THEORY OF COMPUTING, Volume 21 (11), 2025, pp. 1–50
www.theoryofcomputing.org

Towards a Proof of the 2-to-1 Games Conjecture?

Irit Dinur* Subhash Khot† Guy Kindler‡ Dor Minzer§
Muli Safra¶

Received December 8, 2020; Revised March 5, 2024; Published November 26, 2025

Abstract. We propose a combinatorial hypothesis regarding a subspace vs. subspace agreement test, and prove that if correct it leads to a proof of the 2-to-1 Games Conjecture, albeit with imperfect completeness. This paper presents the second installment in a line of work by various subsets of the authors (with additional contributions by Barak, Kothari, and Steurer (ITCS'19)), which led to a proof of the 2-to-2 Games Conjecture.

A preliminary version of this paper appeared in the Proceedings of the 50th ACM Symposium on Theory of Computing (STOC'18) [9].

*Supported by an ISF-UGC grant 1399/14 and BSF grant 2014371.

†Research supported by NSF CCF-1422159, Simons Collaboration on Algorithms and Geometry, and Simons Investigator Award.

‡Supported by ISF grant no. 1692/13.

§This work was completed while the author was in the School of Computer Science, Tel Aviv University and was supported by Clore scholarship.

¶Supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 835152), by an ISF grant 2017/13 and by a BSF grant 2016414.

ACM Classification: F.2.3

AMS Classification: 68Q17

Key words and phrases: probabilistically checkable proofs, Unique Games Conjecture, Grassmann graph

DOI: 10.4086/toc.2025.v021a011

1 Introduction

In recent years, the Unique Games Conjecture [21] and its variants received significant attention. These conjectures have numerous applications to hardness of approximation, and connections to several topics in algorithms, computational complexity, and geometry (see [34, 23, 22] for surveys). However, there is still no consensus regarding the validity of these conjectures. Only recently an approach towards proving the Unique Games Conjecture, or rather a weak form of it, was proposed [27]. Building on previous work from [26], this paper presents an approach, quite orthogonal to that in [27], towards proving the related 2-to-1 Games Conjecture (or rather a variant of it with imperfect completeness). It is the second installment in the series of articles [26, 9, 10, 25] leading up to the proof of the 2-to-1 Games Conjecture.

The main contribution of this paper is the proposal of a combinatorial conjecture concerning a consistency test on the Grassmann graph, which first appeared in [26], and showing that it implies the 2-to-1 Games Conjecture with imperfect completeness. Our combinatorial conjecture was proven to be correct in subsequent work [25].

1.1 Unique Games Conjecture and d-to-1 Games Conjecture

To define Unique Games and d-to-1 games, we begin by defining a more general problem known as the label cover problem (which often goes by the name 2-Prover-1-Round Games in the literature).

Definition 1.1 (Label Cover Problem)**.** A Label Cover instance $G = (A, B, E, \Pi, \Sigma_A, \Sigma_B)$ is given by bipartite graph (A, B, E), two sets of colors Σ_A and Σ_B, and a collection of edge-constraints $\Pi = \{\pi_{uv}\}_{uv\in E}$ such that each edge (u, v) is associated with a constraint $\pi_{uv} \subseteq \Sigma_A \times \Sigma_B$.

Given an assignment $c : A \cup B \to \Sigma_A \cup \Sigma_B$ and an edge (u, v), the constraint π_{uv} is said to be satisfied if $(c(u), c(v)) \in \pi_{uv}$. The goal in the label cover problem is to find an assignment of colors to the vertices $c : A \cup B \to \Sigma_A \cup \Sigma_B$ that satisfies the maximum fraction of constraints.

As we noted earlier, the label cover problem is often referred to as a 2-prover-1-round game. The reason for that is that label cover has an alternative, equivalent active view as a game between a verifier and two non-communicating provers, which is often helpful to consider when doing reductions. In this view the verifier picks a constraint $(u, v) \in E$ at random, asks the “question” u to the first prover and the “question” v to the second prover. The verifier receives “answers” $c(u)$ from the first prover and $c(v)$ from the second prover, and accepts if and only if $(c(u), c(v)) \in \pi_{uv}$. It is easily seen that the maximum acceptance probability of the verifier over all provers’ strategies is equal to the maximum fraction of the constraints that can be satisfied by a coloring to the Label Cover instance. In the rest of the paper we will sometimes switch from the combinatorial, label cover view and the active, 2-prover-1-round game view whenever one is more convenient than the other. In particular, we will refer to label cover instances as 2-prover-1-round games and vice versa.

Unique Games and d-to-1 Games are special types of Label Cover instances, wherein the constraints have a specific structure.

Definition 1.2. (d-to-1 Games) Let $G = (A, B, E, \Pi, \Sigma_A, \Sigma_B)$ be a 2-Prover-1-Round game, and let $d \geqslant 1$ be an integer. A constraint $\pi_{uv} \subseteq \Pi$ is said to be d-to-1 if there is a partition $S_1, ..., S_r$ of Σ_A into sets of size d and an ordering $b_1, \ldots, b_r$ of Σ_B such that

$$\pi_{uv} = \bigcup_{i=1}^{r} S_i \times \{b_i\}$$

(this also implies that $|\Sigma_A| = dr$, $|\Sigma_B| = r$).

We say that $G = (A, B, E, \Pi, \Sigma_A, \Sigma_B)$ is a d-to-1 game if all constraints in Π are d-to-1. A 1-to-1 game is also called a *Unique Game*. In the latter case, $\Sigma_A = \Sigma_B$ and for each edge u, v, the constraint π_{uv} is a perfect matching on $\Sigma_A \times \Sigma_B$.

For this paper, one should consider the number of colors in $|\Sigma_A \cup \Sigma_B|$ as a constant, possibly large, and the size of the constraint graph as the growing input size. A combination of the PCP Theorem [15, 3, 2] and Raz's Parallel Repetition Theorem [31] shows that it is hard to approximate the 2-Prover-1-Round Games problem.

Theorem 1.3. *Let $\delta > 0$ be any positive constant. Then for sufficiently large constants d and $|\Sigma_A| = d\,|\Sigma_B|$, given an instance $G = (A, B, E, \Pi, \Sigma_A, \Sigma_B)$ of a 2-Prover-1-Round Game, it is NP-hard to distinguish between*

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

The game G constructed in Theorem 1.3 has alphabet size which is polynomial in $1/\delta$, and it is a d-to-1 game for d which is also polynomial in $1/\delta$. Theorem 1.3 is used as a canonical hard problem from which numerous hardness of approximation results are proven by reduction, e. g., [1, 7, 17, 18, 16, 13]. However for some problems we do not know how to make similar reductions prove "satisfactory" hardness of approximation results. These include basic problems such as 2-SAT, Vertex Cover, and Max-Cut. At a technical level, the difficulty is that the instances of the 2-Prover-1-Round Games problem given by Theorem 1.3 involve d-to-1 constraints, where d blows up as the desired "soundness" δ approaches 0. It is conceivable that the theorem actually holds with $\delta \to 0$ while keeping d fixed, even with $d = 2$, or if one allows "imperfect completeness," then even with $d = 1$. These are precisely the Unique Games Conjecture and the d-to-1 Games Conjecture proposed in [21].

Conjecture 1.4 (Unique Games Conjecture)**.** For every constant $\delta > 0$ there is a sufficiently large constant $|\Sigma|$, such that 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 a $1 - \delta$ fraction of the constraints of G.
- NO case: no coloring satisfies more than a δ fraction of the constraints of G.

Conjecture 1.5 (d-to-1 Games Conjecture)**.** Let $d \geqslant 2$ be an integer. For every constant $\delta > 0$, for sufficiently large constant $|\Sigma_A| = d\,|\Sigma_B|$, given an instance $G = (A, B, E, \Pi, \Sigma_A, \Sigma_B)$ of a d-to-1-Game, it is NP-hard to distinguish between:

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

Roadmap to the proof of Conjecture 1.5 with imperfect completeness: The main contribution of the current paper is the proposal of a reduction from 3-Lin to 2-to-1 Games along with an analysis of it based on a combinatorial conjecture on the Grassmann graph. Ultimately, this led to the proof of Conjecture 1.5 with imperfect completeness and $d = 2$. Namely, in the variant where the YES case is replaced with the slightly weaker condition that there is a coloring satisfying a $1 - \delta$ fraction of the constraints. Thus, the current paper represents a step towards a proof of the 2-to-1 Games Conjecture, and to get some perspective we outline where the current paper fits in the overall puzzle.

1. First, the article [29] presented an outer PCP which is based on the "covering property," which is a crucial feature in all PCP constructions we discuss (we remark that their motivation was quite different, though).

2. Second, the article [26] presented a PCP construction which attempts to establish a weaker variant of Conjecture 1.5 tailored for several applications (such as hardness of approximating the minimum vertex cover in a graph). That article presented a reduction which is based on the Grassmann graph and a linearity tester on it, and proved the soundness of the reduction assuming some combinatorial conjecture on the Grassmann graph; we refer the reader to [26] for further discussion on that.

3. Third comes the current work. The goal here is to extend the ideas from [26] to get a candidate reduction to 2-to-1 Games with the standard notion of soundness, as in the statement of Conjecture 1.5. Indeed, this is the main contribution of the paper, and we show that the reduction from [26] can be modified to achieve this. Towards this end, we also use the Grassmann graph and a linearity tester on it, however we require a different combinatorial hypothesis on it as well as several more ideas to make the reduction go through. Specifically, we introduce the notion of zoom-outs and show how to incorporate it in the soundness analysis of the reduction.

4. Fourth, motivated by what seemed to be a "moral" connection between our combinatorial hypothesis and edge expansion properties in the Grassmann graph, the paper [9] investigated the structure of small sets in the Grassmann graph that have edge expansion bounded away from 1. In that paper, the authors managed to prove a structural characterization of small sets of vertices in the Grassmann graph that have edge expansion bounded away from 7/8. They conjectured that similar characterization should hold for all small sets with edge expansion bounded away from 1. Following [9], Barak Kothari and Steurer [6]

observed that the "moral" connection above between our combinatorial hypothesis and the characterization of small sets in the Grassmann graph with edge expansion bounded away from 1, is, in fact, not just moral: the two statements are equivalent. Thus, this meant that the proof of the 2-to-1 Games conjecture boils down to the expansion statement. In light of this, the result of [9] already gave some new point of hardness for 2-to-1 Games, and the result of [24] proved that the analogous expansion statement for the closely related Johnson graph holds.

5. Fifth, the expansion hypothesis of [9] was proven in full generality in [25], thereby completing the proof of Conjecture 1.5 with $d = 2$ and imperfect completeness.

1.2 Our main result

We build on the work in [26] and propose a reduction from 3-Lin to 2-to-1 games. We also propose a combinatorial hypothesis regarding the Grassmann agreement test (Test 1.7 below), which is based on the constraints of the Grassmann graph from [26]. We begin by recalling the definition of the Grassmann graph:

Definition 1.6. For a vector space $X = \mathbb{F}_2^n$ over $\mathbb{F}_2$ and a parameter $1 \leqslant \ell < n$, the Grassmann graph $Gr(X, \ell)$ is the graph whose set of vertices is the collection of all linear subspaces of X of dimension ℓ, and two vertices $L, L' \subseteq X$ are adjacent if $\mathsf{dim}(L \cap L') = \ell - 1$.

The Grassmann code is an error correcting code associated with the Grassmann graph. Given a vector space $X = \mathbb{F}_2^n$ and a linear function $f : X \to \mathbb{F}_2$, the Grassmann encoding of f is an assignment F over the vertices of $Gr(X, \ell)$, that to each subspace $L \subseteq X$ of dimension ℓ, assigns $F[L] = f|_L$. The Grassmann code could either be thought of as an extension of the classical plane encoding to high dimensional subspaces [32] (except that the field in our case is small), or as a tensor power of the Hadamard encoding.

We will use the Grassmann encoding in our reduction, and towards that end we require a local tester that, given oracle access to a table F assigning a linear function to each vertex in $Gr(X, \ell)$, checks if it is the table of restrictions of some global linear function f. Our tester will be the following natural consistency check:

Test 1.7. [Grassmann agreement test] Given a table F that assigns to each ℓ-dimensional space $L \in Gr(X, \ell)$ a linear function,

- Choose a random $(\ell - 1)$-dimensional space, $L' \subseteq \mathbb{F}_2^n$, and two ℓ-dimensional spaces $L_1, L_2 \supseteq L'$ independently.
- Accept if $F[L_1]|_{L'} = F[L_2]|_{L'}$

Let $\mathsf{agreement}(F)$ denote the success probability of F in this test, over the choice of L_1, L_2.

In words, the tester picks two subspaces L_1, L_2 that are adjacent in $Gr(X, \ell)$, and checks that their local functions are consistent on the intersection $L_1 \cap L_2$. The tester has several desirable features: first, it only makes 2 queries and has perfect completeness, in the sense that a valid table of restrictions F passes the test with probability 1. Second, it is a 2-to-2 test: fixing a pair of questions, L_1 and L_2, the assignments of L_1 and L_2 that make the test accept align in pairs. This is because each linear function on $L_1 \cap L_2$ has 2 linear extensions to each one of L_1 and L_2, and only these tuples would make the test accept.

The most challenging aspect of the tester is its soundness: suppose that a table F has $\mathsf{agreement}(F) \geqslant \delta$; what can we say about F? Our hypothesis roughly states that any such assignment F must be non-trivially consistent with some global function $f : X \to \mathbb{F}_2$, where $\delta > 0$ is an arbitrarily small constant independent of ℓ and $n = \mathsf{dim}(X)$. The hypothesis is stated formally as Hypothesis 3.6. Our main result is that the hypothesis implies soundness of our reduction, thus implying the 2-to-1 Games Conjecture, albeit with imperfect completeness: instead of being able to satisfy all constraints in the YES case, as in the original conjecture, our reduction generates an instance where not all constraints can be satisfied even in the YES case (although we can get as close to perfect completeness as we want). This is inherent in our construction due to the linearity in the overall reduction.

Theorem 1.8. *Assume Hypothesis 3.6. Then for every constant $\delta > 0$, for a sufficiently large constant $|\Sigma_A|$, given a 2-to-1-Game $G = (A, B, E, \Pi, \Sigma_A, \Sigma_B)$ it is NP-hard to distinguish between:*

- *YES case: there is a coloring satisfying a $1 - \delta$ fraction of the constraints of G. Moreover, one can remove a δ fraction of the vertices and all of the constraints adjacent to them, such that this coloring satisfies all of the remaining constraints.*[1]
- *NO case: no coloring satisfies more than a δ fraction of the constraints of G.*

We note that this theorem also leads to the same conclusions as in [26] regarding NP-hardness of approximating Gap Independent Set and Vertex Cover. Our reduction, as well as the reduction in [26], highlights the importance of agreement tests and can be considered as a motivation for studying their soundness behavior.

Remark 1.9. The Grassmann code and Test 1.7 could be thought of as an analog of the Johnson direct product encoding and the associated direct product tester. In that setting, one encodes a function $f \colon [n] \to \{0, 1\}$ by a table $F \colon \binom{[n]}{k} \to \{0, 1\}^k$ defined as $F[A] = f|_A$. The corresponding tester is the test that picks $B \subseteq [n]$ of size $k/2$, $A, A' \supseteq B$ of size k independently and then checks that $F[A]|_B = F[A']|_B$. This encoding has been well studied in some regimes of parameters, and in particular the soundness of the test is well understood (see for example [20]). The main downside of this encoding is that it is not 2-to-2 (in fact, it is very far from it, being $2^{k/2}$-to-$2^{k/2}$), and hence it is irrelevant for 2-to-1 Games as far as we know.

1.3 The reduction

In this section we sketch the reduction from 3-Lin to the 2-to-1 Games problem that proves Theorem 1.8. We omit some technical details, which are fully described in Section 4.2.

[1] As in [26], where this property is necessary towards applications to the inapproximability of the Vertex Cover and Independent Set problems.

1.3.1 Starting point: Håstad's 3-Lin

An instance of the 3-Lin problem is (X, Eq) where X is a set of n 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 fraction 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 [18] shows that for every positive constant ε, $\mathsf{Gap3Lin}(1-\varepsilon, \frac{1}{2}+\varepsilon)$ is NP-hard.

Theorem 1.8 is proved via a "PCP reduction" from the 3-Lin problem to the 2-to-1 Games problem. The reduction follows a standard framework of composition of an outer and an inner game. The outer game is a standard (smooth) "clause-vs.-variable" parallel repetition of the 3-Lin instance, and the inner game is based on the Grassmann graph and the Grassmann encoding. Below we elaborate on each one of the components.

1.3.2 The outer PCP

We first take the active, 2-prover-1-round game on the instance (X, Eq) of 3-Lin. Namely, we consider the equation versus variable game G wherein the verifier picks an equation $x_1 + x_2 + x_3 = b$ from Eq uniformly, sends the entire equation to the first prover, and a uniformly chosen variable from it to the second prover. It can be shown that if (X, Eq) is $1-\varepsilon$ satisfiable, then there are strategies for the provers that win with probability $1-\varepsilon$, and if (X, Eq) is at most $1/2+\varepsilon$ satisfiable, then no provers' strategy wins with probability exceeding $5/6+\varepsilon$. Our outer PCP game is a smooth parallel repetition of the equation versus variable game. Specifically, taking $k \in \mathbb{N}$ a large parameter and $\beta \in (0,1)$ small (thought of as $\beta = \log\log k/k$, say), the verifier proceeds as follows:

1. The verifier samples equations $e_1, \ldots, e_k$ uniformly from (X, Eq), and lets $U_1, \ldots, U_k$ be the sets of variables in them respectively.
2. For each $i = 1, \ldots, k$ independently, the verifier takes $V_i = U_i$ with probability $1-\beta$, and otherwise the verifier takes $V_i \subseteq U_i$ of size 1 uniformly.
3. The verifier sends $U = U_1 \cup \ldots \cup U_k$ to the first prover and $V = V_1 \cup \ldots \cup V_k$ to the second prover, and expects to get from them assignments $a \in \{0,1\}^U$ and $a' \in \{0,1\}^V$ to their respective sets of variables.
4. The verifier accepts if a satisfies all of the equations $e_1, \ldots, e_k$ and additionally a and a' agree on all of the variables from V. We refer to the former checks as the side conditions, and to the latter checks as consistency checks.

We denote this game by $G_{\mathsf{smooth},k,\beta}$. It is easy to see that if (X, Eq) is $1-\varepsilon$ satisfiable, then the provers can win the game $G_{\mathsf{smooth},k,\beta}$ with probability at least $1-k\varepsilon$. Using the parallel repetition theorem [31, 19], one can show that if (X, Eq) is at most $1/2+\varepsilon$ satisfiable, then no provers' strategy wins the game with probability exceeding $2^{-\Omega_\varepsilon(\beta k)}$; see Lemma 5.4.

1.3.3 The inner PCP

The inner game relies on the Grassmann encoding which was introduced in [26]. The goal here is to check the consistency of the outer PCP game via 2-to-1 constraints, and to do that we must encode the answers of the provers in a different way.

Recall that an answer of the first prover is a string $s \in \{0,1\}^U$, where U is a subset of $3k$ variables from (X, Eq). To encode this string, the prover thinks of the Hadamard encodings, and towards this end considers the space

$$X_U = \left\{x \in \mathbb{F}_2^n \mid x_i = 0 \; \forall i \notin U\right\}. \tag{1.1}$$

A given assignment for the variables of U, namely $s \in \{0,1\}^U$, naturally corresponds to a linear function $f : X_U \to \mathbb{F}_2$ defined as $f(x) = \langle s, x\rangle$, and thus encoding the string s is equivalent to encoding the function f. To encode the function f, we take a parameter ℓ thought of as large (but much smaller than k), and for each ℓ-dimensional subspace $L \subseteq X_U$ we define $F[L] = f|_L$. The table F is the table of restrictions of f to all ℓ-dimensional subspaces, and it is the Grassmann encoding of the function f.

As is often the case in PCP constructions, the table F we described above is the intended type of assignments we wish to have, however a cheating prover may choose an arbitrary assignment F altogether. To remedy this situation we have to introduce constraints/checks on the inner PCP game, but before that we briefly discuss the folding and side conditions. We want to make sure that the first prover's answer encodes an assignment for U that satisfies the equations $e_1, \ldots, e_k$. We do this by identifying some spaces L in the Grassmann encoding in a way that forces $e_1, \ldots, e_k$ to be satisfied. We define

$$H_U = \mathsf{Span}\left\{x_e \mid e \subseteq U\right\},$$

where $x_e \in \mathbb{F}_2^n$ denotes the vector with 1's in the three entries that correspond to variables of e, and 0 elsewhere. We shall identify a pair of subspaces $L_1, L_2 \in Gr(X_U, \ell)$ if $L_1 + H_U = L_2 + H_U$.[2] This makes sense because in the case of an honest prover that answers according to an assignment that satisfies all equations of U, knowing the value of f on L_1 is already enough to deduce it on $L_1 + H_U$ and therefore on L_2, and vice versa. Thus, the number of possible assignments on $L \oplus H_U$ satisfying the side conditions is still 2^ℓ, and in this way we managed to enforce the satisfaction of the side conditions.

