Electronic Colloquium on Computational Complexity, Report No. 124 (2016)

On Independent Sets, 2-to-2 Games and Grassmann Graphs

Subhash Khot [*] Dor Minzer [†] Muli Safra [‡]

Abstract

We present a candidate reduction from the 3-Lin problem to the 2-to-2 Games problem and present a combinatorial hypothesis about Grassmann graphs which, if correct, is sufficient to show the soundness of the reduction in a certain non-standard sense. A reduction that is sound in this non-standard sense implies that it is NP-hard to distinguish whether an n-vertex graph has an independent set of size $\left(1 - \frac{1}{\sqrt{2}}\right) n - o(n)$ or whether every independent set has size $o(n)$, and consequently, that it is NP-hard to approximate the Vertex Cover problem within a factor $\sqrt{2} - o(1)$.

1 Introduction

This paper focusses on hardness of approximation results for the Vertex Cover and the Independent Set problems, which are closely related to the hardness of approximating the 2-to-2 Games problem with a certain non-standard notion of soundness.

1.1 Vertex Cover and Independent Set

Given an n-vertex graph $G = (V, E)$, the Vertex Cover problem asks for a vertex cover of minimum size, namely, a subset $C \subseteq V$ of minimum size that includes at least one endpoint of each edge $e \in E$. This is a classic NP-hard problem and has a greedy 2-approximation algorithm. The algorithm starts with the graph G, initializes $C = \emptyset$, and until the graph has at least one edge remaining, picks an edge, adds both its endpoints to C, removes all edges incident on either of these two endpoints, and repeats. It is easily seen that the final set C is a vertex cover of G and has size at most twice that of the minimum vertex cover. A somewhat better approximation algorithm achieving factor $2 - \Omega\left(\frac{1}{\sqrt{\log n}}\right)$ is known via SDP relaxation [17, 21]. However it is a major open question whether there is a $2 - \delta$ approximation algorithm for some fixed positive constant δ. Surprisingly, as discussed below, there is some evidence to the contrary: Vertex Cover might actually be hard to approximate within a factor $2 - \varepsilon$ for every positive constant ε.

The complement $V \setminus C$ of a vertex-cover C is an independent set, namely, a set of vertices $I \subseteq V$ that has no edge inside it. For constants $0 < \beta < \alpha < 1$, let $\mathsf{GapIS}(\alpha, \beta)$ denote a promise gap-problem where the task is to distinguish whether a given n-vertex graph has an independent set of size at least αn or whether every independent set is of size at most βn. Clearly, if $\mathsf{GapIS}(\alpha, \beta)$ is hard,[1] then it would be hard to approximate Vertex Cover within a factor strictly less than $\frac{1-\beta}{1-\alpha}$.

[*] Department of Computer Science, Courant Institute of Mathematical Sciences, New York University. Supported by the NSF Award CCF-1422159, the Simons Collaboration on Algorithms and Geometry and the Simons Investigator Award.

[†] School of Computer Science, Tel Aviv University.

[‡] School of Computer Science, Tel Aviv University.

[1] The statement that $\mathsf{GapIS}(\alpha, \beta)$ is NP-hard is equivalent to the statement that NP has a Probabilistically Checkable Proof (PCP) that has zero free bit complexity, has completeness at least α and has soundness at most β, see [4, Proposition 5.6, Theorem 8.2]. We avoid the terminology of free bit complexity in this paper.

ISSN 1433-8092

Let ε denote a positive and arbitrarily small constant. We summarize the known NP-hardness results for approximating the Vertex Cover problem, obtained in a sequence of highly influential papers. Building on the PCP Theorem [13, 3, 2], the Parallel Repetition Theorem [31], and the Long Code based PCP framework in [4], Håstad[18] showed that $\mathsf{GapIS}(\frac{1}{4}-\varepsilon, \frac{1}{8}+\varepsilon)$ is NP-hard, implying $\frac{7}{6}-\varepsilon \approx 1.16$ hardness factor for Vertex Cover. Dinur and Safra [11] showed that $\mathsf{GapIS}(p-\varepsilon, 4p^3-3p^4+\varepsilon)$ is NP-hard for $p=\frac{3-\sqrt{5}}{2}$, implying $10\sqrt{5}-21-\varepsilon \approx 1.36$ hardness factor for Vertex Cover. Their paper introduced several techniques, e.g. the Biased Long Code, application of Fourier analytic theorems on Boolean hypercube, and implicitly, the notion of 2-to-2 Games, all of which are indispensable in authors' opinion, for further progress on Vertex Cover.

1.2 d-to-d Games

In this section, we discuss the d-to-d Games and their connection to the Independent Set and Vertex Cover problems.

Definition 1.1. *A 2-Prover-1-Round Game $G=(V,E,\Phi,\Sigma)$ consists of a set of variables V, a set of colors Σ, and a constraint $\Phi(u,v)$ for every (directed) edge $(u,v)\in E$. The goal is to assign colors to variables, say $A: V\to\Sigma$, so as to satisfy the maximum number of the constraints. A constraint $\Phi(u,v)$ is satisfied if $(A(u),A(v))\in\Phi(u,v)$, where by abuse of notation, $\Phi(u,v)\subseteq\Sigma\times\Sigma$ denotes the subset of color-pairs that are deemed satisfactory. The subset $\Phi(u,v)\subseteq\Sigma\times\Sigma$ may in general depend on the edge (u,v).*

Let $d\geqslant 1$ be an integer. A constraint $\Phi(u,v)\subseteq\Sigma\times\Sigma$ is said to be a d-to-d constraint if there are partitions $A_1,\ldots,A_r$ and $B_1,\ldots,B_r$ of Σ into sets of size d such that ($|\Sigma|=rd$)

$$\Phi(u,v)=\bigcup_{i=1}^{r} A_i\times B_i.$$

A 2-Prover-1-Round Game $G=(V,E,\Phi,\Sigma)$ is said to be a d-to-d Game if every constraint $\Phi(u,v)$ is a d-to-d constraint. A 1-to-1 Game is also called a Unique Game. In this case, $\Phi(u,v)$ is simply a perfect matching on $\Sigma\times\Sigma$.

In the above definitions, the number of colors $|\Sigma|$ is thought of as a constant, possibly large, and the size of the constraint graph as the growing input size. Motivated by a hardness of approximation result for the 2-SAT problem, Khot [23] formulated the Unique Games Conjecture:

Conjecture 1.2. *For every constant $\delta>0$, for sufficiently large constant $|\Sigma|$, given an instance $G=(V,E,\Phi,\Sigma)$ of a Unique Game, it is NP-hard to distinguish between*

- *YES case: there is a coloring satisfying $1-\delta$ fraction of the constraints of G.*
- *NO case: no coloring satisfies more than δ fraction of the constraints of G.*

The reduction in [11] implicitly suggests the idea of 2-to-2 Games (though therein, the game is a α-game in the sense of [10] instead of a 2-to-2 game and the notion of soundness is non-standard). Motivated by hardness of approximation result for the Vertex Cover problem, Khot [23] also formulated the d-to-d Conjecture.[2]

[2] Note that the Unique Games Conjecture is, necessarily, made with imperfect completeness whereas the d-to-d Conjecture is made with perfect completeness. Strictly speaking, the conjecture in [23] is a d-to-1 Conjecture. It implies (and in authors' opinion, is morally equivalent to) the d-to-d Conjecture stated here.

Conjecture 1.3. *Fix any integer $d \geqslant 2$. For every constant $\delta > 0$, for sufficiently large constant $|\Sigma|$, given an instance $G = (V, E, \Phi, \Sigma)$ of a d-to-d Game, it is NP-hard to distinguish between*

- *YES case: there is a coloring satisfying all of the constraints of G.*
- *NO case: no coloring satisfies more that δ fraction of the constraints of G.*

Let ε denote a positive and arbitrarily small constant. It was shown that the d-to-d Conjecture implies that $\mathsf{GapIS}\left(1 - \frac{1}{2^{1/d}} - \varepsilon, \varepsilon\right)$ is NP-hard for $d \geqslant 2$ [23]. The result did not apply in case of Unique Games, because the imperfect completeness of Unique Games presented a difficulty. This difficulty was circumvented in [27] where the authors showed that the Unique Games Conjecture implies that $\mathsf{GapIS}(\frac{1}{2} - \varepsilon, \varepsilon)$ is NP-hard and therefore, implies that Vertex Cover is NP-hard to approximate within a factor $2 - \varepsilon$. The Unique Games Conjecture, and to a lesser extent the d-to-d Conjecture, is now a prominent open question in theoretical computer science. It implies hardness of approximation results, often optimal results, for numerous problems and has connections to several areas in algorithms, computational complexity, and geometry, see [36, 26, 25] for surveys on the topic. It is thus worthwhile to investigate possible lines of attack towards proving (or disproving) the Unique Games Conjecture, the d-to-d Conjectures, and their variants. In this paper, we present a line of attack towards proving a variant of the 2-to-2 Conjecture with a certain non-standard notion of soundness, and towards making progress on the Independent Set and Vertex Cover problems.

Unfortunately we have to consider games where the constraints are a mix of 2-to-2 constraints and 1-to-1 constraints and the game satisfies an additional transitivity property. This feature might not be necessary, but we are unable to circumvent it for now.

Definition 1.4. *A Transitive* 2*-to-*2 *Game is a game $G = (V, E, \Phi, \Sigma)$ where*

- *Each constraint $\Phi(u, v)$ is a* 2*-to-*2 *constraint or a* 1*-to-*1 *constraint.*
- *Transitivity: If there is a* 1*-to-*1 *constraint $\Phi(u, v)$ and a* 1*-to-*1 *or a* 2*-to-*2 *constraint $\Phi(v, w)$, then there is also a constraint $\Phi(u, w)$. The constraint $\Phi(u, w)$ is either* 1*-to-*1 *or* 2*-to-*2 *depending on whether $\Phi(v, w)$ is* 1*-to-*1 *or* 2*-to-*2 *respectively. Moreover, the constraint $\Phi(u, w)$ is a composition of constraints $\Phi(u, v)$ and $\Phi(v, w)$, i.e. for every $a, b, c \in \Sigma$,*

$$(a, b) \in \Phi(u, v),\ (b, c) \in \Phi(v, w) \implies (a, c) \in \Phi(u, w).$$

The notion of soundness (NO case) in Conjectures 1.2 and 1.3 states that no coloring satisfies more than a tiny fraction of the constraints. This notion will be referred to as the standard notion of soundness. It has been a folklore among the experts (see [27, Theorem 3.1], where this is stated for the 1-to-1 case) that as far as the Independent Set and Vertex Cover hardness results are concerned, a non-standard notion of soundness for the d-to-d Games suffices. The non-standard notion concerns "(j, δ)-colorings" that we define next. [3]

Definition 1.5. *Let $G(V, E, \Phi, \Sigma)$ be a Transitive* 2*-to-*2 *Game, $\delta > 0$, j be a positive integer, and $X \subseteq V$. A coloring $A \colon X \to \binom{\Sigma}{j}$ is called a (j, δ)-coloring if the following holds (note that one is allowed to assign a set of j colors to every variable in X and the rest of the variables are unassigned):*

- $|X| \geqslant \delta\, |V|$.

[3] Our definition has to take into account the transitivity feature that we unfortunately have to deal with. Also, we restrict ourselves to the case $d = 2$ which is our primary concern.

- *For every $u, v \in X$ such that $\Phi(u, v)$ is a 2-to-2 constraint, there are colors $a \in A(u)$, $b \in A(v)$ such that $(a, b) \in \Phi(u, v)$.*
- *For every $u, v \in X$ such that $\Phi(u, v)$ is a 1-to-1 constraint, the color sets $A(u)$ and $A(v)$ are identical up to the matching $\Phi(u, v)$. More precisely, for every $(a, b) \in \Phi(u, v)$, $a \in A(u)$ if an only if $b \in A(v)$.*

Now we state a variant of Conjecture 1.3 (for $d = 2$) with a non-standard notion of soundness, imperfect completeness, and for transitive 2-to-2 games. This variant is to be thought of as weaker than Conjecture 1.3 in the sense that Conjecture 1.3 (for $d = 2$) implies it (up to an insignificant caveat).

Conjecture 1.6. *For every constant $\delta > 0$ and every positive integer j, for sufficiently large constant $|\Sigma|$, given an instance $G = (V, E, \Phi, \Sigma)$ of a Transitive 2-to-2 Game, it is NP-hard to distinguish between*

- *YES case: there is a $(1, 1 - \delta)$-coloring to the graph G.*
- *NO case: there is no (j, δ)-coloring to the graph G.*

Finally, we note that the result below follows directly from prior works [11, 23, 27]. A proof is presented in Section B for the sake of completeness. The ingredients include the Biased Long Code and analytic theorems of Russo, Margulis and Friedgut on the Boolean hypercube. Some care is required to handle the transitivity feature.

Theorem 1.7. *If Conjecture 1.6 holds, then $\mathsf{GapIS}(1 - \frac{1}{\sqrt{2}} - \varepsilon, \varepsilon)$ is NP-hard for every positive constant ε.*

1.3 Our Results

Roughly speaking, we give a reduction from an NP-hard problem known as 3-Lin to (Transitive) 2-to-2 Game such that the reduction is sound in the sense of Conjecture 1.6 assuming a combinatorial hypothesis. Therefore, correctness of the combinatorial hypothesis would imply Conjecture 1.6 and the corresponding results for GapIS and Vertex Cover via Theorem 1.7.

We now state the results more formally. Let 3-Lin be the following problem. The instance of the problem is (X, Eq) where X is a set of variables taking values over $\mathbb{F}_2$ and Eq is a set of linear equations over $\mathbb{F}_2$ such that every equation depends on three variables in X. The goal is to find an assignment to the variables so as to maximize the number of equations satisfied. Let $\mathsf{Gap3Lin}(c, s)$ denote the promise gap-problem where the task is to distinguish whether a given 3-Lin instance has an assignment satisfying at least c fraction of the equations or whether every assignment satisfies at most s fraction of the equations. A celebrated result of Håstad [19] shows that for every positive constant ε, $\mathsf{Gap3Lin}(1-\varepsilon, \frac{1}{2}+\varepsilon)$ is NP-hard. For our purposes, it is convenient to work with a 3-Lin instance that is regular, i.e. every equation contains three distinct variables, every variable appears in exactly, say 5, equations, and two distinct equations share at most one variable. Starting with Håstad's result, it is a routine exercise to show that $\mathsf{Gap3Lin}(1 - \varepsilon, s^*)$ is NP-hard on regular instances for every positive constant ε and for some absolute constant $s^* < 1$. Our main result is this:

Theorem 1.8. *For every positive integer j and every constant $\delta > 0$, for sufficiently small constant $\varepsilon > 0$, there is a polynomial time reduction mapping a regular instance (X, Eq) of $\mathsf{Gap3Lin}(1-\varepsilon, s^*)$ to an instance $G = (V, E, \Phi, \Sigma)$ of Transitive 2-to-2 Game such that:*

- *YES case: If there is an assignment satisfying at least $1 - \varepsilon$ fraction of the equations in (X, Eq), then there is a $(1, 1 - \delta)$-coloring to G.*

- *NO case: Assuming the combinatorial Hypothesis 2.5, if no assignment satisfies more than s^* fraction of the equations in (X, Eq), then there is no (j, δ)-coloring to G.*

We present the combinatorial hypothesis later, after discussing the Grassmann graph and the motivation behind the hypothesis. The following corollary follows via Theorem 1.7.

Corollary 1.9. *Assuming the combinatorial Hypothesis 2.5,*

- *Conjecture 1.6 is correct.*
- $\mathsf{GapIS}\left(1 - \frac{1}{\sqrt{2}} - \varepsilon, \varepsilon\right)$ *is NP-hard for every positive constant ε.*
- *Vertex Cover is NP-hard to approximate to within a factor $\sqrt{2} - \varepsilon$ for every positive constant ε.*

Remark 1.10. *Our reduction, depending on the correctness of the combinatorial hypothesis, would give $\sqrt{2} - o(1)$ hardness for Vertex Cover, improving on the 1.36 hardness of Dinur and Safra. While the numerical improvement would be interesting, in authors' opinion, a much more interesting feature would be the "gap-location" for the Independent Set problem. Our reduction would show that $\mathsf{GapIS}(\alpha^*, \beta)$ is NP-hard where α^* is a fixed, absolute constant and $\beta \to 0$ is an arbitrarily small constant. Such a result would be remarkable, in authors' opinion, irrespective of whether it gives an improvement in the Vertex Cover hardness factor. The best known result in this direction is that $\mathsf{GapIS}(2^{-k} - o(1),\ 2^{-2^k+1} + o(1))$ is NP-hard for every integer $k \geqslant 2$, by Siu On Chan [7]. Hardness of $\mathsf{GapIS}(\alpha, \beta)$ corresponds to Vertex Cover hardness of $\frac{1-\beta}{1-\alpha}$. An improvement in Vertex Cover hardness would not necessarily yield $\beta \to 0$ while keeping α fixed, which in authors' opinion, is a more fundamental and challenging question.*

Remark 1.11. *In an ongoing work with Dinur and Kindler, we are investigating whether our reduction, with a slight modification, gives soundness in the sense of Conjecture 1.3, i.e., in the NO case, any coloring of the 2-to-2 Game instance satisfies only $o(1)$ fraction of its edges. This could prove the 2-to-2 Games Conjecture, albeit with imperfect completeness. The modification of the reduction amounts to removing some of the constraints in the 2-to-2 Game constructed by the reduction.*

Remark 1.12. *3-Lin is known to have a "Lasserre integrality gap" on random instances with perfect completeness [15, 35, 37]. Our reduction from 3-Lin to 2-to-2 Games and then the reduction from 2-to-2 Games to the Independent Set and Vertex Cover problems could yield similar Lasserre integrality gap for the latter problems. Of course, we do not yet have a soundness analysis for the reduction. As far as integrality gaps are concerned, the initial 3-Lin instance is a random instance, which could perhaps make the soundness analysis more amenable.*

1.4 Overview of the Reduction

A vast majority of hardness of approximation results are proved by constructing special purpose Probabilistically Checkable Proof Systems (PCPs) (e.g. [2, 4, 18, 19, 16, 24, 7]). Sometimes it is more convenient, and certainly helpful to a reader not familiar with PCP terminology, to take a combinatorial view and present a PCP construction, equivalently, as a combinatorial reduction (e.g. [11, 9]). In this paper, we adopt the latter view as far as possible, using PCP terminology wherever helpful or necessary.

A generic and extremely successful framework to construct PCPs, developed in [1, 4, 31, 18, 19], is as follows. Therein a PCP reduction is a "composition" of two modules, an "Inner PCP" and an "Outer PCP".

- The Inner PCP is best thought of as a combinatorial gadget and an analysis of its structural properties. The gadget is often coding-theoretic and amounts to a specific encoding scheme and a probabilistic procedure to test whether a given word is (close to) a codeword and if so, to decode (or "list-decode") that codeword. The choice of the encoding scheme as well as the nature of the tester (e.g. number of queries and the acceptance predicate) are dictated by the target problem for which one desires a hardness of approximation result.

- The Outer PCP is a canonical NP-hard problem known variously as 2-Prover-1-Round Game, 2-CSP, or Label Cover. The problem is known to be very hard to approximate [1, 31], via Raz's Parallel Repetition Theorem.

- The composition amounts to taking several (local) copies of the Inner PCP gadget and "combining" them via the (global) Outer PCP.[4]

With this framework in mind, we give a short and informal overview of our reduction, leaving out several intricate details. We recall that the reduction is intended to construct a 2-to-2 Game, where the goal is to assign "colors" to vertices of a graph and once a color is assigned to a vertex, there are exactly 2 colors to its neighbor that are deemed acceptable. This consideration dictates our choice of Inner PCP and specifically, what we might call "Grassmann encoding". The object being encoded is a linear function on a $\mathbb{F}_2$-vector space. Linearity is inherent to our reduction which dictates, in turn, our choice of Outer PCP as a 2-Prover-1-Round Game "played with" an underlying instance of 3-Lin, making $\mathsf{Gap3Lin}$ a natural starting point for the reduction.

Inner PCP: Grassmann Graph, "Grassmann Encoding", and our Hypotheses

Let $1 \ll \ell \ll n$ be integers. The vertex set $\mathcal{L}$, $|\mathcal{L}| = N$ of the Grassmann graph $G(\{0,1\}^n, \ell)$ consists of all ℓ-dimensional subspaces L of $\{0,1\}^n$, the n-dimensional vector space over $\mathbb{F}_2$. A pair of vertices $L, L' \in \mathcal{L}$ are connected by an edge if and only if $\dim(L \cap L') = \ell - 1$. Given a linear function $f : \{0,1\}^n \to \{0,1\}$ (or equivalently an n-bit string s_f that defines the linear function $x \to \langle s_f, x\rangle$), the Grassmann graph leads to a natural encoding of f by a string of length N over the alphabet $\Sigma = \{1, 2, \ldots, 2^\ell\}$. The encoding writes down, for every ℓ-dimensional subspace L, the linear function $f|_L$, namely, the restriction of f to the subspace L. There are exactly 2^ℓ distinct linear functions on an ℓ-dimensional space which can be numbered with $\Sigma = \{1, 2, \ldots, 2^\ell\}$.

Now suppose that the (global) linear function f is unknown, but for an edge (L, L') in the Grassmann graph, $f|_L = \sigma \in \Sigma$ is known. What do we know about $f|_{L'} = \sigma'$? We note that $\mathsf{dim}(L \cap L') = \ell - 1$ and since σ, σ' are (supposed to be) restrictions of the same global function, it must be the case that they are consistent on $L \cap L'$, i.e. $\sigma|_{L\cap L'} = \sigma'|_{L\cap L'}$. Clearly, for a given σ, there are exactly two possible choices for σ'. More generally, the "acceptable" or "consistent" pairs of functions (σ, σ') on (L, L') respectively are in 2-to-2 correspondence with each other. Let $\Phi(L, L') \subseteq \Sigma \times \Sigma$ denote this set of consistent pairs.

We are naturally led to the following 2-to-2 Game: assign "colors" from Σ (interpreted as linear functions on ℓ-spaces) to the vertices of the Grassmann graph, and be consistent on significant fraction of the "constraints" $\Phi(L, L')$. Of course, one option is to pick a global linear function f and assign $L \to f|_L$; such strategy yields consistency on all edges. Is this essentially the only strategy? Before proceeding, let us mention that there are two notions of consistency that are natural and relevant:

[4] Perhaps a useful analogy here is the text-book reduction from 3SAT to the Traveling Salesperson. Therein, for every variable of of the 3SAT instance, there is a copy of a fixed graph (= TSP-gadget) and then edges are added between these copies using the clauses of the 3SAT instance.

