}\|_2^2 + \frac{4}{9}\|A_{v'}\|_2^2} \\
&\leq \frac{9}{4} \cdot 2^{\Delta+2}\sqrt{\Gamma + 100s_0},
\end{aligned}$$

where we used Equation (17). Thus we may now assume $u = u'$ and $v = v'$. We showed just before Section 5.6.1 that whenever $e' = (u, v) \in E_1$, it holds that $\Delta(D_{uv}, D_v) \leq 2^{\Delta/2+8}\sqrt[4]{\Gamma + 100s_0}$. In Section 5.5.1, we showed that the strategy described there succeeds for (c^*, z^*) with probability at least $1 - 3\Delta(D_{uv}, D_v)$. The claim follows. □

Definition 10. *Let D' be the distribution over edges that picks a pair (c^*, z^*) of the CSP uniformly at random and then associates an edge $e' = (u', v')$ to the pair (c^*, z^*). Formally:*

- *Pick a pair (c^*, z^*) of the CSP uniformly.*
- *Pick an infinite sequence of random tuples $(i, c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1}, w)$, where $i \in [k+1]$ is an index, $c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1} \in \mathcal{C}$ are CSP constraints, and w is a number between 0 and b^2.*
- *Let $e' = (u', v')$ with $u' = (c_1, \ldots, c_{i-1}, c^*, c_{i+1}, \ldots, c_{k+1})$, $v' = (c_1, \ldots, c_{i-1}, z^*, c_{i+1}, \ldots, c_{k+1})$ where $(i, c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1}, w)$ is the first tuple with $w \leq \max\left\{\|A_u\|_2^2, \|A_v\|_2^2\right\}$.*

We will show that an edge $e' \sim D'$ is in E_1 with high probability. From Claim (5.10), it then gives a strategy for the CSP that satisfies more than η fraction of its pairs, reaching a contradiction. Towards this end, we define another distribution D'' on edges and show that it is close to D' and an edge $e \sim D''$ is in E_1 with high probability.

Definition 11. *Let D'' be the distribution over edges that gives an edge $e = (u, v)$ probability proportional to $\max\{\|A_u\|_2^2, \|A_v\|_2^2\}$. Another way to sample an edge $e \sim D''$ is:*

- *Pick an infinite sequence of random tuples $((c^*, z^*), (i, c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1}, w))$, where (c^*, z^*) is a (uniformly) random CSP pair, $i \in [k+1]$ is an index, $c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1} \in \mathcal{C}$ are CSP constraints, and w is a number between 0 and b^2.*

- *Let $e = (u, v)$ with $u = (c_1, \ldots, c_{i-1}, c^*, c_{i+1}, \ldots, c_{k+1})$, $v = (c_1, \ldots, c_{i-1}, z^*, c_{i+1}, \ldots, c_{k+1})$ where $((c^*, z^*), (i, c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{k+1}, w))$ is the first tuple with $w \leq \max\left\{\|A_u\|_2^2, \|A_v\|_2^2\right\}$.*

Claim 5.11. *If $e \sim D''$, then $e \in E_1$ with probability at least $1 - 3\theta$.*

Proof. Let $T = \sum_{e=(u,v)\in E} \max\{\|A_u\|_2^2, \|A_v\|_2^2\}$. Note that $T \geq \sum_{e\in E} N_e^2 = |E|$. The probability that an edge distributed as D'' is not in E_1 is

$$\sum_{e \notin E_1} \frac{\max\{\|A_u\|_2^2, \|A_v\|_2^2\}}{T} \leq \sum_{e\notin E_1} \frac{3N_e^2}{T} \leq \frac{3\theta|E|}{T} \leq 3\theta.$$

□

Next we show that D' and D'' are close. Let $D'|_{\text{CSP}}$ (and $D''|_{\text{CSP}}$ resp.) be a distribution over the CSP pairs (c^*, z^*) obtained by first picking an edge $e = (u, v) \sim D'$ ($e = (u, v) \sim D''$ resp.) and then taking the "projection" to the coordinate on which u contains a constraint c^* and v contains a variable z^*. Clearly, $D'|_{\text{CSP}}$ is uniform on all CSP pairs. However, $D''|_{\text{CSP}}$ is not necessarily uniform. We will show nevertheless that $D''|_{\text{CSP}}$ is close to uniform. Note that this implies in turn that D' and D'' are close since they are identical distributions conditional on the projection being any fixed pair (c^*, z^*).