We now discuss the constraints of the inner PCP game. To get 2-to-1 type constraints (as opposed to 2-to-2 type constraints), on top of the table F that assigns to each space $L + H_U$ a linear function satisfying the side conditions, we also require from the prover a table F' that assigns to each $(\ell - 1)$-dimensional space $L' \subseteq X_U$ a linear function. With the tables F and F' in hand, the constraints in the inner PCP game proceed as follows: sample $L' \subseteq L \subseteq X_U$ uniformly of dimensions $\ell - 1$ and ℓ respectively, query $F[L + H_U]$ and $F'[L']$, and check that $F[L + H_U]|_{L'} = F'[L']$.

It is clear that the intended tables F and F' satisfy all of the constraints of the inner PCP game (being the restrictions of the same linear function f satisfying the side conditions). The

[2] Formally speaking, we move to $Gr(X_U/H_U, \ell)$ instead of $Gr(X_U, \ell)$.

soundness of the inner PCP game, namely the question of what tables F and F' pass the test with non-negligible probability, is much more complicated, and in this paper we make an hypothesis about it. We discuss that hypothesis below, and refer to Hypothesis 3.6 for a formal statement. As discussed earlier, subsequent works proved that Hypothesis 3.6 is correct.

1.3.4 The composition

We now discuss the composition of the inner PCP and outer PCP game. In the combinatorial view, the composition amounts to replacing each question U of the first prover with a copy of the inner PCP game, and then making appropriate version of the constraints of the inner PCP game across these blocks; a formal description of this combinatorial view can be found in Section 4.2. In the 2-prover-1-round game view, which is easier to describe in this introductory section, this composition looks as follows:

1. The verifier picks at random k equations $\{e_1, \ldots, e_k\}$, lets U be the set of $3k$ variables that appear in these equations.

2. 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.

3. The verifier picks an $(\ell - 1)$-dimensional subspace $L' \subseteq X_V$ and an ℓ-dimensional subspace L such that $L' \subseteq L \subseteq X_U$ (the spaces X_U, X_V are as in Eq. (1.1)) and such that $L \cap H_U = \{0\}$.[3]

   The verifier sends (V, L') to the second prover and $(U, L + H_U)$ to the first prover.

4. The first prover answers with a linear function $g_1 : L + H_U \to \mathbb{F}_2$ and the second prover answers with a linear function $g_2 : L' \to \mathbb{F}_2$. The verifier accepts if

$$(g_1)|_{L'} = g_2 \quad \text{and} \quad g_1(e_i) = b_i, \ \forall i = 1, \ldots, k$$

   where b_i is the right-hand side of the equation e_i.

Let us see that the game is indeed a 2-to-1 game. The answer g_1 of the first prover surely determines a unique valid answer g_2 for the second prover. On the other hand, the second prover's answer g_2 extends uniquely to a function $\tilde{g}_2 : L' + H_U \to \mathbb{F}_2$ by setting $\tilde{g}_2(x_{e_i}) = b_i$ for all $i \in [k]$. From there it should be clear that there are only two possible linear functions on $L + H_U \supseteq L' + H_U$ that remain valid for the first prover.

Folding across blocks. The two-player game as described above is likely to be both complete and sound, but our analysis is facilitated by some additional folding. The folding amounts to identifying all of the possible questions to the first prover whose answers, if the prover is honest, determine each other. This does not hurt the completeness and makes it tougher for the provers to cheat, so helps the proof of soundness.

[3] This can be ignored as it happens with overwhelmingly large probability.

More explicitly, we define an equivalence relationship: $(L_1, U_1) \sim (L_2, U_2)$ if $L_1 + H_{U_1} + H_{U_2} = L_2 + H_{U_1} + H_{U_2}$. One can directly see that a linear function for L_1 implies, assuming all of the equations in $U_1 \cup U_2$ hold, the value of the linear function on L_2, and vice versa. Thus, the assignment to a single subspace L_1 implies the assignment to all subspaces L_2 such that $L_1 + H_{U_1} + H_{U_2} = L_2 + H_{U_1} + H_{U_2}$. This motivates us to take the equivalence classes of the relation $\sim$ as the questions to the provers (instead of individual subspaces). We show that this transformation does not hurt the completeness of the PCP construction. This identification is helpful for us in the soundness analysis, since now a strategy of a prover F can be thought of as a collection of assignments to the vertices $L_1 + H_{U_1}$ which is fully consistent inside equivalence classes.[4]

1.4 Agreement tests and our hypothesis

Above, we described our reduction from 3-Lin to 2-to-1 Games. The completeness of this reduction is easy to verify: if the original 3-Lin system is $1 - \varepsilon$ satisfiable, then in the resulting instance of 2-to-1 Games the provers can win with probability at least $1 - k\varepsilon$. The more interesting (and difficult to establish) feature is the soundness of the reduction, namely showing that if (X, Eq) is at most $1/2 + \varepsilon$, then the provers' can win the resulting 2-to-1 Game only with small probability.

The heart of the soundness analysis lies in step 4 of the game constructed by the reduction described above. In that step, the verifier tests the agreement between the assignment one prover gives to a linear space and the assignment the other prover gives to its subspace. We show that when analyzing the soundness of the reduction, the analysis of this test follows provided one has good understanding of the properties of the Grassmann Agreement Test, Test 1.7. Ideally, we would want to deduce that if an assignment F passes Test 1.7 with noticeable probability, then it must be correlated with a global linear function from the success of the Grassmann.

This idealistic hope turns out to be incorrect, and instead one must consider a more convoluted notion of global structure for assignments F that pass Test 1.7 with noticeable probability. This is the content of our hypothesis that is introduced below. As a result of this more complicated notion of global structure, the soundness analysis of the reduction does not follow from it automatically, and indeed we need several more ideas to make the analysis go through. We also discuss these ideas below.

Some background. The Grassmann encoding and the probabilistic test for it that we described above fall within a more general framework of *agreement* tests. In an agreement test, there is a domain X and a collection $\mathcal{S} = \{S \subseteq X\}$ of subsets of X. A function $f : X \to \Sigma$ is encoded by writing down $f|_S$, its restriction to the subset S, for every subset $S \in \mathcal{S}$.

A supposed encoding of f is given by a table $F[\cdot]$. Here $F[\cdot]$ is a table that assigns, to every subset $S \in \mathcal{S}$, a partial function $F[S]$ on it. The intention is that $F[S] = f|_S$ for all $S \in \mathcal{S}$ and for

[4] This issue also arises in [26], where it is solved via a property called "transitivity" of a game, and use it to do a soft form of folding in the soundness analysis. Our solution here also uses a form of "transitivity" of the construction, however it is simpler and goes more along the lines of traditional folding in the PCP literature.

some global function $f : X \to \Sigma$. This encoding is clearly redundant and comes with a natural agreement test: choose two intersecting subsets $S_1 \cap S_2 \neq \emptyset$, and check that $F[S_1]$ agrees with $F[S_2]$ on all points $x \in S_1 \cap S_2$.

In the Grassmann case, the domain X is a vector space, $X = \mathbb{F}_2^n$, and $\mathcal{S}$ is the collection of all possible linear subspaces of X of dimensions ℓ. When $\ell = 2, 3$ and the field is $\mathbb{F}_q$ for larger q rather than $\mathbb{F}_2$ this is almost[5] exactly the lines-table or the planes-table representation of f used in classical PCP constructions.

What kinds of agreement tests have been analyzed before? The two types of agreement tests that have been studied are where the collection $\mathcal{S}$ consists of all subspaces of a certain dimension (see [32, 4, 20]) or where $\mathcal{S}$ consists of all possible k-element sets (see [12, 8, 20, 14]). In all prior cases, the agreement test compares values of two subsets that have a large intersection but also have a large disjoint part. This seemed to be important for the "expansion" of the test that helps the analysis. However, following [26], we consider an agreement test whose intersection between the two queries is almost maximal, as described above in Test 1.7. This very large overlap is important for making the constraints of our test gain the desired property of being 2-to-1 (indeed, if we considered subspaces of dimension ℓ that intersect in dimension $\ell - s$, then the resulting test would be 2^s-to-1).

Zoom-ins and zoom-outs. Let us make two easy observations. First, note that Test 1.7 is 2-to-2 : every value for $F[L_1]$ allows only two possible values for $F[L_2]$ and vice versa. Next, observe that if $F[\cdot]$ was an honest table, assigning each L a function $f|_L$ for some linear function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, then the test accepts with probability 1.

What is the "soundness guarantee" of the test? One is tempted to speculate that if the test passes with probability δ, then[6] the given table $F[\cdot]$ has a "good" consistency with some global linear function f, in the sense that $F[L] = f|_L$ for δ' fraction of the ℓ-dimensional subspaces L, for some δ' depending on δ. The linear function f would then serve as a "decoding" of the given table $F[\cdot]$ and one could even "list-decode," that is make a list of all linear functions f that have a good consistency with $F[\cdot]$, along with an upper bound on the list-size that depends (only) on δ.

The speculation, however, turns out to be false. A counterexample is presented in Section 3.1. In [26], the authors propose to circumvent this counterexample using the idea of a "zoom-in": speculating that while there may not be a global function that agrees with a constant fraction out of the ℓ-dimensional subspaces, there might be one global f that agrees with a constant fraction of the ℓ-dimensional subspaces that contain a certain subspace Q of dimension $q < \ell$. But in [26] a limited context was considered, where the acceptance criterion of the test is more restrictive and non-standard (in terms of the so-called (j, δ)-consistency).

In the present context it turns out that zoom-in alone does not suffice; a counterexample to this effect is presented in Section 3.2. We therefore introduce another idea that we call

[5] A minor difference is that we are considering linear subspaces and not affine subspaces, but this is an unimportant design choice.

[6] Here δ is thought of as a constant, ℓ as a sufficiently large integer after choosing δ, and the global dimension n as a sufficiently large integer after choosing ℓ.

"zoom-out" and propose that a combination of both zoom-in and zoom-out is sufficient to derive a reasonable conclusion. Our main combinatorial hypothesis is that (a strengthening is stated later):

Combinatorial Hypothesis (Informal): There are integers $q, r \geqslant 0$ and $\delta' > 0$ depending only on $\delta > 0$ such that the following holds. Given a table $F[\cdot]$ such that $\mathsf{agreement}(F) \geqslant \delta$, that is for L_1 and L_2 chosen as in Test 1.7,

$$\Pr_{L_1, L_2} [\ F[L_1]|_{L_1 \cap L_2} = F[L_2]|_{L_1 \cap L_2}\ ] \geqslant \delta\,,$$

there exists a subspace Q of dimension q (the "zoom-in" space), a subspace $W \supseteq Q$ of codimension r (the "zoom-out" space), and a linear function $f : W \to \{0, 1\}$ on W, such that $F[L] = f|_L$ for at least δ' fraction of ℓ-dimensional subspaces L such that $Q \subseteq L \subseteq W$.

In short, we propose that the speculation above holds if one restricts to subspaces L such that $Q \subseteq L \subseteq W$ for some "successful" choice of subspaces Q, W that have constant dimension and codimension, respectively.

1.4.1 Using the hypothesis in the soundness analysis

The hypothesis is needed for the soundness analysis of our reduction, where we assume that two provers manage to succeed in the game with high probability and deduce that they can also win in another related "outer PCP" game. To do that, it is not sufficient that *there exists* a successful pair of zoom-in and zoom-out spaces Q and W. The players need to be able to coordinate a successful pair.

Luckily it turns out that our hypothesis does ensure the existence of many zoom pairs. In the statement of the following corollary of the hypothesis the wording in italics is new:

Corollary of Combinatorial Hypothesis (Informal): There are integers $q, r \geqslant 0$ and $\delta' > 0$ depending only on $\delta > 0$ such that the following holds. Given a table $F[\cdot]$ with $\mathsf{agreement}(F) \geqslant \delta$, *for $\alpha(\ell)$ fraction of subspaces Q of dimension q, there exists a subspace $W \supseteq Q$ of codimension r, and a linear function $f : W \to \{0, 1\}$ on W*, such that $F[L] = f|_L$ for at least δ' fraction of ℓ-dimensional subspaces L such that $Q \subseteq L \subseteq W$.

Here $\alpha(\ell)$ is an arbitrary function of ℓ (and δ) and is independent of the global dimension n. We emphasize the quantifiers on Q and W. One is tempted to say "*for $\alpha(\ell)$ fraction of subspaces Q of dimension q and $\alpha(\ell)$ fraction of subspaces $W \supseteq Q$ of codimension r*," but this is false as shown by the counterexample in Section 3.2. This brings us to the main technical challenge that we must overcome to make the reduction go through: how can two non-communicating provers coordinate a zoom-out W which is good for both of them?

1.4.2 Coordinating a good zoom-out

Let us return to the coordination between the analysis at the Inner and Outer PCP levels. Since a constant fraction $\alpha(\ell)$ of the zoom-in spaces Q are "successful," the verifier in the

2-Prover-1-Round Game at the Outer PCP level, can simply send Q as "shared advice" to both the provers and the hypothesis states that the advice is successful with probability $\alpha(\ell)$. This is the way in which the zoom-ins are handled in [26] and we treat zoom-ins in the same way here. Handling zoom-outs is more difficult and is our main technical contribution. In the following, let us fix the zoom-in space Q and let the zoom-out space W contain Q.

To handle the zoom-outs, each prover makes a "list" of all successful zoom-outs $W_1, \ldots, W_M$ from her/his own viewpoint, selects one of these zoom-outs at random, and then hopes to agree with the other prover on a common successful zoom-out. This is because if the provers choose different zoom-outs, the resulting linear functions may have nothing to do with each other and will most likely be inconsistent. For zoom-out coordination to work, firstly, the lists needs to be "short," and secondly, there needs to be a zoom-out space W that is successful for both the provers simultaneously (and hence appears in the lists for both). The latter issue involves delving into the specifics of the PCP composition and is aided by the covering property of the Outer PCP and folding. The former issue, namely upper bounding the list size, will only work as long as the list size is independent of the global dimension n. Naively, there is no such upper bound, since it could be that *every* zoom-out space W of codimension r is successful. However, we circumvent this difficulty by showing that if too many zoom-out subspaces are successful, then there is a larger subspace, of smaller codimension, that is also successful. Thus, we bound the number of maximal successful subspaces.

To be more precise, we prove a statement along the following lines. Suppose that $(W_1, \ldots, W_M)$ are zoom-outs of codimension r that are all successful, in the sense that for each i there is a linear function f_i such that $F[L] = f_i|_L$ for at least a δ fraction of the subspaces L containing Q that are contained in W_i. If M is a large enough function of $\ell, r, 1/\delta$, then there must be a zoom-out W of codimension at most $r-1$ which is also successful. Namely, there is a zoom-out $W \supseteq Q$ of codimension at most $r-1$ containing at least one of the W_i's, and a linear function $f\colon W \to \mathbb{F}_2$, such that $F[L] = f|_L$ for at least $\delta' = \delta'(\delta) > 0$ fraction of L's containing Q that are contained in W. This statement means that, for the appropriate notion of maximality, there are not too many successful maximal zoom-outs (as otherwise, we could take $W_1, \ldots, W_m$ to be all maximal successful zoom-outs and get a contradiction to the above statement).

1.5 Applications

Our main result implies the first gap for Unique Games with constant completeness and vanishing soundness. It is more convenient to derive it from hardness of 2-to-2 games that we now define.

Definition 1.10. A 2-to-2 game instance $G = (U, E, \Sigma, \Pi)$ is given by a graph (U, E), an alphabet Σ and a set of 2-to-2 constraints Π, one constraint for each edge. A constraint π is 2-to-2 if there are 2 partitions $A_1, \ldots, A_k$, $B_1, \ldots, B_k$ of Σ into sets of size 2 (where $k = \frac{1}{2}|\Sigma|$), such that $\pi = \bigcup_{i=1}^{k} A_i \times B_i$.

Our techniques imply the following:

Theorem 1.11. *Assuming Hypothesis 3.6, for every $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that given a 2-to-2 games instance G with alphabet size k it is NP-hard to distinguish between:*

- *YES case: there is an assignment to G satisfying at least a $1 - \varepsilon$ fraction of the constraints.*
- *NO case: every assignment to G satisfies at most an ε fraction of the constraints.*

This theorem implies the following result, whose proof is given in Section A.1.

Corollary 1.12. *Assuming Hypothesis 3.6, for every $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that given a Unique Games instance G of alphabet size k it is NP-hard to distinguish between:*

- *YES case: there is an assignment to G satisfying at least a $\frac{1}{2}$ fraction of the constraints.*
- *NO case: every assignment to G satisfies at most an ε fraction of the constraints.*

We note that many of the known algorithms for Unique Games work well also for 2-to-1 games or for Unique Games with completeness $\frac{1}{2} - \varepsilon$. Therefore, our results can be viewed as evidence towards the truth of the Unique Games Conjecture (Conjecture 1.4). Namely, if one is willing to believe Hypothesis 3.6, new algorithmic ideas must be used in order to refute the Unique Games Conjecture.

Our next application is hardness for Max-Cut-Gain and is based on the reduction of Khot and O'Donnell from [28]. The proof is deferred to Section A.2.

Corollary 1.13. *Assuming Hypothesis 3.6, there exists a constant $c_1 > 0$ such that for every $\varepsilon > 0$, given a graph G it is NP-hard to distinguish between:*

- *YES case: there is a cut in G containing at least a $\frac{1}{2} + \varepsilon$ fraction of the edges.*
- *NO case: every cut in G contains at most a $\frac{1}{2} + c_1 \frac{\varepsilon}{\log(1/\varepsilon)}$ fraction of the edges.*

Our next application is concerned with coloring problems.

Definition 1.14. A graph $G = (U, E)$ is said to be (k, ε) almost colorable if there is $U' \subseteq U$ of size at least $(1 - \varepsilon)|U|$ such that the induced subgraph on U' is k-colorable.

Using the techniques of [11], one can prove the following.

Corollary 1.15. *Assuming Hypothesis 3.6, for every $\varepsilon, \delta > 0$, given a graph G it is NP-hard to distinguish between:*

- *YES case: G is $(4, \varepsilon)$ almost colorable.*
- *NO case: G does not have independent set of fractional size δ.*

The proof follows [11] closely, and hence we omit the details.

Unconditional Results

With the current state of knowledge, this approach already yields non-trivial results for Hypothesis 3.6, and subsequently for 2-to-1 games, 2-to-2 games and Unique Games that we outline below. Using the expansion theorem of [10] and the reduction of [6], one has the following result.

Theorem 1.16. *Hypothesis 3.6 holds for every $\delta > 1/4$.*

Using this along with simple adaptations of the reduction presented in this paper, one has the following results.

Theorem 1.17. *For every $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that given a 2-to-1 game G with alphabet size k, it is NP-hard to distinguish between:*

- *YES case: there is an assignment satisfying at least a $1 - \varepsilon$ fraction of the constraints of G.*
- *NO case: every assignment satisfies at most a $\frac{1}{2} + \varepsilon$ fraction of the constraints of G.*

This result should be compared to the best currently known explicit gaps for 2-to-1 games due to [5], who proved it is NP-hard to distinguish between satisfiable 2-to-1 games instances and $\left(\frac{23}{24} + \varepsilon\right)$-satisfiable instances.

Theorem 1.18. *For every $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that given a 2-to-2 game G with alphabet size k, it is NP-hard to distinguish between:*

- *YES case: there is an assignment satisfying at least a $1 - \varepsilon$ fraction of the constraints of G.*
- *NO case: every assignment satisfies at most a $\frac{1}{4} + \varepsilon$ fraction of the constraints of G.*

Using the same reduction as in the proof of Corollary 1.12, one has the following corollary.

Corollary 1.19. *For every $\varepsilon > 0$ there exists $k \in \mathbb{N}$ such that given a Unique Games instance G with alphabet size k, it is NP-hard to distinguish between:*

- *YES case: there is an assignment satisfying at least a $\frac{1}{2}$ fraction of the constraints of G.*
- *NO case: every assignment satisfies at most a $\frac{1}{4} + \varepsilon$ fraction of the constraints of G.*

This result improves the soundness of the result of O'Donnell and Wright [30] who proved it is NP-hard to distinguish $\frac{1}{2}$-satisfiable Unique Games instances from $\left(\frac{3}{8} + \varepsilon\right)$-satisfiable instances.