- (Standard Consistency): An assignment $A : \mathcal{L} \to \Sigma$ is said to be δ-consistent if it is consistent on δ fraction of the edges, i.e. for δ fraction of the edges (L, L'), $(A[L], A[L']) \in \Phi(L, L')$.
- (Non-Standard Consistency): An assignment $F : \mathcal{L} \to \binom{\Sigma}{j} \cup \{\emptyset\}$ (i.e. every vertex either gets j colors or does not get any color) is said to be (j, δ)-consistent if there is a subset $\mathcal{S} \subseteq \mathcal{L},\ |\mathcal{S}| \geqslant \delta|\mathcal{L}|$ such that (a) for all $L \in \mathcal{S},\ F[L] \neq \emptyset$ (b) for all edges (L, L') inside $\mathcal{S}$, there are colors $\sigma \in F[L], \sigma' \in F[L']$ with $(\sigma, \sigma') \in \Phi(L, L')$.

One is tempted to speculate that (parameters ℓ, n are thought of as arbitrarily large with $\ell \ll n$):

Speculation (1): For every δ, there is δ' such that given any δ-consistent (in the standard sense) assignment A, there is a global linear function $f : \{0,1\}^n \to \{0,1\}$ such that for δ' fraction of the vertices L, it holds that $A[L] = f|_L$.

Speculation (2): For every $j \geqslant 1, \delta > 0$, there is δ', such that given any (j, δ)-consistent (in the non-standard sense) assignment F, there is a global linear function $f : \{0,1\}^n \to \{0,1\}$ such that for δ' fraction of the vertices L, it holds that $f|_L \in F[L]$.

In coding-theoretic terms, in both the speculations, the assignments A or F are regarded as "received words". If the desired global linear function f exists, it then serves as a "decoding" of the received word. We however know that Speculation (2) is false in the case $j \geqslant 3$. A counter-example appears in Section 2. Since Speculation (1) implies Speculation (2),[5] Speculation (1) is also false. We believe that Speculation (2) is correct in the case $j = 1$ and seems to present interesting challenges. We show, in Section 6, that it follows from our Hypothesis 2.7, via a "ℓ-space vs b-space" linearity test.[6] The linearity test and its analysis using Fourier method are presented in Section D. Hypothesis 2.7 states that in the Grassmann graph, a set of constant density contains a connected component of constant density inside it. In addition, we propose Hypothesis 2.8 stating that the Grassmann graph is a "small set vertex expander".[7] These hypotheses might be a good starting point for further investigation.

We now state our main hypothesis informally. We would like to somehow salvage Speculation (2). We hypothesize that (see Hypothesis 2.5 for a formal statement) given a (j, δ)-assignment to the Grassmann graph $G(\{0,1\}^n, \ell)$, there exists a q-dimensional subspace Q such that if one "zooms-into" the subgraph induced on ℓ-spaces L that contain Q (this subgraph is isomorphic to $G(\{0,1\}^{n-q}, \ell - q)$), then indeed there is a global linear function that is consistent with the given assignment on δ' fraction of vertices in the induced subgraph. Here q, δ' depend on j, δ. We in fact hypothesize that the zoom-in is successful in this sense for α fraction of q-dimensional subspaces Q where α depends on j, δ, ℓ. This hypothesis is sufficient to prove Conjecture 1.6. The "zoom-in" is a new feature in the context of Inner PCPs and our Outer PCP needs to have an appropriate mechanism to handle it.

Outer PCP: 2-Prover-1-Round Game

We present the Outer PCP as a 2-Prover-1-Round Game. Usually, this game is constructed from a hard instance of 3SAT, e.g. [19, 18], in which case it is compatible with a "Long Code" based Inner PCP.

[5] The Grassmann graph is dense in the sense that a set of density δ contains at least δ^2 fraction of the edges. Given a (j, δ)-assignment, one can pick a random assignment from its j-list for every vertex that has been assigned and satisfy $\frac{\delta^2}{j}$ fraction of the edges in expectation. Hence existence of a (j, δ)-consistent assignment implies existence of a $\frac{\delta^2}{j}$-consistent assignment.

[6] This test is in the spirit of "line vs point" and "plane vs plane" low degree test in [33, 3, 32]. However our analysis is Fourier-based instead of algebraic and combinatorial.

[7] The Grassmann graph $G(\{0,1\}^n, \ell)$ is not a "small set edge expander". It has sets of sub-constant size with edge expansion $\leqslant \frac{1}{2}$, e.g. fix a non-zero point $x \in \{0,1\}^n$ and consider the set of all ℓ-spaces containing x.

However in our case, to be compatible with the "Grassmann Code" based Inner PCP, the game needs to be constructed from a hard instance of a linear constraint satisfaction problem, $\mathsf{Gap3Lin}$ being the natural choice (this has been done previously, e.g. [22, 28], with a "Hadamard Code" based Inner PCP).

Let (X, Eq) be an instance of $\mathsf{Gap3Lin}(1-\varepsilon, s^*)$ where ε can be chosen to be arbitrarily small and $s^* < 1$ is an absolute constant. The 2-Prover-1-Round Game is a game between a verifier and two non-communicating provers, where the provers wish to convince the verifier that the instance (X, Eq) has a $(1-\varepsilon)$-satisfying assignment. Fix a parameter k, thought of as a large integer, and a "smoothness" parameter β, say $\beta = k^{-\frac{3}{4}}$ for the sake of concreteness. The game proceeds as follows:

- The verifier picks at random k equations $\{e_1, \dots, e_k\}$, lets U to be the set of $3k$ variables that appear in these equations, and sends U to the first prover as a question.
- The verifier picks a subset of variables $V \subseteq U$ by including in V, independently for $1 \leqslant i \leqslant k$, (a) all three variables from the equation e_i with probability $1-\beta$ and (b) one of the three variables chosen at random from the equation e_i with probability β. Note that the size of V is, w.h.p., close to its expected size $3k - 2\beta k$, so V is nearly the same as U. The verifier sends V to the second prover as a question.
- The provers answer with bit-strings $s_U, |s_U| = |U|$ and $s_V, |s_V| = |V|$ respectively, supposedly giving the assignment to the set of variables they received.
- The verifier accepts if and only if $s_U|_V = s_V$ (i.e. if the two provers agree on the shared variables V) and s_U satisfies the k equations (this is known as a "side condition").

The parameter k is a constant, so the size of the game is polynomial in the size of the $\mathsf{Gap3Lin}$ instance. Instead of viewing the game as "active" verification, one can write down the description of the game as a graph, with possible questions as its vertices and possible question-pairs asked to the provers as its edges. The game is then viewed as a "passive" optimization problem: assigning colors (= bit-strings of appropriate length) to the vertices, so as to satisfy constraints on the edges. The following statements show that approximating the provers' optimal strategy (which, in the passive view, is same as a coloring that maximizes the fraction of the edge-constraints satisfied) is a very hard problem, and hence can be used as a canonical hard problem for further reduction.

(Completeness): It is clear that if the instance (X, Eq) has a $(1-\varepsilon)$-satisfying assignment, the provers can answer according to this (global) assignment. The k equations chosen by the verifier are all satisfied with probability $\geqslant 1 - k\varepsilon$, in which case the verifier accepts.

(Soundness): On the other hand, it follows from the Parallel Repetition Theorem [31, 20, 30, 12] that if every assignment to the instance (X, Eq) is at most s^*-satisfying, then any strategy of the provers can make the verifier accept with probability at most $2^{-\Omega(\beta k)}$.

Composition of Inner and Outer PCP

We "compose" the Inner PCP and the Outer PCP, constructing an instance $G_{2:2}$ of a Transitive 2-to-2 Game as in Definition 1.4 and Theorem 1.8. Only the questions U to the first prover in the Outer PCP appear explicitly in the construction whereas the questions V to the second prover are only implicitly used. The composition, at a high level, is rather straightforward. However, incorporating the "side conditions" from the Outer PCP and ensuring the "transitivity" of the 2-to-2 Game $G_{2:2}$ present serious difficulties. Both of these issues are skipped altogether from this overview. Also, in the actual reduction, there are more constraints in $G_{2:2}$ that described here.

In the 2-to-2 Game $G_{2:2}$, for every question U to the first prover, there is a copy of the Grassmann graph $G(\{0,1\}^U, \ell)$. A vertex L in this graph is to be assigned a color (or a j-list of colors) from the alphabet $\Sigma, |\Sigma| = 2^\ell$, the colors being interpreted as linear functions on L. The intention is as follows. Suppose that in the Outer PCP, the prover intends to answer with a bit string $s, |s| = |U| = 3k$, a supposed assignment to the variables in U. The string s is thought of as the linear "inner product" function $f_s : \{0,1\}^U \to \{0,1\},\ f_s(x) = \langle s, x\rangle$. The assignment of colors to the vertices of the Grassmann graph is then precisely the encoding of the linear function f_s, i.e. a vertex L is assigned the color $f_s|L$. We add the 2-to-2 constraints for all edges (L, L') of this Grassmann graph as in the Inner PCP.

To summarize, the 2-to-2 Game $G_{2:2}$ has a block of vertices for every question U to the first prover and along with the edge-constraints inside it, the block is exactly a copy of the Grassmann graph/encoding/game. Now we describe the edges across two different blocks. Let U, U' be two distinct questions to the first prover and V be a question to the second prover such that $V \subseteq U$ as well as $V \subseteq U'$, i.e. the verifier can potentially ask the question-pair (U, V) as well as the question-pair (U', V). Obviously the space $\{0,1\}^V$ is contained in both $\{0,1\}^U$ and $\{0,1\}^{U'}$. There are two types of edges[8] between the block of U and the block of U'.

- For any $L \subseteq \{0,1\}^V, \mathsf{dim}(L) = \ell$, L is contained in both U and U', and hence there are vertices u, u' in their blocks corresponding to L. We add a 1-to-1 constraint between u, u'. Note that (a) the colorings to the blocks of U, U' are supposed to be the encodings of the "inner product" functions f_{s_U} and $f_{s_{U'}}$ respectively (b) the assignments s_U and $s_{U'}$ are supposed to be restrictions of some global assignment to U and U' respectively and hence are supposed to agree with an assignment s_V on V. Therefore, the linear function $f_{s_U}|_L$, i.e. the intended color of u, and the linear function $f_{s_{U'}}|_L$, i.e. the intended color of u', must be the same, i.e. $f_{s_U}|_L = f_{s_V}|_L = f_{s_{U'}}|_L$. This defines the 1-to-1 constraint between u, u'.

- Similarly, for any $L, L' \subseteq \{0,1\}^V, \mathsf{dim}(L) = \mathsf{dim}(L') = \ell, \mathsf{dim}(L \cap L') = \ell - 1$, L is contained in $\{0,1\}^U$ and L' is contained in $\{0,1\}^{U'}$, and hence there are vertices u, u' in the blocks of U, U' respectively, corresponding to L, L' respectively. We add a 2-to-2 constraint between u, u'. As before, the intended color of u is $f_{s_U}|_L = f_{s_V}|_L$ and the intended color of u' is $f_{s_{U'}}|_{L'} = f_{s_V}|_{L'}$. Since $\mathsf{dim}(L \cap L') = \ell - 1$, there is a 2-to-2 correspondence between the functions $f_{s_V}|_L$ and $f_{s_V}|_{L'}$, which defines the 2-to-2 constraint between u, u'.

This completes the informal description of the reduction. The actual reduction, with several additional details, is presented in Section 4. The transitivity of the game $G_{2:2}$ is proved in Section A.

Advice, Covering Property, and Soundness Analysis

Let us attempt the soundness analysis at a high level, showing the need for an additional "advice feature" in the Outer PCP, as well as a certain "covering property". Given a (j, δ)-coloring to the game $G_{2:2}$, the soundness analysis derives prover strategies in the Outer PCP with a good success probability. This implies conversely that if the Outer PCP is chosen beforehand to have low enough soundness ($= 2^{-\Omega(\beta k)}$), then the game $G_{2:2}$ has no (j, δ)-coloring, completing the proof of Theorem 1.8.

Accordingly, suppose there is a (j, δ)-coloring to the game $G_{2:2}$ and assume for simplicity that a δ fraction of vertices in every block have been (j-list-)colored. Fix a question U to the first prover. Towards deriving her answer, she looks at the (j, δ)-coloring of her copy of the Grassmann graph $G(\{0,1\}^U, \ell)$. Our

[8] In the actual reduction, potentially, there are edges between blocks of U, U', even when there is no question V that appears along with both U, U'.

main hypothesis (Hypothesis 2.5) implies that for a good fraction of q-dimensional subspaces $Q \subseteq \{0,1\}^U$, after zooming-into the subgraph induced on vertices (= ℓ-spaces) that contain Q, there is a global linear function $f = f_s : \{0,1\}^U \to \{0,1\}$ that has good agreement with the given coloring. The prover returns the string s as her answer. Though she does not know which zoom-in space Q works, she can chooses Q randomly, and it is hypothesized to work with a good probability. We emphasize that the "decoded" global linear function f_s, and hence her answer, in general depends on the choice of the zoom-in space Q.

Now let V be the question to the second prover. Since $\{0,1\}^V$ is contained in $\{0,1\}^U$, the Grassmann graph $G(\{0,1\}^V, \ell)$ is an induced subgraph of the Grassmann graph $G(\{0,1\}^U, \ell)$. The second prover wishes to derive his answer from the coloring to his graph $G(\{0,1\}^V, \ell)$. Let S be the subset of vertices in the graph $G(\{0,1\}^U, \ell)$ that are colored, with density of S being δ. The prover can only use the coloring to the set $S \cap G(\{0,1\}^V, \ell)$, which might however have negligible density in $G(\{0,1\}^V, \ell)$. If so, he has no information to derive his answer from, and in the worst case, this could happen for almost every question V asked to him, for a fixed question U to the first prover. The purpose of the "smoothness" parameter in the Outer PCP is to precisely avoid this issue. Provided that $\beta\sqrt{k} \cdot 2^\ell \to 0$ (which happens with $\beta = k^{-\frac{3}{4}}$ and k large enough), V has expected size $|U| - 2\beta k$ that is close enough to $|U| = 3k$ that one has the following guarantee: for a fixed question U to the first prover and for any subset S of density δ in $G(\{0,1\}^U, \ell)$, for almost every question V to the second prover, the density of $S \cap G(\{0,1\}^V, \ell)$ in $G(\{0,1\}^V, \ell)$ is $\approx \delta$. This guarantee is referred to as the "covering property", a special case of which is defined and used in [28]. The property is stated formally as Lemmas 4.6, 4.7 and proved in Section C.

We assume therefore that for the question V to the second prover, δ fraction of vertices of his Grassmann graph $G(\{0,1\}^V, \ell)$ are colored. In a similar manner as the first prover, he wishes to zoom-into a q-dimensional subspace $Q' \subseteq \{0,1\}^V$, decode a global linear function $f_{s'} : \{0,1\}^V \to \{0,1\}$, return s' as the answer, and hope that $s' = s|_V$, i.e., that his answer is consistent with the first prover's answer. Strictly speaking, he outputs a short list of all $f_{s'}$ that have agreement with coloring to his Grassmann graph and hopes that one of them is consistent with the first prover's function f_s. It is reasonable to expect this consistency because both provers are using the same coloring: if f_s is consistent with coloring to $S \subseteq G(\{0,1\}^U, \ell)$, then its restriction $f_s|_V$ is consistent with $S \cap G(\{0,1\}^V, \ell)$ for a good fraction of questions V to the second prover, and then $f_s|_V = f_{s'}$ appears in the decoded list of the second prover. There is one catch however. The decoded global functions depend in general on the zoom-in space, so the two provers must "agree" on the zoom-in space, i.e. manage to choose $Q = Q'$, without communication. We resolve this issue by letting the verifier in the Outer PCP choose a random q-dimensional subspace $Q \subseteq \{0,1\}^V$ and send it to both provers as extra "advice" along with their questions. We make sure that this advice does not compromise or hurt the soundness of the Outer PCP.

2 The Grassmann Graph and related Hypotheses

In this section we introduce the Grassmann graph and related hypotheses that are relevant towards the soundness analysis of our PCP construction (at the Inner PCP level). The Grassmann graph leads to an encoding of a linear function on a high-dimensional $\mathbb{F}_2$-vector space and a "2-to-2 test" to check the encoding. The linear function is encoded by writing down its restriction to all ℓ-dimensional subspaces, the restrictions themselves being linear functions on the respective subspaces. Given a supposed encoding, i.e. an assignment of a linear function to every ℓ-dimensional subspace, one can test that the given linear functions on a pair of ℓ-dimensional subspaces are consistent on their intersection. For the test to have the 2-to-2 property, the test is performed only on a pair of ℓ-dimensional subspaces for which their intersection is $(\ell-1)$-dimensional. Naturally, the following "decoding" question arises: Given an assignment to the ℓ-spaces that demonstrates some consistency, is there a global linear function that explains some (or almost all) of the consistency? We hypothesize that the answer is affirmative, but subtle. Also, towards the analysis of our reduction, we have to deal with more general "(j, δ)-assignments" where a δ fraction of the ℓ-spaces are each assigned a list of j linear functions on it.

2.1 The Grassmann Graph

Let $\mathbb{F}$ be a field of size p, V be a linear space of dimension n over $\mathbb{F}$ and $1 \leqslant \ell \leqslant n-1$ a positive integer. The Grassmann graph $G(V, \ell)_p$ is defined as follows:

- The vertices are all ℓ-dimensional subspaces of V.
- The edges are pairs of vertices L, L' such that $\mathsf{dim}(L \cap L') = \ell - 1$.

The Grassmann graph has been moderately studied in the literature, mainly in the context of distance-regular graphs [6]. Classical theorems such as Erdös-Ko-Rado and Kruskal-Katona are also known to hold [8]. Here are some known facts regarding the Grassmann graph, though we do not necessarily need them.

Fact 2.1. *Suppose* $1 \leqslant \ell \leqslant \frac{n}{2}$.

1. The number of vertices in the graph $G(V, \ell)_p$ is the ℓ^{th} p-nomial coefficient (sometimes referred to as "Gaussian binomial coefficient")

$$\begin{bmatrix} n \\ \ell \end{bmatrix}_p \overset{def}{=} \prod_{i=0}^{\ell-1} \frac{p^n - p^i}{p^\ell - p^i}.$$

2. The graph is regular with degree $d = \frac{p^n - p^\ell}{p^\ell - p^{\ell-1}} \begin{bmatrix} \ell \\ \ell-1 \end{bmatrix}_p$. For $p = 2$, this is $\Theta(2^n)$.

3. The eigenvalues of the (adjacency matrix of the) graph are

$$\lambda_j = p^{j+1} \begin{bmatrix} \ell - j \\ 1 \end{bmatrix} \begin{bmatrix} n - \ell - j \\ 1 \end{bmatrix} - \begin{bmatrix} j \\ 1 \end{bmatrix}_p.$$

with multiplicities $\begin{bmatrix} n \\ j \end{bmatrix}_p - \begin{bmatrix} n \\ j-1 \end{bmatrix}_p$, for $j = 0, ..., \ell$ [9].

For $p = 2$ the eigenvalues are approximately $(1 + 2^{j-\ell}) 2^{n+1-j}$, and so the normalized eigenvalues are $\Theta(2^{-j})$.

We will only be interested in the case $p = 2$ and the subscript p will be omitted henceforth. The following observation will be useful.

Fact 2.2. *Given the Grassmann graph $G(V, \ell)$, $\mathsf{dim}(V) = n$ and a q-dimensional subspace $Q \subseteq V$, $0 \leqslant q \leqslant \ell - 1$, let Zoom_Q denote the subset of vertices $L \in G(V, \ell)$ such that $Q \subseteq L$. Then the induced subgraph on the subset Zoom_Q is isomorphic to the (lower order) Grassmann graph $G(V', \ell')$ with $\mathsf{dim}(V') = n - q$, $\ell' = \ell - q$. A natural isomorphism is by letting $V' = V/Q$ to be the quotient space.*

[9] $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is defined to be 0.

2.2 (j, δ)-Assignments and Zooming-in

Let $[2^\ell]$ denote the set of linear functions on an ℓ-dimensional space. We would need to consider the so-called "(j, δ)-assignments" to the vertices of the Grassmann graph.

Definition 2.3. *The density of a set $S \subseteq G(V, \ell)$ is its fractional size, i.e.* $\mathsf{Density}(S) = \frac{|S|}{|\{L \in G(V,\ell)\}|} = \frac{|S|}{\left[{n \atop \ell}\right]}$.

Definition 2.4. *Let $S \subseteq G(V, \ell)$ and let $F : S \to \binom{[2^\ell]}{j}$ assign, to each ℓ-space $L \in S$, a set $F[L]$ of j linear functions on L. The assignment F is said to be (j, δ)-edge-consistent, or simply a (j, δ)-assignment, if*

- *S has density at least δ.*
- *If $L_1, L_2 \in S$ are connected by an edge, then there is a pair $a_1 \in F[L_1], a_2 \in F[L_2]$ such that a_1, a_2 agree on $L_1 \cap L_2$.*

We consider the scenario where j and δ are given constants, then ℓ is allowed to be sufficiently large, and finally the global dimension n is allowed to be sufficiently large compared to ℓ. One would hope that a (j, δ)-assignment implies the existence of a global linear function $g\colon V \to \{0, 1\}$ that "explains" some of the consistency. Specifically, is there a global linear function g such that $g|_L \in F[L]$ for $\delta' = \delta'(j, \delta)$ fraction of L? The answer turns out to be negative as seen from the following example.