Towards showing that $D''|_{\text{CSP}}$ is close to uniform, we will define yet another distribution D over edges and show that D and D'' are close and that $D|_{\text{CSP}}$ is close to uniform.

Definition 12. *Let D be the distribution over all edges that gives an edge $e = (u, v)$ probability proportional to $\|A_u\|_2^2$. Equivalently, D is the distribution that picks $u \in U$ with probability proportional to $\|A_u\|_2^2$ and then picks a random edge incident on u (among the $(k + 1) \cdot d$ edges incident on u).*

Claim 5.12. *$\Delta(D, D'') \leq 4\theta$.*

Proof. Let

$$S = \sum_{e=(u,v)\in E} \|A_u\|_2^2, \qquad T = \sum_{e=(u,v)\in E} \max\{\|A_u\|_2^2, \|A_v\|_2^2\}, \qquad T \geq |E|.$$

We have,

$$S \leq T \leq \sum_{e\in E_1} \max\{\|A_u\|_2^2, \|A_v\|_2^2\} + \sum_{e\notin E_1} 3N_e^2 \leq (1+\theta) \sum_{e\in E_1} \|A_u\|_2^2 + 3\theta|E| \leq (1+\theta)S + 3\theta T.$$

In particular, $S \geq \frac{1-3\theta}{1+\theta} T \geq (1 - 4\theta)T \geq (1 - 4\theta)|E| \geq \frac{1}{2}|E|$. Now,

$$2 \cdot \Delta(D, D'') = \sum_e \left| \frac{\|A_u\|_2^2}{S} - \frac{\max\{\|A_u\|_2^2, \|A_v\|_2^2\}}{T} \right|.$$

We split the sum into $e \notin E_1$ and $e \in E_1$ and show that both are small. We start with the sum over $e \notin E_1$. We analyze the expression:

$$\left| \frac{T\|A_u\|_2^2 - S \max\{\|A_u\|_2^2, \|A_v\|_2^2\}}{ST} \right|$$

If $T\|A_u\|_2^2 \geq S\max\{\|A_u\|_2^2, \|A_v\|_2^2\}$, then, using $T \leq 2S$, we obtain the expression

$$\frac{T\|A_u\|_2^2 - S\max\{\|A_u\|_2^2, \|A_v\|_2^2\}}{ST} \leq \frac{2S\|A_u\|_2^2 - S\max\{\|A_u\|_2^2, \|A_v\|_2^2\}}{S^2} \leq \frac{\|A_u\|_2^2}{S}.$$

If $T\|A_u\|_2^2 < S\max\{\|A_u\|_2^2, \|A_v\|_2^2\}$, then, using $S \leq T$, we obtain the expression

$$\frac{S\max\{\|A_u\|_2^2, \|A_v\|_2^2\} - T\|A_u\|_2^2}{ST} \leq \frac{\max\{\|A_u\|_2^2, \|A_v\|_2^2\} - \|A_u\|_2^2}{S} \leq \frac{\|A_v\|_2^2}{S}$$

Overall,

$$\sum_{e \notin E_1} \left| \frac{\|A_u\|_2^2}{S} - \frac{\max\{\|A_u\|_2^2, \|A_v\|_2^2\}}{T} \right| \leq \sum_{e \notin E_1} \frac{\|A_u\|_2^2 + \|A_v\|_2^2}{S} \leq \frac{1}{S} \cdot \sum_{e \notin E_1} N_e^2 \leq \frac{1}{S} \cdot \theta |E| \leq 2\theta.$$

Noting that for $e \in E_1$, $\max\{\|A_u\|_2^2, \|A_v\|_2^2\} \leq (1+\theta)\|A_u\|_2^2$, the sum over $e \in E_1$ can be upper bounded as:

$$\sum_{e \in E_1} \left| \frac{\|A_u\|_2^2}{S} - \frac{\|A_u\|_2^2}{T} \right| + \sum_{e \in E_1} \left| \frac{\|A_u\|_2^2}{T} - \frac{\max\{\|A_u\|_2^2, \|A_v\|_2^2\}}{T} \right| \leq \frac{T-S}{T} + \frac{\theta S}{T} \leq 5\theta.$$