2 Preliminaries

2.1 Linear subspaces

Let X be an n-dimensional vector space. For two subspaces $L_1, L_2 \subseteq X$ we denote by $L_1 + L_2 = \mathsf{Span}(L_1, L_2) = \{x_1 + x_2 \mid x_i \in L_i\}$. Similarly, for a vector v and a subspace L let $v + L$ be the subspace spanned by v and L.

Definition 2.1 (Side condition)**.** A side condition is a pair (H, h) where $H \subseteq X$ and $h : H \to \mathbb{F}_2$ is a linear function. For a subspace $H \subseteq Y \subseteq X$, a linear function $f : Y \to \mathbb{F}_2$ is said to respect the side condition (H, h) if $f|_H = h$. We will assume that $\mathsf{dim}(H) \leqslant \frac{n}{2}$.

Intuitively, the side conditions should be thought of as a collection of pre-determined values a linear function should have. As explained in the introduction, side conditions naturally arise as equations in a 3-Lin instance chosen during the reduction.

Definition 2.2. Let $L, H \subseteq X$ be subspaces. If $L \cap H = \{0\}$ then any linear function $f : L \to \mathbb{F}_2$ extends uniquely to a linear function $\tilde{f} : L + H \to \mathbb{F}_2$ that respects the side condition (H, h), called the (H, h)-extension of f, and defined by

$$\tilde{f}(z) = \tilde{f}(x + y) := f(x) + h(y)$$

where $z = x + y$ is the unique way to write $z \in L + H$ as a sum of $x \in L$ and $y \in H$.

2.2 The Grassmann graph over $\mathbb{F}_2$

For a vector space $X = \mathbb{F}_2^n$ and a non-zero integer $\ell < n$ we denote by $Gr(X, \ell)$ the collection of all linear subspaces of X of dimension ℓ. The set $Gr(X, \ell)$ is called the Grassmann graph, where we connect two subspaces $L_1, L_2 \in Gr(X, \ell)$ if their intersection is of dimension at least $(\ell - 1)$. A table $F[\cdot]$ is an *assignment* for $Gr(X, \ell)$ if it assigns a linear function on L to each $L \in Gr(X, \ell)$, namely $F[L] : L \to \mathbb{F}_2$.

We repeat the definition of the agreement of a table F, which measures the success probability of the Grassmann Agreement Test on it. It is just

$$\mathsf{agreement}(F) := \Pr_{L_1, L_2} [\, F[L_1]|_{L'} = F[L_2]|_{L'} \,]$$

where $L' = L_1 \cap L_2$ is chosen uniformly from $Gr(X, \ell - 1)$ and then L_1, L_2 are chosen independently from $Gr(X, \ell)$ conditioned on containing L'. Note that this random selection process induces a distribution on the edges of the Grassmann graph, making it an edge-weighted graph.

Given a table F we can also consider other agreement parameters. For any $t < \ell$, let $\mathcal{D}_t$ be the distribution that selects a random subspace L' of dimension t and then two subspaces $L_1, L_2 \supseteq L'$ independently and let

$$\mathsf{agreement}_t(F) := \Pr_{(L_1, L_2) \sim \mathcal{D}_t} [F[L_1]_{L'} = F[L_2]_{L'}] \,. \tag{2.1}$$

So $\mathsf{agreement}(F)$ is shorthand for $\mathsf{agreement}_{\ell-1}(F)$.

2.3 Quotient vector space

For a d-dimensional subspace $Q \subseteq X$, the quotient space X/Q is a vector space whose elements are all subspaces $v + Q$. It is easy to check that this is a $(\mathsf{dim}(X) - \mathsf{dim}(Q))$-dimensional vector space. The zero element is the d-dimensional space $0 + Q = Q$ and all non-zero elements are $(d + 1)$-dimensional spaces $v + Q$, where $v \notin Q$.

For a given q-dimensional subspace $Q \subseteq X$ there is a canonical mapping $\varphi : X \to X/Q$ sending $x \in X$ to $x + Q$. The mapping extends to subspaces by mapping the points $x \in L$ to the subspace $\{x + Q \mid x \in L\} = L + Q$. This naturally partitions the spaces in $Gr(X, \ell)$ into equivalence classes where the class of $K \in Gr(X/Q, \ell - q)$ consists of all spaces L such that $L + Q = K$. Moreover,

Claim 2.3. *Let $Q \subseteq X$ be a q-dimensional space, $q < \ell$, and let $\mathcal{L}_Q = \{L \in Gr(X, \ell) \mid L \supseteq Q\}$. There is a bijection $\varphi : Gr(X/Q, \ell - q) \to \mathcal{L}_Q$.* □

Proof. Every element in $Gr(X/Q, \ell - q)$ is by definition, a subspace that is spanned by $\ell - q$ elements of X/Q, e. g., $v_1 + Q, v_2 + Q, \dots, v_{\ell-q} + Q$. We map it to $L := \mathsf{Span}(v_1, \dots, v_{\ell-q}, Q) \subseteq X$. Clearly $\mathsf{dim}(L) = \ell$ and $L \supseteq Q$. The mapping is clearly independent of the choice of basis, and is injective. To see that it is onto, for every subspace $L \supseteq Q$ of dimension ℓ let us choose some basis $b_1, \dots, b_q, v_1, \dots, v_{\ell-q}$ so that $Q = \mathsf{Span}(b_1, \dots, b_q)$. We map it to the space spanned by $v_1 + Q, \dots, v_{\ell-q} + Q$ that belongs to $Gr(X/Q, \ell - q)$. This is indeed true because any linear dependence among the $v_i + Q$ would translate to a linear dependence in $L + Q \subseteq X$. □

We will later have to focus on a certain subset of the spaces in $Gr(X, \ell)$, which we call a zoom, and is as follows. Let $Q \subseteq W \subseteq X$ be subspaces, such that $\mathsf{dim}(Q) \leqslant \ell \leqslant \mathsf{dim}(W)$. Define

$$\mathsf{Zoom}[Q, W] = \{L \in Gr(X, \ell) \mid Q \subseteq L \subseteq W\} .$$

It follows from the above that

Claim 2.4. *There is a bijection between $\mathsf{Zoom}[Q, W]$ and $Gr(W/Q, \ell)$.*

Claim 2.5. *Let $f : X \to \mathbb{F}_2$ be a linear function such that for each $x \in Q$, $f(x) = 0$. Then there is a unique function $\tilde{f} : X/Q \to \mathbb{F}_2$ so that*

$$\forall x \in X, \quad \tilde{f}(x + Q) = f(x) .$$

Proof. For each $y \in X/Q$ choose some $x \in X$ such that $y = x + Q$ and define $\tilde{f}(y) = f(x)$. The definition doesn't depend on the choice of x because if $x_1 + Q = y = x_2 + Q$ then $x_1 + x_2 \in Q$, so by linearity of f, $0 = f(x_1 + x_2) = f(x_1) + f(x_2)$. □

3 The Grassmann agreement test

In this section we describe the hypothesis regarding the Grassmann agreement test, together with some examples which motivate it.

Let X be an n-dimensional vector space over $\mathbb{F}_2$, let $1 \leqslant \ell < n$, and let $F[\cdot]$ be an assignment over $Gr(X, \ell)$. We now must introduce several parameters: we have parameters, δ, δ', q, r and C that are all regarded as constants: δ will be the probability that the assignment F passes the test, and q, r, C, δ' will be parameters from our hypothesis that are related to the global structure of F. There is ℓ, the dimension of subspaces that is thought of as a large enough integer given the previous parameters, and n, the dimension of X, should be sufficiently large given the other constants and ℓ.

Given a linear function $f : X \to \mathbb{F}_2$, the construction leads to an encoding of f that writes down for each $L \in Gr(X, \ell)$ the restriction $f|_L$ of f to L. In other words, f is encoded by a table F such that $F[L] = f|_L$ for each L. Since $f|_L$ is a linear function on L and there are precisely 2^ℓ linear functions on L, one can describe $f|_L$ using a symbol from the alphabet $[2^\ell] = \{1, \dots, 2^\ell\}$. This encoding scheme $f \mapsto F$ has a relative distance $\approx 1 - 2^{-\ell}$. The Grassmann agreement test (Test 1.7) is a natural attempt to check whether a supposed encoding is indeed a valid one.

Let us start with two easy but important observations

- The test is clearly 2-to-2: for any assignment for L_1, which is a linear function $f_1 : L_1 \to \mathbb{F}_2$, let $f' = (f_1)|_{L'}$ be its restriction to $L' = L_1 \cap L_2$. Let $f_2 : L_2 \to \mathbb{F}_2$ be the assignment for L_2. For the test to accept, it must be that $f_2|_{L'} = f'$. Since this determines f_2 on a subspace of codimension 1, there are exactly two possible values for f_2 that cause the test to accept.

- The test has "perfect completeness," namely if $F[\cdot]$ is a valid encoding of some (global) linear function $f : X \to \mathbb{F}_2$, $F[L] = f|_L$ for all L, then the test accepts with probability 1.

A natural inverse question arises from the "soundness" point of view: given a table $F[\cdot]$ such that $\mathsf{agreement}(F) \geqslant \delta$ for some constant $\delta > 0$, is it necessarily the case that $F[\cdot]$ is "globally consistent" with some linear function $f : X \to \mathbb{F}_2$? Indeed, we quote below a nice (and useful) inverse theorem for an agreement test with smaller intersection size, namely when $\mathsf{dim}(L_1 \cap L_2) = \ell/10$, which is analogous to the "line versus point" and "plane versus plane" low-degree tests studied in the literature [33, 4, 32]. Recall from Eq. (2.1) that $\mathsf{agreement}_t(F)$ is the probability that two subspaces that intersect on a t-dimensional space agree on their intersection.

Lemma 3.1 (Agreement test with intersection size $t = \ell/10$)**.** *Let $F[\cdot]$ be an assignment for $Gr(X, \ell)$, and let $t = \ell/10$. If $\mathsf{agreement}_t(F) \geqslant \delta$ then there is a global linear function $g : X \to \mathbb{F}_2$ such that* $\Pr_L [F[L] = g|_L] \geqslant \delta' = \frac{\delta^3}{300}$.

This lemma was proven in [26, Theorem D.1] and earlier (for assignments that are not necessarily linear) in [20]. Here is a formal derivation of this lemma from [26], who considered the very similar ℓ-space versus t-space agreement test,

Proof. We can define another table $A[\cdot]$ that assigns, to every t-dimensional subspace B, a linear function $A[B]$ on it, by letting $A[B] = F[L]|_B$ for a randomly chosen ℓ-dimensional subspace L containing B. It follows that, for a random pair $B \subseteq L$ of t-dimensional and ℓ-dimensional subspaces respectively, $A[B]$ and $F[L]$ are consistent on B with probability at least δ. In other words, the tables $A[\cdot], F[\cdot]$ pass the "ℓ-space versus t-space" test, as defined in [26, Section D], with probability at least δ. By [26, Theorem D.1], there is a global linear function $g : U \to \{0, 1\}$ that agrees with $F[\cdot]$ on $\frac{\delta^3}{300}$ fraction of ℓ-dimensional subspaces. □

In analogy, one is tempted to speculate that a similar theorem holds also for our test,

Speculation 3.2 (False)**.** For every $\delta > 0$, there exists a $\delta' > 0$, such that for any assignment F for $Gr(X, \ell)$, if $\mathsf{agreement}(F) > \delta$ then there exists a global linear function $f : X \to \mathbb{F}_2$, such that $F[L] = f|_L$ for δ' fraction of the subspaces $L \in Gr(X, \ell)$.

It turns out that the speculation above is false, as shown by the examples below. We propose to "salvage" the speculation using the idea of "zoom-in" and "zoom-out," leading to our main combinatorial hypothesis.

3.1 Zooming-in

First counterexample

Here is a counterexample to Speculation 3.2.

For each $x \in X$ let $f_x : X \to \mathbb{F}_2$ be a distinct linear function. Randomly order the elements of $X - \{0\}$, and for each L assign $F[L] = f_x|_L$ where $x \in L$ is the smallest (according to this order) element in L.

Clearly there is no single linear function that agrees with a $\delta > 0$ fraction of the spaces L, for $\delta \gg 2^{-\ell}$. However, we claim that $\mathsf{agreement}(F) \geqslant \Omega(1)$. Indeed, on choosing L' and then $L_1, L_2 \supseteq L'$, there is constant probability that the smallest element $x \in L'$ is also smallest in both L_1 and L_2, in which case both $F[L_1]$ and $F[L_2]$ agree with f_x and the test accepts.

Remark 3.3. There are several known variants on this example, one of which is presented in [26]. This example was already described earlier in [8] in the analogous setting of "direct product" tests, which is a similar agreement testing question, except that the global function is not linear, and it is given via restrictions to ℓ-element sets and not ℓ-dimensional subspaces. Works on direct product or agreement testing [20, 14] resolved this issue using an idea which we describe next. This is called local structure versus global structure in [20, 14] and a zoom-in in [26].

Zooming-in

In [26], the authors circumvent the counterexample above by moving to a localized part of the space, which is called a "zoom-in." Since we will propose another idea called a "zoom-out," let us introduce a piece of notation that handles both.

Definition 3.4 (Zooming)**.** For subspaces $Q \subseteq W$, thought of as having a constant dimension and codimension respectively inside the global space X, let

$$\mathsf{Zoom}[Q, W] = \{L \mid L \subseteq X,\ \dim(L) = \ell,\ Q \subseteq L \subseteq W\}$$

denote the set of ℓ-dimensional subspaces that are between Q and W.[7]

[7] It is easily seen that the subgraph of the Grassmann graph induced on the subset of vertices $\mathsf{Zoom}[Q, W]$ is isomorphic to a lower order Grassmann graph $G(X', \ell')$ where $X' = W/Q$ is the quotient space, $\dim(X') = \dim(W) - \dim(Q)$, and $\ell' = \ell - \dim(Q)$.

The "zoomed-in" set of all ℓ-dimensional subspaces containing Q is then $\mathsf{Zoom}[Q, X]$ whereas the set of all ℓ-dimensional subspaces, namely $Gr(X, \ell)$, is $\mathsf{Zoom}[\{0\}, X]$. We note that when zooming in on a space Q of dimension smaller than ℓ, then while the resulting set $\mathsf{Zoom}[Q, X]$ is localized in some sense, it is still global in another: the ℓ-dimensional spaces contained in it cover all of X. This will not be the case when one applies a zoom-out, discussed below.

In the counterexample above, $F[\cdot]$ is an assignment for which $\mathsf{agreement}(F) = \Omega(1)$, but there is no global linear function that is $\Omega(1)$-consistent with $F[\cdot]$. However suppose we fix some $z^* \in X, z^* \neq 0$, and focus on the set $\mathsf{Zoom}[\mathsf{Span}(z^*), X]$ consisting of all ℓ-dimensional subspaces containing z^*.

- It is not difficult to see that if z^* is in the first $\approx 2^{n-\ell}$ elements in the linear order then the set of spaces in which z^* is the minimal element has constant density inside the "focus set" $\mathsf{Zoom}[\mathsf{Span}(z^*), X]$. Indeed, taking $L \subseteq X$ of dimension ℓ uniformly, we have that L contains expectedly one of the first $2^{n-\ell}$ elements in the linear order. Thus, conditioned on containing z^*, with constant probability it does not contain any other points among the first $2^{n-\ell}$.
- The assignment $F[\cdot]$ on spaces in which z^* is minimal is precisely the restriction of the global linear function f_{z^*}. That is after zooming in, the assignment $F[\cdot]$ does have good consistency with the global linear function f_{z^*}. The zoom-in space $\mathsf{Span}(z^*)$ is said to be "successful" in this sense.
- The fraction of successful zoom-ins $z^* \neq 0$ is $\approx 2^{-\ell}$ fraction of all the points $z^* \in X$.

These observations lead to the speculation below stating that an assignment $F[\cdot]$ with

$$\mathsf{agreement}(F) \geqslant \delta$$

does have good consistency with a global linear function after zooming into some constant dimensional subspace and moreover a "reasonable" fraction (that may depend on ℓ) of the zoom-ins are successful (the global linear function may depend on the choice of the zoom-in space).

Speculation 3.5 (False)**.** For every $\delta > 0$, there is an integer $q \geqslant 0$, a $\delta' > 0$, and a function $\alpha(\cdot) > 0$ of an integer parameter, such that the following holds. Given an assignment F for $Gr(X, \ell)$ with $\mathsf{agreement}(F) \geqslant \delta$, for $\alpha(\ell)$ fraction of q-dimensional subspaces Q, there exists a global linear function $f_Q \colon X \to \{0, 1\}$, such that $F[L] = f_Q|_L$ for δ' fraction of the vertices in $\mathsf{Zoom}[Q, X]$.

3.2 Hyperplane example and zooming out

We now present a counterexample to Speculation 3.5 and then propose a "fix" using the idea of a "zoom-out."

Hyperplane Example

We assume that the dimension of the global space X is $\gg 2^\ell$. Let $m = \gamma 2^\ell$ for a small constant γ ($\gamma = \frac{1}{20}$ works). Let $W_1, \dots, W_m$ be hyperplanes (= subspaces of codimension one) in X that are in general position, meaning the intersection of any k of them, for $1 \leqslant k \leqslant m$, has codimension k. Let $f_i \colon W_i \to \{0,1\}$ be linear functions on these hyperplanes such that for $1 \leqslant i \neq j \leqslant m$, f_i and f_j are different on $W_i \cap W_j$ (random functions will satisfy this w.h.p.). We define the assignment F for $Gr(X,\ell)$ by letting $F[L] = (f_i)|_L$ if i is the only index such that $L \subseteq W_i$. If L is not contained in any W_i, or it is contained in more than one, choose $F[L]$ at random.

Let L_1, L_2 be chosen according to the test distribution, namely L_1, L_2 are random ℓ-spaces that intersect on an $(\ell-1)$-dimensional subspace. These subspaces can be chosen by first choosing a random space R of dimension $\ell+1$ and then choosing two ℓ-spaces inside R.[8] Let E_i be the event that R is contained in W_i, let $S_i \subseteq E_i$ be the event that R is in W_i and not in any other W_j, and let $S = \bigcup_i S_i$. Clearly,

$$\mathsf{agreement}(F) \geqslant \Pr[S]$$

because in this case $F[L_1]$ and $F[L_2]$ are both consistent with the same f_i. We claim that $\Pr[S] \geqslant \frac{\gamma}{2} = \Omega(1)$. Indeed, $\Pr[E_i] \approx 2^{-\ell-1}$ and $\Pr[E_i \wedge E_j] \approx 2^{-2(\ell+1)}$. By inclusion-exclusion principle, the probability that R is contained in precisely one of the W_i, $1 \leqslant i \leqslant m$, is at least

$$\Pr[S] \geqslant m \cdot 2^{-\ell-1} - \binom{m}{2} \cdot 2^{-2(\ell+1)} \geqslant \frac{\gamma}{2} - \frac{\gamma^2}{8} \geqslant \frac{\gamma}{4}\,.$$

Finally, it is clear that there is no single linear function that agrees with F on more than $\exp(-\ell)$ fraction of the ℓ-dimensional subspaces. Moreover, we next show that zooming in won't work, exhibiting a counterexample to Speculation 3.5: There is no constant q and a q-dimensional subspace Q, for which $F[\cdot]$ is consistent with some global linear function on $\Omega(1)$ fraction of ℓ-dimensional subspaces containing Q.

Zoom-in not sufficient for consistency. Fix any q-dimensional subspace Q and let $\mathsf{Zoom}[Q,X]$ be the set of all ℓ-dimensional subspaces containing Q. These will be the only subspaces under consideration henceforth. Let $f : X \to \mathbb{F}_2$ be any global linear function. We look at the consistency between f and $F[\cdot]$ on the subspaces in $\mathsf{Zoom}[Q,X]$. Let $\mathcal{L}_i$ be the set of those $L \in \mathsf{Zoom}[Q,X]$ that are contained in W_i. Clearly, if $Q \not\subseteq W_i$, then $\mathcal{L}_i = \emptyset$ and otherwise, $\mathcal{L}_i$ has some fixed size depending on $q, \ell, \mathsf{dim}(X)$. Let $\mathcal{T}_i \subseteq \mathcal{L}_i$ be the set of those $L \in \mathsf{Zoom}[Q,X]$ that are contained in W_i but not in any other W_j, $j \neq i$.