2.2.1 Subspace Example

Fix $Z \subseteq V$ to be a subspace of dimension $n - \ell$ and pick a set of vertices $S \subseteq G(V, \ell)$ as

$$S = \{ L \in G(V, \ell) \mid \mathsf{dim}(L \cap Z) = 2 \} .$$

It is not difficult to see that S has constant density which can be computed to be ≈ 0.20. We will exhibit a $(3, \approx 0.20)$-assignment $F[\cdot]$ which has no non-trivial consistency with any global linear function. For each $z \in Z \setminus \{0\}$, choose $f_z : V \to \{0, 1\}$ to be a global linear function arbitrarily. The assignment $F : S \to \binom{[2^\ell]}{3}$ is now defined as

$$F[L] = \{f_z|_L \mid z \in (L \cap Z) \setminus \{0\}\}.$$

In words, L is assigned three linear functions that are restrictions to L of the three global linear functions f_z for $z \in (L \cap Z) \setminus \{0\}$. Note that $\mathsf{dim}(L \cap Z) = 2$ and hence $|F[L]| = 3$. Now we show that if $L_1, L_2 \in S$ have an edge connecting them, then the assignments $F[L_1], F[L_2]$ are consistent. Indeed, when $\mathsf{dim}(L_1) = \ell$, $\mathsf{dim}(L_1 \cap L_2) = \ell - 1$, $\mathsf{dim}(L_1 \cap Z) = 2$, we have

$$1 \leqslant \mathsf{dim}(L_1 \cap L_2 \cap Z) \leqslant 2.$$

In particular, there exists $z \in (L_1 \cap L_2 \cap Z) \setminus \{0\}$. By design, we have $f_z|_{L_1} \in F[L_1]$ and $f_z|_{L_2} \in F[L_2]$ and moreover that $f_z|_{L_1}, \ f_z|_{L_2}$ agree on $L_1 \cap L_2$, both being restrictions of the same global function f_z. Finally, we note that since the choice of functions f_z is arbitrary, no global linear function has a non-trivial consistency with the assignment $F[\cdot]$. For the sake of concreteness, one can let f_z be the linear "inner product" function $x \to \langle z, x \rangle$. It is not difficult to see that for any global linear function $f : V \to \{0, 1\}$,

$$\Pr_{L \in G(V,\ell)} [f|_L \in F[L]] \leqslant 2^{-\Omega(\ell)}.$$

2.2.2 Zooming-in

We now consider a plausible way to circumvent the above example. For a Grassmann graph $G(V,\ell)$ and a subspace $Q \subseteq V$, let

$$\mathsf{Zoom}_Q = \{L \mid L \in G(V,\ell),\ Q \subseteq L\}$$

be the subset of vertices that contain Q (we intend to take $\mathsf{dim}(Q) \ll \ell$). Similarly, for a subset $S \subseteq G(V,\ell)$, let

$$\mathsf{Zoom}_Q[S] = \{L \mid L \in S,\ Q \subseteq L\}$$

be the subset of vertices of S that contain Q. Returning to the subspace example above, let $F[\cdot]$ be the $(3, \approx 0.20)$-assignment to the Grassmann graph $G(V,\ell)$ as therein. Let S be the subset of density ≈ 0.20 to which $F[\cdot]$ actually assigns a list of 3 linear functions. We noted that no global linear function has a non-trivial consistency with $F[\cdot]$. To be specific, for any global linear function $f : V \to \{0,1\}$,

$$\Pr_{L \in G(V,\ell)} [f|_L \in F[L]] \leqslant 2^{-\Omega(\ell)}.$$

We observe however that there exists a one-dimensional subspace Q such that after zooming-into Q (i.e. conditioning on the ℓ-spaces containing Q), there does exist a global linear function with good consistency with $F[\cdot]$. Indeed, let $z \in Z \setminus \{0\}$ be an arbitrary point, f_z be the global linear function associated with z and let $Q = \mathsf{Span}\{z\}$. For any subspace $L \in \mathsf{Zoom}_Q[S]$, we have $z \in Q \subseteq L$ and hence $f_z|_L \in F[L]$. Thus the global linear function f_z is consistent with the assignment $F[\cdot]$ on every $L \in \mathsf{Zoom}_Q[S]$. Moreover, $\mathsf{Zoom}_Q[S]$ when regarded as a subset of Zoom_Q has a constant density, say C (in fact its density is higher than the original density of S in $G(V,\ell)$ which is ≈ 0.20). Thus

$$\Pr_{L \in G(V,\ell)} [f_z|_L \in F[L] \mid Q \subseteq L] \geqslant C$$

where the probability is conditional on the ℓ-spaces containing Q. Further, if the point z were chosen at random from the global space V, with probability $\approx 2^{-\ell}$, we have $z \in Z \setminus \{0\}$ and then zooming-into $Q = \mathsf{Span}\{z\}$ gives a global linear function with good consistency. To summarize our specific example,

Given a $(3, \approx 0.20)$-assignment $F[\cdot]$ to $G(V,\ell)$, for $\approx 2^{-\ell}$ fraction of one-dimensional subspaces $Q \subseteq V$, zooming-into Q gives a global linear function that is $\Omega(1)$-consistent with $F[\cdot]$.

We now hypothesize that something to this effect always holds for any (j,δ)-assignment to a Grassmann graph $G(V,\ell)$ when one is allowed to zoom-into a q-dimensional subspace Q with constant q and the zoom-in succeeds for a non-negligible fraction (that may depend arbitrarily on ℓ) of q-dimensional subspaces Q. Our main hypothesis appears below, followed by its variants and special cases.

2.3 Our Hypotheses

The main hypothesis

Hypothesis 2.5. *For every integer $j \geqslant 1$ and constant $\delta > 0$, there exist an integer $q \geqslant 0$, a constant $C > 0$, and a function $\alpha(\cdot) > 0$ of an integer parameter such that for all sufficiently large integers ℓ, for all sufficiently large integers n, the following holds: Let $F[\cdot]$ be a (j,δ)-assignment to the Grassmann graph $G(V,\ell)$ with $\mathsf{dim}(V) = n$. Then for at least $\alpha(\ell)$ fraction of the q-dimensional subspaces $Q \subseteq V$, there exists a global linear function $g_Q : V \to \{0,1\}$ such that (note the conditional probability)*

$$\Pr_{L \in G(V,\ell)} [g_Q|_L \in F[L] \mid Q \subseteq L] \geqslant C. \tag{1}$$

Upper bound on list-decoding size

In Hypothesis 2.5, given a (j,δ)-assignment $F[\cdot]$, a global linear function g_Q satisfying Equation (1) is viewed as a "decoded" global linear function. Naturally, such decoded function is only useful when there are not many functions satisfying Equation (1), hence one would like to obtain an upper bound on the number of such functions. Indeed, a reasonable upper bound, stated below, follows from the work of Blinovsky [5]. We include a proof in Section E.2 for the sake of completeness.

Theorem 2.6. *Let $F[\cdot]$ assign to every $L \in G(V,\ell)$, a list $F[L]$ of at most j linear functions on L. Let Q be any q-dimensional subspace of V. Then there are at most $\frac{jC}{C^2 - j\cdot 2^{q-\ell}}$ global functions g for which*

$$\Pr_{L\in G(V,\ell)} [g|_L \in F[L] \mid Q \subseteq L] \geqslant C.$$

Hypotheses about connectivity of Grassmann graph

We are far from understanding the case of general j in the Hypothesis 2.5 and even the case $j = 1$ presents interesting challenges. We show in Section 6 that Hypothesis 2.5 in the case $j = 1$ follows from the Hypothesis 2.7 below without the need for zoom-in. A linearity test is developed therein that could be of independent interest.

Hypothesis 2.7. *For every constant $\delta > 0$, there exists a constant $\varepsilon > 0$ such that for all sufficiently large integers ℓ, for all sufficiently large integers n, the following holds: Let S be any set of vertices in the Grassmann graph $G(V,\ell)$, $\mathsf{dim}(V) = n$ with density at least δ. Then the induced subgraph on S contains a connected component of density at least ε.*

The hypothesis below seems like a natural related question, stating that the Grassmann graph is a "small set vertex expander". This could be a good starting point for further investigations.

Hypothesis 2.8. *There is a function $\lambda(\varepsilon)$ of a positive parameter ε such that*

- *$\lambda(\varepsilon) \to \infty$ as $\varepsilon \to 0$.*
- *For any $\varepsilon > 0$, for all large enough integers ℓ, for all large enough integers n, the following holds: Let S be a set of vertices in the Grassmann graph $G(V,\ell)$, $\mathsf{dim}(V) = n$ with density ε and let $\Gamma(S)$ denote the vertex neighborhood of S, i.e. $\Gamma(S) = \{L' \mid \exists\, L \in S$ such that $(L, L') \in G(V,\ell)\}$. Then*

$$|\Gamma(S)| \geqslant \lambda(\varepsilon)\cdot |S|\,.$$

Hypothesis with side condition

As is standard, while composing the "Inner PCP" with the "Outer PCP", we require that the decoded global linear function g_Q in Equation (1), Hypothesis 2.5 itself respects certain linear "side condition". We state a variant of Hypothesis 2.5 that takes into account the side condition and show that this variant follows easily from Hypothesis 2.5.

Definition 2.9. *A pair $(\{h_1,...,h_r\},(b_1,...,b_r))$, where $\{h_i \in \{0,1\}^n\}_{i=1}^r$ are linearly independent and $b_i \in \{0,1\}$, is called a* side condition *for a function $g\colon \{0,1\}^n \to \{0,1\}$. We say that g respects the side condition if $g(h_i) = b_i$, for every i.*

Note that when g is a linear function respecting the side condition $(\{h_1, \ldots, h_r\}, (b_1, \ldots, b_r))$, the value of g on the space $H = \mathsf{Span}\{h_1, ..., h_r\}$ is fixed. We will often simplify notation and say g respects the side condition H, when $(b_1, \ldots, b_r)$ is clear from the context. Note that the vertices of the Grassmann graph $G(V, \ell)$ are ℓ-dimensional subspaces of V. Now we instead think of the vertex set as

$$\{ L \oplus H \mid L \in G(V, \ell)\},$$

restricted to only those L such that $L \cap H = \{0\}$ and moreover, if $L \oplus H = L' \oplus H$, then the two vertices are identified together[10]. Note that $\mathsf{dim}(L \oplus H) = \ell + r$. There is an edge between $L \oplus H$ and $L' \oplus H$ if and only their intersection has dimension $\ell + r - 1$. It can be easily seen that the resulting graph is isomorphic to a "lower order" Grassmann graph $G(V', \ell)$, where $V' \subseteq V$ is a complementing space to H (i.e. $V' \oplus H = V,\ V' \cap H = \{0\},\ \mathsf{dim}(V') = n - r$).

A "(j, δ)-assignment respecting the side condition" is an assignment $F[\cdot] : S \to \binom{[2^\ell]}{j}$ to a set of vertices S such that

- S has density at least δ.
- For each vertex $L \oplus H \in S$, $F[L \oplus H]$ is a list of j linear functions on $L \oplus H$ that respect the side condition. Note that since the side condition already specifies the values of a linear function on H, the number of linear functions on $L \oplus H$ that respect the side condition is exactly 2^ℓ.
- For any $L \oplus H, L' \oplus H \in S$ that are connected by an edge, there are linear functions $a \in F[L \oplus H],\ a' \in F[L' \oplus H]$ that agree on the intersection $L \oplus H \cap L' \oplus H$.

We now state the variant of Hypothesis 2.5 that takes into account the side condition. A (rather self-evident) proof that Hypothesis 2.10 follows from Hypothesis 2.5 appears in Section E.1.

Hypothesis 2.10. *For every integer $j \geqslant 1$ and constant $\delta > 0$, there exist an integer $q \geqslant 0$, a constant $C > 0$, and a function $\alpha(\cdot) > 0$ of an integer parameter such that for all sufficiently large integers ℓ, for all sufficiently large integers n, the following holds: Let $F[\cdot]$ be a (j, δ)-assignment respecting the side condition $(\{h_i\}_{i=1}^r,\ \{b_i\}_{i=1}^r)$ to the Grassmann graph $G(V, \ell)$ with $\mathsf{dim}(V) = n$ and $r \leqslant \frac{n}{3}$. Then for at least $\alpha(\ell)$ fraction of the q-dimensional subspaces $Q \subseteq V$, there exists a global linear function $g_Q : V \to \{0, 1\}$ that respects the side condition such that (note the conditional probability)*

$$\Pr_{L \in G(V, \ell)} [g_Q|_{L \oplus H} \in F[L \oplus H] \mid Q \subseteq L] \geqslant C. \tag{2}$$

3 The Outer PCP

Our Outer PCP is a carefully constructed 2-Prover-1-Round Game from a regular instance of the 3-Lin problem. Recall (see the paragraph before Theorem 1.8) that an instance (X, Eq) of the 3-Lin problem consists of a set of $\mathbb{F}_2$-valued variables X and a set of equations Eq, each equation containing three (distinct) variables. The instance is regular if every variable appears in exactly, say 5, equations, and two distinct equations share at most one variable. Starting with a 3-Lin instance given by Håstad's reduction [19], a standard sequence of transformations can turn the instance into a regular one, while preserving the near-perfect completeness and keeping the soundness bounded away from 1. To summarize:

[10] In our application, we will have $r = \frac{n}{3} \gg \ell$, so almost all ℓ-dimensional spaces $L \subseteq V$ satisfy $L \cap H = \{0\}$.

Theorem 3.1. *There exists an absolute constant $\frac{1}{2} < s^* < 1$ such that for every constant $\varepsilon > 0$, the* $\mathsf{Gap3Lin}(1-\varepsilon,\ s^*)$ *problem on regular instances is NP-hard.*

Let (X, Eq) be an instance of $\mathsf{Gap3Lin}(1-\varepsilon, s^*)$ as in Theorem 3.1. We intend to construct a 2-Prover-1-Round Game that is used as our Outer PCP. Instead of taking a passive view of 2-Prover-1-Round Game as a constraint satisfaction problem as in Definition 1.1, it is more intuitive to take an equivalent active view in terms of two provers and a probabilistic verifier. The two provers wish to convince the verifier that the 3-Lin instance is near-satisfiable. Since our construction has multiple subtle features, we present it incrementally, adding one feature at a time. The construction is along the lines of [28], "smoothness" and "covering" features are as therein and there is an additional "advice" feature.

3.1 Equation vs Variable Game

We start with a standard "equation vs variable" game that the reader might be already familiar with. In this game, the verifier chooses an equation $e \in \mathsf{Eq}$ uniformly at random, sends it to the first prover, chooses a variable x randomly from the three variables occurring in the equation e and sends it to the second prover. The provers are expected to provide a $\mathbb{F}_2$-value for each of the variables they receive. The verifier accepts if and only if the first prover provides a satisfying assignment to e and if both provers give x the same value.

Completeness: Suppose there is an assignment to (X, Eq) that satisfies $1-\varepsilon$ fraction of the equations. The provers can answer according to this assignment and the verifier accepts with probability at least $1-\varepsilon$.

Soundness: Suppose no assignment to (X, Eq) satisfies more than s^* fraction of the equations. The strategy of the second prover is simply an assignment to all the variables. This assignment fails to satisfy $1-s^*$ fraction of the equations. For every equation that fails, the second prover either has to give inconsistent answer to at least one of its variables or answer with an unsatisfying assignment to the equation. Thus the provers cannot make the verifier accept with probability more than $1-\frac{1-s^*}{3}$ (i.e. bounded away from 1).

3.2 Smooth Equation vs Variable Game

We modify the equation vs variable game slightly and call it a smooth game.[11] Let $\beta \in (0,1)$ be a smoothness parameter. The verifier sends an equation e to the first prover as before. To the second prover however, the verifier sends a random variable x occurring in e with probability β, and sends the equation e with probability $1-\beta$ (hence asking the same question to both the provers).

Completeness: As before, the completeness is at least $1-\varepsilon$.

Soundness: The new game is effectively a trivial game with probability $1-\beta$ and is same as the equation vs variable game with probability β. Hence the soundness is at most $1-\Omega(\beta)$, where the Ω-notation hides the dependence on s^* (which is an absolute constant anyways).

3.3 Smooth Equation vs Variable Game with Advice

Our application requires a further modification of the smooth game. Roughly speaking, the provers are also provided extra "advice" that acts like publicly shared randomness . Nevertheless, this advice cannot considerably help the provers.

[11] Smoothness refers to the property of a game wherein for a fixed question and two distinct answers to the first prover, w.h.p. over the choice of the question to the second prover, the second prover's answers need to be distinct for the verifier to accept. The game described is smooth provided $\beta \ll 1$.

As before, the verifier picks an equation e at random, say $x_{i_1} + x_{i_2} + x_{i_3} = b_i$, and sends it to the first prover. With probability $1 - \beta$, the second prover receives the equation e as well, and otherwise a single variable from the equation e chosen at random. Let $V \subseteq \{x_{i_1}, x_{i_2}, x_{i_3}\}$ be the set of variables sent to the second prover (so $|V|$ is 1 or 3). The verifier chooses an advice vector $a \in \{0,1\}^V$ at random. If $|V| = 3$, define $a^* = a$, and if $|V| = 1$, let a^* be obtained from a by padding with 0 in place of $\{x_{i_1}, x_{i_2}, x_{i_3}\} \setminus V$. The verifier sends the first prover the vector a^* and the second prover the vector a. As before, the provers are expected to provide a value for each of the variable they receive.

Call this game $G_{\beta,1}$. The extra advice could give the first prover a hint as to which variables the second prover receives. For example, if the first prover's advice vector is $a^* = (0,0,1)$, she knows that the second prover has received either all three variables or (just) the variable x_{i_3}. However, when the first prover receives the vector $(0,0,0)$, she does not know whether the second prover has received all three variables along with advice $a = (0,0,0)$ or a single variable, whose identity she does not know, along with advice $a = (0)$. It is clear from this discussion that:

Completeness: The completeness of game $G_{\beta,1}$ is at least $1 - \varepsilon$.

Soundness: The soundness of game $G_{\beta,1}$ is at most $1 - \Omega(\beta)$.

We further generalize to the game $G_{\beta,q}$ for any integer $q \geqslant 0$ where instead of sampling and sending the provers one pair (a^*, a) respectively, the verifier samples independently, q pairs $(a_1^*, a_1), \ldots, (a_q^*, a_q)$, and sends the list $[a_1^*, \ldots, a_q^*]$ to the first prover and the list $[a_1, \ldots, a_q]$ to the second prover. It is not difficult to see that:

Completeness: The completeness of game $G_{\beta,q}$ is at least $1 - \varepsilon$.

Soundness: The soundness of game $G_{\beta,q}$ is at most $1 - \Omega\left(\frac{\beta}{2^q}\right)$. Intuitively, the verifier rejects when the second prover is sent a single variable (which happens with probability β) along with the advice-list $[(0), \ldots, (0)]$ (which happens with probability $\frac{1}{2^q}$).

Remark 3.2. *The soundness of the $\left(\frac{2^q}{\beta}\right)$-fold parallel repetition game $G_{\beta,q}^{\otimes 2^q/\beta}$ is upper bounded by an absolute constant less than* 1. *Intuitively, in $\frac{2^q}{\beta}$ "trials", with constant probability, there is a "coordinate" on which the second prover receives a single variable along with the advice-list* $[(0), \ldots, (0)]$, *and then the verifier rejects with a constant probability.*

3.4 The Final Game (Outer PCP)

Finally, our Outer PCP is a k-fold parallel repetition of the game $G_{\beta,q}$, i.e. the game $G_{\beta,q}^{\otimes k}$.

Completeness: The completeness of game $G_{\beta,q}^{\otimes k}$ is at least $1 - k\varepsilon$.

Soundness: The soundness of game $G_{\beta,q}^{\otimes k}$ is at most $2^{-\Omega(\beta k/2^q)}$. The game can be considered as $\frac{\beta k}{2^q}$-fold parallel repetition of the game $G_{\beta,q}^{\otimes 2^q/\beta}$ which has constant soundness as per Remark 3.2. One can then apply the parallel repetition theorem for projection games with no dependency on the answer size as in [30, 12].

Remark 3.3. *Let U, V be the questions sent to the first and the second prover in the game $G_{\beta,q}^{\otimes k}$, not taking into account the "advice" yet. Thus U is a set of $3k$ variables and $V \subseteq U$ with expected size $\mathbb{E}[|V|] = 3k - 2\beta k$. With a careful look, it can be seen that the advice-list for the first prover is a list $[x_1, \ldots, x_q]$ with $\forall\, 1 \leqslant i \leqslant q,\ x_i \in \{0,1\}^U$. Similarly, the advice-list for the second prover is a list $[y_1, \ldots, y_q]$ with $\forall\, 1 \leqslant i \leqslant q,\ y_i \in \{0,1\}^V$. Moreover, if one regards the space $\{0,1\}^V$ as a subspace of $\{0,1\}^U$ in a natural manner, then $\forall\, 1 \leqslant i \leqslant q,\ x_i = y_i$. Thus the advice is to be interpreted as a list of q points in $\{0,1\}^V$ that is sent to both provers.*

4 The Main Reduction

In this section we present our reduction towards proving Theorem 1.8. The soundness analysis of the reduction is presented in Section 5.

4.1 Setting of the Parameters

Let (X, eq) be an instance of regular $\mathsf{Gap3Lin}(1-\varepsilon, s^*)$ as in Theorem 3.1. We will use the game $G_{\beta,q}^{\otimes k}$ in Section 3.4 as the "Outer PCP". Since there are several parameters involved, we specify the (tedious) order in which the parameters are chosen.

Let (j, δ) be the parameters required of the reduction in Theorem 1.8. Depending on (j, δ), let $q, C, \alpha(\cdot)$ be as given in Hypothesis 2.10 and let ℓ be an integer large enough so that Hypothesis 2.10 holds for all sufficiently large integers k ($= n$ therein). The soundness analysis of the reduction shows (modulo Hypothesis 2.10) that a (j, δ)-coloring to the 2-to-2 Game yields a prover strategy in the Outer PCP with success probability roughly $\frac{\delta \cdot \alpha(\ell) \cdot C^2}{j}$. Conversely, by setting the parameters β, k appropriately, the soundness of the Outer PCP, $2^{-\Omega(\beta k/2^q)}$, is ensured to be small enough beforehand so that the 2-to-2 Game does not have a (j, δ)-coloring. In addition, a certain smoothness or covering parameter $\beta\sqrt{k} \cdot 2^\ell$ also needs to be sufficiently small. One can choose $\frac{1}{k} \ll \beta \ll \frac{1}{\sqrt{k}}$ and k large enough so that both the soundness of the Outer PCP and the covering parameter are small enough. Finally the completeness parameter $1-\varepsilon$ for the $\mathsf{Gap3Lin}$ instance is chosen to be close enough to 1 so that the Outer PCP as well as the 2-to-2 Game have completeness $1 - k\varepsilon \geqslant 1 - \delta$.

4.2 The Reduction

Consider the game $G_{\beta,q}^{\otimes k}$ and ignore the advice for now. Let $\mathcal{U}$ and $\mathcal{V}$ denote the sets of questions asked to the first and the second prover respectively. Specifically, $\mathcal{U}$ is the set of all k-tuples of equations, $U = (e_1, \ldots, e_k)$ from the regular $\mathsf{Gap3Lin}$ instance (X, eq). For our purposes, it will be convenient to retain only those "legitimate" $U = (e_1, \ldots, e_k)$ such that (a) the equations $e_1, \ldots, e_k$ are distinct and do not share variables and (b) for any pair of variables $x \in e_i$ and $y \in e_j$, $i \neq j$, x, y do not appear together in any equation in the instance (X, eq). Due to regularity of the instance (X, eq), every variable appears in a constant number of equations, and hence the fraction of U that are not legitimate is negligible, i.e. $O(\frac{k^2}{|X|})$, and dropping these does not affect our analysis. We assume henceforth that $\mathcal{U}$ consists of only the legitimate tuples U.

The verifier in the game $G_{\beta,q}^{\otimes k}$ picks a k-tuple $U = (e_1, \ldots, e_k) \in \mathcal{U}$ uniformly at random and then constructs a k-tuple V such that independently for $1 \leqslant i \leqslant k$, the i^{th} element of V is the equation e_i with probability $1-\beta$ and is a variable in the equation e_i with probability β. Thus the set of questions $\mathcal{V}$ to the second prover consists of "mixed" tuples. In the following, we will work only with the set $\mathcal{U}$ and the role of the set $\mathcal{V}$ will be implicit.