□

Now we show that $D|_{\text{CSP}}$ is close to uniform. Let D_U be the distribution on U that picks $u \in U$ with probability proportional to $\|A_u\|_2^2$. Let $D_{\mathcal{C}}$ be the distribution on CSP constraints that picks $u \sim D_U$ and then picks a random constraint c^* in u. It is enough to show that $D_{\mathcal{C}}$ is close to uniform. Let $D_U^1, \ldots, D_U^{k+1}$ be the marginals of D_U on each of the $k+1$ coordinates so that $D_{\mathcal{C}} = \frac{1}{k+1}\sum_{i=1}^{k+1} D_U^i$. Note that:

- $\forall u \in U, \|A_u\|_2^2 \leq b^2$.
- $\sum_{u \in U} \|A_u\|_2^2 = \frac{|U|}{|E|}\sum_{e \in E} \|A_u\|_2^2 = \frac{|U| \cdot S}{|E|} \geq \frac{1}{2}|U|$ (this uses a calculation in the proof of Claim 5.12).
- Hence $\forall u \in U, D_U(u) = \frac{\|A_u\|_2^2}{\sum_{u \in U}\|A_u\|_2^2} \leq \frac{2b^2}{|U|}$.

This implies that the entropy of D_U is at least $H(D_U) \geq \log|U| - 2\log b - 1$. Using the sub-additivity and concavity of entropy,

$$H(D_{\mathcal{C}}) = H\left(\frac{1}{k+1}\sum_{i=1}^{k+1} D_U^i\right) \geq \frac{1}{k+1}\sum_{i=1}^{k+1} H(D_U^i) \geq \frac{H(D_U)}{k+1} \geq \log|\mathcal{C}| - \frac{2\log b + 1}{k+1}.$$

Thus when k is sufficiently large, $H(D_{\mathcal{C}})$ is close to its maximum possible value of $\log|\mathcal{C}|$ and therefore $\Delta(D_{\mathcal{C}}, \mathsf{Uniform}) \leq \theta$ as desired.

This implies, as argued before, $\Delta(D|_{\text{CSP}}, \mathsf{Uniform}) \leq \theta$ and $\Delta(D''|_{\text{CSP}}, \mathsf{Uniform}) \leq 30\theta$ using Claim 5.12. Since $D'|_{\text{CSP}} = \mathsf{Uniform}$, we have $\Delta(D'|_{\text{CSP}}, D''|_{\text{CSP}}) \leq 30\theta$, which implies that $\Delta(D', D'') \leq 30\theta$. The last argument uses the observation that conditional on the projection being (c^*, z^*), D' and D'' are identical. Combining with Claim 5.10, Claim 5.11, and choosing θ, Γ, s_0 small enough, we get a strategy for the CSP that succeeds with probability exceeding η. This completes the soundness anlaysis.

5.7 Discretization

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

Since N, 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

[ABS10] S. Arora, B. Barak, and D. Steurer. Subexponential algorithms for unique games and related problems. In *Proc. 51st IEEE Symp. on Foundations of Computer Science*, 2010.

[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.

[ALM+98] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

[AS98] S. Arora and S. Safra. Probabilistic checking of proofs: a new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.

[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.

[BFLS91] L. Babai, L. Fortnow, L. A. Levin, and M. Szegedy. Checking computations in polylogarithmic time. In *Proc. 23rd ACM Symp. on Theory of Computing*, pages 21–32, 1991.

[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.

[CT91] T. M. Cover and J. Thomas. *Elements of Information Theory.* Wiley, 1991.

[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.

[Hol09] T. Holenstein. Parallel repetition: Simplification and the no-signaling case. *Theory of Computing*, 5(1):141–172, 2009.

[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.

[KM10] S. Khot and D. Moshkovitz. Hardness of approximately solving linear equations over reals. Technical report, ECCC Report TR10-053, 2010.

[KT02] J. Kleinberg and E. Tardos. Approximation algorithms for classification problems with pairwise relationships: metric labeling and markov random fields. *Journal of the ACM*, 49(5):616–639, 2002.

[Pol02] D. Pollard. *A user's guide to measure theoretic probability.* Cambridge University Press, 2002.

[Rag08] P. Raghavendra. Optimal algorithms and inapproximability results for every csp? In *Proc. 40th ACM Symp. on Theory of Computing*, pages 245–254, 2008.

[Raz98] R. Raz. A parallel repetition theorem. In *SIAM Journal on Computing*, volume 27, pages 763–803, 1998.

[Zwi98] U. Zwick. Approximation algorithms for constraint satisfaction problemsinvolving at most three variables per constraint. In *Proc. of 9th SODA*, pages 201–210, 1998.