Firstly, since $F[\cdot]$ is defined randomly outside $\cup_{i=1}^m \mathcal{T}_i$, non-trivial consistency between f and $F[\cdot]$, if any, has to be on $\cup_{i=1}^m \mathcal{T}_i$. Secondly, since $F[\cdot]$ is defined according to distinct functions f_i, f_j on $\mathcal{T}_i, \mathcal{T}_j$ respectively, the consistency between f and $F[\cdot]$ is non-negligible on at most one $\mathcal{T}_i$. Suppose that Q is contained in exactly k of the W_i's, say $W_1, \dots, W_k$. We consider two cases depending on how large k is, and show that the consistency between f and $F[\cdot]$ is at most $\approx 2^{-\frac{\ell}{2}}$ in both the cases, exhibiting the counterexample.

[8] This distribution gives a slightly larger probability for $L_1 = L_2$ than the test distribution, but this difference is negligible so we ignore it.

- (Case when $k \leqslant 2^{\frac{\ell}{2}}$). We note that w.l.o.g. $\mathcal{L}_{k+1} = \cdots = \mathcal{L}_m = \emptyset$ and the density of $\cup_{i=1}^{k} \mathcal{L}_i$ inside $\mathsf{Zoom}[Q, X]$ is at most $k \cdot 2^{q-\ell} \leqslant 2^{q-\frac{\ell}{2}}$. Since other than approximately $2^{-\ell}$ fraction of consistency that emerges from the random assignment, the consistency between f and $F[\cdot]$ has to be subspaces in $\cup_{i=1}^{k} \mathcal{T}_i \subseteq \cup_{i=1}^{k} \mathcal{L}_i$, the consistency is upper bounded by $2^{q-\frac{\ell}{2}} + O(2^{-\ell}) = O(2^{-\frac{\ell}{2}})$.
- (Case when $k \geqslant 2^{\frac{\ell}{2}}$). Since by symmetry, all $\mathcal{T}_i$, $i = 1, \ldots, k$, have the same size and are pairwise disjoint, and the consistency is non-negligible on at most one $\mathcal{T}_i$, the consistency is upper bounded by $\frac{1}{k} + O(2^{-\ell}) \leqslant O(2^{-\frac{\ell}{2}})$.

Zooming-out

We propose to circumvent the counterexample above using an idea of a "zoom-out" wherein one focusses on ℓ-dimensional subspaces that are contained in a subspace of constant codimension. Indeed, in the example above, we can choose any of the hyperplanes W_i and focus on $\mathsf{Zoom}[\{0\}, W_i]$, the subset of those L that are contained in W_i. By definition, there does exist a global linear function, namely f_i, that is consistent with $F[\cdot]$ on the subspaces in $\mathsf{Zoom}[\{0\}, W_i]$ which are not contained in any other W_j. As noted, this has density at least $\frac{\gamma}{2}$ inside $\mathsf{Zoom}[\{0\}, W_i]$, and hence f_i is $\frac{\gamma}{2}$-consistent with $F[\cdot]$ on $\mathsf{Zoom}[\{0\}, W_i]$. We observe in addition that only the hyperplanes $W_1, \ldots, W_m$ (and further subspaces of them of constant codimension if one wishes) furnish a "successful" zoom-out.

3.3 Our main hypothesis

It is certainly possible to combine the counterexamples in Section 3.1 and Section 3.2 so that both the zoom-in and the zoom-out are needed to circumvent the combined example. [9] Our main hypothesis, stated next, proposes that there always exist a zoom-in and a zoom-out on subspaces of constant dimension and codimension respectively that together are successful.

Hypothesis 3.6. For every $\delta > 0$, there exist integers $r, q \geqslant 0$ and $C > 0$, such that for all sufficiently large integers ℓ, for all sufficiently large integers n, the following holds. Let $F[\cdot]$ be an assignment for $Gr(X, \ell)$, $\mathsf{dim}(X) = n$, such that $\mathsf{agreement}(F) \geqslant \delta$. Then there exist subspaces $Q \subseteq W \subseteq X$ such that $\mathsf{dim}(Q) = q$ and $\mathsf{dim}(W) = n - r$, and a global linear function $g_{Q,W} : W \longrightarrow \mathbb{F}_2$ such that (note the conditional probability)

$$\Pr_{L \in Gr(X,\ell)} \left[\, g_{Q,W}|_L = F[L] \;\middle|\; Q \subseteq L \subseteq W \,\right] \geqslant C \,. \tag{3.1}$$

In the reduction below we have a 2-Prover-1-Round game, where players use a zoom combination as in Hypothesis 3.6 to choose a global function, which they return as answer.

[9] One can take zoom-outs $W_1, \ldots, W_k$ as in the second example, and for each $i = 1, \ldots, k$ and a point $x \in W_i$ consider a randomly chosen linear function $f_x : W_i \to \mathbb{F}_2$. With this setting, for each L contained in only a single W_i, we may define $F[L] = f_{i,x}|_L$, where x is the minimal element in L according to some fixed linear order, and define $F[L]$ arbitrarily otherwise.

There are some issues with this that we can already explain at this point: first, the hypothesis provides a linear function over W instead of over the entire space. This is handled by noting that because W has constant codimension, a linear function defined on it cannot have too many extensions over the entire space. The second issue is that, as it turns out, for the players to have large-enough success probability there must be a constant fraction of spaces Q for which a W exists such that Q and W are a successful zoom pair. The second issue is solved by the following lemma, which shows that Hypothesis 3.6 implies a non-negligible amount of good Zoom-in.[10]

Lemma 3.7. *Assume Hypothesis 3.6. For every $\delta > 0$, there exist integers $r, q \geqslant 0$, $C > 0$, and a function $\alpha\colon \mathbb{N} \to (0,1)$ such that for all sufficiently large integers ℓ, for all sufficiently large integers n, the following holds. Let $F[\cdot]$ be an assignment for $Gr(X,\ell)$ with $\mathsf{agreement}(F) \geqslant \delta$. Then for at least $\alpha(\ell)$ fraction of the q-dimensional subspaces $Q \subseteq V$, there exists a subspace W, $Q \subseteq W \subseteq V$ of codimension r, and a global linear function $g_{Q,W} : W \to \mathbb{F}_2$ such that (note the conditional probability)*

$$\Pr_{L\in Gr(X,\ell)}\left[\, g_{Q,W}|_L = F[L] \;\middle|\; Q \subseteq L \subseteq W \,\right] \geqslant C\,.$$

Proof. The proof strategy is as follows: apply the Hypothesis sequentially, finding a single Q at a time. Each time a Q is found, we "erase" the assignment on $L \supseteq Q$ by assigning a random linear function for $F[L]$ on every $L \supseteq Q$ and continue. The two main points are (a) each newly found Q is only contained in a small fraction of the L's, hence as long as we have not found enough Q, the erased spaces L do not decrease $\mathsf{agreement}(F)$, and (b) on the other hand the values assigned at random function essentially like erasure, in that they cannot contribute much to the subspaces Q that are found in subsequent steps.

Let r, q, C be from Hypothesis 3.6 for $\delta/2$, $\tilde{F} = F$. We prove the lemma for parameters $r, q, C/2$ and $\alpha(\ell) = \delta 2^{-\ell^2-2}$.

Denote by $\mathcal{Q}$ the set of Q found so far, $N = |\mathcal{Q}|$, and by $\mathcal{L}$ the set of spaces L that contain some $Q \in \mathcal{Q}$, namely whose assignment was rerandomized. At each step, as long as $\mathsf{agreement}(\tilde{F}) \geqslant \delta/2$, we apply Hypothesis 3.6 to obtain Q, W and $g_{Q,W}$ such that Equation (3.1) holds. We then define

$$\mathcal{L} \leftarrow \mathcal{L} \cup \{\, L \mid Q \subseteq L, \mathsf{dim}(L) = \ell \,\}\ ,$$

and reassign the spaces of $\mathcal{L}$ on $\tilde{F}$ in a manner to be described later. Notation: For any $i \leqslant n$ let $\begin{bmatrix} n \\ i \end{bmatrix} = \left|Gr(\mathbb{F}_2^n, i)\right|$ denote the number of i-dimensional subspaces in an n-dimensional space.

Claim 3.8. *At the end of the process, $N \geqslant \delta 2^{-\ell^2-2}\begin{bmatrix} n \\ q \end{bmatrix}$.*

Proof. Note each Q causes at most

$$\frac{\begin{bmatrix} n \\ \ell-q \end{bmatrix}}{\begin{bmatrix} n \\ \ell \end{bmatrix}} \leqslant \frac{2^{n(\ell-q)}}{2^{\ell(n-\ell)}} = 2^{\ell^2-qn}$$

[10] A similar Hypothesis to the above could have been made in [26]. The proof for the equivalence between the two is even simpler.

fraction of the L spaces to be added to $\mathcal{L}$. Therefore overall during the process, $\mathcal{L}$ contains at most $N \cdot 2^{\ell^2 - qn}$ fraction of the spaces in $Gr(X, \ell)$. Hence with probability at most $2N \cdot 2^{\ell^2 - qn}$ one of the subspaces L_1, L_2 considered in the test is reassigned. When the process is stuck, $\mathsf{agreement}(\tilde{F}) < \delta/2$, and so it must be the case that

$$\begin{aligned} 2N \cdot 2^{\ell^2 - qn} &\geqslant \delta/2 \\ \Longrightarrow N \geqslant \delta 2^{qn - \ell^2 - 2} &\geqslant \delta 2^{-\ell^2 - 2} \begin{bmatrix} n \\ q \end{bmatrix}. \end{aligned}$$

□

Next we argue that each newly found Q, has $W, g_{Q,W}$ for which Equation (3.1) holds for F, albeit with $C/2$. This is achieved by claiming that at most $2^{1-\ell}$ of the consistency with $\tilde{F}$ comes from spaces that were reassigned. The latter is shown by considering two cases: if at most $2^{-\ell}$ fraction of $Q \subseteq L \subseteq W$ are in $\mathcal{L}$, it is obvious. Otherwise

Claim 3.9. *With probability $1 - o_n(1)$, for every Q of dimension q, W of dimension r such that at least $2^{-\ell}$ fraction of $\{ L \mid Q \subseteq L \subseteq W \}$ is in $\mathcal{L}$, for every linear function $g_{Q,W} \colon W \to \mathbb{F}_2$,*

$$\Pr_{\substack{L \in_R \mathcal{L} \\ Q \subseteq L \subseteq W}} \left[g_{Q,W}|_L \equiv \tilde{F}[L] \right] \leqslant 2^{1-\ell} .$$

Proof. Fix such $Q, W, g_{Q,W}$, and denote $A = \{ L \mid Q \subseteq L \subseteq W \} \cap \mathcal{L}$. For each $L \in A$ define the indicator random variable Z_L which is 1 iff $g_{Q,W}|_L \equiv \tilde{F}[L]$. Then its expectation is $2^{-\ell}$, and, using Chernoff bound, the required probability is bounded by

$$\Pr_{L \in_R A} \left[\frac{1}{|A|} \sum_{L \in A} Z_L \geqslant 2^{1-\ell} \right] \leqslant \Pr_{L \in_R A} \left[\left| \frac{1}{|A|} \sum_{L \in A} Z_L - 2^{-\ell} \right| \geqslant 2^{-\ell} \right] \leqslant 2^{-\frac{1}{3} 2^{-2\ell} |A|} .$$

Note that $|A| \geqslant 2^{-\ell} \left| \{ L \mid Q \subseteq L \subseteq W \} \right| \geqslant 2^{-\ell} 2^{(\ell - q)(n - r - \ell)}$. Therefore, using a union bound over $Q, W, g_{Q,W}$, the probability there exists such bad triplet is at most

$$2^{qn} 2^{rn} 2^{n-r} 2^{-\frac{1}{3} 2^{-2\ell} 2^{(\ell - q)(n - r - \ell)}} = o_n(1) .$$

□

In particular, the previous claim implies there exist a reassignment (and we will pick such one in the reassignment phase) of vertices on $\mathcal{L}$, such that on each newly found Q, W, at most $2^{1-\ell}$ of the agreement with $\tilde{F}$ comes from $\mathcal{L}$. Hence at least $C - 2^{1-\ell} \geqslant C/2$ of the agreement comes outside $\mathcal{L}$, i. e., Equation (3.1) holds with $C/2$. □

3.4 A list decoding bound

In this section we bound the number of successful zoom-outs. This bound will later be used for showing that the provers can coordinate their answer with non-negligible probability as explained in Section 1.4.2.

Fix $X = \mathbb{F}_2^n$, $r \ll \ell \ll n$, and suppose $F[\cdot]$ is a table that assigns to every vertex $L \in Gr(X, \ell)$ a linear function $F[L]$ on L. We would like to have different thresholds for when a subspace W used for zoom-out is considered successful, depending on the codimension of W.

Definition 3.10. Fix numbers $\tau_0, \tau_1, \ldots, \tau_r \geqslant 0$ and let $\vec{\tau} = (\tau_0, \ldots, \tau_r)$. For a subspace $W \subseteq X$ and a linear function $g : W \longrightarrow \mathbb{F}_2$, we say that the pair (g, W) $\vec{\tau}$-occurs in F if, letting $i = \mathsf{dim}(X) - \mathsf{dim}(W) \leqslant r$,

$$\Pr_{L \in Gr(W,\ell)} [F[L] = g|_L] \geqslant \tau_i \,.$$

Furthermore, the pair (g, W) is *maximal* if there is no $W' \supsetneq W$ and linear function $g' : W' \longrightarrow \mathbb{F}_2$ such that $(g')|_W = g$ such that (g', W') $\vec{\tau}$-occurs in F.

Definition 3.11 (List of maximal pairs)**.** For an assignment F for $Gr(X, \ell)$ and a set of parameters $\vec{\tau} = (\tau_0, \ldots, \tau_r)$ define $\mathsf{LIST}^{\vec{\tau}}(F)$ to be the collection of maximal pairs $\vec{\tau}$-occuring in F,

$$\mathsf{LIST}^{\vec{\tau}}(F) = \left\{(g, W) \,|\, (g, W) \text{ is a maximal pair for } F\right\} \,.$$

The following is the main lemma of this section,

Lemma 3.12 (List size bound)**.** *For all $r \in \mathbb{N}$, $\tau_r > 0$, for sufficiently large ℓ and sufficiently large n compared to ℓ the following holds. For every F, there are numbers $\tau_i = 10^{-9}(\tau_{i+1})^{12}$, $0 \leqslant i < r$, such that the set of maximal pairs*

$$\mathsf{LIST} = \{(g, W) \,|\, (g, W) \textit{ is a maximal pair}\}$$

has size bounded by $\frac{(r+1)2^{8r^2\ell}}{(\tau_r)^{\mathsf{exp}(r)}}$.

The proof of Lemma 3.12 relies on a sunflower type statement, and on an analysis of an agreement test with smaller intersection size, similar to Lemma 3.1.

Lemma 3.13 (Sunflower Lemma for linear spaces)**.** *Let $\mathcal{Y} \subseteq Gr(X, r)$ be a collection of N subspaces, and let m be an integer such that $(m \cdot 2^r)^r \leqslant N$. Then there exist m subspaces $Y_1, \ldots, Y_m \in \mathcal{Y}$ that form a sunflower, namely denoting $Y = \bigcap_{i=1}^m Y_i$, we have $Y_i \cap Y_j = Y$ for all $1 \leqslant i \neq j \leqslant m$.*

The proof of Lemma 3.13 is similar to the proof of the usual sunflower lemma (if one settles for worse parameters, it would follow immediately from the usual lemma). As we are currently not aware of a reference for it, we include it here.

Proof. We apply induction over r. The base case, $r = 1$, follows since the intersection of any two different 1-dimensional subspaces is $\{0\}$, and thus any m distinct such subspaces form a sunflower.

Now assume the lemma holds for $r-1$, and let $\mathcal{Y} \subseteq Gr(X,r)$ be as in the lemma. Take $\mathcal{Z} = \{Z_1,\dots,Z_k\} \subseteq \mathcal{Y}$ to be a maximal set of subspaces such that the intersection between any two of them is $\{0\}$. Note that if $k \geqslant m$ then $Z_1,\dots,Z_m$ form a sunflower as we desire, so assume from now on that $k \leqslant (m-1)$. Denoting $S = \bigcup_{i=1}^{k} Z_i \setminus \{0\}$, we thus have that $|S| < 2^r(m-1)$. Moreover, because of the maximality of $\mathcal{Z}$ we have that $Y \cap S \neq \{0\}$ for any $Y \in \mathcal{Y} \setminus \mathcal{Z}$. It follows by the pigeonhole principle that there is a point $x \in S$ which is contained by at least

$$\frac{N-m+1}{2^r(m-1)} \geqslant \frac{(m\cdot 2^r)^r - m + 1}{2^r(m-1)} \geqslant \left(m \cdot 2^{r-1}\right)^{r-1} \tag{3.2}$$

subspaces in $\mathcal{Y}$.

Now fix $P : X \to X$ to be any linear projection which satisfies $\mathsf{Ker}(P) = \mathsf{Span}(x)$, and note that for a subspace $Y \in Gr(X,r)$ which contains x we have $PY \in Gr(PX, r-1)$, and also $P^{-1}(PY) = Y$. Hence different r-dimensional subspaces which contain x are mapped by P to distinct $(r-1)$-dimensional subspaces. Since by the bound in Eq. (3.2) there are at least $\left(m \cdot 2^{r-1}\right)^{r-1}$ subspaces $Y \in \mathcal{Y}$ that contain x, we thus have that the set $\mathcal{Y}' = \{PY \mid x \in Y \in \mathcal{Y}\}$ contains at least $\left(m \cdot 2^{r-1}\right)^{r-1}$ distinct elements from $Gr(PX, r-1)$.

Applying the inductive hypothesis to $\mathcal{Y}'$, we obtain a sunflower $Y_1',\dots,Y_m'$ in $\mathcal{Y}'$. Letting $Y_i = P^{-1}(Y_i')$ for $i = 1,\dots,m$, one easily verifies that $Y_1,\dots,Y_m$ is a sunflower in $\mathcal{Y}$ as required. □

Corollary 3.14. *Let $W_1,\dots,W_N \subseteq X$ be distinct subspaces of codimension r. Let m be an integer such that $(m \cdot 2^r)^r \leqslant N$. Then there is some subspace $W \subseteq X$ of codimension $r - s < r$ such that W contains m of the W_i's, say $W_1,\dots,W_m$, and such that for all $1 \leqslant i \neq j \leqslant m$, $W_i \cap W_j$ has codimension $2s$ in W.*

Proof of Corollary 3.14. Let $W_1,\dots,W_N$ be (distinct) subspaces of X of codimension r. Let us write $W_i = (Y_i)^{\perp} = \{w \in X \mid \langle w, y\rangle = 0,\ \forall y \in Y_i\}$ for appropriate r-dimensional subspaces Y_i and note that the Y_i are distinct since the W_i are distinct. From the sunflower lemma for linear spaces we obtain the subsequence $Y_1,\dots,Y_m$ such that for $Y = \bigcap_{i=1}^{m} Y_i$ and all $i \neq j$ we have $Y_i \cap Y_j = Y$. Let $W = Y^{\perp}$ and denote $s = r - \mathsf{dim}(Y)$ so that $r - s = \mathsf{codim}(W)$. Clearly, $\mathsf{dim}(Y) < r$ so $s > 0$. Also, we have that $W_i \subseteq W$ and $W_i \cap W_j = (Y_i + Y_j)^{\perp}$ for $1 \leqslant i, j \leqslant m$. It follows that the subspace $(Y_i + Y_j)^{\perp}$ has codimension $\mathsf{dim}(Y_i + Y_j) = 2r - (r - s) = r + s$ in X and codimension $r + s - (r - s) = 2s$ in W. □

Lemma 3.15 (List size bound)**.** *Let $F[\cdot]$ be a table that assigns, to every ℓ-dimensional subspace L of an n-dimensional space V, a linear function $F[L]$ on L. Suppose ℓ is a sufficiently large integer, $b = \frac{\ell}{10}$ and $n \geqslant 2\ell$. Let $\beta > 2^{-\ell/2}$, let $g_1,\dots,g_m$ be the list of all global linear functions on V that have β-agreement with $F[\cdot]$, namely for every $1 \leqslant i \leqslant m$, $F[L] = g_i|_L$ for at least β fraction of subspaces $L \subseteq V$ and moreover every such global linear function appears in the list. Then $m \leqslant \frac{\beta}{\beta^2 - 2^{-\ell}}$ and the probability*

$$\Pr_{\substack{L,L',\\ \mathsf{dim}(L\cap L')=b}} \left[F[L]|_{L\cap L'} = F[L']|_{L\cap L'} \ \wedge\ F[L] \notin \{g_1|_L,\dots,g_m|_L\} \ \wedge\ F[L'] \notin \{g_1|_{L'},\dots,g_m|_{L'}\}\right]$$

is at most $10\sqrt[3]{\beta}$.