We are now ready to describe the Transitive 2-to-2 Game $G_{2:2}(V(G_{2:2}), E(G_{2:2}), \Sigma, \Phi)$ that our reduction constructs. For any $U \in \mathcal{U}$, we regard U as the tuple of k equations $(e_1, \ldots, e_k)$ as well as the set of $3k$ variables appearing in these equations, say $(x_{11}, x_{12}, x_{13}, \ldots, x_{k1}, x_{k2}, x_{k3})$. For each equation e_i, define a vector $v_i \in \{0,1\}^U$ that has 1 on coordinates corresponding to variables in e_i and 0 on the rest. Denote $H_U = \mathsf{Span}\{v_1, ..., v_k\}$ referred to as the space of side conditions. Let $b_1, \ldots, b_k \in \{0,1\}$ be the "right hand sides" of the equations, i.e. the equation e_i is $x_{i1} + x_{i2} + x_{i3} = b_i$. Define

$$\mathcal{L}_U \stackrel{def}{=} \left\{ L \subseteq \{0,1\}^U \,\middle|\, \mathsf{dim}(L) = \ell,\ L \cap H_U = \{0\} \right\}.$$

Note that for $L \in \mathcal{L}_U$, its intersection with H_U is trivial and hence $\mathsf{dim}(L \oplus H_U) = \ell + k$. Also, $|U| = 3k$, $\mathsf{dim}(H_U) = k$ and $\mathsf{dim}(L) = \ell$. The fraction of ℓ-spaces $L \subseteq \{0,1\}^U$ such that $L \cap H_U \neq \{0\},\ L \notin \mathcal{L}_U$ is negligible ($\approx 2^{\ell - 2k}$, see Fact E.4).

Vertices of $G_{2:2}$: The game $G_{2:2}$ has a block of vertices $\mathsf{Block}[U]$ for every $U \in \mathcal{U}$ defined as

$$\mathsf{Block}[U] = \{\, L \oplus H_U \mid L \in \mathcal{L}_U \,\}\,.$$

The vertex set of $G_{2:2}$ is the (disjoint) union of all blocks:

$$V(G_{2:2}) = \bigcup_{U \in \mathcal{U}} \mathsf{Block}[U].$$

Colors of $G_{2:2}$: The set of colors Σ has size $|\Sigma| = 2^\ell$. For a vertex $L \oplus H_U$, its color set Σ is identified with

$$\{\psi : L \oplus H_U \to \{0,1\} \mid \psi \text{ is linear, } \forall\, 1 \leqslant i \leqslant k,\ \psi(v_i) = b_i\}.$$

In words, the vertex $L \oplus H_U$ is to be assigned a linear function $\psi : L \oplus H_U \to \{0,1\}$ that "respects the side conditions", meaning $\psi(v_i) = b_i$ for $1 \leqslant i \leqslant k$. Since the values of ψ are already determined on H_U, there are exactly 2^ℓ eligible linear functions ψ.

Edges and Constraints of $G_{2:2}$: Towards defining the edges and constraints of the game $G_{2:2}$, we stress a notational (and perhaps conceptual) point. X is the set of all variables in the $\mathsf{Gap3Lin}$ instance, so $U \subseteq X$ and $\{0,1\}^U$ is a subspace of $\{0,1\}^X$ in a natural manner. Every subspace under consideration can be considered as a subspace of $\{0,1\}^X$ and we can freely take the intersections or direct sums of subspaces. For instance if U_1, U_2 are two sets of variables and $L_1 \subseteq \{0,1\}^{U_1}$, $L_2 \subseteq \{0,1\}^{U_2}$ are subspaces, we can consider both L_1, L_2 as subspaces of $\{0,1\}^{U_1 \cup U_2}$ (which in turn is a subspace of $\{0,1\}^X$) and then the subspaces $L_1 \cap L_2, L_1 \oplus L_2$ make sense.

We are ready to define the edges and the constraints of $G_{2:2}$. For $U, U' \in \mathcal{U}$ (allowing the possibility that $U = U'$), we describe the edges between their respective blocks.[12] There is an edge between vertices $L \oplus H_U,\ L' \oplus H_{U'}$ if either of the two conditions holds. Either

$$\mathsf{dim}(L \oplus H_U \oplus H_{U'}) = \mathsf{dim}(L' \oplus H_U \oplus H_{U'}) = \mathsf{dim}(L \oplus L' \oplus H_U \oplus H_{U'}), \tag{3}$$

in which case, the constraint is 1-to-1, or

$$\mathsf{dim}(L \oplus H_U \oplus H_{U'}) = \mathsf{dim}(L' \oplus H_U \oplus H_{U'}) = \mathsf{dim}(L \oplus L' \oplus H_U \oplus H_{U'}) - 1, \tag{4}$$

in which case, the constraint is 2-to-2. This definition is, admittedly, rather mysterious and we try to clarify it somewhat. We recommend reading the proofs of Lemmas 4.2, 4.3, 4.4 to start having some intuition. We first consider the 1-to-1 constraints.

1**-to-**1 **Constraints:** From Lemma 4.2, we always have

$$\mathsf{dim}(L \oplus H_U \oplus H_{U'}) = \mathsf{dim}(L' \oplus H_U \oplus H_{U'}).$$

If, in addition, this dimension is same as that of $L \oplus L' \oplus H_U \oplus H_{U'}$ which contains both the spaces above, then all the three spaces must be identical, i.e. $L \oplus H_U \oplus H_{U'} = L' \oplus H_U \oplus H_{U'} = L \oplus L' \oplus H_U \oplus H_{U'} = Z$,

[12] Here U, U' are thought of as sets of variables that are nearly identical. There might be edges between the blocks of U, U' that differ significantly, but those edges are merely "accidental" and do not have much relevance towards the soundness of the reduction.

say. From Lemma 4.4, there is a 1-to-1 correspondence between linear functions on $L \oplus H_U$ (that respect side condition on H_U) and linear functions on $L \oplus H_U \oplus H_{U'} = Z$ (that respect side condition on both $H_U, H_{U'}$), and the same holds between $L' \oplus H_{U'}$ and $L' \oplus H_U \oplus H_{U'} = Z$. This gives a 1-to-1 correspondence between linear functions on $L \oplus H_U$ and $L' \oplus H_{U'}$ (respecting the relevant side conditions) which is regarded as the 1-to-1 constraint on the "coloring" of $L \oplus H_U$ and $L' \oplus H_{U'}$.

2**-to-**2 **Constraints:** As before, from Lemma 4.2, we always have

$$\mathsf{dim}(A = L \oplus H_U \oplus H_{U'}) = \mathsf{dim}(A' = L' \oplus H_U \oplus H_{U'}) = d \text{ (say)}.$$

Now suppose that $Z = L \oplus L' \oplus H_U \oplus H_{U'}$, $\mathsf{dim}(Z) = d + 1$. Since $Z = A \oplus A'$, it follows that $\mathsf{dim}(A \cap A') = d - 1$. Thus, it is possible to choose a basis I for $A \cap A'$ and $v \in L$, $v' \in L'$ so that $I \cup \{v\}$ is a basis for A, $I \cup \{v'\}$ is a basis for A', and $I \cup \{v, v'\}$ is a basis for Z. In the following, all linear functions considered are supposed to respect the side condition on H_U or $H_{U'}$ or both, depending on whether the relevant space contains H_U, $H_{U'}$ or both.

Every linear function f on $A \cap A' = \mathsf{Span}(I)$ has exactly two extensions f_1, f_2 to $A = \mathsf{Span}(I \cup \{v\})$, depending on their value on v, and has exactly two extensions f'_1, f'_2 to $A' = \mathsf{Span}(I \cup \{v'\})$, depending on their value on v'. Moreover by Lemma 4.4, linear functions on A are in one-to-one correspondence with those on $L \oplus H_U$. Denote by $\tilde{f}_1, \tilde{f}_2$ the mates of f_1, f_2 respectively via this correspondence. Similarly, linear functions on A' are in one-to-one correspondence with those on $L' \oplus H_{U'}$ and let $\tilde{f}'_1, \tilde{f}'_2$ be the mates of f'_1, f'_2. This gives a 2-to-2 constraint between $L \oplus H_U$ and $L' \oplus H_{U'}$ that matches the pair $(\tilde{f}_1, \tilde{f}_2)$ with the pair $(\tilde{f}'_1, \tilde{f}'_2)$.

Remark 4.1. *Another useful way to describe the constraint, both in the* 1*-to-*1 *and* 2*-to-*2 *case, is as follows: if there is a space Z that includes both $L \oplus H_U$ and $L' \oplus H_{U'}$ and has an assignment β that respects side conditions on $H_U, H_{U'}$, then $\beta|_{L \oplus H_U}, \beta|_{L' \oplus H_{U'}}$ are colorings to $L \oplus H_U$ and $L' \oplus H_{U'}$ respectively that satisfy the constraint. In both cases above Z happens to be $L \oplus L' \oplus H_U \oplus H_{U'}$, but we will have occasion to use an even larger space Z in certain proofs.*

Auxiliary Lemmas

Lemma 4.2. *Let $U, U' \in \mathcal{U}$ and $\mathsf{Eq}[U], \mathsf{Eq}[U']$ denote the sets of equations (k in number) in U, U' respectively. Then, for $L \in \mathcal{L}_U$,*

$$\mathsf{dim}(L \oplus H_U \oplus H_{U'}) = \ell + 2k - |\mathsf{Eq}[U] \cap \mathsf{Eq}[U']|.$$

Proof. Let $\mathsf{Eq}[U'] = \{e'_1, \ldots, e'_k\}$ and recall that $H_{U'} = \mathsf{Span}(v_{e'_1}, \ldots, v_{e'_k})$. Let C denote the "current space" that is initialized to $C = L \oplus H_U$ and has dimension $\ell + k$. We consider equations $e'_1, \ldots, e'_k \in \mathsf{Eq}[U']$ one by one, and check whether "adding" $v_{e'_i}$ to the current space increases its dimension. If the equation $e'_i \in \mathsf{Eq}[U]$, then $v_{e'_i} \in H_U$ already, and hence $\mathsf{dim}(C \oplus \mathsf{Span}(v_{e'_i})) = \mathsf{dim}(C)$. Otherwise $e'_i \notin \mathsf{Eq}[U]$ and shares at most one variable with $U \cup (U' \setminus e'_i)$. This is where we use the fact that U, U' are "legitimate" tuples in the sense described in the first paragraph of current section. Thus $v_{e'_i}$ is linearly independent of $L \oplus H_U \oplus_{j \neq i} \mathsf{Span}(v_{e'_j})$. Hence $C \oplus \mathsf{Span}(v_{e'_i})$ has dimension 1 larger than that of C. Carrying the argument for $i = 1, \ldots, k$ shows that in the end $C = L \oplus H_U \oplus H_{U'}$ and $\mathsf{dim}(C)$ is as desired. □

Lemma 4.3. *Let $U, U' \in \mathcal{U}$ and $L \in \mathcal{L}_U$, $L' \in \mathcal{L}_{U'}$. Then*

$$\mathsf{dim}(L \oplus H_U \oplus H'_U) = \mathsf{dim}(L' \oplus H_U \oplus H'_U).$$

Proof. From Lemma 4.2, both the dimensions are equal to $\ell + 2k - |\mathsf{Eq}[U] \cap \mathsf{Eq}[U']|$. □

Lemma 4.4. *Let $U, U' \in \mathcal{U}$ and $L \in \mathcal{L}_U$. Then any linear function on $L \oplus H_U$ that respects the side condition on H_U, has a unique extension to $L \oplus H_U \oplus H_{U'}$ that respects the side condition on both H_U and $H_{U'}$.*

Proof. Let f be a linear function on $L \oplus H_U$ that respects the side condition on H_U. Clearly, it has at most one extension to $L \oplus H_U \oplus H_{U'}$ that respects the side condition on both H_U and $H_{U'}$, so the main point is to show that there indeed is such an extension. Similar to the proof of Lemma 4.2, let C denote the current space, g denote the current linear function on C, so that initially $C = L \oplus H_U, g = f$ and at each step, g respects the side condition on H_U and the side condition due to equations $e'_1, \ldots, e'_{i-1}$ considered so far. Consider the equation e'_i. If $e'_i \in \mathsf{Eq}[U]$ then $C \oplus \mathsf{Span}(v_{e'_i}) = C$ and we keep g unchanged and proceed next. If $e'_i \notin \mathsf{Eq}[U]$, then as in the proof of Lemma 4.2, $v_{e'_i}$ is linearly independent of $L \oplus H_U \oplus_{j \neq i} \mathsf{Span}(v_{e'_j})$. Hence $C \oplus \mathsf{Span}(v_{e'_i})$ has dimension 1 larger than that of C and the function g can be safely extended to vector $v_{e'_i}$ as required. To be precise, one sets $g(v_{e'_i}) = b'_i$ where b'_i is the 'right hand side" of the equation e'_i and then extends g linearly to $C \oplus \mathsf{Span}(v_{e'_i})$. Carrying the argument for $i = 1, \ldots, k$, completes the proof. □

4.3 Completeness

It is easily observed that the reduction satisfies the completeness condition as in Theorem 1.8. Let $\sigma \colon X \to \{0,1\}$ be an assignment to the $\mathsf{Gap3Lin}$ instance (X, Eq) that satisfies $1-\varepsilon$ fraction of the equations. Let Eq' denote the set of the equations satisfied so that $|\mathsf{Eq}'| \geqslant (1-\varepsilon)|\mathsf{Eq}|$. Let $\mathcal{U}' \subseteq \mathcal{U}$ be the subset of k-tuples of equations U such that all its k equations are satisfied, i.e. $U \subseteq \mathsf{Eq}'$. Clearly, $|\mathcal{U}'| \geqslant (1-k\varepsilon)|\mathcal{U}| \geqslant (1-\delta)|\mathcal{U}|$ by choosing ε sufficiently small.

For every $U \in \mathcal{U}'$, let $\sigma[U]$ denote the linear function on $\{0,1\}^U$ that maps $x \in \{0,1\}^U$ to $\langle \sigma|_U, x \rangle$. Since σ satisfies all equations inside U, the linear function $\sigma[U]$ respects the side condition H_U. Now assign to every vertex $L \oplus H_U$ in $\mathsf{Block}[U]$, the linear function $\sigma[U]|_{L \oplus H_U}$. We show that this assignment satisfies all constraints whose both endpoints have been assigned. Indeed if $(L \oplus H_U,\ L' \oplus H_{U'})$ is a constraint such that both endpoints are assigned, then the constraint is satisfied since all spaces are assigned using the same global assignment σ. Thus the 2-to-2 Game has a $(1, 1-\delta)$-assignment.

4.4 Covering Property

We need a certain covering property towards the soundness analysis. While this property was introduced in [28], we need a more general notion. The covering property, the zoom-in required in Hypothesis 2.10, and the "advice" in the 2-Prover-1-Round game in Section 3.4 (the Outer PCP) are all used in a coordinated manner in the soundness analysis.

Let U be the set of $3k$ variables in a fixed set of k equations. We recall that in the Outer PCP game, the verifier chooses $V \subseteq U$ randomly by choosing from each equation independently (a) with probability β, one of the variables from the equation and (b) with probability $1-\beta$, all three variables from the equation. We consider $\{0,1\}^V$ as a subspace of $\{0,1\}^U$ in a natural manner. Slightly rephrasing a result from [28], the statistical distance between the following two distributions over one-dimensional subspaces of $\{0,1\}^U$ is small, i.e. at most $O(\beta\sqrt{k})$.[13]

[13] The intuition is as follows. A one-dimensional subspace is same as a non-zero point. A random point in $\{0,1\}^U$ (and in $\{0,1\}^V$) has negligible chance of being zero, so we might as well consider the distribution of (a) a random point in $\{0,1\}^U$ and

- Choose a random one-dimensional subspace $P \subseteq \{0,1\}^U$.
- Choose $V \subseteq U$ as described above, choose a random one-dimensional subspace $P' \subseteq \{0,1\}^V$ and regard it as a subspace of $\{0,1\}^U$.

We will need an analogous statement regarding two distributions over ℓ-dimensional subspaces of $\{0,1\}^U$. We define the two distributions below and prove the subsequent lemmas in Section C.

Definition 4.5. *Let U be a fixed set of k equations and $V \subseteq U$ be chosen as above with parameter β. Let $\ell \geqslant 1$ be an integer. Let $\mathcal{L}$, $\mathcal{L}'$ be distributions over ℓ-dimensional subspaces of $\{0,1\}^U$ sampled as follows.*

- *$\mathcal{L}$: Choose a uniformly random ℓ-dimensional subspace of $\{0,1\}^U$.*
- *$\mathcal{L}'$: Choose $V \subseteq U$ as above, choose a uniformly random ℓ-dimensional subspace of $\{0,1\}^V$ and regard it as a subspace of $\{0,1\}^U$.*

Lemma 4.6. *Suppose $2^\ell \beta \leqslant \frac{1}{8}$. Let $\mathcal{L}$, $\mathcal{L}'$ be distributions over ℓ-dimensional subspaces over $\{0,1\}^U$ sampled as in Definition 4.5. Then the statistical distance between $\mathcal{L}$, $\mathcal{L}'$ is bounded as*

$$\mathsf{SD}(\mathcal{L},\ \mathcal{L}') \leqslant \beta\sqrt{k} \cdot 2^{\ell+4}.$$

Lemma 4.7. *Let $0 \leqslant q \leqslant \ell - 1$ be an integer. Let Q be q-dimensional subspace of $\{0,1\}^U$. Let $\mathcal{L}_Q$ and $\mathcal{L}'_Q$ be distributions $\mathcal{L}$ and $\mathcal{L}'$ conditioned on the event that a sampled ℓ-subspace L contains Q. Suppose $2^\ell \beta \leqslant \frac{1}{8}$. Then for at least $1 - \sqrt{\beta}\, k^{\frac{1}{4}}$ fraction of Q,*

$$\mathsf{SD}(\mathcal{L}_Q,\ \mathcal{L}'_Q) \leqslant \sqrt{\beta}\, k^{\frac{1}{4}} \cdot 2^{\ell+5}. \qquad (5)$$

5 Soundness Analysis

In this section, given a (j,δ)-assignment to the game $G_{2:2}$ constructed in Section 4, we show how to extract a provers' strategy in the Outer PCP game ($G^{\otimes k}_{\beta,q}$ as in Section 3.4) that succeeds with probability $p = p(j,\delta,\ell)$. If the soundness of the Outer PCP game is chosen to be smaller than p to begin with, it implies that the game $G_{2:2}$ has no (j,δ)-assignment, proving Theorem 1.8.

We recall that the first prover (the "larger" prover) receives as a question, a set U of $3k$ variables (in k equations) and an advice-list $[x_1,\ldots,x_q]$ of q points in $\{0,1\}^U$ (in fact in $\{0,1\}^V$ as stated next). The second prover (the "smaller" prover) receives as a question a subset $V \subseteq U$ of variables and an advice-list $[x_1,\ldots,x_q]$ of the same q points in $\{0,1\}^V$. We will extract the provers' strategies in the next two subsections, and show in the last subsection that these strategies succeed with a good probability.

5.1 Strategy for the First (Larger) Prover

We recall that a typical vertex in the game $G_{2:2}$ is denoted as $L \oplus H_U$. Specifically, for $U \in \mathcal{U}$, the block of vertices corresponding to U is[14]

$$\mathsf{Block}[U] = \{L \oplus H_U \mid L \subseteq \{0,1\}^U,\ \mathsf{dim}(L) = \ell,\ L \cap H_U = \{0\}\},$$

(b) a random point in $\{0,1\}^V$ after choosing V and then "lifting it up" by appending 0 in the coordinates $U \setminus V$. We note that $|U \setminus V| \approx 2\beta k$. A point chosen from the second distribution has $\approx 2\beta k$ more zeroes than that from the first distribution. However the imbalance between the number of zeroes and ones in a typical point in $\{0,1\}^U$ is $\approx \sqrt{k}$, so when $\beta k \ll \sqrt{k}$, a deviation of $2\beta k$ zeroes is nearly imperceptible.

[14] We emphasize that $\mathsf{Block}[U]$ contains the vertex $L \oplus H_U$ for essentially all ℓ-dimensional subspaces $L \subseteq \{0,1\}^U$.

and the set of vertices of $G_{2:2}$ is the union of all blocks of vertices over $U \in \mathcal{U}$. Let $F[\cdot]$ be the given (j, δ)-assignment to the game $G_{2:2}$. Let us emphasize that this means:

- For at least δ fraction of the vertices $L \oplus H_U$, a list $F[L \oplus H_U]$ of j linear functions on $L \oplus H_U$ (that respect the side condition on H_U) is given. The remaining vertices are unassigned and do not play any role in the analysis.

- If there is a 2-to-2 constraint between vertices $L \oplus H_U$ and $L' \oplus H_{U'}$, both of which are assigned, then there are linear functions $f \in F[L \oplus H_U],\ f' \in F[L' \oplus H_{U'}]$ that satisfy the constraint.

- If there is a 1-to-1 constraint between vertices $L \oplus H_U$ and $L' \oplus H_{U'}$, both of which are assigned, there is a one-to-one correspondence between the lists $F[L \oplus H_U],\ F[L' \oplus H_{U'}]$, via the same one-to-one correspondence that defines the 1-to-1 constraint.

By an averaging argument, for at least $\frac{\delta}{2}$ fraction of the tuples U, at least $\frac{\delta}{2}$ fraction of vertices in $\mathsf{Block}[U]$ are assigned. Call such a tuple U good and let $\mathcal{U}_{\mathsf{good}}$ be the set of good tuples with $|\mathcal{U}|_{\mathsf{good}} \geqslant \frac{\delta}{2} \cdot |\mathcal{U}|$.

Let the question to the first prover be $U \in \mathcal{U}$ along with the advice-list $[x_1, \ldots, x_q]$ of points in $\{0,1\}^U$. If $U \notin \mathcal{U}_{\mathsf{good}}$, the prover gives up, so let us assume $U \in \mathcal{U}_{\mathsf{good}}$, and let $\mathsf{Assigned}[U] \subseteq \mathsf{Block}[U]$ denote the set of vertices in its block that have been assigned, $|\mathsf{Assigned}[U]| \geqslant \frac{\delta}{2} \cdot |\mathsf{Block}[U]|$. Since this is a $(j, \frac{\delta}{2})$-assignment respecting the side condition on H_U, Hypothesis 2.10 states that for some $q, \alpha(\cdot), C$ that depend on $(j, \frac{\delta}{2})$, for at least $\alpha(\ell)$ fraction of the q-dimensional subspaces $Q \subseteq \{0,1\}^U$, there exists a global linear function $g_Q : \{0,1\}^U \to \{0,1\}$ that respects the side condition on H_U and

$$\Pr_{L \subseteq \{0,1\}^U,\ \mathsf{dim}(L)=\ell} [g_Q|_{L \oplus H_U} \in F[L \oplus H_U] \mid Q \subseteq L] \geqslant C. \tag{6}$$

We call such a choice of Q "lucky" and let $\mathcal{Q}_{\mathsf{lucky}}$ be the set of all lucky q-dimensional subspaces of $\{0,1\}^U$. We note that the parameter q was chosen beforehand to exactly match with that arising in Hypothesis 2.10. Moreover, call a q-dimensional subspace Q "smooth" if it satisfies Condition (5) in Lemma 4.7[15], and let $\mathcal{Q}_{\mathsf{smooth}}$ be the set of all smooth q-dimensional subspaces of $\{0,1\}^U$.

The prover looks at the advice-list and lets $Q = \mathsf{Span}(x_1, \ldots, x_q)$. If $Q \notin \mathcal{Q}_{\mathsf{lucky}}$ or $Q \notin \mathcal{Q}_{\mathsf{smooth}}$, the prover gives up. Otherwise the prover picks a global linear function $g_Q : \{0,1\}^U \to \{0,1\}$ respecting the side condition and satisfying Equation (6) (if there is more than one, one of them is picked arbitrarily), and outputs g_Q as the answer. Strictly speaking, the linear functions g_Q amounts to a function $x \to \langle \sigma_Q, x \rangle$ on $\{0,1\}^U$ for some $\sigma_Q \in \{0,1\}^U$ and the prover answers that σ_Q is the assignment to the $3k$ received variables (σ_Q satisfies the k equations as g_Q respects the side condition). However, it is more convenient to view the function g_Q itself as the answer.

We note that a uniformly random q-subspace of $\{0,1\}^U$ is lucky with probability $\geqslant \alpha(\ell)$ (by Hypothesis 2.10) and is smooth with probability $\geqslant 1 - \beta k^{\frac{1}{4}} \geqslant 1 - \frac{\alpha(\ell)}{2}$ when the parameters β, k are chosen appropriately. Thus with probability at least $\frac{\alpha(\ell)}{2}$, the space Q dictated by the advice-list is both lucky and smooth.

[15] To recall, the condition is that the distributions $\mathcal{L}_Q$ and $\mathcal{L}'_Q$ are close in statistical distance; the former distribution chooses a random ℓ-subspace of $\{0,1\}^U$ containing Q and the latter distribution chooses a random question $V \subseteq U$ to the second prover and then a random ℓ-subspace of $\{0,1\}^V$ containing Q.

5.2 A Strategy for the Second (Smaller) Prover

Let the question to the second prover be V along with the advice-list $[x_1, \ldots, x_q]$ of points in $\{0,1\}^V$. Let $Q = \mathsf{Span}(x_1, \ldots, x_q)$ be the q-dimensional subspace of $\{0,1\}^V$. Let $\mathcal{L}_V$ be the set of all ℓ-dimensional subspaces of $\{0,1\}^V$ (though when the prover decides on an answer, only the subspaces containing Q are relevant):

$$\mathcal{L}_V = \left\{ L \mid L \subseteq \{0,1\}^V,\ \mathsf{dim}(L) = \ell \right\}.$$

The prover first obtains an assignment $F_V[\cdot]$ to $\mathcal{L}_V$. Fix $L \in \mathcal{L}_V$. The prover examines every k-tuple of equations U such that $V \subseteq U$, i.e. every question that could have been asked to the first prover, when the question of the second prover is V. Note that $L \subseteq \{0,1\}^V \subseteq \{0,1\}^U$ and hence there is a vertex $L \oplus H_U$ of the game $G_{2:2}$ in $\mathsf{Block}[U]$. If the vertex $L \oplus H_U$ has been assigned, then the prover defines

$$F_V[L] = \{f|_L \mid f \in F[L \oplus H_U]\},$$

i.e. restrictions of all functions in $F[L \oplus H_U]$ to L. In general there are several U that contain V, so a priori, there is ambiguity in the definition of $F_V[\cdot]$. The claim below shows however that the definition is unambiguous.

Claim 5.1. *$F_V[\cdot]$ is well defined. That is, if $V \subseteq U$, $V \subseteq U'$ and if $L \oplus H_U$, $L \oplus H_{U'}$ are both assigned, then the restrictions of $F[L \oplus H_U]$ and $F[L \oplus H_{U'}]$ to L are identical.*

Proof. Notice that $L \oplus H_U, L \oplus H_{U'}$ have a 1-to-1 constraint between them in the game $G_{2:2}$. By Definition 1.5, a (j, δ)-assignment must assign identical sets of j assignments to the vertices that have a 1-to-1 constraint between them. □

Once F_V is defined, the prover zooms-into Q, and chooses at random any linear function $h_Q \colon \{0,1\}^V \to \{0,1\}$ that satisfies (if one exists and if so, we show that the list-size is bounded)

$$\Pr_{L \in \mathcal{L}_V} [h_Q|_L \in F_V[L]] \mid Q \subseteq L] \geqslant \frac{C}{4}. \tag{7}$$

The prover outputs $h_Q : \{0,1\}^V \to \{0,1\}$ as the answer. Again, strictly speaking, the linear function h_Q amounts to a function $y \to \langle \tau_Q, y \rangle$ on $\{0,1\}^V$ for some $\tau_Q \in \{0,1\}^V$ and the prover answers that τ_Q is the assignment to the variables he received. However, it is more convenient to view the function h_Q itself as the answer.

5.3 The Success Probability of the Provers

We now show that the provers' strategy succeeds with a good probability. Let $U, V, [x_1, \ldots, x_q], x_i \in \{0,1\}^V$ be the provers' questions and $Q = \mathsf{Span}(x_1, \ldots, x_q)$. We already observed that with probability at least $\frac{\delta}{2}$, $U \in \mathcal{U}_{\mathsf{good}}$ and with probability $\frac{\alpha(\ell)}{2}$, Q is both lucky and smooth (from the first prover's perspective). Assume that all these properties hold. Then the answer of the first prover is a global function $g_Q : \{0,1\}^U \to \{0,1\}$ that satisfies the side condition on H_U, and

$$\Pr_{L \subseteq \{0,1\}^U,\ \mathsf{dim}(L) = \ell} [g_Q|_{L \oplus H_U} \in F[L \oplus H_U] \mid Q \subseteq L] \geqslant C.$$

Since Q is smooth, by Lemma 4.7, the uniform distribution on ℓ-spaces in $\{0,1\}^U$ containing Q is $(\sqrt{\beta} k^{\frac{1}{4}} \cdot 2^{\ell+5})$-close in statistical distance to the distribution that chooses a question $V \subseteq U$ to the second prover and then chooses uniformly an ℓ-space in $\{0,1\}^U$ containing Q. By setting the parameters β, k appropriately, we can assume that this statistical distance is at most $\frac{C}{2}$ and conclude from the above inequality:

$$\Pr_{V,\, L \in \mathcal{L}_V} [g_Q|_{L \oplus H_U} \in F[L \oplus H_U] \mid Q \subseteq L] \geqslant \frac{C}{2}.$$

By an averaging argument, with probability at least $\frac{C}{4}$ over the choice of question V to the second prover,

$$\Pr_{L \in \mathcal{L}_V} [g_Q|_{L \oplus H_U} \in F[L \oplus H_U] \mid Q \subseteq L] \geqslant \frac{C}{4}.$$

Fix any such good choice of question V. Note that the assignment $F_V[L]$ to the ℓ-spaces L of the second prover is precisely the restriction of the assignment $F[L \oplus H_U]$ to the ℓ-spaces of the first prover. Letting $h_Q^* : \{0,1\}^V \to \{0,1\}$ to be the restriction of $g_Q : \{0,1\}^U \to \{0,1\}$ to $\{0,1\}^V$, we can rewrite the inequality above as:

$$\Pr_{L \in \mathcal{L}_V} \left[h_Q^*|_L \in F_V[L] \mid Q \subseteq L\right] \geqslant \frac{C}{4}.$$

Thus the function h_Q^* satisfies Condition (7) and is a legitimate candidate for the second prover's answer. By Theorem 2.6, the number of functions h_Q satisfying Condition (7) is at most $\frac{jC/4}{(C/4)^2 - j \cdot 2^{q-\ell}} \leqslant \frac{8j}{C}$ for a large enough choice of ℓ. When the second prover does pick h_Q^* as the answer, both the provers' answers are consistent (h_Q^* being a restriction of g_Q) and the provers succeed. Their overall success probability is at least

$$\frac{\delta}{2} \cdot \frac{\alpha(\ell)}{2} \cdot \frac{C}{4} \cdot \frac{C}{8j} = \frac{\delta\, \alpha(\ell)\, C^2}{128\, j}.$$

6 The Case $j = 1$ of Hypothesis 2.5

In this section it is shown that Hypothesis 2.5, in the case $j = 1$, follows from Hypothesis 2.7 without the need of zoom-ins.

Let $f : S \to [2^\ell]$, $\mathsf{Density}(S) = \delta$ be the given $(1, \delta)$-assignment to the Grassmann graph $G(V = \{0,1\}^n, \ell)$, meaning, for any two ℓ-spaces $L_1, L_2 \in S$ such that $\mathsf{dim}(L_1 \cap L_2) = \ell - 1$, we have the consistency $f[L_1]_{L_1 \cap L_2} = f[L_2]_{L_1 \cap L_2}$. We intend to show, using Hypothesis 2.7, that there is a global linear function $g : \{0,1\}^n \to \{0,1\}$ such that $g|_L = f[L]$ for a fraction $C = C(\delta)$ of the ℓ-spaces L.

Here is the idea. Fix an integer $b = \frac{\ell}{10}$. Using Hypothesis 2.7, we conclude rather easily, that for a constant fraction of pairs $L_1, L_2 \in S$ such that $\mathsf{dim}(L_1 \cap L_2) = b$, we still have $f[L_1]_{L_1 \cap L_2} = f[L_2]_{L_1 \cap L_2}$. This enables us to assign linear functions to b-dimensional spaces that have a good agreement with the given assignment to ℓ-dimensional spaces. In other words, this assignment passes the "ℓ-space vs b-space linearity test"[16] with good probability. Using a Fourier analytic approach,[17] we are able to show a soundness guarantee for the "ℓ-space vs b-space linearity test", implying the existence of a desired global linear function. A formal proof appears below. The analysis of the linearity test is presented in Section D which might be of independent interest.

[16] Analogous to the "line vs point low degree test".

[17] As opposed to the rather involved algebraic (and/or combinatorial) analysis of the "line vs point" and "plane vs plane" low degree test in [3, 32].

Let $b = \frac{\ell}{10}$. For a b-dimensional space $B \subseteq V$, let Zoom_B be the (lower order) Grassmann graph induced on the set of vertices $\{L | L \in G(V, \ell), B \subseteq L\}$ (see Fact 2.2). Let $\mathsf{Zoom}_B[S] = S \cap \mathsf{Zoom}_B$ be the set of vertices in S that contain B, and let $\mathsf{Density}(\mathsf{Zoom}_B[S])$ be its density inside Zoom_B. Clearly,

$$\mathop{\mathbb{E}}_{B \subseteq V,\ \mathsf{dim}(B)=b} [\mathsf{Density}(\mathsf{Zoom}_B[S])] = \mathsf{Density}(S) = \delta.$$

By an averaging argument, $\mathsf{Density}(\mathsf{Zoom}_B[S]) \geqslant \frac{\delta}{2}$ for at least $\frac{\delta}{2}$ fraction of b-spaces B; denote by $\mathcal{B}$ the set of all such "good" b-spaces. Fix any $B \in \mathcal{B}$. Note that there are 2^b different $\mathbb{F}_2$-valued linear functions on B. Partition $\mathsf{Zoom}_B[S]$ into classes $\mathcal{C}_1, \ldots, \mathcal{C}_{2^b}$ according to the restriction of $f[L]|_B$ for $L \in \mathsf{Zoom}_B[S]$. We observe that for any edge (L, L') of the Grassmann graph inside $\mathsf{Zoom}_B[S]$, the linear functions $f[L], f[L']$ agree on $L_1 \cap L_2 \supseteq B$ and hence the edge is inside one of the partitions $\mathcal{C}_i$. Since $\mathsf{Density}(\mathsf{Zoom}_B[S]) \geqslant \frac{\delta}{2}$, Hypothesis 2.7 implies that there is a connected component $\mathcal{C}$ of density $\geqslant \varepsilon$ in $\mathsf{Zoom}_B[S]$ and as observed, $\mathcal{C} \subseteq \mathcal{C}_{i_0}$ for some $1 \leqslant i_0 \leqslant 2^b$. Let $h[B]$ denote the linear function on B that equals the common function $f[L]|_B$ over $L \in \mathcal{C}_{i_0}$. This gives an assignment $h : \mathcal{B} \to [2^b]$ of linear functions to b-spaces. From the discussion, if a pair $(B, L), B \subseteq L$ of a b-space and a ℓ-space is chosen at random from $V = \{0, 1\}^n$, then

$$\Pr_{B \subseteq L \subseteq V} [f[L]|_B = h[B]] \geqslant \frac{\delta}{2} \cdot \varepsilon,$$

where $B \in \mathcal{B}$ with probability at least $\frac{\delta}{2}$ and then L is in the "large" connected component of $\mathsf{Zoom}_B[S]$ with probability at least ε. That is, the "ℓ-space vs b-space test" succeeds with probability $\geqslant \frac{\delta}{2} \cdot \varepsilon$. Theorem D.1 now implies that there is a global linear function $g : V \to \{0, 1\}$ that agrees with at least C fraction of the L-spaces in $G(V, \ell)$.

7 Acknowledgements

We thank Prahladh Harsha, Johan Håstad, and Sushant Sachdeva for a counter-example to an earlier (rather different) construction. We thank Dana Moshkovitz and Ran Raz for related discussions over the years. A big thanks to Irit Dinur and Guy Kindler for discussions that clarified several of authors' ideas and led to a better presentation. The authors acknowledge the hospitality of the French Roast cafe (78 West 11th Street, New York) while working on the paper.

References

[1] S. Arora, L. Babai, J. Stern, and Z. Sweedyk. The hardness of approximate optima in lattices, codes, and systems of linear equations. *J. Comput. Syst. Sci.*, 54(2):317–331, 1997.

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

[3] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998.

[4] M. Bellare, O. Goldreich, and M. Sudan. Free bits, PCPs, and nonapproximability – towards tight results. *SIAM Journal on Computing*, 27(3):804–915, 1998.

[5] V. M. Blinovsky. Bounds for codes in the case of list decoding of finite volume. *Problems of Information Transmission*, 22(1):719, 1986.

[6] A. Brouwer, A. Cohen, and A. Neumaier. *Distance-Regular Graphs*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics. Springer Berlin Heidelberg, 2012.

[7] S. O. Chan. Approximation resistance from pairwise independent subgroups. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 447–456, 2013.

[8] A. Chowdhury and B. Patks. Shadows and intersections in vector spaces. *Journal of Combinatorial Theory, Series A*, 117(8):1095 – 1106, 2010.

[9] I. Dinur. The PCP theorem by gap amplification. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA, May 21-23, 2006*, pages 241–250, 2006.

[10] I. Dinur, E. Mossel, and O. Regev. Conditional hardness for approximate coloring. *SIAM J. Comput.*, 39(3):843–873, 2009.

[11] I. Dinur and S. Safra. On the Hardness of Approximating Minimum Vertex Cover. *Annals of Mathematics*, 162(1):439–485, 2005.

[12] I. Dinur and D. Steurer. Analytical approach to parallel repetition. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 624–633, 2014.

[13] U. Feige, S. Goldwasser, L. Lovász, S. Safra, and M. Szegedy. Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, Mar. 1996.

[14] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[15] D. Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theor. Comput. Sci.*, 259(1-2):613–622, 2001.

[16] V. Guruswami, J. Håstad, and M. Sudan. Hardness of approximate hypergraph coloring. *SIAM J. Comput.*, 31(6):1663–1686, 2002.

[17] E. Halperin. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. *SIAM J. Comput.*, 31(5):1608–1623, 2002.

[18] J. Håstad. Clique is hard to approximate within $n^{1-\varepsilon}$. In *37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October, 1996*, pages 627–636, 1996.

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

[20] T. Holenstein. Parallel repetition: simplifications and the no-signaling case. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 411–419, 2007.

[21] G. Karakostas. A better approximation ratio for the vertex cover problem. *ACM Trans. Algorithms*, 5(4), 2009.

[22] S. Khot. Improved inaproximability results for maxclique, chromatic number and approximate graph coloring. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 600–609, 2001.

[23] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 17th Annual IEEE Conference on Computational Complexity, Montréal, Québec, Canada, May 21-24, 2002*, page 25, 2002.

[24] S. Khot. Ruling out PTAS for graph min-bisection, dense k-subgraph, and bipartite clique. *SIAM J. Comput.*, 36(4):1025–1071, 2006.

[25] S. Khot. Inapproximability of NP-complete problems, discrete fourier analysis, and geometry. In *Proceedings of the International Congress of Mathematicians 2010*, pages 2676–2697, 2010.

[26] S. Khot. On the unique games conjecture. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity, CCC 2010, Cambridge, Massachusetts, June 9-12, 2010*, pages 99–121, 2010.

[27] S. Khot and O. Regev. Vertex cover might be hard to approximate to within $2 - \varepsilon$. *J. Comput. Syst. Sci.*, 74(3):335–349, May 2008.

[28] S. Khot and M. Safra. A two-prover one-round game with strong soundness. *Theory of Computing*, 9(28):863–887, 2013.

[29] G. Margulis. Probabilistic characteristics of graphs with large connectivity. *Prob. Peredachi Inform*, 10:101–108, 1974.

[30] A. Rao. Parallel repetition in projection games and a concentration bound. *SIAM Journal on Computing*, 40(6):1871–1891, 2011.

[31] R. Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, June 1998.

[32] R. Raz and S. Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing, El Paso, Texas, USA, May 4-6, 1997*, pages 475–484, 1997.

[33] R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, Feb. 1996.

[34] L. Russo. An approximate zero-one law. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 61(1):129–139, 1982.

[35] G. Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 593–602, 2008.

[36] L. Trevisan. On Khot's unique games conjecture. *Bulletin of the AMS*, 49(1):91–111, 2012.

[37] M. Tulsiani. CSP gaps and reductions in the lasserre hierarchy. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 303–312, 2009.

Appendix

A Transitivity of $G_{2:2}$

In this section, we show that the game $G_{2:2}$ constructed in Section 4 is transitive as per Definition 1.4.

Lemma A.1. *Suppose $L_1 \oplus H_{U_1}, L_2 \oplus H_{U_2}$ have a 1-to-1 constraint between them in $G_{2:2}$, and $L_2 \oplus H_{U_2}, L_3 \oplus H_{U_3}$ have a 1-to-1 or a 2-to-2 constraint between them. Then there is a constraint between $L_1 \oplus H_{U_1}, L_3 \oplus H_{U_3}$, and it is 1-to-1 or a 2-to-2 depending on whether the constraint between $L_2 \oplus H_{U_2},\ L_3 \oplus H_{U_3}$ is 1-to-1 or 2-to-2 respectively.*

Proof. Since there is 1-to-1 constraint between $L_1 \oplus H_{U_1}, L_2 \oplus H_{U_2}$, we have

$$L_1 \oplus H_{U_1} \oplus H_{U_2} = L_2 \oplus H_{U_1} \oplus H_{U_2}. \tag{8}$$

We first consider the case when the constraint between $L_2 \oplus H_{U_2},\ L_3 \oplus H_{U_3}$ is also 1-to-1. This gives

$$L_2 \oplus H_{U_2} \oplus H_{U_3} = L_3 \oplus H_{U_2} \oplus H_{U_3}. \tag{9}$$

Combining the above equations gives ("add" H_{U_1} to Equation (9) and do a "substitution" using Equation (8))

$$L_1 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3} = L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}. \tag{10}$$

Now we would like to "remove" H_{U_2} from both the sides so as to obtain $L_1 \oplus H_{U_1} \oplus H_{U_3} = L_3 \oplus H_{U_1} \oplus H_{U_3}$ and implying that there is a 1-to-1 constraint between $L_1 \oplus H_{U_1}$ and $L_3 \oplus H_{U_3}$. This "removal" can be done for the following reason. Write $H_{U_2} = A \oplus B,\ A \cap B = \{0\}$ where (a) A is the span of all vectors v_e such that the equation e occurs in U_2, but also occurs in U_1 or U_3, and hence $A \subseteq H_{U_1} \oplus H_{U_3}$. (b) B is the span of all vectors v_e such that the equation e occurs in U_2, but not in U_1 nor in U_3. Any such equation e shares at most one variable with U_1 and at most one variable with U_3 and no variable with $U_2 \setminus e$. Hence there is a variable that is "private" to e, meaning it does not occur in $U_1 \cup U_3 \cup (U_2 \setminus e)$. In particular, the "private" variables of the equations contributing to B are distinct. Thus the intersection of B and $L_1 \oplus L_3 \oplus H_{U_1} \oplus H_{U_3} \subseteq \{0,1\}^{U_1} \oplus \{0,1\}^{U_3}$ is $\{0\}$. To summarize, we can write Equation (10) as

$$L_1 \oplus (H_{U_1} \oplus H_{U_3} \oplus A) \oplus B = L_3 \oplus (H_{U_1} \oplus H_{U_3} \oplus A) \oplus B,$$

which simplifies to

$$(L_1 \oplus H_{U_1} \oplus H_{U_3}) \oplus B = (L_3 \oplus H_{U_1} \oplus H_{U_3}) \oplus B,$$

and we can now safely "remove" B, using Fact E.5.

We now consider the case when the constraint between $L_2 \oplus H_{U_2},\ L_3 \oplus H_{U_3}$ is 2-to-2. We have Equation (8) as before, but instead of Equation (9), we now have

$$\mathsf{dim}(L_2 \oplus H_{U_2} \oplus H_{U_3}) = \mathsf{dim}(L_3 \oplus H_{U_2} \oplus H_{U_3}) = \mathsf{dim}(L_2 \oplus L_3 \oplus H_{U_2} \oplus H_{U_3}) - 1. \tag{11}$$

We claim that one can "add" H_{U_1} to all three "sums" in Equation (11). Arguing as earlier, one can write $H_{U_1} = A' \oplus B',\ A' \cap B' = \{0\}$ where $A' \subseteq H_{U_2} \oplus H_{U_3}$ and B' is linearly independent of $\{0,1\}^{U_2} \oplus \{0,1\}^{U_3}$. Thus "adding" H_{U_1} to all three "sums", increases the dimension of each "sum" by precisely $\mathsf{dim}(B')$. Thus

$$\mathsf{dim}(L_2 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) = \mathsf{dim}(L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) = \mathsf{dim}(L_2 \oplus L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) - 1.$$

Using Equation (8) and "substituting", we get

$\mathsf{dim}(L_1 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) = \mathsf{dim}(L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) = \mathsf{dim}(L_1 \oplus L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3}) - 1.$

Now, arguing as earlier again, we "remove" H_{U_2} from all three "sums". One can write $H_{U_2} = A \oplus B,\ A \cap B = \{0\}$ where $A \subseteq H_{U_1} \oplus H_{U_3}$ and B intersects $L_1 \oplus L_3 \oplus H_{U_1} \oplus H_{U_3}$ only at $\{0\}$. Thus "removing" H_{U_2} from all three "sums", decreases the dimension of each "sum" by precisely $\mathsf{dim}(B)$. Thus

$$\mathsf{dim}(L_1 \oplus H_{U_1} \oplus H_{U_3}) = \mathsf{dim}(L_3 \oplus H_{U_1} \oplus H_{U_3}) = \mathsf{dim}(L_1 \oplus L_3 \oplus H_{U_1} \oplus H_{U_3}) - 1,$$

implying that there is a 2-to-2 constraint between $L_1 \oplus H_{U_1}$ and $L_3 \oplus H_{U_3}$. □

Lemma A.2. *Let $s_1 = L_1 \oplus H_{U_1}, s_2 = L_2 \oplus H_{U_2}, s_3 = L_3 \oplus H_{U_3}$ be vertices in $G_{2:2}$ such that there is 1-to-1 constraint between (s_1, s_2) and a constraint between (s_2, s_3). Then the constraint between (s_1, s_3) (as guaranteed by Lemma A.1) is a composition of the constraints between (s_1, s_2) and (s_2, s_3).*

Specifically, if linear functions (respecting relevant side conditions) f on $L_1 \oplus H_{U_1}$, g on $L_2 \oplus H_{U_2}$, and h on $L_3 \oplus H_{U_3}$ are such that (f, g) satisfy (s_1, s_2) and (g, h) satisfy (s_2, s_3), then (f, h) satisfy (s_1, s_3).