Proof. The upper bound on m is as in [26, Theorem 2.6]. Now assume, on the contrary, that the probability in the statement of the lemma is at least $10\sqrt[3]{\beta}$. We define another table $F^*[\cdot]$ where $F^*[L] = F[L]$ if $F[L] \notin \{g_1|_L, \ldots, g_m|_L\}$ (let $\mathcal{L}^*$ denote the set of such L) and otherwise $F^*[L]$ is defined as a random linear function on L. The assumption implies that for $10\sqrt[3]{\beta}$ fraction of pairs $(L, L'), \mathsf{dim}(L \cap L') = b$, $F^*[L], F^*[L']$ are consistent on $L \cap L'$. By Lemma 3.15, there exists a global linear function $g : V \to \{0, 1\}$ that agrees with $F^*[\cdot]$ on at least 3β fraction of subspaces $L \subseteq V$. Since $F^*[\cdot]$ is defined at random outside $\mathcal{L}^*$, this agreement must essentially be on $\mathcal{L}^*$ (one could have used a Chernoff bound and taken a union bound over all global linear functions beforehand). However, $F[\cdot]$ and $F^*[\cdot]$ agree on $\mathcal{L}^*$ and hence g agrees with $F[\cdot]$ at β fraction of $L \subseteq V$. This is a contradiction since g is distinct from $g_1, \ldots, g_m$; indeed for any $L \in \mathcal{L}^*$ such that $F[L] = g|_L$, we have $g|_L \notin \{g_1|_L, \ldots, g_m|_L\}$. □

We are now ready to prove Lemma 3.12, and we begin with an overview of the proof.

Overview of the proof of Lemma 3.12. First, starting with a list of M zoom-outs that are good for F of codimension at most r, we find a fixed codimension $0 \leqslant r' \leqslant r$ such that there are at least M/r successful zoom-outs for F of codimension r'. We then apply Corollary 3.14 to find a sunflower type structure among these zoom-outs, and for simplicity of discussion we assume that the M/r zoom-outs we have already form a generic collection. By that, we mean that any two zoom-outs W, W' in the collection satisfy that $W \cap W'$ has codimension $2r'$.

Using this information, we conclude that $\mathsf{agreement}_t(F) \geqslant \Omega_\delta(1)$ for $t = \ell/10$. In words, table F satisfies that sampling L, L' of dimension ℓ that intersect in dimension $\ell/10$, the probability that $F[L]$ and $F[L']$ agree on $L \cap L'$ is constant. Once we establish this, the proof of the lemma readily follows from Lemma 3.1, and we next explain why $\mathsf{agreement}_t(F) \geqslant \Omega_\delta(1)$.

Choose L and L' as above, and consider a zoom-out W_i in the list. Note that the probability that both L, L' are contained in W_i and F agrees with the global function f_i of W_i on both L, L' is (roughly) $2^{-r'(2\ell-t)}\xi^2$. Here, ξ is the agreement of f_i with F inside W_i. In that case we would have that $F[L]|_{L\cap L'} = F[L']|_{L\cap L'}$ as desired. Thus, if we the number of the zoom-outs N/r exceeds $2^{r'(2\ell-t)}$, we would expect L, L' to be commonly contained in some W_i and have F agree with f_i on both of them, in which case we would be done. We show that this is essentially true. To make this argument work, one needs to take care of the potential over-counting that arises, and to do that we use the inclusion-exclusion principle. This is the part of the argument in which we use the fact that the collection of zoom-outs we have is generic. Indeed, intuitively the point of the genericness is it that the events that $L + L' \subseteq W_i$ and $L + L' \subseteq W_j$ for distinct W_i and W_j in the collection are almost independent, which is useful when doing inclusion-exclusion.

We now move on to the formal argument.

Proof of Lemma 3.12. Note that $\tau_0 \geqslant (\tau_r)^{\exp(r)}$. Assume, towards a contradiction, that there are more than $(r + 1)2^{8r^2\ell}/\tau_0$ maximal pairs. Then there is some $0 < r' \leqslant r$ for which the list LIST' defined as $\mathsf{LIST}' = \{(g, W) \in \mathsf{LIST} \mid \mathsf{codim}(W) = r'\}$ has size at least $2^{8r^2\ell}/\tau_0$. Set $\tau = \tau_{r'}$.

Note that it could happen that both (g, W) and (g', W) belong to LIST' for $g \neq g'$. However, this can happen for at most $O(1/\tau) \leqslant O(1/\tau_0)$ distinct linear functions due to Lemma 3.15. So there are at least $N := 2^{8r^2\ell}$ distinct subspaces in LIST' which we number $W_1, \ldots, W_N$ (ignoring the rest).

Applying Corollary 3.14, there exists a subspace $W \subseteq X$ of codimension $r' - s$, $1 \leqslant s < r'$, that contains (by re-indexing) subspaces $W_1, \ldots, W_m$ such that for all $1 \leqslant i \neq j \leqslant m$, W_i has codimension s inside W and $W_i \cap W_j$ has codimension $2s$ inside W. Corollary 3.14 gives a lower bound $m \geqslant \frac{N^{\frac{1}{r}}}{2^r} \geqslant 2^{4r\ell}$. We assume that $m = \gamma \cdot 2^{s\cdot(2\ell-b)}$ (ignoring the rest) where $\gamma = \frac{\tau^2}{2}$. We will prove that there is a linear function f such that

$$\Pr_{L \in Gr(W,\ell)}[F[L] = f|_L] \geqslant \tau^{12}/2000\,. \tag{3.3}$$

Moreover, we will prove that

$$\exists (g_i, W_i) \in \mathsf{LIST}', i \in [m], \qquad f|_{W_i} = g_i\,. \tag{3.4}$$

Let us first show how Equations (3.3) and (3.4) together imply the lemma. From Eq. (3.3) we deduce that (f, W) occurs in F. Indeed the codimension of W is some $i < r'$, so $\Pr_{L\in Gr(W,\ell)}[F[L] = f|_L] \geqslant \tau^{12}/2000 > \tau_i$. This means that either $(f, W) \in \mathsf{LIST}$ or there is some $(f', W') \in \mathsf{LIST}$ such that $W' \supseteq W$ and $(f')_W = f$. Either way this contradicts the fact that $(W_i, g_i) \in \mathsf{LIST}$ is maximal (since $W' \supseteq W \supseteq W_i$ and $(f')|_{W_i} = f|_{W_i} = g_i$).

To prove Eq. (3.3) we will first show that F restricted to $Gr(W, \ell)$ passes a linearity agreement test with sufficient probability. Fix $b = \ell/10$. Let L, L' be chosen uniformly at random conditioned on their intersection having dimension b. We will prove that

$$\Pr_{L,L'}[\,F[L]|_{L\cap L'} = F[L']|_{L\cap L'}\,] \geqslant \tau^4/4 \tag{3.5}$$

which, using Lemma 3.1, implies Eq. (3.3).

To prove Eq. (3.5) we note that the pair (L, L') can be chosen by first choosing a random $(2\ell - b)$-dimensional subspace R and then choosing random ℓ-dimensional subspaces $L, L' \subseteq R$ with b-dimensional intersection. Formally, it is more convenient to choose R as the span of $2\ell - b$ randomly chosen vectors in W (the probability this is not of dimension $2\ell - b$ is exponentially small in n and can be ignored).The choice of L, L' after choosing R is essentially independent; choosing them independently, it does hold that $\mathsf{dim}(L \cap L') = b$ except with probability $2^{-\Omega(b)}$. Fix an index $1 \leqslant i \leqslant m$. It will be convenient to define events $\mathcal{E}_i, \mathcal{P}_i, \mathcal{S}_i$ such that $\mathcal{P}_i \subseteq \mathcal{E}_i$, $\mathcal{S}_i \subseteq \mathcal{E}_i$ as follows.

- Let $\mathcal{E}_i$ be the event that both $L, L' \subseteq W_i$. Then

$$\Pr[\mathcal{E}_i] = \Pr[R \subseteq W_i] = 2^{-s(2\ell-b)}\,.$$

- Let $\mathcal{P}_i$ be the event that both $L, L' \subseteq W_i$, but for $1 \leqslant j \neq i \leqslant m$, W_j does not contain both L, L' (i. e., the pair (L, L') is *private* to W_i). Then

$$\Pr[\mathcal{E}_i \wedge \neg\mathcal{P}_i] \leqslant \sum_{j\neq i} \Pr\left[\mathcal{E}_i \cap \mathcal{E}_j\right] \leqslant (m-1)2^{-2s(2\ell-b)} \leqslant \gamma 2^{-s(2\ell-b)}\,.$$

where we have estimated $\Pr[\mathcal{E}_i \cap \mathcal{E}_j] = \Pr[R \subseteq W_i \cap W_j] = 2^{-2s(2\ell-b)}$. We note that the events $\mathcal{P}_i, 1 \leqslant i \leqslant m$, are disjoint.

- Let $\mathcal{S}_i$ be the event that both $L, L' \subseteq W_i$ and moreover that $F[L] = g_i|_L$, $F[L'] = g_i|_{L'}$ (and in particular $F[L], F[L']$ are consistent on $L \cap L'$). We noted that $R \subseteq W_i$ with probability $2^{-s(2\ell-b)}$ and then denoting by $p(R)$, the fraction of ℓ-dimensional subspaces $L \subseteq R$ for which $F[L] = g_i|_L$,

$$\Pr\left[\mathcal{S}_i\right] = 2^{-s(2\ell-b)} \cdot \mathop{\mathbb{E}}_{R \subseteq W_i}\left[p(R)^2\right] \geqslant 2^{-s(2\ell-b)} \cdot \mathop{\mathbb{E}}_{R \subseteq W_i}\left[p(R)\right]^2 \geqslant 2^{-s(2\ell-b)} \cdot \tau^2 \,.$$

- Combining the above,

$$\begin{aligned}\Pr\left[\mathcal{S}_i \wedge \mathcal{P}_i\right] &= \Pr[\mathcal{S}_i \setminus (\mathcal{S}_i \wedge \neg\mathcal{P}_i)] \\ &\geqslant \Pr[\mathcal{S}_i] - \Pr[\mathcal{E}_i \wedge \neg\mathcal{P}_i] \\ &\geqslant (\tau^2 - \gamma) \cdot 2^{-s(2\ell-b)} \geqslant \frac{\tau^2}{2} \cdot 2^{-s(2\ell-b)} = \frac{\tau^2}{2} \cdot \Pr\left[\mathcal{E}_i\right].\end{aligned}$$

The probability that $F[L], F[L']$ are consistent on $L \cap L'$ is at least ($\mathcal{S}_i \wedge \mathcal{P}_i$ are disjoint)

$$\Pr\left[\vee_{i=1}^m \mathcal{S}_i \wedge \mathcal{P}_i\right] = \sum_{i=1}^m \Pr\left[\mathcal{S}_i \wedge \mathcal{P}_i\right] \geqslant m \cdot \frac{\tau^2}{2} \cdot 2^{-s(2\ell-b)} = \gamma \cdot \frac{\tau^2}{2} = \frac{\tau^4}{4}\,, \tag{3.6}$$

thus establishing Eq. (3.5). By Lemma 3.1, there is a global linear function $f : W \longrightarrow \{0,1\}$ that agrees with $F[\cdot]$ on $\frac{\tau^{12}}{2000}$ fraction of ℓ-dimensional subspaces, thus implying Eq. (3.3).

We now proceed to show Eq. (3.4), namely that for some $i \in [m]$, $f|_{W_i} = g_i$. This will conclude the proof of the lemma.

Let $f_1, \ldots, f_k$ be the list of all global linear functions on W that have $10^{-9}\tau^{12}$ agreement with $F[\cdot]$ (So the pairs (f_j, W) occur in F for all $j = 1, \ldots, k$). So far, we have established that this list is non-empty. We note that the event $\vee_{i=1}^m \mathcal{S}_i \wedge \mathcal{P}_i$ implies the event that $F[L], F[L']$ are consistent on $L \cap L'$. By Lemma 3.15, $k \leqslant \frac{2\cdot 10^9}{\tau^{12}}$ and

$$\Pr\left[\left(\vee_{i=1}^m \mathcal{S}_i \wedge \mathcal{P}_i\right) \ \wedge (F[L] \notin \{f_1|_L, \ldots, f_k|_L\}) \ \wedge (F[L'] \notin \{f_1|_{L'}, \ldots, f_k|_{L'}\})\right] \leqslant \frac{\tau^4}{100}\,. \tag{3.7}$$

From Equations (3.6) and (3.7) and noting that the roles of L, L' are symmetric,

$$\Pr\left[\left(\vee_{i=1}^m \mathcal{S}_i \wedge \mathcal{P}_i\right) \wedge (F[L] \in \{f_1|_L, \ldots, f_k|_L\})\right] \geqslant \frac{\tau^4}{16}\,.$$

We conclude by the union bound that

$$\sum_{i=1}^m \Pr\left[\mathcal{S}_i \wedge \mathcal{P}_i \wedge (F[L] \in \{f_1|_L, \ldots, f_k|_L\})\right] \geqslant \frac{\tau^4}{16}\,.$$

Replacing the event $\mathcal{S}_i \wedge \mathcal{P}_i$ by its implication $\mathcal{E}_i \wedge (F[L] = g_i|_L)$ and further relaxing to implication $g_i|_L \in \{f_1|_L, \dots, f_k|_L\}$, we have

$$\sum_{i=1}^{m} \Pr\left[\mathcal{E}_i \wedge (g_i|_L \in \{f_1|_L, \dots, f_k|_L\})\right] \geqslant \frac{\tau^4}{16}\,.$$

Noting that $m = \gamma \cdot 2^{s(2\ell - b)}$ and $\Pr\left[\mathcal{E}_i\right] = 2^{-s(2\ell-b)}$, we rewrite as

$$\frac{1}{m}\sum_{i=1}^{m} \Pr\left[g_i|_L \in \{f_1|_L, \dots, f_k|_L\} \mid \mathcal{E}_i\right] \geqslant \frac{1}{\gamma} \cdot \frac{\tau^4}{16} = \frac{\tau^2}{8}\,.$$

In the above inequality, $\mathcal{E}_i$ is the event that both $L, L' \subseteq W_i$, but L' has no role, so we can rewrite as

$$\frac{1}{m}\sum_{i=1}^{m} \Pr\left[g_i|_L \in \{f_1|_L, \dots, f_k|_L\} \mid L \subseteq W_i\right] \geqslant \frac{\tau^2}{8}\,.$$

Now we can finish the proof. By an averaging argument it follows there is $i \in [m]$ such that

$$\Pr\left[g_i|_L \in \{f_1|_L, \dots, f_k|_L\} \mid L \subseteq W_i\right] \geqslant \frac{\tau^2}{8}\,.$$

Clearly, it must then be the case that g_i is identically equal to one of the functions $f_1|_{W_i}, \dots, f_k|_{W_i}$, since otherwise g_i would agree with any $f_j|_{W_i}$ for at most $2^{-\ell}$ fraction of $L \subseteq W_i$ and then one could take a union bound over $1 \leqslant j \leqslant k$. This shows Eq. (3.4) and completes the proof. □

3.5 Main lemmas

In this section we collect all that we have proved, into two lemmas that will be used in the proof of soundness.

Definition 3.16 (Q - List). Let F be an assignment for $Gr(X, \ell)$, and let $\vec{\tau}$ be a parameter vector. For a subspace $Q \in Gr(X, q)$, let us say that the pair (f, W) $\vec{\tau}$-occurs with respect to Q in F if $\mathsf{codim}(W) = i$ and

$$\Pr_{Q \subseteq L \subseteq W}\left[F[L] = f|_L\right] \geqslant \tau_i\,,$$

and let us say the pair is maximal if there is no (f', W') that τ-occurs w.r.t. Q in F such that $W' \supsetneq W$ and $f'|_W = f$. Let the Q-list of F be

$$\mathsf{LIST}_Q^{\vec{\tau}}(F) = \{(f, W) \mid (f, W) \text{ is maximal}\}\,.$$

Lemma 3.17. *Assume Hypothesis 3.6. For all $\delta > 0$ there are $C > 0$ and $q, r \in \mathbb{N}$, a function $\alpha\colon \mathbb{N} \to (0, 1)$, and set $\tau_r = C$ and $\tau_i = 10^{-9}(\tau_{i+1})^{12}$, for $0 \leqslant i < r$ such that for all sufficiently large $\ell > 0$, for all sufficiently large n, the following holds. Let X be an n-dimensional vector space over $\mathbb{F}_2$ and let F be an assignment over $Gr(X, \ell)$.*

- *If* $\mathsf{agreement}(F) \geqslant \delta$ *then* $\Pr_{Q\in Gr(X,q)}\left[\mathsf{LIST}^{\vec{\tau}}_{Q}(F) \neq \emptyset\right] \geqslant \alpha(\ell)$.
- *For all* $Q \in Gr(X,q)$, $\left|\mathsf{LIST}^{\vec{\tau}}_{Q}(F)\right| \leqslant 2^q \cdot 2^{8r^2\ell}/C^{\exp(r)}$.

Before proving the lemma, we next show that if the given assignment $F[\cdot]$ is invariant (defined shortly below) with respect to a subspace H and a linear function $h : H \to \mathbb{F}_2$ then for any Q and any (f, W) in the Q-list of F, it must be that $f_H = h$.

Definition 3.18 (Invariant assignment)**.** Let F be an assignment over $Gr(X,\ell)$ and let (H,h) be a side condition. F is (H,h)-invariant if for every space $K = L + H$, where $L \in Gr(X,\ell)$ and $H \cap L = \{0\}$, there is a linear function $f : K \to \mathbb{F}_2$ such that $f|_H = h$ and for every L' such that $L' + H = K$, $f|_{L'} = F[L']$.

We can think of an invariant assignment as follows. Partition the set of subspaces $Gr(X,\ell)$ into equivalence classes according to the value of $L + H$. Each equivalence class $K = L + H \in Gr(X/H,\ell)$ is assigned a function $\tilde{F}[K] : K \to \mathbb{F}_2$, and then to compute $F[L]$ one computes $K = L + H$, looks up the function $\tilde{F}[K]$ and outputs its restriction to $L \subseteq K$.

Lemma 3.19. *Suppose F is an assignment for $Gr(X,\ell)$, that is (H,h)-invariant. Then for any Q and any $(g, W) \in \mathsf{LIST}^{\vec{\tau}}_{Q}(F)$, it must be that $W \supseteq H$ and $f|_H = h$.*

Next we turn to proving the lemmas.

Proof of Lemma 3.17. The first item follows immediately from Lemma 3.7, which relies on the main hypothesis, since $\tau_i \geqslant C$ for all $0 \leqslant i \leqslant r$. The second item essentially follows from Lemma 3.12, but first we must set it up. Let φ be the canonical bijection from $\{L \in Gr(X,\ell) \,|\, L \supseteq Q\}$ to $Gr(X/Q,\ell - q)$ which maps every $L \supseteq Q$ to L/Q. Define an assignment F_Q for $Gr(X/Q,\ell-q)$ from F as follows.

For every L such that $F[L]|_Q = 0$ we define $F_Q(L/Q) = g$ where $g = \widetilde{F[L]} : L/Q \to \mathbb{F}_2$ is the function such that $g(x+Q) = F[L](x)$ for all $x \in L$, guaranteed by Claim 2.5.

For spaces L with $F[L]|_Q \neq 0$, we shift it as follows. Let $p_1, \dots, p_{2^q} : X \to \mathbb{F}_2$ be arbitrary linear functions that are distinct on Q (i. e., their restriction to Q attains all possible linear functions on Q). For L such that $F[L]|_Q = p_j$ define $F'[L] = F[L] + p_j|_L$. F' now has the property that for every $L \supseteq Q$, $F'[L]|_Q = 0$, and we set $F_Q[L] := \widetilde{F'[L]}$ for all $L \supseteq Q$.

Let $(f, W) \in \mathsf{LIST}^{\vec{\tau}}_{Q}(F)$, and let p_j be such that $(f + p_j)|_Q = 0$. Let $\widetilde{f+p_j} : X/Q \to \mathbb{F}_2$ be the linear function guaranteed in Claim 2.5 to obey

$$\forall x \in X, \quad \widetilde{f+p_j}(\varphi(x)) = (f+p_j)(x)\,.$$

For every $Q \subseteq L \subseteq W$ such that $F[L] = f|_L$ we get $F_Q[L] = \widetilde{F'[L]} = \widetilde{f+p_j}|_{L/Q}$, so we conclude that $(\widetilde{f+p_j}, W) \in \mathsf{LIST}^{\vec{\tau}}(F_Q)$.

On the other hand, each $(g, W) \in \mathsf{LIST}^{\vec{\tau}}(F_Q)$ can come from at most 2^q different pairs (f_j, W), one per restriction to Q. So we conclude that

$$\left|\mathsf{LIST}^{\vec{\tau}}_{Q}(F)\right| \leqslant 2^q \cdot \left|\mathsf{LIST}^{\vec{\tau}}(F_Q)\right|$$

and the right hand side is bounded by Lemma 3.12. □

Proof of Lemma 3.19. By assumption, for $i = \mathsf{codim}(W)$,

$$\Pr_{Q \subseteq L \subseteq W}[F[L] = g|_L] \geqslant \tau_i \,. \tag{3.8}$$

Let $H' = W \cap H$. We show first that $g|_{H'} = h|_{H'}$, namely g respects the side condition $(H', h|_{H'})$. Assume on the contrary that $g|_{H'} \neq h|_{H'}$. Note that almost all ℓ-spaces L such that $Q \subseteq L \subseteq W$ satisfy that $L \cap H' = \{0\}$, and we focus on these spaces henceforth. Partition them into equivalence classes by the mapping $\varphi : Gr(X, \ell) \to Gr(X/H', \ell)$ through $\varphi(L) = L + H'$.

Fix some class $\tilde{L}$, let $L_1 \in \varphi^{-1}(\tilde{L})$ and let $Y = L_1 + H' \subsetneq L + H$. Since F is (H, h)-invariant, there is a function $f : Y \to \mathbb{F}_2$ such that $f|'_H = h|_{H'}$ *and* for any $L' \in \varphi^{-1}(\tilde{L})$ also $F[L'] = f|_{L'}$. Since $g_{Q,W}|_{H'} \neq h$ it must be that $g_{Q,W}|_Y \neq f$.

Therefore, for a random L such that $L + H' = Y$, one could have the agreement $g_{Q,W}|_L = f|_L = F[L]$ with probability at most $2^{q-\ell}$. Since ℓ could have been chosen large enough, one contradicts Equation (3.8).

It is clear that $g_{Q,W}$ can be extended uniquely to $g : W' + H \to \mathbb{F}_2$ so that $g|_H = h$ and $g|_{W'} = g_{Q,W}$. This is a function that satisfies the full side condition. Clearly the partition of subspaces L into classes $L + H' \in Gr(X/H', \ell)$ is a refinement of the partition into classes $L + H \in Gr(X/H, \ell)$. For every class $L + H$, the (H, h) invariance of F implies that there is a single function f such that $f|_L = F[L]$ for all L in the class. This implies that for all of the sub-classes $L' + H' \subseteq L + H$ the same must hold. So the agreement on each class is again either negligible or perfect. Hence Equation (3.8) implies that (the same statement with W replaced by $\tilde{W} = W + H$ and $\mathsf{codim}(\tilde{W}) < i$)

$$\Pr_{Q \subseteq L \subseteq \tilde{W}}[g|_L = F[L]] \geqslant \tau_i/2 > \tau_{i-1} \,. \tag{3.9}$$

This contradicts the maximality of (g, W) in $\mathsf{LIST}^{\vec{\tau}}_Q(F)$, so it must have been that $W \supseteq H$ in the first place. □

4 The reduction

In this section we elaborate on the reduction from 3-Lin to 2-to-1 Games problem that proves Theorem 1.8. The reduction is sketched in the introduction as a two prover game. The full reduction is described here formally in Section 4.2. We then discuss some properties of the construction that are important for proving completeness and soundness.

4.1 Starting point: The Gap3Lin problem

An instance of the 3-Lin 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 fraction 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 [18] shows that for every $\varepsilon > 0$, $\mathsf{Gap3Lin}(1-\varepsilon, \frac{1}{2}+\varepsilon)$ 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 the following.

Theorem 4.1. *There is an absolute constant $s^* < 1$ such that for every $\varepsilon > 0$, $\mathsf{Gap3Lin}(1-\varepsilon, s^*)$ is NP-hard on regular instances.*

4.2 The full reduction

In this section we construct the instance of the 2-to-1 game G_{folded}. The construction follows the reduction described in the introduction, but is written formally as a constraint graph and not a two-player game.

We first describe a constraint graph that is not folded, $G_{\mathsf{unfolded}} = (A, B, E, \Pi, \Sigma_A, \Sigma_B)$, and then change it into the final instance $G_{\mathsf{folded}} = (\tilde{A}, B, \tilde{E}, \tilde{\Pi}, \Sigma_A, \Sigma_B)$ by identifying sets of vertices in A. Details follow.

The vertices. Let $\mathcal{U}$ be the set of all k-tuples of equations $U = (e_1, \dots, e_k)$ from the regular $\mathsf{Gap3Lin}$ instance (X, eq) that are "legitimate," namely such that (a) the equations $e_1, \dots, e_k$ 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.

For every $U \in \mathcal{U}$ we denote by $X_U \subseteq \mathbb{F}_2^n$ the linear subspace whose elements have support in U. Similarly, let $\mathcal{V}$ be the collection of all sets V of up to $3k$ variables. Let $X_V \subseteq \mathbb{F}_2^n$ be the linear subspace whose elements have support in V. For each $U \in \mathcal{U}$ let

$$H_U = \mathsf{Span}\,\{x_e \mid e \in U\}$$

where $x_e \in \mathbb{F}_2^n$ is the vector that is zero on all but three coordinates, corresponding to the variables participating in the equation e. For each $U \in \mathcal{U}$ we will have a block of vertices corresponding to the elements of $Gr(X_U, \ell)$ and let

$$A = \{(U, L) \mid U \in \mathcal{U},\ L \in Gr(X_U, \ell),\ L \cap H_U = \{0\}\}\ ,$$

and similarly

$$B = \{(V, L') \mid V \in \mathcal{V},\ L' \in Gr(X_V, \ell - 1)\}\ .$$

The edges. The edges are described through a random process that outputs a pair $(U, L), (V, L')$. The probability of outputting a certain pair is the weight on the corresponding edge.

1. Choose a k-tuple $U = (e_1, \ldots, e_k) \in \mathcal{U}$ uniformly at random and then construct 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.
2. Choose a random $L' \in Gr(X_V, \ell - 1)$ and a random $L \in Gr(X_U, \ell)$ such that $L' \subseteq L$.
3. Output $(U, L), (V, L')$.

Constraints. We let $\Sigma_A = \{0, 1\}^\ell$ and $\Sigma_B = \{0, 1\}^{\ell-1}$. $\sigma \in \Sigma_A$ is interpreted as a linear function $\sigma : L \to \mathbb{F}_2$ and $\sigma \in \Sigma_B$ is interpreted as a linear function $\sigma : L' \to \mathbb{F}_2$. This can be done, say, by fixing an arbitrary basis for each subspace L, L'. The constraint between (U, L) and (V, L') accepts pairs (σ, σ') iff $\sigma|_{L'} = \sigma'$. It is clear that this is a $2:1$ constraint.

This completes the construction of G_{unfolded}. At this point the construction does not yet make sense because it does not take into account the constraints imposed by the equations in the 3-Lin instance at all. This will be achieved next, by folding.

Folding. We now turn to define $G_{\mathsf{folded}} = (\tilde{A}, B, \tilde{E}, \tilde{\Pi}, \Sigma_A, \Sigma_B)$. We partition the set A into equivalence classes, $A = C_1 \sqcup C_2 \sqcup \cdots$. This is done in a way that ensures that a linear function σ assigned to any $(U_0, L_0) \in C$ uniquely determines a linear function assigned to every other $(U, L) \in C$. For each $U \in \mathcal{U}$ define the linear function

$$h_U : H_U \to \mathbb{F}_2, \qquad h_U(x_e) = b_e, \ \forall e \in U \tag{4.1}$$

where $b_e \in \{0, 1\}$ is the RHS of the equation e. Given $(U_0, L_0) \in A$, we define $C(U_0, L_0)$ by

$$C(U_0, L_0) = \{(U, L) \in A \,|\, L + H_U + H_{U_0} = L_0 + H_U + H_{U_0}\} \ .$$

The following lemma implies that each C is an equivalence class.

Lemma 4.2. *For (U, L) there is an ℓ-dimensional subspace $R \subseteq \{0, 1\}^n$ such that*

$$C(U, L) = \{(U', L') \in A \,|\, L' + H_{U'} = R + H_{U'}\} \,.$$

Furthermore, for any two vertices $(U_1, L_1), (U_2, L_2) \in A$, either $C(U_1, L_1) = C(U_2, L_2)$ or $C(U_1, L_1) \cap C(U_2, L_2) = \emptyset$

We defer the proof of the lemma to Subsection 4.3, after we've established some properties of our folding. Now define

$$\tilde{A} = \{C(U, L) \,|\, (U, L) \in A\} \ .$$

We further define the set of edges $\tilde{E}$ by a random process: choose a random edge $((U, L), (V, L')) \in E$ and then output $(\, C(U, L) \,, \, (V, L') \,) \in \tilde{A} \times B$.

An assignment for C will be interpreted as follows. For each C we fix some R as guaranteed by Lemma 4.2 (there may be more than one R per a given class and we fix one arbitrarily for each C), and then the assignment σ is interpreted as a linear function over R, $\sigma : R \to \mathbb{F}_2$. In order to define the constraint on the edge between C and (V, L') we explain how an assignment to $\tilde{A}$ is unfolded into an assignment for A.

Recall from Definition 2.2 that if $L \cap H = \{0\}$ then any linear function $f : L \longrightarrow \mathbb{F}_2$ has a unique (H, h)-extension, $\tilde{f} : L + H \longrightarrow \mathbb{F}_2$, where $\tilde{f}(z) = f(x) + h(y)$ and where $z = x + y$ is the unique way to write z as a sum of $x \in L$ and $y \in H$. This allows us to make the following definition.

Definition 4.3 (Unfolding)**.** Let R be the representative of an equivalence class C. An assignment for C is a linear function $\sigma : R \longrightarrow \mathbb{F}_2$. For every $(U, L) \in C$ we unfold $\sigma = \tilde{\mathcal{A}}(C)$ by defining, $\mathcal{A}(U, L) = \tilde{\sigma}_U|_L$, where $\tilde{\sigma}_U$ is the (H_U, h_U) extension of σ, and where $h_U : H_U \longrightarrow \mathbb{F}_2$ is the function from Eq. (4.1).

The constraints in G_{folded} are defined to be the aggregate of the constraints in G_{unfolded}: a constraint between C and (V, L') accepts a pair of assignments (σ, σ') if the unfolding of σ satisfies *all* of the constraints between members of C and the assignment σ' for (V, L'). The weights are summed up as well. By definition there is at least one $(U, L) \in C$ such that G_{unfolded} has an edge between (U, L) and (V, L'), but there could be more than one.

It is clear that the constraints are *at most* $2 : 1$ but it might seem that some or many of the constraints are completely unsatisfiable. The following claim implies that this is not the case.

Claim 4.4. *Let $C \in \tilde{A}$, let $\sigma = \tilde{\mathcal{A}}(C)$, and for each $(U, L) \in C$ let $\mathcal{A}(U, L) = \tilde{\sigma}_U|_L$ be its unfolding. Then for any subspace D and for any $(U_1, L_1), (U_2, L_2) \in C$ such that $D \subseteq L_1 \cap L_2$, $A(U_1, L_1)|_D = A(U_2, L_2)|_D$.*

Proof. The claim would follow if we show for every U_1, U_2, denoting $H = H_{U_1} + H_{U_2}$, there is a unique linear function $h : H \longrightarrow \mathbb{F}_2$ such that $h|_{H_{U_1}} = h_{U_1}$ and $h|_{H_{U_2}} = h_{U_2}$. Indeed, in that case let $g : R + H \longrightarrow \mathbb{F}_2$ be the (H, h) extension of $\sigma : R \longrightarrow \mathbb{F}_2$ (g exists since $R \cap H = \{0\}$) and observe that $g|_{L_i} = A(U_i, L_i)$ for $i = 1, 2$, so they must coincide on $D = L_1 \cap L_2$.

To show the existence of h we must check there are no contradictions between the requirements. This follows by inspecting the intersection structure of U_1, U_2. Suppose U_1 is the set of variables of equations $e_1, \ldots, e_k$ and U_2 is the set of variables of $e'_1, \ldots, e'_k$. The equations can be reordered so that for all $i \neq j$, $e_i \cap e'_j = \emptyset$, and for all i, either $e_i = e'_i$, or $|e_i \cap e'_i| = 1$ or $e_i \cap e'_i = \emptyset$. This means that the collection $\left\{x_e \mid e = e_i \text{ or } e = e'_i\right\}$ consists of linearly independent vectors (some repeating) and the function h is well defined. □

4.3 Properties of the construction due to folding

In this subsection we observe some important properties of the construction that come from the folding, leading up to a proof of Lemma 4.2.

We begin with the notion of singled out coordinates. Consider a set of variables U coming from k equations, and a vector $x \in \{0, 1\}^U$. For simplicity, we order U according the some ordering of the k equations so that the variables of the first equation e_1 come first, then the variables of the second equation e_2, and so on. Note that if x is constant on all of the variables of some equation – say of e_1 and that $(x_1, x_2, x_3) = (1, 1, 1)$ – then the value of an assignment on x does not give any new information on the assignment to the variables of the first equation.

Indeed, the value of a linear function on x only concerns the sum of these values, which is supposed to be a pre-determined value determined by e_1. Thus, x potentially gives information about these variables only when x_1, x_2, x_3 are not all equal, and in that case by adding $(1,1,1)$ and up to re-ordering of the variables, one may assume that $x_1 = 1$ and $x_2 = x_3 = 0$, and so the value of a linear function on x can be thought of as telling us information about the value of the first variable.

With this in mind, the set of singled out coordinates is the set of coordinates on which we potentially get information, defined formally as follows:

Definition 4.5 (Singled-out coordinates)**.** Fix $(U, L) \in A$. For each $i \in U$ there is a single equation containing i whose variables are contained in U. For each point $x \in L$ let

$$I_U(x) = \{i \in [n] \mid x_i \neq x_j = x_k, \text{ where } \{i,j,k\} \text{ are}\\ \text{the variables of the unique equation containing } i\}\,.$$

Further, we consider all the coordinates that are singled out in a class,

$$I_U(L) = \bigcup_{x \in L} I_U(x), \quad \text{and} \quad I(C) = \bigcup_{(U,L)\in C} I_U(L)\,. \tag{4.2}$$

The following claim asserts that if (U_1, L_1) and (U_2, L_2) are in the same equivalence class, then their sets of singled out coordinates are the same. Intuitively, this says that the information a linear function on (U_1, L_1) gives on the variables is the same as the information a linear function on (U_2, L_2) gives.

Claim 4.6. *If $(U_1, L_1), (U_2, L_2) \in C$ then $I_{U_1}(L_1) = I_{U_2}(L_2)$.*

The claim implies that $I(C) = I_U(L)$ for any $(U, L) \in C$ (making the union in Eq. (4.2) degenerate).

Proof. We show that if $i \in I_{U_1}(L_1)$ then $i \in I_{U_2}(L_2)$, and the claim will follow from symmetry. So let $i \in I_{U_1}(L_1)$, let $e = \{i,j,k\}$ be the equation in U_1 that contains it, and let $x \in L_1$ be such that $i \in I_{U_1}(x)$, that is $x_i \neq x_j = x_k$. By assumption $x \in L_1 + H_1 \subseteq L_1 + H_1 + H_2 = L_2 + H_1 + H_2$ so we can write $x = x^1 + x^2$ where $x^1 \in H_1$ and $x^2 \in L_2 + H_2$. Since $x^1 \in H_1$, we know that $x^1_i = x^1_j = x^1_k$ (since $ijk \in U_1$ and all equations are disjoint). This means, since $x^2 = x - x^1$, that $(x^2)_i \neq (x^2)_j = (x^2)_k$. We will show that $i \in I_{U_2}(x^2) \subseteq I_{U_2}(L_2)$. If $\{ijk\} \subseteq U_2$ then this holds. If not, then it must be that $(x^2)_{i,j,k} = 100$ (the option of $(x^2)_{i,j,k} = 011$ is ruled out because U_2 cannot have equations containing j and k without containing i, j, k. This is because we dropped the illegitimate k-tuples of equations.). Since $(x^2)_i = 1$, there must be some equation $\{i, j', k'\}$ that is contained in U_2, such that $j', k' \notin \{i,j,k\}$. It remains to observe that $(x^2)_{j',k'} = 00$. This is because $j', k' \notin U_1$ implies that $x_{j',k'} = 00 = (x^1)_{j',k'}$. Therefore, $(x^2)_{j',k'} = 00$ so $(x^2)_i \neq (x^2)_{j'} = (x^2)_{k'}$ and thus $i \in I_{U_2}(x^2)$. □

Definition 4.7. A vector $z \in L \subseteq X_U$ is called reduced w.r.t. $C = C(U, L)$ if in every equation $e \subseteq U$ with variables i, j, k such that $e \not\subseteq I(C)$, $z_{i,j,k}$ has at most one 1 coordinate.

Clearly for every $x \in X_U$ there is some $y \in H_U$ such that $r = x - y$ is reduced. Moreover,

Claim 4.8. *For every $(U, L) \in A$ there is some $(U, R) \in \mathcal{C}(U, L)$ such that R is a subspace all of whose vectors are reduced.*

Proof. Let $b_1, \ldots, b_\ell$ be a basis for L, and let $r_i = b_i + y$ for $y \in H_U$ and r_i reduced. Let $R = \mathsf{Span}(r_1, \ldots, r_\ell)$. Since $L \cap H = 0$ we get $\mathsf{dim}(R) = \mathsf{dim}(L)$. Finally, note also that if r, r' are reduced, then $r + r'$ is reduced as well: let e be an equation with variables indexed i, j, k in U. Since r and r' have at most one 1 coordinate in $\{i, j, k\}$, $r + r'$ has at most two 1 coordinates in $\{i, j, k\}$. But if it has two, the variable with 0 coordinate is singled out in $r + r'$, while the two other variables are singled out in r and r', which implies that $e \subseteq I(\mathcal{C}(U, L))$. □

We can now prove Lemma 4.2.

Proof. Fix (U, L) and let $\mathcal{C} = \mathcal{C}(U, L)$. By definition

$$\mathcal{C}(U, L) = \{(U', L') \mid L + H_U + H_{U'} = L' + H_U + H_{U'}\} .$$

Let R be a reduced space such that $(U, R) \in \mathcal{C}(U, L)$, namely $R + H_U = L + H_U$. Such a space R exists by the previous claim. We must prove

$$\mathcal{C}(U, L) = \{(U', L') \in A \mid L' + H_{U'} = R + H_{U'}\} .$$

The $\supseteq$ direction is clear because $L' + H_{U'} = R + H_{U'}$ implies

$$L' + H_{U'} + H_U = R + H_{U'} + H_U = L + H_{U'} + H_U .$$

To prove $\subseteq$ we will show that every $(U', L') \in \mathcal{C}(U, L)$ indeed satisfies that $L' + H_{U'} = R + H_{U'}$. Let R' be a reduced space such that $R' + H_{U'} = L' + H_{U'}$. We first claim that $R + H_{\mathcal{C}} = R' + H_{\mathcal{C}}$, where $H_{\mathcal{C}} = \mathsf{Span}\{x_e \mid e \subseteq I(\mathcal{C})\}$. Let $H = H_U + H_{U'}$. We know

$$R + H = L + H = L' + H = R' + H ,$$

so for any vector $r \in R$ we can write $r = r' + y$ where $r' \in R'$, $y \in H$. Since both r, r' are reduced, we claim that y in fact must be in $H_{\mathcal{C}} = \{x_e \mid e \subseteq I(\mathcal{C})\}$. Since $I(\mathcal{C}) = I_U(L) = I_{U'}(L')$ by Claim 4.6, clearly $H_{\mathcal{C}} \subseteq H_U \cap H_{U'}$. Moreover, let $e = \{i, j, k\} \subseteq U \cup U'$ be an equation such that $y_{ijk} \neq 000$, so y must have at least two 1's in coordinates ijk (because if $\{ijk\} \subseteq U$ there can be at most one equation from U' that intersects $\{ijk\}$ non-trivially, and this equation must have exactly one of i, j, or k. Similarly if $\{ijk\} \subseteq U'$.). Since r, r' are reduced and $y = r + r'$ has weight more than 1 on $\{ijk\}$, it follows that they single out two distinct indices from $\{ijk\}$ so $\{ijk\} \subseteq I(\mathcal{C})$. Thus $R \subseteq R' + H_{\mathcal{C}}$ which proves, by symmetry, that $R + H_{\mathcal{C}} = R' + H_{\mathcal{C}}$.

This proves the first part of the lemma because $L' + H_{U'} = R' + H_{U'} = R + H_{U'}$.