Proof. In the following, whenever we construct a linear function on a certain space, it will always respect the side condition contained in that space. Since (g, h) satisfy the constraint (s_2, s_3), there is a linear function β on $W = L_2 \oplus L_3 \oplus H_{U_2} \oplus H_{U_3}$ that respects side conditions on H_{U_2} and H_{U_3} and

$$g = \beta|_{L_2 \oplus H_{U_2}}, \qquad h = \beta|_{L_3 \oplus H_{U_3}}.$$

Let $Z = W \oplus H_{U_1}$ and extend the linear function β on W uniquely to a linear function γ on Z so as to respect the side condition H_{U_1}. This is possible because every equation in U_1 that does not appear in U_2 or U_3 has a "private variable", as in the proof of the previous lemma. We note that

$$\gamma|_{L_2 \oplus H_{U_2}} = (\gamma|_W)|_{L_2 \oplus H_{U_2}} = \beta|_{L_2 \oplus H_{U_2}} = g,$$

$$\gamma|_{L_3 \oplus H_{U_3}} = (\gamma|_W)|_{L_3 \oplus H_{U_3}} = \beta|_{L_3 \oplus H_{U_3}} = h.$$

Since there is a 1-to-1 constraint (s_1, s_2), we have

$$L_1 \oplus H_{U_1} \oplus H_{U_2} = L_2 \oplus H_{U_1} \oplus H_{U_2}.$$

It therefore holds that

$$Z = L_2 \oplus L_3 \oplus H_{U_1} \oplus H_{U_2} \oplus H_{U_3} \supseteq L_2 \oplus H_{U_1} \oplus H_{U_2} = L_1 \oplus H_{U_1} \oplus H_{U_2} \supseteq L_1 \oplus H_{U_1}.$$

Since γ is an assignment on Z, by Remark 4.1, $\gamma|_{L_1 \oplus H_{U_1}}$ and $\gamma|_{L_2 \oplus H_{U_2}} = g$ satisfy the constraint (s_1, s_2). However (f, g) is supposed to satisfy this 1-to-1 constraint, and hence we must have $f = \gamma|_{L_1 \oplus H_{U_1}}$. Now we have $f = \gamma|_{L_1 \oplus H_{U_1}}$ and $h = \gamma|_{L_3 \oplus H_{U_3}}$ and by Remark 4.1, (f, h) satisfy the constraint (s_1, s_3). □

B Reduction from 2-to-2 Games to the Independent Set Problem

In this section, we present a reduction from the (Transitive) 2-to-2 Games problem to the Independent Set problem, proving Theorem 1.7. The reduction is along the lines of [11, 23, 27], using the Biased Long Code and analytic theorems of Russo, Margulis and Friedgut, introduced in [11]. Some care is required to handle the transitivity feature.

B.1 Biased Long Code

While the Biased Long Code can be viewed as an encoding scheme, it is more convenient to take a combinatorial view and treat it as a weighted Kneser graph. The valid codewords then amount to certain large (in fact the largest) independent sets in this graph. The analysis of the Biased Long Code amounts to a structural theorem about independent sets of moderately large (= linear) size.

Definition B.1. *For a bias parameter $p \in (0,1)$ and alphabet Σ, the vertex set of weighted Kneser graph $G_p[\Sigma]$ is $\mathcal{P}(\Sigma)$, the family of all subsets of Σ. The weight of a vertex $A \subseteq \Sigma$ is $\mu_p(A) = p^{|A|}(1-p)^{|\Sigma|-|A|}$. The edge set is $\{\,(A,B) \mid A, B \subseteq \Sigma,\ A \cap B = \emptyset\}$.*

It can be shown easily that the largest independent sets in $G_p[\Sigma]$ have weight p. These are precisely the sets $I_{\sigma_0} = \{F \mid \sigma_0 \in F\}$ for any fixed $\sigma_0 \in \Sigma$.

Definition B.2. *For a set family $\mathcal{F} \subseteq \mathcal{P}(\Sigma)$, let $\mu_p(\mathcal{F})$ denote its weight under μ_p. Let $A \sim \mu_p$ denote the process of picking a set $A \subseteq \Sigma$ according to the distribution μ_p. For a fixed element $\sigma \in \Sigma$, let $\mathsf{Infl}_\sigma(\mathcal{F})$ denote its influence on the family $\mathcal{F}$ defined as*

$$\mathsf{Infl}_\sigma(\mathcal{F}) = \Pr_{A \sim \mu_p} [\,\textit{Exactly one of the pair } A \textit{ and } A\Delta\{\sigma\} \textit{ is in } \mathcal{F}].$$

The average sensitivity of a family $\mathsf{as}_p(\mathcal{F})$ is the sum of all influences, i.e.

$$\mathsf{as}_p(\mathcal{F}) = \sum_{\sigma \in \Sigma} \mathsf{Infl}_\sigma(\mathcal{F}).$$

B.2 The Reduction

Let $G = (V, E, \Phi, \Sigma)$ be the instance of a Transitive 2-to-2 Game as in Conjecture 1.6. The parameters j and δ therein will be chosen later. The Independent Set instance $G' = (V', E')$ is defined as follows. Set the parameter $p = 1 - \frac{1}{\sqrt{2}} - \delta$. The vertex set of the instance is

$$V' = \{\,(x, A) \mid x \in V,\ A \subseteq \Sigma\}\,.$$

The weight of the vertex (x, A) is $\frac{1}{|V|} \cdot \mu_p(A)$, so that the total weight of all the vertices is 1. The edge set is

$$E' = \{\,((x_1, A_1), (x_2, A_2)) \mid (x_1, x_2) \in E \ \wedge\ \forall\, \sigma_1 \in A_1,\ \sigma_2 \in A_2, (\sigma_1, \sigma_2) \notin \Phi(x_1, x_2)\}\,.$$

In words, there is a cloud of vertices for every $x \in V$. For every constraint $(x_1, x_2) \in E$, there are cross edges between the respective clouds. There is an edge between $(x_1, A_1), (x_2, A_2)$ if there is no pair of colors in the sets A_1, A_2 that satisfy the constraint on (x_1, x_2). [18]

B.3 Completeness

Let $\mathcal{C} : X \to \Sigma$ be a $(1, 1-\delta)$-coloring of the game $G = (V, E, \Phi, \Sigma)$ where $X \subseteq V$, $|X| = (1-\delta)|V|$. The coloring satisfies all the constraints inside X. Consider the set of vertices in $G'(V', E')$,

$$I \overset{def}{=} \{\,(x, A) \mid x \in X,\ \mathcal{C}(x) \in A\}\,.$$

[18] One could add edges inside each cloud according to the Kneser graph. The reduction does not need it though.

Clearly, the set I includes a weight p of the vertices inside the cloud for every $x \in X$. Hence the weight of I is $(1-\delta)p \geqslant 1 - \frac{1}{\sqrt{2}} - 2\delta$. We observe that I is an independent set. For every pair of vertices $(x, A), (x', A') \in I$, we show that there is no edge between them in G'. Since the coloring $\mathcal{C}$ satisfies the constraint (x, x'), we have $(\mathcal{C}(x), \mathcal{C}(x')) \in \Phi(x, x')$. By definition of the set I, we have $\mathcal{C}(x) \in A$, $\mathcal{C}(x') \in A'$. Thus A, A' contain a consistent pair of colors, so there is no edge between (x, A) and (x', A').

B.4 Soundness

We begin by stating two auxiliary lemmas towards the soundness analysis. The relevance of the 2-to-2-ness of the constraints and the choice of $p \approx 1 - \frac{1}{\sqrt{2}}$ is apparent from the statements of these lemmas. Let Σ and Γ be alphabets such that $|\Gamma| = \frac{|\Sigma|}{2}$ and $\pi : \Sigma \to \Gamma$ be a 2-to-1 map. For $F \subseteq \Sigma$, its projection $\pi(F) \subseteq \Gamma$ is defined naturally as $\{\pi(\sigma) | \sigma \in F\}$. For a family $\mathcal{F} \subseteq \mathcal{P}(\Sigma)$, the projected family $\pi(\mathcal{F}) \subseteq \mathcal{P}(\Gamma)$ is defined naturally as $\{\pi(F) | F \in \mathcal{F}\}$. For a subset $H \subseteq \Gamma$, the set $\pi^{-1}(H)$, $|\pi^{-1}(H)| = 2|H|$ is defined naturally as $\{\sigma | \sigma \in \Sigma,\ \pi(\sigma) \in H\}$.

Lemma B.3. *For $q \in (0, 1)$, $\mu_{1-(1-q)^2}(\pi(\mathcal{F})) \geqslant \mu_q(\mathcal{F})$.*

Proof. For every $H \subseteq \Gamma$, we define $\pi^{\uparrow}(H) = \{F \subseteq \Sigma | \pi(F) = H\}$. We observe that

- $\mu_{1-(1-q)^2}(H) = \mu_q(\pi^{\uparrow}(H))$.
- The families $\pi^{\uparrow}(H)$ over all $H \subseteq \Gamma$ is a disjoint partition of the family $\mathcal{P}(\Sigma)$.

The lemma follows by noting that

$$\mu_{1-(1-q)^2}(\pi(\mathcal{F})) = \sum_{H \in \pi(\mathcal{F})} \mu_{1-(1-q)^2}(H) = \sum_{H \in \pi(\mathcal{F})} \mu_q(\pi^{\uparrow}(H)) \geqslant \sum_{H \in \pi(\mathcal{F})} \mu_q(\pi^{\uparrow}(H) \cap \mathcal{F}) = \mu_q(\mathcal{F}).$$

□

Lemma B.4. *Let $\mathcal{F} \subseteq \mathcal{P}(\Sigma)$, $\mathcal{F}' \subseteq \mathcal{P}(\Sigma')$ be two families, each of weight strictly larger than $\frac{1}{2}$ under the distribution μ_q with $q = 1 - \frac{1}{\sqrt{2}}$. Let $\pi : \Sigma \to \Gamma$, $\pi' : \Sigma' \to \Gamma$ be 2-to-1 maps (so $|\Sigma| = |\Sigma'| = 2|\Gamma|$). Then there exist $F \in \mathcal{F}$, $F' \in \mathcal{F}'$ such that $\pi(F) \cap \pi'(F') = \emptyset$.*

Proof. We note that $1 - (1-q)^2 = \frac{1}{2}$ and from Lemma B.3, $\mu_{\frac{1}{2}}(\pi(\mathcal{F})) \geqslant \mu_q(\mathcal{F}) > \frac{1}{2}$ and similarly $\mu_{\frac{1}{2}}(\pi'(\mathcal{F}')) > \frac{1}{2}$. Thus $\pi(\mathcal{F})$ and $\pi'(\mathcal{F}')$ are families, each containing more than half (in the usual counting sense) of the sets from $\mathcal{P}(\Gamma)$. Hence there must exist $H \in \pi(\mathcal{F})$, $H' \in \pi'(\mathcal{F}')$ that are complements of each other and in particular $H \cap H' = \emptyset$. □

We now present the soundness analysis. Given a maximal independent set I of weight at least ε in G', we show how to construct a (j, δ)-coloring for $G = (V, E, \Phi, \Sigma)$ where j, δ depend only on ε. For every $x \in G$, consider the part of I inside the cloud of x,

$$\mathcal{F}_x = \{ A \mid A \subseteq \Sigma,\ (x, A) \in I \}.$$

Claim B.5. *The family $\mathcal{F}_x$ is monotone.*

Proof. Otherwise, there are $A \subseteq B$ such that $A \in \mathcal{F}_x$ and $B \notin \mathcal{F}_x$. Then $I \cup \{(x, B)\}$ is an independent set larger than I, contradicting the maximality of I. □

Since the independent set I has weight ε, by an averaging argument, there is a set $X \subseteq V$, $|X| \geqslant \frac{\varepsilon}{2} \cdot |V|$ such that I includes a weight $\geqslant \frac{\varepsilon}{2}$ of vertices from the cloud of x, i.e. $\mu_p(\mathcal{F}_x) \geqslant \frac{\varepsilon}{2}$ for $x \in X$.

Theorem B.6 (Russo - Margulis [34, 29])**.** *Suppose $\mathcal{F}$ is a monotone family. Then $\mu_q(\mathcal{F})$ is an increasing function of q and*

$$\frac{d\mu_q(\mathcal{F})}{dq} = \mathsf{as}_q(\mathcal{F}).$$

Claim B.7. *There exists $p' \in (p, p + \delta)$ such that*

$$\mathop{\mathbb{E}}_{x \in X}\left[\mathsf{as}_{p'}(\mathcal{F}_x)\right] \leqslant \frac{1}{\delta}.$$

Proof. By Lagrange's mean value theorem, it follows that there exists $p' \in (p, p + \delta)$ such that

$$\mathop{\mathbb{E}}_{x \in X}\left[\mathsf{as}_{p'}(\mathcal{F}_x)\right] = \frac{d}{dq}\Big(\mathop{\mathbb{E}}_{x \in X}\left[\mu_q(\mathcal{F}_x)\right]\Big)\Big|_{q=p'} = \frac{\mathbb{E}_{x \in X}\left[\mu_{p+\delta}(\mathcal{F}_x)\right] - \mathbb{E}_{x \in X}\left[\mu_p(\mathcal{F}_x)\right]}{\delta} \leqslant \frac{1}{\delta}.$$

□

From Claim B.7 and an averaging argument, there is a set $X' \subseteq X$, $|X'| \geqslant \frac{|X|}{2}$ such that for all $x \in X'$, $\mathsf{as}_{p'}(\mathcal{F}_x) \leqslant \frac{2}{\delta}$. A theorem of Friedgut states that families with bounded average sensitivity are well-approximated by "juntas".

Definition B.8 (Junta)**.** *A family $\mathcal{F} \subseteq \mathcal{P}(\Sigma)$ is called a j-junta, if there exists $J \subseteq \Sigma, |J| = j$ such that the membership of a set A in $\mathcal{F}$ is determined by only $A \cap J$.*

Theorem B.9 (Friedgut [14])**.** *There exists a constant $C(q) \geqslant 1$ such that for every $\mathcal{F} \subseteq P(\Sigma)$ and an accuracy parameter $\eta > 0$, there exists $\mathcal{F}' \subseteq P(\Sigma)$ that is a j-junta and*

- $j = C(q)^{\mathsf{as}_q(\mathcal{F})/\eta}$.
- $\mu_q(\mathcal{F} \Delta \mathcal{F}') \leqslant \eta$.

Fix $x \in X'$ and set $\eta \overset{def}{=} \frac{\varepsilon}{20}$. Since $\mathsf{as}_{p'}(\mathcal{F}_x) \leqslant \frac{2}{\delta}$, it follows from Friedgut's Theorem that $\mathcal{F}_x$ is η-close to a k-junta with $k = C(p')^{2/(\delta\eta)}$. Let $J_x \subseteq \Sigma$ denote the set of elements on which the junta depends. Clearly, the set-family that is a junta on J_x and is closest to $\mathcal{F}_x$ is the "majority vote" on each setting of J_x, namely

$$[\mathcal{F}_x]_{\frac{1}{2}} \overset{def}{=} \left\{ F \cup F' \;\middle|\; F \subseteq \Sigma \setminus J_x,\ F' \subseteq J_x,\ \Pr_{A \subseteq \Sigma \setminus J_x,\ A \sim \mu_{p'}}\left[A \cup F' \in \mathcal{F}_x\right] > \frac{1}{2} \right\}.$$

The following claim shows that the family $[\mathcal{F}_x]_{\frac{3}{4}}$ is also close to $\mathcal{F}_x$ (and will be more useful to work with):

$$[\mathcal{F}_x]_{\frac{3}{4}} \overset{def}{=} \left\{ F \cup F' \;\middle|\; F \subseteq \Sigma \setminus J_x,\ F' \subseteq J_x \ \Pr_{A \subseteq \Sigma \setminus J_x,\ A \sim \mu_{p'}}\left[A \cup F' \in \mathcal{F}_x\right] > \frac{3}{4} \right\}.$$

Claim B.10. $\mu_{p'}(\mathcal{F}_x \Delta [\mathcal{F}_x]_{\frac{3}{4}}) \leqslant 5\eta$.

Proof. Let $\mathcal{F}^* \subseteq \mathcal{P}(J_x)$ be the family of subsets $F \subseteq J_x$ such that

$$\frac{1}{2} < \Pr_{A \subseteq \Sigma \setminus J_x,\ A \sim \mu_{p'}} [A \cup F \in \mathcal{F}_x] \leqslant \frac{3}{4}.$$

Notice that for each such F (a) at least $\frac{1}{4}$ (weighted) fraction of its extensions to Σ are not in $\mathcal{F}_x$ (b) each extension is in $[\mathcal{F}_x]_{\frac{1}{2}}$ (c) no extension is in $[\mathcal{F}_x]_{\frac{3}{4}}$. Hence

$$\frac{1}{4} \cdot \Pr_{F \subseteq J_x,\ F \sim \mu_{p'}} [F \in \mathcal{F}^*] \leqslant \mu_{p'}(\mathcal{F}_x \Delta [\mathcal{F}_x]_{\frac{1}{2}}) \leqslant \eta.$$

It follows that

$$\mu([\mathcal{F}_x]_{\frac{1}{2}} \Delta [\mathcal{F}_x]_{\frac{3}{4}}) = \Pr_{F \subseteq J_x,\ F \sim \mu_{p'}} [F \in \mathcal{F}^*] \leqslant 4\eta.$$

We finish the proof using the triangle inequality,

$$\mu_{p'}(\mathcal{F}_x \Delta [\mathcal{F}_x]_{\frac{3}{4}}) \leqslant \mu_{p'}(\mathcal{F}_x \Delta [\mathcal{F}_x]_{\frac{1}{2}}) + \mu_{p'}([\mathcal{F}_x]_{\frac{1}{2}} \Delta [\mathcal{F}_x]_{\frac{3}{4}}) \leqslant 5\eta.$$

□

Claim B.11. $[\mathcal{F}_x]_{\frac{3}{4}} \neq \emptyset$.

Proof. Using the triangle inequality, the previous claim, and that $\eta = \frac{\varepsilon}{20}$,

$$\mu_{p'}([\mathcal{F}_x]_{\frac{3}{4}}) \geqslant \mu_{p'}(\mathcal{F}_x) - \mu_{p'}(\mathcal{F}_x \Delta [\mathcal{F}_x]_{\frac{3}{4}}) \geqslant \frac{\varepsilon}{2} - 5\eta > 0.$$

□

Definition B.12. *The extended junta $EJ(x)$ of x is defined by*

$$EJ(x) = J_x \ \cup \ \left\{ \sigma \in \Sigma \mid \mathsf{Infl}_\sigma(\mathcal{F}_x) \geqslant 2^{-10k} \right\}.$$

We note that for $x \in X'$, $\mathsf{as}_{p'}(\mathcal{F}_x) \leqslant \frac{2}{\delta}$ and since the average sensitivity is the sum of all influences, $|EJ(x)| \leqslant j = k + \frac{2 \cdot 2^{10k}}{\delta}$. Our coloring to the game $G = (V, E, \Phi, \Sigma)$ will assign, to every $x \in X'$, a set of at most j colors $EJ(x)$. We now show that this is indeed a (j, δ)-coloring. Firstly,

$$|X'| \geqslant \frac{|X|}{2} \geqslant \frac{\varepsilon}{4}|V| \geqslant \delta |V|$$

as δ will be chosen accordingly. Secondly, we need to show that every constraint (x_1, x_2) inside X' is satisfied in the sense of Definition 1.5. Fix any such constraint. It is a 2-to-2 or a 1-to-1 constraint. Our main soundness lemma below takes care of the 2-to-2 case, and the 1-to-1 case then follows directly from the transitivity of the game.

Lemma B.13. *Suppose $x_1, x_2 \in X'$ are such that (x_1, x_2) is a 2-to-2 constraint. Then there exist consistent colors for x_1, x_2 in their respective extended juntas. I.e. there exist $\sigma_1 \in EJ(x_1),\ \sigma_2 \in EJ(x_2)$ such that $(\sigma_1, \sigma_2) \in \Phi(x_1, x_2)$.*

Proof. It will be convenient to think of the 2-to-2 constraint in terms of a pair of 2-to-1 maps $\pi_1 : \Sigma_1 \to \Gamma, \pi_2 : \Sigma_2 \to \Gamma$. Here $\Sigma_1 = \Sigma_2 = \Sigma$ are the same alphabet, but it will be convenient to think of them as separate. A coloring (a_1, a_2) to vertices (x_1, x_2) satisfies the 2-to-2 constraint if and only if $\pi_1(a_1) = \pi_2(a_2)$. Assume towards a contradiction that there is no pair of consistent colors in the extended juntas for x_1 and x_2. The assumption can be stated as

$$\pi_1(EJ(x_1)) \cap \pi_2(EJ(x_2)) = \emptyset.$$

Note in particular that $J_{x_2} \subseteq EJ(x_2)$ and hence

$$\pi_1(EJ(x_1)) \cap \pi_2(J_{x_2}) = \emptyset. \tag{12}$$

Our goal is to exhibit $F_1 \in \mathcal{F}_{x_1},\ F_2 \in \mathcal{F}_{x_2}$ such that $\pi_1(F_1) \cap \pi_2(F_2) = \emptyset$. We will zero-in on such $F_1,\ F_2$ in progressive manner. We consider the case of F_1, the other case being similar. We zero-in on a sequence of sets

$$A_1 \subseteq B_1 \subseteq B_1 \subseteq F_1,$$

that are contained, respectively, in progressively expanding "universe of focus"

$$J_{x_1} \ \subseteq \ \pi_1^{-1}(\pi_1(J_{x_1})) \ \subseteq \ \pi_1^{-1}(\pi_1(J_{x_1})) \cup \pi_1^{-1}(\pi_2(J_{x_2})) \ \subseteq \ \Sigma_1.$$

We clarify that the set $\pi_1^{-1}(\pi_1(J_{x_1}))$ is a superset of J_{x_1} and can have size up to $2|J_{x_1}|$ since that map π_1 is 2-to-1. The weights (sizes) of set-families are with respect to $\mu_{p'}$, unless stated otherwise.

- Recalling the definition of $[\mathcal{F}_{x_1}]_{\frac{3}{4}}$ and using Claim B.11, there is $A_1 \subseteq J_{x_1}$ such that at least $\frac{3}{4}$ of its extensions outside J_{x_1} are in $\mathcal{F}_{x_1}$.
- We let $B_1 = A_1 \cup \left(\pi_1^{-1}(\pi_1(J_{x_1})) \setminus J_{x_1}\right)$. Due to monotonicity of $\mathcal{F}_{x_1}$, at least $\frac{3}{4}$ of extensions of B_1 outside $\pi_1^{-1}(\pi_1(J_{x_1}))$ are in $\mathcal{F}_{x_1}$.
- We now retain B_1 as is, but consider it as subset of enlarged universe $\pi_1^{-1}(\pi_1(J_{x_1})) \cup \pi_1^{-1}(\pi_2(J_{x_2}))$. The elements added to the enlarged universe, namely $\pi_1^{-1}(\pi_2(J_{x_2}))$ are outside of $EJ(x_1)$ (using Equation (12)), hence have influence at most 2^{-10k}, and are at most $2k$ in number. The fraction of extensions of B_1 outside $\pi_1^{-1}(\pi_1(J_{x_1})) \cup \pi_1^{-1}(\pi_2(J_{x_2}))$ remains at least

$$\frac{3}{4} - 2^{-10k} \cdot p'^{-2k-1}(1-p')^{-2k-1} \geqslant \frac{5}{8}.$$

Using a similar argument for x_2, to summarize, there exist

$$B_1 \subseteq D_1 = \pi_1^{-1}(\pi_1(J_{x_1})) \cup \pi_1^{-1}(\pi_2(J_{x_2})),\ \ B_2 \subseteq D_2 = \pi_2^{-1}(\pi_1(J_{x_1})) \cup \pi_2^{-1}(\pi_2(J_{x_2}))$$

such that at least $\frac{5}{8}$ of their extensions outside D_1 and D_2 respectively are in $\mathcal{F}_{x_1}$ and $\mathcal{F}_{x_2}$ respectively. Note that

$$\pi_1(B_1) \cap \pi_2(B_2) = \emptyset. \tag{13}$$

We are almost done. Denote

$$\mathcal{F}_1 = \{S_1 \subseteq \Sigma_1 \setminus D_1 \mid B_1 \cup S_1 \in \mathcal{F}_{x_1}\}, \quad \mathcal{F}_2 = \{S_2 \subseteq \Sigma_2 \setminus D_2 \mid B_2 \cup S_2 \in \mathcal{F}_{x_2}\},$$

so that $\mu_{p'}(\mathcal{F}_1) \geqslant \frac{5}{8}$ and due to monotonicity, letting $q = 1 - \frac{1}{\sqrt{2}} \geqslant p'$, $\mu_q(\mathcal{F}_1) \geqslant \frac{5}{8}$, and similarly $\mu_q(\mathcal{F}_2) \geqslant \frac{5}{8}$. Applying Lemma B.4 to $\mathcal{F}_1, \mathcal{F}_2$ along with 2-to-1 maps $\pi_1 : \Sigma_1 \setminus D_1 \to \Gamma \setminus \pi_1(D_1)$ and $\pi_2 : \Sigma_2 \setminus D_2 \to \Gamma \setminus \pi_2(D_2)$ (we have $\pi_1(D_1) = \pi_2(D_2) = J_{x_1} \cup J_{x_2}$), there exist $F_1^* \subseteq \Sigma_1 \setminus D_1,\ F_2^* \subseteq \Sigma_2 \setminus D_2$ such that

$$\pi_1(F_1^*) \cap \pi_2(F_2^*) = \emptyset. \tag{14}$$

Finally, letting $F_1 = B_1 \cup F_1^*$ and $F_2 = B_2 \cup F_2^*$, and using Equations (13),(14), we conclude that $F_1 \in \mathcal{F}_{x_1},\ F_2 \in \mathcal{F}_{x_2},\ \pi_1(F_1) \cap \pi_2(F_2) = \emptyset$ as desired. □

Finally, we handle the 1-to-1 constraints inside X'. Let $G_{X'}$ be the subgraph of $G(V, E, \Phi, \Sigma)$ induced on X'. The transitivity of the game, as per Definition 1.4, then implies that $G|_{X'}$ can be partitioned into cliques $\mathcal{C}_1, \dots, \mathcal{C}_m$ such that

- All constraints inside a clique $\mathcal{C}_r$ are 1-to-1, i.e. matchings on $\Sigma \times \Sigma$. For any $x, y, z \in \mathcal{C}_r$, the matchings between $(x, y), (y, z)$ can be composed to derive the matching between (x, z).
- For $r \neq s$, either there is no edge between $\mathcal{C}_r, \mathcal{C}_s$ or there is a complete bipartite graph between $\mathcal{C}_r, \mathcal{C}_s$ with all constraints being 2-to-2. In the latter case, for any $x, y \in \mathcal{C}_r$ and $z \in \mathcal{C}_s$, the 2-to-2 constraint (x, z) is a composition of the 1-to-1 constraint (x, y) and the 2-to-2 constraint (y, z).

These considerations show that all vertices inside a clique C_r play an essentially identical role. Therefore, in a maximal independent set I, for any $x, y \in \mathcal{C}_r$, the families $\mathcal{F}_x, \mathcal{F}_y \subseteq \mathcal{P}(\Sigma)$ are identical up to the permutation of Σ that defines the 1-to-1 constraint (x, y), and hence the color-sets $EJ(x), EJ(y)$ are identical up to the (same) permutation. This shows that 1-to-1 constraints are satisfied in the sense of Definition 1.5.

C Covering Property

Proof of Lemma 4.7 from Lemma 4.6

Let $U, V, k, \beta, \ell, \mathcal{L}, \mathcal{L}', q, Q, \mathcal{L}_Q, \mathcal{L}'_Q$ be as in Definition 4.5 and Lemmas 4.6, 4.7. Let $\mathcal{Q}, \mathcal{Q}'$ be distributions over q-dimensional subspaces of $\{0,1\}^U$ that are analogous to $\mathcal{L}, \mathcal{L}'$ respectively (i.e. as in Definition 4.5, with parameter q instead of ℓ). It is easily observed that an equivalent way to sample from $\mathcal{Q}$ (resp. $\mathcal{Q}'$) is to sample an ℓ-space L from $\mathcal{L}$ (resp. $\mathcal{L}'$) and then sample a uniformly random q-dimensional subspace of L. We stress that $\mathcal{Q}$ and $\mathcal{L}$ are uniform distributions on q-dimensional and ℓ-dimensional subspaces of $\{0,1\}^U$

respectively. We have the sequence of arguments

$$\begin{aligned}
\mathop{\mathbb{E}}_{Q\sim\mathcal{Q}}\left[\mathsf{SD}(\mathcal{L}_Q,\mathcal{L}'_Q)\right] &= \sum_{Q}\Pr\left[\mathcal{Q}=Q\right]\sum_{L\supseteq Q}\Big|\Pr\left[\mathcal{L}_Q=L\right]-\Pr\left[\mathcal{L}'_Q=L\right]\Big| \\
&= \sum_{Q}\sum_{L\supseteq L}\Big|\Pr\left[\mathcal{Q}=Q\right]\cdot\Pr\left[\mathcal{L}_Q=L\right]-\Pr\left[\mathcal{Q}=Q\right]\cdot\Pr\left[\mathcal{L}'_Q=L\right]\Big| \\
&\leqslant \sum_{Q}\sum_{L\supseteq Q}\Big|\Pr\left[\mathcal{Q}=Q\right]\cdot\Pr\left[\mathcal{L}_Q=L\right]-\Pr\left[\mathcal{Q}'=Q\right]\cdot\Pr\left[\mathcal{L}'_Q=L\right]\Big| + \\
&\quad \sum_{Q}\sum_{L\supseteq Q}\Big|\Pr\left[\mathcal{Q}'=Q\right]\cdot\Pr\left[\mathcal{L}'_Q=L\right]-\Pr\left[\mathcal{Q}=Q\right]\cdot\Pr\left[\mathcal{L}'_Q=L\right]\Big| \\
&= \sum_{L}\Big|\Pr\left[\mathcal{L}=L\right]-\Pr\left[\mathcal{L}'=L\right]\Big| + \\
&\quad \sum_{Q}\Big|\Pr\left[\mathcal{Q}'=Q\right]-\Pr\left[\mathcal{Q}=Q\right]\Big|\sum_{L\supseteq Q}\Pr\left[\mathcal{L}'_Q=L\right] \\
&= \mathsf{SD}(\mathcal{L},\mathcal{L}')+\mathsf{SD}(\mathcal{Q},\mathcal{Q}'),
\end{aligned}$$

where we used triangle inequality and the fact that sampling $Q\in\mathcal{Q}$ and then $L\sim\mathcal{L}_Q$ is equivalent to sampling $L\sim\mathcal{L}$ and similarly, sampling $Q\in\mathcal{Q}'$ and then $L\sim\mathcal{L}'_Q$ is equivalent to sampling $L\sim\mathcal{L}'$. Now Lemma 4.6 upper bounds both $\mathsf{SD}(\mathcal{L},\mathcal{L}')$ and $\mathsf{SD}(\mathcal{Q},\mathcal{Q}')$ by $\beta\sqrt{k}\cdot 2^{\ell+4}$ and hence

$$\mathop{\mathbb{E}}_{Q\sim\mathcal{Q}}\left[\mathsf{SD}(\mathcal{L}_Q,\mathcal{L}'_Q)\right]\leqslant\beta\sqrt{k}\cdot 2^{\ell+5}.$$

Lemma 4.7 now follows by Markov inequality.

Proof of Lemma 4.6

We recall that $U,|U|=3k$ is a set of $3k$ variables in k equations. A subset $V\subseteq U$ is chosen by choosing independently for each equation, one of the variables in the equation with probability β and all three variables in the equation with probability $1-\beta$. The expected size of V is $3k-2\beta k$ and except with probability $2^{-\Omega(k)}$, we have $|V|\geqslant 2k$.

We note that choosing a uniformly random ℓ-subspace L of $\{0,1\}^U$ (resp. $\{0,1\}^V$) is equivalent to choosing uniformly a sequence of points $x_1,\ldots,x_\ell$ in $\{0,1\}^U$ (resp. $\{0,1\}^V$) that are linearly independent and letting $L=\mathsf{Span}(x_1,\ldots,x_\ell)$. Since a uniformly random and independent sequence of points $x_1,\ldots,x_\ell$ in $\{0,1\}^U$ (resp. in $\{0,1\}^V$) is linearly independent except with probability $\leqslant 2^{\ell-\mathsf{dim}(U)}$ (resp. $\leqslant 2^{\ell-\mathsf{dim}(V)}$, see Fact E.3), we might as well focus on such sequences of points. It is thus enough to bound the statistical distance between distributions $\mathcal{D},\mathcal{D}'$ over $(\{0,1\}^U)^\ell$ sampled as:

- $\mathcal{D}$: Choose uniformly and independently $x_1,\ldots,x_\ell\in\{0,1\}^U$.
- $\mathcal{D}'$: Choose $V\subseteq U$, choose uniformly and independently $x'_1,\ldots,x'_\ell\in\{0,1\}^V$ and regard them as points in $\{0,1\}^U$ (by appending 0 in coordinates $U\setminus V$).

We now observe that since the process of choosing $V\subseteq U$ is independent over the k equations, $\mathcal{D}=\mathcal{S}^k$ and $\mathcal{D}'=\mathcal{S}'^k$ where $\mathcal{S},\mathcal{S}'$ are the "basic" distributions exactly as above, but with $k=1,|U|=3$. A bound on the statistical distance between $\mathcal{D},\mathcal{D}'$ now follows in the same manner as in [28, Lemma 3.1], by bounding the Hellinger distance between $\mathcal{S}, \mathcal{S}'$, using the multiplicativity of the Hellinger distance to bound the Hellinger distance between $\mathcal{D}, \mathcal{D}'$ and finally, bounding the Hellinger distance in terms of the statistical distance. We observe how a bound on Hellinger distance between $\mathcal{S}, \mathcal{S}'$ also follows already from the proof of [28, Lemma 3.1]. Re-writing the sampling process for $\mathcal{S}, \mathcal{S}'$ for convenience (this is the special case $k = 1, |U| = 3$):

- $\mathcal{S}$: Choose uniformly at random $x_1, \ldots, x_\ell \in \{0,1\}^3$.
- $\mathcal{S}'$: With probability $1 - \beta$, choose uniformly at random $x_1, \ldots, x_\ell \in \{0,1\}^3$. Otherwise:
  Choose uniformly at random $b_1, \ldots, b_\ell \in \{0,1\}$. Output with probability $\frac{\beta}{3}$ each,

$$b_1 00, \ldots, b_\ell 00, \quad \text{or} \quad 0b_1 0, \ldots, 0b_\ell 0, \quad \text{or} \quad 00b_1, \ldots, 00b_\ell.$$

The distributions $\mathcal{S}, \mathcal{S}'$ are over $(\{0,1\}^3)^\ell$ which is equivalent to Σ^3 where $\Sigma = \{0,1\}^\ell$ is the concatenation of the first (second, third, respectively) bit of each of the ℓ strings of length three. Note that $0^\ell \in \Sigma$. Denoting the uniform distribution over Σ by $\mathsf{Uniform}(\Sigma)$, it is seen that

$$\mathcal{S} = (\mathsf{Uniform}(\Sigma), \mathsf{Uniform}(\Sigma), \mathsf{Uniform}(\Sigma)),$$

i.e. three independent and uniform copies of Σ, whereas,

$$\begin{aligned} \mathcal{S}' \;=\; & (1-\beta)\,(\mathsf{Uniform}(\Sigma), \mathsf{Uniform}(\Sigma), \mathsf{Uniform}(\Sigma)) + \\ & \frac{\beta}{3}\left(\mathsf{Uniform}(\Sigma), 0^\ell, 0^\ell\right) + \frac{\beta}{3}\left(0^\ell, \mathsf{Uniform}(\Sigma), 0^\ell\right) + \frac{\beta}{3}\left(0^\ell, 0^\ell, \mathsf{Uniform}(\Sigma)\right). \end{aligned}$$

With this viewpoint, the Hellinger distance between $\mathcal{S}, \mathcal{S}'$ is calculated to be at most $4\beta^2|\Sigma|^2$ in the proof of [28, Lemma 3.1]. The statistical distance between $\mathcal{D}, \mathcal{D}'$ is then at most $16\beta\sqrt{k} \cdot |\Sigma|$.

D "ℓ-space vs b-space" Linearity Test

In this section, we present and analyze "ℓ-space vs b-space" linearity test. The analysis is Fourier analytic and, as is standard, it is convenient to think of boolean values as $\{-1,1\}$ and replace addition over $\mathbb{F}_2$ by product of the signs $\{-1,1\}$. A function $f : \Omega = \{-1,1\}^n \to \{-1,1\}$ is linear if $f(x)f(y) = f(x \cdot y)$ for all $x, y \in \Omega$ and $x \cdot y$ denotes the coordinatewise product of x, y.

The ℓ-space vs b-space Linearity Test

For $\Omega = \{-1,1\}^n$, let $\mathcal{B}$ and $\mathcal{L}$ denote the set of all b-dimensional and ℓ-dimensional subspaces of Ω. Let A, F be tables that assign, for $B \in \mathcal{B},\ L \in \mathcal{L}$ respectively, linear functions $A[B] : B \to \{-1,1\},\ F[L] : L \to \{-1,1\}$ on the respective subspaces. The test picks a pair (B, L) uniformly at random with $B \subseteq L, B \in \mathcal{B}, L \in \mathcal{L}$ and accepts if and only if

$$F[L]|_B \equiv A[B].$$

Our result is the following:

Theorem D.1. *Let $\Omega, \mathcal{B}, \mathcal{L}$ and parameters $n, \ell, 1 \leqslant b \leqslant \frac{\ell}{4}$ be as in the description of the test above. Let A, F be tables that assign linear functions to $B \in \mathcal{B}$ and $L \in \mathcal{L}$ respectively. Suppose the tables pass the linearity test with probability at least $\frac{1}{2^b} + \varepsilon$ where $\varepsilon \geqslant 2^{2-b/4}$, i.e.*

$$\Pr_{B \subseteq L, B \in \mathcal{B}, L \in \mathcal{L}} [F[L]|_B \equiv A[B]] \geqslant \frac{1}{2^b} + \varepsilon.$$

Then there exists a global linear function $g : \Omega \to \{-1,1\}$ that agrees with at least $\frac{\varepsilon^3}{300}$ fraction of the ℓ-spaces, that is

$$\Pr_{L \in \mathcal{L}} [F[L] \equiv g|_L] \geqslant \frac{\varepsilon^3}{300}.$$

The rest of the section is devoted to proving Theorem D.1. We start by viewing the entire table $A[\cdot]$ as a function $f : \{-1,1\}^{nb} \to \{-1,1\}^b$ as follows. In notation, for $(v_1, \dots, v_b) \in (\{-1,1\}^n)^b = \{-1,1\}^{nb}$,

$$f(v_1, \dots, v_b) = (A[\mathsf{Span}(v_1, \dots, v_b)](v_1), \dots, A[\mathsf{Span}(v_1, \dots, v_b)](v_b)).$$

In words, to evaluate $f(v_1, \dots, v_b)$, one considers the b-space $B = \mathsf{Span}(v_1, \dots, v_b)$, and the linear function $A[B]$ on B. The linear function assigns, in particular, $\{-1,1\}$-values to the vectors $v_1, \dots, v_b$ respectively. The list of these b values is defined to be $f(v_1, \dots, v_b)$. Since the output of f is a string of length b, we can think of f as a collection of $\{-1,1\}$-valued functions, $f_1, \dots, f_b$, one for each output coordinate. In notation, $f_i : \{-1,1\}^{nb} \to \{-1,1\}$ is defined as

$$f_i(v_1, \dots, v_b) = A[\mathsf{Span}(v_1, \dots, v_b)](v_i).$$

We must make a couple of clarifying remarks. First, when the input vectors $\{v_1, ..., v_b\}$ are linearly dependent, then their span B has dimension less than b and $A[B]$ is undefined. However the fraction of such inputs is negligible (at most 2^{b-n}) and on those inputs f can be defined arbitrarily without affecting the analysis. Second, since the same b-space may have different bases, f has many symmetries, e.g. $f_1(v_1, ..., v_b) = f_2(v_2, v_1, ..., v_b)$. We will use these symmetries, but not in any explicit manner.

The Gowers' Test

The main idea behind the analysis is to use a "Gowers' Test" as an auxiliary tool. We can relate the acceptance probability of the ℓ-space vs b-space test to that of the acceptance probability of the Gowers' test. The Gowers' test allows us to conveniently switch from local considerations to global considerations. Let $\vec{1}$ denote a b-dimensional vector with all coordinates 1.

Definition D.2. [Gowers' Test] *Given $h : \{-1,1\}^{nb} \to \{-1,1\}^b$, pick $x, y, z \in \{-1,1\}^{nb}$ randomly and check if*

$$h(x)h(y)h(z)h(x \cdot y \cdot z) = \vec{1}.$$

Represent a function $h : \{-1,1\}^{nb} \to \{-1,1\}^b$ as $h = (h_1, \dots, h_b)$ where h_i are the coordinatewise functions. For $T \subseteq [b]$, let $h_T = \prod_{i \in T} h_i$ be the product functions. The lemma below expresses the probability of h passing the Gowers' test in terms of the Fourier coefficients of products of functions h_T.

Lemma D.3. *The probability that $h \colon \{-1,1\}^{nb} \to \{-1,1\}^b$ passes the Gowers' Test is:*

$$\Pr_{x,y,z \in \{-1,1\}^{nb}} \left[h(x)h(y)h(z)h(x \cdot y \cdot z) = \vec{1} \right] = \frac{1}{2^b} + \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \emptyset} \sum_{S \subseteq [nb]} \widehat{h}_T^4(S).$$

Proof. For the test to pass, it must pass on every coordinate. Thus,

$$\begin{aligned}
&\Pr_{x,y,z\in\{-1,1\}^{nb}}\left[h(x)h(y)h(z)h(x\cdot y\cdot z)=\vec{1}\right]\\
=&\mathop{\mathbb{E}}_{x,y,z\in\{-1,1\}^{nb}}\left[\prod_{i=1}^{b}\frac{1+h_i(x)h_i(y)h_i(z)h_i(x\cdot y\cdot z)}{2}\right]\\
=&\frac{1}{2^b}+\frac{1}{2^b}\sum_{T\subseteq[b],T\neq\emptyset}\mathop{\mathbb{E}}_{x,y,z\in\{-1,1\}^{nb}}\left[h_T(x)h_T(y)h_T(z)h_T(x\cdot y\cdot z)\right]\\
=&\frac{1}{2^b}+\frac{1}{2^b}\sum_{T\subseteq[b],T\neq\emptyset}\sum_{S\subseteq[nb]}\widehat{h}_T^4(S).
\end{aligned}$$

□

The main trick is that Lemma D.3 is applied globally as well as locally and then the information gained from the two applications is combined. Globally, the lemma is applied to the function $f:\{-1,1\}^{nb}\to\{-1,1\}^b$ that (essentially) represents the entire assignment $\{A[B]|B\in\mathcal{B}\}$. Locally, for a fixed ℓ-space L, the lemma is applied to the function $g:\{-1,1\}^{\ell b}\to\{-1,1\}^b$ that represents, in a similar manner, the assignment $\{A[B]|B\subseteq L\}$ (i.e. only the assignment to b-spaces that are contained in L). We present the local application first.