For the "furthermore" part, suppose that for $i = 1, 2$, $\mathcal{C}_i$ is a class with a reduced space R_i. We show that if $\mathcal{C}_1 \cap \mathcal{C}_2 \neq \emptyset$, then $\mathcal{C}_1 = \mathcal{C}_2$. So let $(U, L) \in \mathcal{C}_1 \cap \mathcal{C}_2$. Then $R_1 + H_U = L + H_U = R_2 + H_U$. We see that $(U, R_2) \in \mathcal{C}_1$ and by Claim 4.6 $I(\mathcal{C}_1) = I_U(R_1) = I_U(R_2) = I(\mathcal{C}_2)$. By arguments identical to those in the first part of the proof, this implies that $R_1 + H_{12} = R_2 + H_{12}$ where $H_{12} = \mathsf{Span}\{x_e \mid e \subseteq I(\mathcal{C}_1) = I(\mathcal{C}_2)\}$. Now every $(U_1, L_1) \in \mathcal{C}_1$ obeys $L_1 + H_{U_1} = R_1 + H_{U_1}$ and since $H_{12} \subseteq H_{U_1}$, we can plug in $R_1 + H_{12} + H_{U_1} = R_2 + H_{12} + H_{U_1} = R_2 + H_{U_1}$ showing that $(U_1, L_1) \in \mathcal{C}_2$, so $\mathcal{C}_1 \subseteq \mathcal{C}_2$ and by symmetry $\mathcal{C}_1 = \mathcal{C}_2$. □

4.4 Covering property

We state the covering property as described in [26, Lemmas 4.6, 4.7]. It shows that two distributions over ℓ-dimensional subspaces of X_U are close in statistical distance.

Definition 4.9. Let $U \in \mathcal{U}$ and let $\ell \geqslant 1$ be an integer. Let $\mathcal{L}$, $\mathcal{L}'$ be distributions over ℓ-dimensional subspaces of X_U sampled as follows.

- $\mathcal{L}$: Choose a uniformly random ℓ-dimensional subspace of X_U.
- $\mathcal{L}'$: Choose a random $V \subseteq U$ as in the edge distribution of G_{unfolded}, and then choose a uniformly random ℓ-dimensional subspace of X_V.

Lemma 4.10. *Suppose $2^\ell \beta \leqslant \frac{1}{8}$. Let $\mathcal{L}$, $\mathcal{L}'$ be distributions over ℓ-dimensional subspaces over X_U sampled as in Definition 4.9. 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 4.11. *Let $0 \leqslant q \leqslant \ell - 1$ be an integer. Let Q be q-dimensional subspace of X_U and let $\mathcal{L}_Q$ and $\mathcal{L}'_Q$ be the 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,*

$$\mathsf{SD}(\mathcal{L}_Q,\ \mathcal{L}'_Q) \leqslant \sqrt{\beta}\, k^{\frac{1}{4}} \cdot 2^{\ell+5}. \tag{4.3}$$

We call such subspaces Q smooth.

5 Completeness and soundness of the construction

5.1 Completeness

Let us now prove the completeness of the construction, stated in the following lemma.

Lemma 5.1 (Completeness). *Suppose there is an assignment for the variables of the 3-Lin instance, (X, Eq), satisfying a $1 - \varepsilon$ fraction of the equations. Then there is an assignment to at least a $1 - k\varepsilon$ fraction of the vertices in G_{folded} such that all constraints induced on these variables are satisfied. This assignment satisfies a $1 - k\varepsilon$ fraction of the constraints.*

We prove the lemma in the remaining of this section. We first (easily) define an assignment for G_{unfolded}, $\mathcal{A} : A \to \{0,1\}^\ell$ and $\mathcal{B} : B \to \{0,1\}^{\ell-1}$ that satisfies $1 - k\varepsilon$ of the constraints. We will then convert $\mathcal{A}$ to $\tilde{\mathcal{A}}$, so that $\tilde{\mathcal{A}}, \mathcal{B}$ becomes an assignment for G_{folded}, and show that it satisfies many of the constraints of G_{folded}. This part involves some unusually non-trivial arguments.

Let $a : X \to \mathbb{F}_2$ be the assignment for (X, Eq) promised in the lemma, and let $\mathsf{Eq}' \subseteq \mathsf{Eq}$ be the set of equations that are *not* satisfied by a. We will also view a as a linear function $a : \mathbb{F}_2^n \to \mathbb{F}_2$ by setting $a(z_1, \ldots, z_n) := \langle a, z \rangle$ (here $\langle \cdot, \cdot \rangle$ denotes inner product over $\mathbb{F}_2$).

For every U, if all of the equations in U are not in Eq', then we assign $\mathcal{A}(U, L) := a|_L$. If U involves some unsatisfied equation, it seems tempting to leave it unassigned since these amount to a $k\varepsilon$ fraction of $\mathcal{U}$, at most. This would be fine for G_{unfolded} however when we move to G_{folded}, it turns out likely that nearly every equivalence class C contains some U that has an equation in Eq'. So we cannot afford to ignore these. Instead, we assign all (U, L) except those for which $I_U(L)$ contains some equation from Eq'. Similarly we assign all (V, L') except those for which $I_V(L')$ contains some equation from Eq'.

Let (U, L) be such that $I_U(L)$ contains no equation from Eq'. Every $x \in L$ can be written uniquely as a sum $x = x_I + y$ where

$$x_I \in \mathsf{Span}\,\{x_i \mid i \in I(C)\} \quad \text{and} \quad y \in \mathsf{Span}\,\{x_e \mid e \subseteq U, e \not\subset I(C)\} \subseteq H_U\,.$$

(Uniqueness is because these spaces intersect at $\{0\}$: the equations $e \in U$ are pairwise disjoint and for each $e \not\subset I(C)$ there can be at most a single $i \in I \cap e$.)

We define $\sigma : L \to \mathbb{F}_2$ by setting for each $x \in L$,

$$\sigma(x) := \langle a, x_I \rangle + h(y)\,,$$

where $h : H \to \mathbb{F}_2$ is the linear function defined by the RHS of the equations in U, that is $h(x_e) = b_e$. We define $\mathcal{A}(U, L) = \sigma$. The assignment to (V, L') is defined similarly. This completes the description of the assignment $(\mathcal{A}, \mathcal{B})$ for G_{unfolded}.

Claim 5.2. *For (U, L) for which $\mathcal{A}(U, L)$ is defined, all of the constraints of G_{unfolded} involving (U, L) are satisfied.*

Proof. Note that if (U, L) is assigned then so is (V, L') since $V \subseteq U$ and $L' \subseteq L$ implies that $I_U(L) \supseteq I_V(L')$. It follows directly from the definition that the assignments are consistent. □

We convert $\mathcal{A}$ to $\tilde{\mathcal{A}}$ using the assignment for the representatives: If $C \in \tilde{A}$ is such that $I(C)$ contains all three variables from some $e \in \mathsf{Eq}'$ we will *not* assign it. For every other C, we let $\tilde{\mathcal{A}}(C) := \mathcal{A}(U_0, L_0)$ where (U_0, L_0) is the representative of C. It remains to check two things,

1. For each assigned C, the unfolding of $\tilde{\mathcal{A}}(C)$ coincides with $\mathcal{A}$ for every $(U, L) \in C$. This implies, together with Claim 5.2, that for every assigned C, all of the constraints incident on it are satisfied.

2. When choosing a random edge in G_{unfolded}, the probability that both endpoints are assigned is at least $1 - k\varepsilon$.

To see the first item, fix C and a representative (U_0, L_0) and let us unfold $\sigma : L_0 \to \mathbb{F}_2$ to some (U, L). To do this we look at $g_U : L_0 + H_{U_0} + H_U \to \mathbb{F}_2$ the unique linear function that extends σ. Tracing the definitions we see that the restriction of g_U to L is equal to $\mathcal{A}(U, L)$.

To see the second item, we observe that a random (U, L) is assigned as long as $I_U(L)$ contains no equation from Eq', but $I_U(L) \subseteq U$ and U itself contains no equation from Eq' with probability at least $1 - k\varepsilon$.

5.2 Soundness

Lemma 5.3 (Soundness). *Assume Hypothesis 3.6. For every $\delta > 0$ there exists large enough $\ell \ll k$, such that given an assignment $\tilde{\mathcal{A}}, \mathcal{B}$ for G_{folded} that satisfies a δ fraction of the constraints, there is an assignment for the 3-Lin instance $(\mathsf{X}, \mathsf{Eq})$ that satisfies more than an s^* fraction of the equations.*

Proof. Let $\tilde{\mathcal{A}}, \mathcal{B}$ be assignments that satisfy at least δ fraction of the constraints. Let $\mathcal{A}$ be the unfolding of $\tilde{\mathcal{A}}$, and from now on we consider the assignment $\mathcal{A}, \mathcal{B}$ for G_{unfolded}.

For each $U \in \mathcal{U}$, let $F_U[\cdot]$ be an assignment for $Gr(X_U, \ell)$ defined by $F_U[L] = \mathcal{A}(U, L)$. Let $h_U : H_U \to \mathbb{F}_2$ be the function defined by $h(x_e) = b_e$ for each $e \in U$. It is clear by unpacking the definitions that the table $F_U[\cdot]$ is invariant under (H_U, h_U) (explicitly: If $L_1 + H_U = L_2 + H_U$ then both are equal to $R + H_U$ and $\tilde{\sigma} : R + H_U \to \mathbb{F}_2$ must equal the (H_U, h_U) extension of $\mathcal{A}(U, L_1)$ as well as the (H_U, h_U) extension of $\mathcal{A}(U, L_2)$.).

We also define, for each V, an assignment $F_V[\cdot]$ for $Gr(X_V, \ell)$ as follows. For each $L \in Gr(X_V, \ell)$, we define $F_V[L] := \mathcal{A}(U, L)$ where $U \supseteq V$ can be chosen arbitrarily: From Claim 4.4 we see that this definition does not depend on the choice of U because $\mathcal{A}$ is folded.

Let $C > 0$, q, and r be the numbers promised in Lemma 3.17 for agreement $\delta^2/8$. Set $\tau_r = C$ and $\tau_i = 10^{-9}(\tau_{i+1})^{12}$, for $0 \leqslant i < r$. Furthermore, set $\eta_r = \tau_0/2$ and $\eta_i = 10^{-9}(\eta_{i+1})^{12}$. Let $\vec{\tau} = (\tau_0, \ldots, \tau_r)$ and let $\vec{\eta} = (\eta_0, \ldots, \eta_r)$.

For every U and every subspace $Q \subseteq X_U$ let $\mathsf{LIST}^{\vec{\tau}}_Q(F_U)$ be the list of maximal pairs, as per Definition 3.16. Analogously, for every V and every subspace $Q \subseteq X_V$ let $\mathsf{LIST}^{\vec{\eta}}_Q(F_V)$ be the list of corresponding maximal pairs.

Outer PCP game. We prove soundness, following [26], by going through an outer PCP game as follows. Consider the following two-player game that is based on the initial 3-Lin instance $(\mathsf{X}, \mathsf{Eq})$ and the parameters $q \in \mathbb{N}$ and $\beta > 0$.

- The verifier chooses U, V as in the edge distribution of G_{unfolded}. That is U is selected uniformly in $\mathcal{U}$ and V comes from replacing each equation with probability β by a single variable.
- The verifier chooses $Q \in Gr(X_V, q)$ uniformly and sends (U, Q) to the first player, and (V, Q) to the second player.
- Player 1 answers with $a \in \{0,1\}^U$ and player 2 answers with $b \in \{0,1\}^V$. The verifier accepts iff $a|_V = b$ and a satisfies all of the linear equations on U.

This game is essentially a parallel repetition of the 3-Lin instance. If the players were not given Q then clearly if $(\mathsf{X}, \mathsf{Eq})$ is far from satisfiable, then the players cannot succeed with probability more than $\exp(-\beta k)$. The subspace Q does reveal some information to the players, but not too much (essentially, Q reveals less than $(1-\exp(-q))$ fraction of the question-pairs, and this leaves sufficiently many hard question pairs). In particular, the following result is established in [26, Section 3.4].

Lemma 5.4 (Soundness of outer PCP [26]). *If every assignment for* $(\mathsf{X}, \mathsf{Eq})$ *satisfies at most an* $s^* < 1$ *fraction of the equations, then the players have no strategy that succeeds with probability better than* $\exp(-\beta k/\exp(q))$. □

We note that if $\beta \gg 1/k$, then the soundness approaches 0 as k approaches infinity. On the other hand, the covering property from Section 4.4 is only effective for $\beta \ll \frac{1}{\sqrt{k}}$. Any $\frac{1}{k} \ll \beta \ll \frac{1}{\sqrt{k}}$ will work for us, and for the sake of concreteness we pick $\beta = \frac{\log\log k}{k}$.

Our proof of soundness will proceed by extracting from $\tilde{\mathcal{A}}, \mathcal{B}$ a strategy for the players in this game that succeeds with probability greater than an ε. We will use the following auxiliary lemma.

Lemma 5.5. *Let* U *be a fixed question to the first prover, let* $Q \subseteq \mathbb{F}_2^n$ *have dimension* q *and let* $W \subseteq X_U$ *have codimension* r. *Let* $V \subseteq U$ *be a random question to the second prover in the outer PCP game conditioned on* Q. *Then, with probability at least* $1 - 2^{3q+r+1}\beta$ *we have* $\mathsf{dim}(W \cap X_V) = |V| - r$.

Proof. Let V_i, U_i denote the set of variables in V, U that appear in the ith equation the verifier chooses. Note that the value of V_i depends only on the projection of the vectors in Q onto the coordinates in U_i. There are at most 2^{3q} options for this projection and each one is attained with probability at least $(1-\beta)2^{-3q}$, so we get $\Pr_V[V_i \neq U_i \mid Q] \leqslant \frac{2^{3q}}{1-\beta}\beta \leqslant 2^{3q+1}\beta$. Next think of $W \subseteq X_U$ as a subspace defined by equations $\langle u_j, x\rangle = 0$ for $j = 1, \dots, r$. The event that $\mathsf{dim}(W \cap X_V) < |V| - r$ is the event that $\{e_a\}_{a \in U \setminus V} \cup \{u_1, \dots, u_r\}$ is linearly dependent, where $e_a \in \mathbb{F}_2^n$ is the vector which is 1 only at coordinate a. For that to happen there needs to be a linear combination of the u_j whose non-zero coordinates are all in $U \setminus V$. The number of linear combination of u_j is 2^r, and for each one of them the probability its support is in $U \setminus V$ is at most $2^{3q+1}\beta$ by our earlier observation. The result now follows from the union bound. □

Strategies of the players. The first player, upon getting question (U, Q), chooses a random element of $(f, W) \in \mathsf{LIST}_Q^{\vec{\tau}}(F_U)$, where $f : W \longrightarrow \mathbb{F}_2$ is a linear function and $Q \subseteq W \subseteq X_U$ has codimension at most r. The player chooses r bits to randomly extend f to a linear function on all of X_U, and outputs that as answer. More accurately, the player outputs a Boolean assignment for U that corresponds to this linear function. If $\mathsf{LIST}_Q(F_U) = \emptyset$ the player outputs a random assignment for U that satisfies the equations of U.

The second player does the same: upon getting question (V, Q) it randomly selects an element from $\mathsf{LIST}_Q^{\vec{\eta}}(F_V)$, randomly extends it to X_V and outputs that as the answer. If $\mathsf{LIST}_Q^{\vec{\eta}}(F_V) = \emptyset$ the player outputs a random assignment for V.

Analyzing the success probability of the provers. We consider the following events that depend on the choice of U and Q:

1. Let E_1 be the event that $\mathsf{agreement}(F_U) \geqslant \delta^2/8$. Then $\Pr[E_1] \geqslant \delta/2$.

2. Let E_2 be the event $Q \in Gr(X_U, q)$ is smooth, namely Eq. (4.3) holds for Q. Then for every U, sampling $Q \in Gr(X_U, q)$ uniformly we have, by Lemma 4.11, that $\Pr_{Q \in Gr(X_U, q)}[E_2] \geqslant$

$1 - \sqrt{\beta}k^{1/4}$. Recall that Q is sampled by taking $V \subseteq U$ as in the outer PCP game and then sampling $Q \in Gr(X_V, q)$ uniformly. Applying Lemma 4.10 (with ℓ there being q) this distribution is $\beta\sqrt{k}2^{q+4}$-close to uniform over $Q \in Gr(X_U, q)$, hence we conclude that sampling Q as in the outer PCP game we have $\Pr[E_2] \geqslant 1 - \sqrt{\beta}k^{1/4} - \beta\sqrt{k}2^{q+4} \gg 1 - \delta\alpha(\ell)/8$.

3. Let E_3 be the event that $\mathsf{LIST}^{\vec{\tau}}_{Q}(F_U) \neq \emptyset$. Then $\Pr[E_3|E_1] \geqslant 3\alpha(\ell)/4$.

Altogether this implies that $\Pr[E_1 \wedge E_3] \geqslant \delta/2 \cdot 3\alpha(\ell)/4$, and $\Pr[\neg E_2] \leqslant \delta\alpha(\ell)/8$ so

$$\Pr_{U,Q}[E_1 \wedge E_2 \wedge E_3] \geqslant \Pr[E_1 \wedge E_3] - \Pr[\neg E_2] \geqslant \delta\alpha(\ell)/4\,. \tag{5.1}$$

Before we prove the three items leading to Eq. (5.1) let us see how it implies soundness. Assume that $E_1 \wedge E_2 \wedge E_3$ holds and suppose the first player chooses $(f, W) \in \mathsf{LIST}^{\vec{\tau}}_{Q}(F_U)$ and answers according to it. Recall that the player also tosses $i = \mathsf{codim}(W)$ additional random coins to complete f to a function on all of X_U. Henceforth, we fix Q and U.

By Lemma 3.19, $W \supseteq H_U$ and $f|_{H_U} = h_U$ (i. e., f satisfies the side condition), so regardless of the $i \leqslant r$ coin tosses, the answer of the first player satisfies all of the equations of U. We now prove that there is a good chance that the second player's answer is consistent with f. Since $(f, W) \in \mathsf{LIST}^{\vec{\tau}}_{Q}(F_U)$ we know that

$$\Pr_{Q \subseteq L \subseteq W}[F_U[L] = f|_L] \geqslant \tau_i\,,$$

meaning that $\Pr_{Q \subseteq L}[F_U[L] = f|_L, L \subseteq W] \geqslant \tau_i \Pr_{Q \subseteq L}[L \subseteq W]$. By Lemma 4.11 and since E_2 holds, we get

$$\Pr_{V, Q \subseteq L \subseteq X_V}[F_U[L] = f|_L, L \subseteq W] \geqslant \tau_i \Pr_{Q \subseteq L}[L \subseteq W] - 2^{\ell+5}\sqrt{\beta}k^{1/4},$$

which is at least $0.9\tau_i 2^{-i(\ell-q)}$ as $\Pr_{Q \subseteq L}[L \subseteq W] \geqslant 0.99 \cdot 2^{-i(\ell-q)}$ and τ_i are both much larger compared to $2^{\ell}\sqrt{\beta}k^{1/4}$. By Lemma 5.5 we further conclude that

$$\Pr_{V, Q \subseteq L \subseteq X_V}[F_U[L] = f|_L, L \subseteq W, \mathsf{dim}(W \cap X_V) = |V| - i] \geqslant 0.9\tau_i 2^{-i(\ell-q)} - 2^{3q+i+1}\beta \geqslant 0.8\tau_i 2^{-i(\ell-q)}\,.$$

For each fixed V the probability on the left hand side is at most $2^{-i(\ell-q)}$, so we get from an averaging argument that with probability at least $0.3\tau_i$ over the choice of V we have that $\mathsf{dim}(W \cap X_V) = |V| - i$ and $\Pr_{Q \subseteq L \subseteq X_V}[F_U[L] = f|_L, L \subseteq W] \geqslant 0.5\tau_i 2^{-i(\ell-q)}$, implying

$$\Pr_{Q \subseteq L \subseteq W \cap X_V}[F_V[L] = f|_L] \geqslant \frac{1}{2}\tau_i\,.$$

In that case $(f|_{W \cap X_V}, W \cap X_V)$ $\vec{\eta}$-occurs w.r.t. Q in F_V for the parameter vector $\vec{\eta}$ as chosen above. In particular, there is some $(f_1, W_1) \in \mathsf{LIST}^{\vec{\eta}}_{Q}(F_V)$ such that $W_1 \supseteq W \cap X_V$ and $f_1|_{W \cap X_V} = f|_{W \cap X_V}$. The second player chooses (f_1, W_1) with probability at least $1/M$ where $M = \left|\mathsf{LIST}^{\vec{\eta}}_{Q}(F_V)\right| \leqslant$

$2^q 2^{8r^2\ell}\eta_r^{-\exp(r)}$; in the last inequality we used Lemma 3.17. If this happens, then with probability at least $2^{-i} \geqslant 2^{-r}$ over the i random choices of the first player, the players win. We have shown that

$$\Pr\left[\text{The players win} \mid E_1 \wedge E_2 \wedge E_3\right] \geqslant 2^{-r} \cdot 0.3\tau_i \cdot 2^{-q}2^{-8r^2\ell}\eta_r^{\exp(r)} > 0\,,$$

which depends on δ and ℓ but are independent of k. So for large enough k this contradicts the soundness of Lemma 5.4.