Fix an ℓ-space L. Locally, L can be identified with $\{-1,1\}^\ell$ and the linear function $F[L]$ on it can be identified with a Fourier character χ_S for some $S\subseteq[\ell]$. The assignment $\{A[B]|B\subseteq L\}$ can be represented, in a similar manner as before, by a function $g:\{-1,1\}^{\ell b}\to\{-1,1\}^b$, $g=(g_1,\ldots,g_b)$ where for $(w_1,\ldots,w_b)\in(\{-1,1\}^\ell)^b=\{-1,1\}^{\ell b}$,

$$g_i(w_1,\ldots,w_b)=A[\mathsf{Span}(w_1,\ldots,w_b)](w_i).$$

We note that g really is the restriction of f to L^b. As before, for $T\subseteq[b]$, let $g_T=\prod_{i\in T}g_i$ be the product functions. We now relate the probability that g passes the Gowers' test with the probability that the linearity test passes for the fixed L, i.e. the probability that $F[L]|_B=A[B]$ for a random $B\subseteq L$. Let $1-\gamma$ be the probability that random vectors $w_1,\ldots,w_b\in\{-1,1\}^\ell$ are linearly independent, so that $\gamma\leqslant 2^{b-\ell}$ is negligible. Thus choosing a random b-dimensional subspace of L is essentially same as choosing b random vectors from $L=\{-1,1\}^\ell$. We now have

$$\begin{aligned}
(1-\gamma)\cdot\Pr_{B\subseteq L}\left[F[L]|_B=A[B]\right]&\leqslant\Pr_{w_1,\ldots,w_b\in\{-1,1\}^\ell}\left[\wedge_{i=1}^b F[L](w_i)=A[\mathsf{Span}(w_1,\ldots,w_b)](w_i)\right]\\
&=\Pr_{w_1,\ldots,w_b\in\{-1,1\}^\ell}\left[\wedge_{i=1}^b \chi_S(w_i)=g_i(w_1,\ldots,w_b)\right]\\
&=\mathop{\mathbb{E}}_{w_1,\ldots,w_b\in\{-1,1\}^\ell}\left[\prod_{i=1}^{b}\frac{1+\chi_S(w_i)g_i(w_1,\ldots,w_b)}{2}\right]\\
&=\frac{1}{2^b}+\frac{1}{2^b}\sum_{T\subseteq[b],T\neq\emptyset}\mathop{\mathbb{E}}_{w_1,\ldots,w_b\in\{-1,1\}^\ell}\left[g_T(w_1,\ldots,w_b)\prod_{i\in T}\chi_S(w_i)\right]\\
&=\frac{1}{2^b}+\frac{1}{2^b}\sum_{T\subseteq[b],T\neq\emptyset}\widehat{g}_T(S_T),
\end{aligned}$$

where $S_T \subseteq [\ell b]$ is defined as $(S_T(1), \ldots, S_T(b))$ and $S_T(i) \subseteq [\ell]$ equals S if $i \in T$ and equals $\emptyset$ if $i \notin T$. Hence noting that $\gamma \leqslant 2^{b-\ell}$,

$$\Pr_{B \subseteq L}[F[L]|_B = A[B]] \leqslant \frac{1}{2^b} + 2 \cdot 2^{b-\ell} + \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \emptyset} \widehat{g}_T(S_T).$$

Now we take average of this inequality over the choice of $L \in \mathcal{L}$ and note that the L.H.S. then equals the probability that the linearity test accepts (which is $\geqslant \frac{1}{2^b} + \varepsilon$). This gives

$$\frac{\varepsilon}{2} \leqslant \varepsilon - 2 \cdot 2^{b-\ell} \leqslant \mathop{\mathbb{E}}_{L \in \mathcal{L}} \left[\frac{1}{2^b} \sum_{T \subseteq [b], T \neq \emptyset} \widehat{g}_T(S_T) \right].$$

We keep in mind that g and S depend on the choice of L. Using convexity of the function $x \to x^4$, we get

$$\frac{\varepsilon^4}{16} \leqslant \mathop{\mathbb{E}}_{L \in \mathcal{L}} \left[\frac{1}{2^b} \sum_{T \subseteq [b], T \neq \emptyset} \widehat{g}_T^4(S_T) \right].$$

Applying Lemma D.3 to $g : \{-1,1\}^\ell \to \{-1,1\}^b$, we get

$$\frac{\varepsilon^4}{16} \leqslant \mathop{\mathbb{E}}_{L \in \mathcal{L}} \Big[\Pr[g \text{ passes Gowers' test}] \Big].$$

Now we relate the R.H.S. to the probability that f passes the Gowers' test, using the fact that g really is the restriction of f to L^b. Let $x = (x_1, \ldots, x_b), y = (y_1, \ldots, y_b), z = (z_1, \ldots, z_b)$ where x_i, y_i, z_i are either in L or in the global space $\{-1,1\}^n$, as understood from the context. We would like to argue as

$$\begin{aligned}
\frac{\varepsilon^4}{16} &\leqslant \mathop{\mathbb{E}}_{L \in \mathcal{L}} \Big[\Pr[g \text{ passes Gowers' test}] \Big] \\
&= \Pr_{L \in \mathcal{L},\, x_i, y_i, z_i \in L} \Big[g(x)g(y)g(z)g(x \cdot y \cdot z) = \vec{1} \Big] \\
&\approx \Pr_{x_i, y_i, z_i \in \{-1,1\}^n} \Big[f(x)f(y)f(z)f(x \cdot y \cdot z) = \vec{1} \Big] \\
&= \Pr[f \text{ passes Gowers' test}].
\end{aligned}$$

This is an almost correct argument, except that the distribution $\mathcal{D}$ of $x_i, y_i, z_i \in \{-1,1\}^n$ is slightly different from the distribution $\mathcal{D}'$ of $L \in \mathcal{L},\ x_i, y_i, z_i \in L = \{-1,1\}^\ell$ (i.e. first choosing L at random and then choosing x_i, y_i, z_i from inside L). The distributions are identical however if conditioned on the $3b$ vectors x_i, y_i, z_i being linearly independent. The probability of this happening is at least $1 - 2^{3b-n}$ and $1 - 2^{3b-\ell}$ depending on the space they are chosen from. It follows that the statistical distance between the distributions is at most $3 \cdot 2^{3b-\ell}$ and the argument above is correct up to that much error. It follows that (provided $\varepsilon \geqslant 4 \cdot 2^{-b/4}$)

$$\frac{\varepsilon^4}{24} \leqslant \Pr[f \text{ passes Gowers' test}].$$

Applying Lemma D.3 to f,

$$\frac{1}{2^b} + \frac{1}{2^b} \sum_{T \subseteq [b], T \neq \emptyset} \sum_{S \subseteq [nb]} \widehat{f}_T^4(S) \geqslant \frac{\varepsilon^4}{24}.$$

Noting that the sum of squares of Fourier coefficients of a boolean function equals 1, we see that there exists $T \subseteq [b], T \neq \emptyset$ and $S \subseteq [nb]$ such that $\widehat{f}_T^2(S) \geqslant \frac{\varepsilon^4}{32}$. We are almost done, by inspecting the coefficient $\widehat{f}_T(S)$. Let $S = (S_1, \ldots, S_b)$, $S_i \subseteq [n]$, and denote $B = \mathsf{Span}(v_1, \ldots, v_b)$ below. By definition of Fourier coefficients and of the functions f, f_T,

$$\begin{aligned}
\widehat{f}_T(S) &= \mathop{\mathbb{E}}_{v_1,\ldots,v_b \in \{-1,1\}^n} \left[f_T(v_1, \ldots, v_b) \cdot \prod_{i=1}^{b} \chi_{S_i}(v_i) \right] \\
&= \mathop{\mathbb{E}}_{v_1,\ldots,v_b \in \{-1,1\}^n} \left[\prod_{i \in T} A[B](v_i) \cdot \prod_{i=1}^{b} \chi_{S_i}(v_i) \right] \\
&= \mathop{\mathbb{E}}_{\substack{B,\ \mathsf{dim}(B)=b,\ v_1,\ldots,v_b \in B, \\ \mathsf{Rank}(v_1,\ldots,v_b)=b}} \left[\prod_{i \in T} A[B](v_i) \cdot \prod_{i=1}^{b} \chi_{S_i}(v_i) \right] \pm 2^{b-n},
\end{aligned}$$

where while choosing $v_1, \ldots, v_b \in \{-1,1\}^n$, they are assumed to be linearly independent (introducing the negligible error term 2^{b-n}) and then their choice is same as first choosing a random b-space B and then letting $v_1, \ldots, v_b$ be a random basis of B. Regard $B = \{-1,1\}^b$ and $A[B]$ as the linear function $\chi_{S[B]}$ for $S[B] \subseteq [b]$. The global function $\chi_{S_i}(v_i)$ where $S_i \subseteq [n], v_i \in \{-1,1\}^n$, after restricting to $v_i \in B$, amounts to a linear function on B, say $\chi_{S_i \downarrow B}$ with $S_i \downarrow B \subseteq [b]$. Thus

$$\widehat{f}_T(S) = \mathop{\mathbb{E}}_{\substack{B,\ \mathsf{dim}(B)=b,\ v_1,\ldots,v_b \in B, \\ \mathsf{Rank}(v_1,\ldots,v_b)=b}} \left[\prod_{i \in T} \chi_{S[B] \Delta S_i \downarrow B}(v_i) \cdot \prod_{i \in [b] \setminus T} \chi_{S_i \downarrow B}(v_i) \right] \pm 2^{b-n}.$$

Let us look at the expectation for a fixed B. Call B good if

$$\forall\, i \in T,\ S_i \downarrow B = S[B], \qquad \forall\, i \in [b] \setminus T,\ S_i \downarrow B = \emptyset, \tag{15}$$

and let $\mathcal{B}'$ be the set of such good B. For a good B, the expectation equals 1 and from Lemma E.2, the expectation is bounded by 2^{-b+1} in magnitude for a bad B. Thus

$$\widehat{f}_T(S) = \Pr_B \left[B \in \mathcal{B}' \right] \pm 2^{-b+1} \pm 2^{b-n}.$$

Since $|\widehat{f}_T(S)| \geqslant \frac{\varepsilon^2}{6}$, it follows that $\Pr_B \left[B \in \mathcal{B}' \right] \geqslant \frac{\varepsilon^2}{10}$ (since $\varepsilon \geqslant 2^{2-b/4} \geqslant 2^{3-b/2}$). Now we show that in fact for some $S^* \subseteq [n]$, for all $i \in T$, $S_i = S^*$ and for all $i \in [b] \setminus T$, $S_i = \emptyset$. This is because if this were not the case, for a random b-space B, Condition (15) holds with probability at most 2^{-b}, upper bounding $\Pr_B \left[B \in \mathcal{B}' \right]$ by 2^{-b}, a contradiction. It follows that $\chi_{S^*} : \{-1,1\}^n \to \{-1,1\}$ is a global linear function that agrees with the given linear function $A[B]$ on $\geqslant \frac{\varepsilon^2}{10}$ fraction of B, $\mathsf{dim}(B) = b$.

Agreement with ℓ-spaces

What we have concluded so far is that if tables F, A pass the ℓ-space vs b-space linearity test with probability $\geqslant \frac{1}{2^b} + \varepsilon$, then there is a global linear function $g : \{-1,1\}^n \to \{-1,1\}^n$ that agrees with $A[B]$ for $\geqslant \frac{\varepsilon^2}{10}$ fraction of b-spaces B. Theorem D.1 however demands a good agreement with $F[L]$ for ℓ-spaces L. This is easy to fix. Let

$$\mathcal{B}^* = \left\{ B \mid \Pr_{L : B \subseteq L} \left[F[L]|_B = A[B] \right] \geqslant \frac{\varepsilon}{2} \right\}.$$

Since the linearity test succeeds with probability $\geqslant \frac{1}{2^b} + \varepsilon$, by an averaging argument, $|\mathcal{B}^*| \geqslant \frac{\varepsilon}{2} \cdot |\mathcal{B}|$. Now modify the table $A[\cdot]$ to table $A'[\cdot]$ so that $A'[B] = A[B]$ for $B \in \mathcal{B}^*$ and $A'[B]$ is a random linear function on B otherwise. Clearly, the tables F, A' still pass the linearity test with probability $\geqslant 2^{-b} + \frac{\varepsilon}{2}$ and by the analysis so far, there is a global linear function g that agrees with $A'[B]$ for $\geqslant \frac{\varepsilon^2}{40}$ fraction of $B \in \mathcal{B}$. Since $A'[B]$ for $B \notin \mathcal{B}^*$ was defined at random, their contribution to consistency with g is negligible, i.e. at most 2^{-b}. Thus we have

$$\Pr_B [g|_B = A[B] \ \wedge \ B \in B^*] \geqslant \frac{\varepsilon^2}{80}.$$

For every $B \in \mathcal{B}^*$, by definition, $A[B]$ is consistent with $F[L]$ for at least $\frac{\varepsilon}{2}$ fraction of L containing B. Hence,

$$\Pr_{B \subseteq L} [g|_B = A[B] \ \wedge \ B \in B^* \ \wedge F[L]_B = A[B]] \geqslant \frac{\varepsilon^3}{160}.$$

In particular,

$$\mathop{\mathbb{E}}_L \left[\Pr_{B \subseteq L} [F[L]|_B = g|B] \right] \geqslant \frac{\varepsilon^3}{160}.$$

This implies immediately that $g|_L = F[L]$ for at least $\frac{\varepsilon^3}{160} - 2^b \geqslant \frac{\varepsilon^3}{300}$ fraction of L, since for L not satisfying this, the inside probability is at most 2^{-b}.

E Missing Proofs

E.1 Hypothesis 2.5 implies Hypothesis 2.10

Let $G(V, \ell)$, $\mathsf{dim}(V) = n$ and the side condition $\{h_i\}_{i=1}^r, \{b_i\}_{i=1}^r$, $r \leqslant \frac{n}{3}$ be as in Hypothesis 2.10. Let $H = \mathsf{Span}(h_1, \dots, h_r)$, $\mathsf{dim}(H) = r$. Let $W[H]$ be any "complementing space" to H so that $V = H \oplus W[H]$, $H \cap W[H] = \{0\}$. We identify linear functions on subspaces of V that respect the side condition H with their restrictions to $W[H]$, move to the "lower order" Grassmann graph $G(W[H], \ell)$, apply Hypothesis 2.5, and then "pull-back" the function on $W[H]$ guaranteed by Hypothesis 2.5 to a function on V that respects the side condition.

Formally, let $q, \alpha(\cdot), C$ be as in Hypothesis 2.5 given $(j, \frac{\delta}{2})$ and n sufficiently large. Let $Q \subseteq V$ be a random q-dimensional space. With probability at least $1 - 2^{r+q-n}$ (see Fact E.4), we have $Q \cap H = \{0\}$ and we condition on this event henceforth.

Claim E.1. $W[Q] \overset{def}{=} (Q \oplus H) \cap W[H]$ *is a random q-dimensional subspace of* $W[H]$.

Proof. Firstly, the dimension consideration shows that

$$\mathsf{dim}(W[Q]) = \mathsf{dim}(W[H]) + \mathsf{dim}(Q \oplus H) - \mathsf{dim}(W[H] \oplus Q \oplus H) = (n - r) + (q + r) - n = q.$$

Also, it is easily seen that each q-dimensional subspace of $W[H]$ has equally many pre-images under the mapping $Q \to (Q \oplus H) \cap W[H]$. □

Let $F[\cdot]$ be the (j, δ)-assignment to $G(V, \ell)$ respecting the side condition H and S be the set of its vertices that have been assigned. We "move" to the lower order Grassmann graph $G(W[H], \ell)$ and define a (j, δ)-assignment $\tilde{F}[\cdot]$ to it as (denoting the set of its vertices assigned as $\tilde{S}$)

$$\tilde{S} = \{ L \in G(W[H], \ell) \mid L \oplus H \in S \}.$$

$$\tilde{F}[L] = F[L \oplus H]|_L.$$

By Hypothesis 2.10, with probability at least $\alpha(\ell)$ over the choice of Q, there exists $g_Q \colon W[H] \to \{0,1\}$ such that

$$\Pr_{L:Q \subseteq L \subseteq W[H]} \left[g_Q|_L \in \tilde{F}[L] \right] \geqslant C.$$

Define g'_Q to be the unique extension of g_Q to V respecting the side condition. Since spaces $L \subseteq W[H]$ can be pulled back to $L \oplus H$, g'_Q satisfies Equation (2) of Hypothesis 2.10 as required.

E.2 Proof of Theorem 2.6

Denote by N the size of $\mathcal{L} \overset{def}{=} \{ L \in G(V,\ell) \mid Q \subseteq L \}$ and let $f_1, ..., f_m$ be all functions agreeing with $F[\cdot]$ on at least C fraction of $L \in \mathcal{L}$. We construct a bipartite graph, where the left side consists of $f_1, ..., f_m$ and the right side consists of pairs $\{ (L,\sigma) \mid L \in \mathcal{L}, \sigma \in F[L] \}$. We connect f_i and (L,σ) by an edge if $f_i|_L \equiv \sigma$. Then the degree of each f_i is at least $C \cdot N$ and the number of vertices on the right side is at most jN. Let us remove edges if necessary so that the degree of each f_i is exactly $C \cdot N$.

Denote by $d(L,\sigma)$ the degree of (L,σ) and let us count the number of triples $\{f_i, f_j, (L,\sigma)\}$ where $i \neq j$ and $(f_i,(L,\sigma))$, $(f_j,(L,\sigma))$ are both edges in the bipartite graph. Using Cauchy-Schwartz and noting that the number of vertices on the right side is at most jN and $\sum_{L \in \mathcal{L}, \sigma \in F[L]} d(L,\sigma) = CmN$, the number of such triples is lower bounded as

$$\sum_{\substack{L \in \mathcal{L},\\ \sigma \in F[L]}} \binom{d(L,\sigma)}{2} = \sum_{\substack{L \in \mathcal{L},\\ \sigma \in F[L]}} \frac{d(L,\sigma)^2}{2} - \frac{d(L,\sigma)}{2} \geqslant jN \cdot \frac{\left(\frac{CmN}{jN}\right)^2}{2} - \frac{CmN}{2} = \frac{C^2 m^2 N}{2j} - \frac{CmN}{2}.$$

On the other hand, since any distinct pair of functions f_i, f_j agree on at most $2^{q-\ell}$ fraction of $L \in \mathcal{L}$, the number of such triples is at most $\binom{m}{2} 2^{q-\ell} N \leqslant \frac{m^2 2^{q-\ell} N}{2}$. Combining the two bounds gives $m \leqslant \frac{jC}{C^2 - j2^{q-\ell}}$.

E.3 Auxiliary Lemmas and Facts

Lemma E.2. *Let $s_1, \ldots, s_b \in \mathbb{F}_2^b$ such that at least one of them is non-zero. Let $v_1, \ldots, v_b \in \mathbb{F}_2^b$ be chosen at random. Then the following conditional expectation is bounded as:*

$$\left| \mathop{\mathbb{E}}_{v_1,\ldots,v_b} \left[\prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle} \mid \mathsf{Rank}(v_1, \ldots, v_b) = b \right] \right| \leqslant 2^{-b+1}.$$

Proof. Note that without the conditioning, the expectation is clearly zero. The point is to prove the upper bound conditional on the event that $v_1, \ldots, v_b$ are linearly independent (and hence form a basis of $\mathbb{F}_2^b$. Assume w.l.o.g. that s_1 is non-zero. Let

$$\mathcal{A} = \{ A = (v_1, \ldots, v_b) \mid \mathsf{Rank}(A) = b \},$$

so that we are interested in the expectation

$$\mathop{\mathbb{E}}_{A} \left[\prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle} \mid A \in \mathcal{A} \right].$$

Let

$$\mathcal{A}' = \{A = (v_1, \ldots, v_b) \mid \mathsf{Rank}(A) = b,\ \forall\, 2 \leqslant i \leqslant b,\ \langle s_1, v_i \rangle = 0\}.$$

It is easily seen that $|\mathcal{A}'| \leqslant 2^{-b+1} \cdot |\mathcal{A}|$. Imagine choosing $v_2, v_3, \ldots, v_b$ so that every v_i is outside the span of the previously chosen ones. If we require (in addition) that every v_i also lies in the hyperplane defined by the equation $\langle s_1, x \rangle = 0$, then at each step, this happens with probability at most $\frac{1}{2}$, showing the desired upper bound on $|\mathcal{A}'|$. Hence the two expectations

$$\mathop{\mathbb{E}}_{A}\left[\prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle} \mid A \in \mathcal{A}\right], \qquad \mathop{\mathbb{E}}_{A}\left[\prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle} \mid A \in \mathcal{A} \setminus \mathcal{A}'\right]$$

differ by at most 2^{-b+1}. We show that the latter is zero. For fixed $\alpha_2, \alpha_3, \ldots, \alpha_b \in \mathbb{F}_2$, consider the following bijection on $\mathcal{A} \setminus \mathcal{A}'$ (that adds to the first vector, a linear combination of others):

$$(v_1, v_2, v_3, \ldots, v_b) \to (v_1 + \sum_{i=2}^{b} \alpha_i v_i,\ v_2, v_3, \ldots, v_b).$$

The quantity of interest changes as follows:

$$\prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle} \to (-1)^{\sum_{i=2}^{b} \alpha_i \langle s_1, v_i \rangle} \cdot \prod_{i=1}^{b} (-1)^{\langle s_i, v_i \rangle}.$$

Now take expectation of L.H.S. over the choice of $A = (v_1, \ldots, v_b) \in \mathcal{A} \setminus \mathcal{A}'$ and expectation of R.H.S. over the choice of $A \in \mathcal{A} \setminus \mathcal{A}'$ as well as over a random choice of $\alpha_2, \ldots, \alpha_b$. The two expectations are equal (due to bijectivity) and the expectation of the L.H.S. is what we are interested in. Since $\langle s_1, v_i \rangle \neq 0$ for some $2 \leqslant i \leqslant b$, the expectation over the R.H.S. is zero and we are done. □

Fact E.3. *Let V be an n-dimensional vector space over $\mathbb{F}_2$, and $1 \leqslant \ell \leqslant n-1$. Let $x_1, \ldots, x_\ell \in V$ be chosen randomly and independently. Then $x_1, \ldots, x_\ell$ are linearly independent with probability $\geqslant 1 - 2^{\ell - n}$.*

Proof. If $x_1, \ldots, x_\ell$ are linearly dependent, then for some $1 \leqslant i \leqslant \ell$, x_i is in the span of $x_1, \ldots, x_{i-1}$. Hence the probability that these ℓ vectors are linearly dependent can be upper-bounded as

$$\sum_{i=1}^{\ell} \Pr_{x_1, \ldots, x_i \in V} [x_i \in \mathsf{Span}\{x_1, ..., x_{i-1}\}] \leqslant \sum_{i=1}^{\ell} \frac{2^{i-1}}{2^n} \leqslant 2^{\ell - n}.$$

□

Fact E.4. *Let V be an n-dimensional vector space over $\mathbb{F}_2$, $H \subseteq U$ be a subspace of dimension r, and $1 \leqslant \ell \leqslant n - \ell$. Let $x_1, \ldots, x_\ell \in V$ be chosen randomly and independently. Then*

$$\Pr_{x_1, \ldots, x_\ell \in V} [\mathsf{Span}(\{x_1, ... x_\ell\}) \cap H = \{0\}] \geqslant 1 - 2^{r + \ell - n}.$$

Proof. If $\mathsf{Span}(\{x_1, \ldots, x_\ell\}) \cap H \neq \{0\}$, then for some $1 \leqslant i \leqslant \ell$, x_i is in the span of $H \cup \{x_1, \ldots, x_{i-1}\}$. Hence the probability that $\mathsf{Span}(\{x_1, \ldots, x_\ell\}) \cap H \neq \{0\}$ can be upper-bounded as

$$\sum_{i=1}^{\ell} \Pr_{x_1, \ldots, x_\ell \in V} [x_i \in H \oplus \mathsf{Span}(\{x_1, ..., x_{i-1}\})] \leqslant \sum_{i=1}^{\ell} \frac{2^{r+i-1}}{2^n} \leqslant 2^{r + \ell - n}.$$

□

Fact E.5. *Let A, A', B be subspaces of a vector space V over $\mathbb{F}_2$ such that $A \oplus B = A' \oplus B$ and $(A \oplus A') \cap B = \{0\}$. Then $A = A'$.*

Proof. By symmetry, it suffices to show that $A \subseteq A'$. Let $a \in A$. Then $a \in A \oplus B = A' \oplus B$ and so there are $a' \in A'$, $b \in B$ such that $a = a' \oplus b$. Hence $b = a \oplus a' \in A \oplus A'$, and b must be 0. □