It remains to prove the three items leading to Eq. (5.1).

Proof of items 1,2 and 3 Let $p(U)$ denote the fraction of constraints satisfied by $\mathcal{A}, \mathcal{B}$ after picking U. So $\mathbb{E}_U\left[p(U)\right] \geqslant \delta$. By an averaging argument, for at least $\frac{\delta}{2}$ fraction of the tuples U, we have $p(U) \geqslant \frac{\delta}{2}$. We will show

Claim 5.6. $\mathsf{agreement}(F_U) \geqslant p(U)^2 - O(\sqrt{\beta}k^{\frac{1}{4}} \cdot 2^{\ell})$.

So with probability $\geqslant \delta/2$ over the choice of U we have $p(U) \geqslant \delta/2$, which implies that E_1 holds. This implies item 1. Item 2 is argued above. Lastly, for item 3, if E_1 holds then by Lemma 3.17 there is probability at least $\alpha(\ell)$ over the choice of $Q \in Gr(X_U, q)$ that $\mathsf{LIST}^{\vec{\tau}}_{Q}(F_U) \neq \emptyset$. Using Lemma 4.10 (with ℓ there being q), the distribution of Q in the outer PCP game is $\beta\sqrt{k}2^q$-close to uniform over $Gr(X_U, q)$, implying $\Pr\left[E_3|E_1\right] \geqslant \alpha(\ell) - \beta\sqrt{k}2^q \geqslant 3\alpha(\ell)/4$.

Proof of Claim 5.6. For a fixed U, consider the distribution over $(V, L') \in B$, conditioned on U. The probability $p(U)$ is equal to the probability that we choose $V \subseteq U$ and then L' a subspace of V, and then a random space $L \supseteq L'$ such that $L \subseteq X_U$. By the covering property (Lemma 4.10), if Q is smooth (which is implied by E_2) then L' is distributed nearly uniformly in $Gr(X_U, \ell - 1)$ and we will pretend it is exactly uniformly (but we include this statistical distance in the final result). Define a randomized assignment for L' by selecting a random $V \subseteq U$ conditioned on $L' \subseteq X_V$, and setting $F'_U[L']$ to be $\mathcal{B}(V, L')$.

Let L, L' be chosen by first choosing a random $L \in Gr(X_U, \ell)$ and then a random $L' \subseteq L$ of dimension $\ell - 1$. We claim that

$$\Pr_{L \supseteq L'}\left[F_U[L]|_{L'} = F'_U[L']\right] \geqslant p(U) - O(\sqrt{\beta}k^{\frac{1}{4}} \cdot 2^{\ell})\,. \tag{5.2}$$

To see that, let $\mathcal{D}$ be the distribution on L, L', V of choosing a random edge in G_{unfolded} conditioned on U. Then we know that $p(U) = \Pr_{\mathcal{D}}\left[F_U(L) = \mathcal{B}(V, L')\right]$. Let $\mathcal{D}'$ be the distribution where we first choose $L' \in Gr(X_U, \ell - 1)$ and then $L \supseteq L'$ and finally V such that $V \subseteq U$ and $X_V \supseteq L'$. The covering property implies that $\mathcal{D} \approx \mathcal{D}'$ and the error is negligible so we ignore it. Also, the answer $\mathcal{B}(V, L')$ when $(L, L') \sim \mathcal{D}'$ is distributed exactly as in the definition of $F'_U[L']$. So we have

$$\begin{aligned} p(U) = \Pr_{\mathcal{D}}\left[F_U(L) = \mathcal{B}(V, L')\right] &\geqslant \Pr_{\mathcal{D}'}\left[F_U(L) = \mathcal{B}(V, L')\right] - O(\sqrt{\beta}k^{\frac{1}{4}} \cdot 2^{\ell}) \\ &= \Pr_{\mathcal{D}'}\left[F_U(L) = F'_U[L']\right] - O(\sqrt{\beta}k^{\frac{1}{4}} \cdot 2^{\ell})\,, \end{aligned}$$

proving Equation (5.2).

For each $L' \in Gr(X_U, \ell - 1)$ let $p(L') = \Pr_{L \supseteq L'}\left[F_U[L]|_{L'} = F'_U[L']\right]$ be the probability that the two-function test succeeds conditioned on L'. Also let $q(L') = \Pr_{L_1, L_2 \supseteq L'}\left[F_U[L_1]|_{L'} = F_U[L_2]|_{L'}\right]$. Clearly $q(L') \geqslant p(L')^2$ because whenever two spaces $L_1, L_2 \supseteq L'$ agree with $F[L']$ they agree with each other. So the probability that the agreement test passes equals,

$$\mathbb{E}_{L'}[q(L')] \geqslant \mathbb{E}_{L'}[p(L')^2] \geqslant [\mathbb{E}_{L'} p(L')]^2 = p(U)^2 - O(\sqrt{\beta} k^{\frac{1}{4}} \cdot 2^{\ell})\,.$$

□

□

5.3 Proof of (Main) Theorem 1.8

Fix $\delta > 0$. For this δ, Hypothesis 3.6 guarantees a global linear function as long as ℓ and k are large enough and with constants $r, q \in \mathbb{N}$ and $C > 0$. We choose k large enough so that soundness holds, and then choose the completeness parameter ε in the initial 3-Lin instance small enough so that $1 - k\varepsilon \geqslant 1 - \delta$, so completeness holds as well. The reduction from the 3-Lin instance to G_{folded} together with the completeness and soundness lemmas (Lemmas 5.1 and 5.3) prove Theorem 1.8.

A Missing proofs

A.1 Proof of Corollary 1.12

Fix $\varepsilon > 0$, and take a 2-to-1 games instance $G = (U, E, \Sigma, \Pi)$ from Theorem 1.8. Define $2k = |\Sigma|$, and construct a Unique Games instance $G' = (U \cup U', E', \Phi, \Sigma)$ as follows. The vertex set consists of U, U' that are two copies of U. For each $u \in U$ we denote by u, u' its copies. The edge set is $E' = \{(u, v), (u', v') \mid u \in U, v \in V\}$. To define the constraints, let us write each constraint $\pi \in \Pi$ slightly differently.

The constraint π on edge (a, b) can be stated as the existence of two partitions $[2k] = \left\{\alpha_1, \alpha'_1, \ldots, \alpha_k, \alpha'_k\right\}$, $[2k] = \left\{\beta_1, \beta'_1, \ldots, \beta_k, \beta'_k\right\}$ such that

$$\pi = \left\{(\alpha_i, \beta_i), (\alpha'_i, \beta_i), (\alpha'_i, \beta_i), (\alpha'_i, \beta'_i) \,\middle|\, i = 1, \ldots, k\right\}\,.$$

We define the constraint on (u, v) to be $\phi(u, v) = \left\{(\alpha_i, \beta_i), (\alpha'_i, \beta'_i) \,\middle|\, i = 1, \ldots, k\right\}$ and on (u', v') to be $\phi(u', v') = \left\{(\alpha'_i, \beta_i), (\alpha_i, \beta'_i) \,\middle|\, i = 1, \ldots, k\right\}$.

Completeness: If A is an assignment to G satisfying $1 - \varepsilon$ of the constraints, then assign G' according to it: for $u \in U$, assign $B(u) = B(u') = A(u)$. Then clearly if A satisfies an edge (u, v) then either (u, v) or (u', v') is satisfied by B in G', and so B satisfies $\frac{1}{2}(1 - \varepsilon)$ fraction of the constraints of G'.

Soundness: Let B be an assignment for G' satisfying a δ fraction of the constraints. Consider the two assignments to G, A, A' defined by $A(u) = B(u)$, $A'(u) = B(u')$ for every $u \in U$. Since

B satisfies a δ fraction of the edges of G, it must satisfy either at least a δ fraction of the edges (u, v) or at least a δ fraction of the edges (u', v'). In the former we get that A satisfies at least a δ fraction of the constraints in G and in the latter we get that A' satisfies at least a δ fraction of the constraints in G. In any case, since G is at most ε satisfiable, we get that $\delta \leqslant \varepsilon$.

With the above completeness and soundness, we take G' a fully satisfiable Unique Games instance on a disjoint set of vertices which has a 2ε fraction of the constraints that G has, and produce the game $G \cup G'$. Thus, in the completeness case one can satisfy at least

$$\frac{1}{1+2\varepsilon}\left(\frac{1}{2}-\varepsilon\right)+\frac{2\varepsilon}{1+2\varepsilon} \geqslant \frac{1}{2}$$

of the constraints, and in the soundness one can satisfy at most a 3ε fraction of the constraints.

A.2 Proof of Corollary 1.13

The reduction is identical to the reduction of [28] but starting with Cor. 1.12. In the YES case, with probability $x \geqslant \frac{1}{4} - O(\varepsilon)$ the test checks two consistent long-code assignment and those contribute at least $\frac{1}{2} + \varepsilon$ to the cut size. With probability $1 - x$ the test checks two inconsistent long-code assignments, and those contribute $\frac{1}{2}$ to the cut size (this is verified easily by observing that $\langle \chi_i, A\chi_j \rangle = 0$ for every $i \neq j$ where A is the Davie-Reeds operator from [28]). Therefore the maximum cut size is at least

$$x\left(\frac{1}{2}+\varepsilon\right)+(1-x)\frac{1}{2} \geqslant \frac{1}{2}+\Omega(\varepsilon)\,.$$

The analysis of the NO case is identical to [28].

Acknowledgements

We thank Noga Alon, Yuval Filmus, Dana Moshkovitz, and Ran Raz for several helpful discussions. We thank Boaz Barak and Pravesh Kothari for helpful comments.

References

[1] SANJEEV ARORA, LÁSZLÓ BABAI, JACQUES STERN, AND Z. SWEEDYK: The hardness of approximate optima in lattices, codes, and systems of linear equations. *J. Comput. System Sci.*, 54(2):317–331, 1997. [doi:10.1006/jcss.1997.1472] 3

[2] SANJEEV ARORA, CARSTEN LUND, RAJEEV MOTWANI, MADHU SUDAN, AND MARIO SZEGEDY: Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998. [doi:10.1145/278298.278306] 3

[3] SANJEEV ARORA AND SHMUEL SAFRA: Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998. [doi:10.1145/273865.273901] 3

[4] SANJEEV ARORA AND MADHU SUDAN: Improved low-degree testing and its applications. *Combinatorica*, 23(3):365–426, 2003. [doi:10.1007/s00493-003-0025-0] 11, 18

[5] PER AUSTRIN, RYAN O'DONNELL, LI-YANG TAN, AND JOHN WRIGHT: New NP-hardness results for 3-coloring and 2-to-1 label cover. *ACM Trans. Comput. Theory*, 6(1):2:1–20, 2014. [doi:10.1145/2537800] 15

[6] BOAZ BARAK, PRAVESH K. KOTHARI, AND DAVID STEURER: Small-set expansion in short-code graph and the 2-to-2 conjecture. In *Proc. 10th Innovations in Theoret. Comp. Sci. Conf. (ITCS'19)*, pp. 9:1–12. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.ITCS.2019.9] 4, 15

[7] MIHIR BELLARE, ODED GOLDREICH, AND MADHU SUDAN: Free bits, PCPs, and non-approximability – Towards tight results. *SIAM J. Comput.*, 27(3):804–915, 1998. [doi:10.1137/S0097539796302531] 3

[8] IRIT DINUR AND ELAZAR GOLDENBERG: Locally testing direct products in the low error range. In *Proc. 49th FOCS*, pp. 613–622. IEEE Comp. Soc., 2008. [doi:10.1109/FOCS.2008.26] 11, 19

[9] IRIT DINUR, SUBHASH KHOT, GUY KINDLER, DOR MINZER, AND MULI SAFRA: Towards a proof of the 2-to-1 games conjecture? In *Proc. 50th STOC*, pp. 376–389. ACM Press, 2018. [doi:10.1145/3188745.3188804] 1, 2, 4, 5

[10] IRIT DINUR, SUBHASH KHOT, GUY KINDLER, DOR MINZER, AND MULI SAFRA: On non-optimally expanding sets in Grassmann graphs. *Israel J. Math.*, 243(1):377–420, 2021. [doi:10.1007/s11856-021-2164-7] 2, 15

[11] IRIT DINUR, SUBHASH KHOT, WILL PERKINS, AND MULI SAFRA: Hardness of finding independent sets in almost 3-colorable graphs. In *Proc. 51st FOCS*, pp. 212–221. IEEE Comp. Soc., 2010. [doi:10.1109/FOCS.2010.84] 14

[12] IRIT DINUR AND OMER REINGOLD: Assignment testers: Towards combinatorial proofs of the PCP theorem. *SIAM J. Comput.*, 36(4):975–1024, 2006. [doi:10.1137/S0097539705446962] 11

[13] IRIT DINUR AND SAMUEL SAFRA: On the hardness of approximating minimum vertex cover. *Ann. Math.*, 162(1):439–485, 2005. Accessible at JSTOR. 3

[14] IRIT DINUR AND DAVID STEURER: Analytical approach to parallel repetition. In *Proc. 46th STOC*, pp. 624–633. ACM Press, 2014. [doi:10.1145/2591796.2591884] 11, 19

[15] URIEL FEIGE, SHAFI GOLDWASSER, LÁSZLÓ LOVÁSZ, SHMUEL SAFRA, AND MARIO SZEGEDY: Interactive proofs and the hardness of approximating cliques. *J. ACM*, 43(2):268–292, 1996. [doi:10.1145/226643.226652] 3

[16] VENKATESAN GURUSWAMI, JOHAN HÅSTAD, AND MADHU SUDAN: Hardness of approximate hypergraph coloring. *SIAM J. Comput.*, 31(6):1663–1686, 2002. [doi:10.1137/S0097539700377165] 3

[17] JOHAN HÅSTAD: Clique is hard to approximate within $n^{1-\epsilon}$. *Acta Math.*, 182:105–142, 1999. Preliminary version in FOCS'96. [doi:10.1007/BF02392825] 3

[18] JOHAN HÅSTAD: Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001. [doi:10.1145/502090.502098] 3, 7, 33

[19] THOMAS HOLENSTEIN: Parallel repetition: Simplifications and the no-signaling case. *Theory of Computing*, 5(8):141–172, 2009. Preliminary version in STOC'07. [doi:10.4086/toc.2009.v005a008] 7

[20] RUSSELL IMPAGLIAZZO, VALENTINE KABANETS, AND AVI WIGDERSON: New direct-product testers and 2-query PCPs. *SIAM J. Comput.*, 41(6):1722–1768, 2012. [doi:10.1137/09077299X] 6, 11, 18, 19

[21] SUBHASH KHOT: On the power of unique 2-prover 1-round games. In *Proc. 17th IEEE Conf. on Comput. Complexity (CCC'02)*, pp. 25–25. IEEE Comp. Soc., 2002. [doi:10.1109/CCC.2002.1004334] 2, 3

[22] SUBHASH KHOT: Inapproximability of NP-complete problems, discrete Fourier analysis, and geometry. In *Proceedings of the International Congress of Mathematicians 2010*, pp. 2676–2697. World Scientific, 2010. Accessible at author's website. [doi:10.1142/9789814324359_0163] 2

[23] SUBHASH KHOT: On the unique games conjecture. In *Proc. 25th IEEE Conf. on Comput. Complexity (CCC'10)*, pp. 99–121. IEEE Comp. Soc., 2010. [doi:10.1109/CCC.2010.19] 2

[24] SUBHASH KHOT, DOR MINZER, DANA MOSHKOVITZ, AND MULI SAFRA: Small set expansion in the Johnson graph. *Theory of Computing*, 21(2):1–43, 2025. [doi:10.4086/toc.2025.v021a002, ECCC:TR18-078] 5

[25] SUBHASH KHOT, DOR MINZER, AND MULI SAFRA: Pseudorandom sets in Grassmann graph have near-perfect expansion. *Ann. Math.*, 198(1):1–92, 2023. [doi:10.4007/annals.2023.198.1.1] 2, 5

[26] SUBHASH KHOT, DOR MINZER, AND MULI SAFRA: On independent sets, 2-to-2 games, and Grassmann graphs. *Theory of Computing*, 21(10):1–56, 2025. Preliminary version in STOC'17. [doi:10.4086/toc.2025.v021a010] 2, 4, 5, 6, 8, 10, 11, 13, 18, 19, 23, 27, 38, 40, 41

[27] SUBHASH KHOT AND DANA MOSHKOVITZ: Candidate hard unique game. In *Proc. 48th STOC*, pp. 63–76. ACM Press, 2016. [doi:10.1145/2897518.2897531] 2

[28] SUBHASH KHOT AND RYAN O'DONNELL: SDP gaps and UGC-hardness for Max-Cut-Gain. *Theory of Computing*, 5(4):83–117, 2009. [doi:10.4086/toc.2009.v005a004] 14, 45

[29] SUBHASH KHOT AND MULI SAFRA: A two-prover one-round game with strong soundness. *Theory of Computing*, 9(28):863–887, 2013. [doi:10.4086/toc.2013.v009a028] 4

[30] RYAN O'DONNELL AND JOHN WRIGHT: A new point of NP-hardness for unique games. In *Proc. 44th STOC*, pp. 289–306. ACM Press, 2012. [doi:10.1145/2213977.2214005] 15

[31] RAN RAZ: A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998. [doi:10.1137/S0097539795280895] 3, 7

[32] RAN RAZ AND SHMUEL SAFRA: A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proc. 29th STOC*, pp. 475–484. ACM Press, 1997. [doi:10.1145/258533.258641] 5, 11, 18

[33] RONITT RUBINFELD AND MADHU SUDAN: Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996. [doi:10.1137/S0097539793255151] 18

[34] LUCA TREVISAN: On Khot's Unique Games Conjecture. *Bull. AMS*, 49(1):91–111, 2012. AMS link. 2

AUTHORS

Irit Dinur
Professor
Department of Applied Mathematics and Computer Science
The Weizmann Institute of Science
Rehovot, Israel
irit.dinur@weizmann.ac.il
http://www.wisdom.weizmann.ac.il/~dinuri/

Subhash Khot
Professor
Courant Institute of Mathematical Sciences
New York University
NY, USA
khot@cs.nyu.edu
https://cs.nyu.edu/~khot/

Guy Kindler
Professor
Rachel and Selim Benin School of Computer Science and Engineering
Hebrew University of Jerusalem
Jerusalem, Israel
wgkindler@gmail.com
https://www.cs.huji.ac.il/~gkindler/

Dor Minzer
Associate Professor
Department of Mathematics
Massachusetts Institute of Technology
MA, USA
dminzer@mit.edu
https://sites.google.com/view/dorminzer/home

Muli Safra
Professor
Department of Computer Science
Tel-Aviv University
Tel-Aviv, Israel
safra@mail.tau.ac.il
https://sites.google.com/site/mulisafra/home

ABOUT THE AUTHORS

Irit Dinur is a professor at the Weizmann Institute of Science. She has a Ph. D. from Tel Aviv university under the guidance of Muli Safra, after which she spent several years as a postdoctoral fellow at the Institute for Advanced Study, Princeton and at the University of California, Berkeley. She is interested broadly in theoretical computer science and mathematics, and more specifically in complexity theory, probabilistically checkable proofs, hardness of approximation, and most recently in the growing area of high dimensional expansion. She has a wife and three kids. She is the recipient of the 2019 Gödel Prize.

Subhash Khot is a Professor in the Computer Science Department at New York University. He has a Ph. D. from Princeton University, completed in 2003 under the supervision of Sanjeev Arora. His teacher and mentor at Vyankatrao Highschool, Mr. Vaman G. Gogate, played a decisive role in directing his attention to mathematics. If not for Mr. Gogate's guidance and some fortuitous turn of events, the chance of someone from the remote town of Ichalkaranji pursuing mathematical research was strictly nil. Subsequently, Mr. Gogate also mentored Amit Deshpande and Raghav Kulkarni (TCS), Abhijit Gadde (Physics), and Dinesh Bharadia (EE). He kept in touch with all his past students and continued to provide guidance on all aspects of life. He passed away in 2020 on Sept 5, which coincidentally is celebrated as the Teachers' Day in India.

Guy Kindler is a professor of computer science at the Hebrew University of Jerusalem. He has a Ph. D. from Tel Aviv University under the guidance of Muli Safra, after which he spent several years as a postdoctoral fellow at the Institute for Advanced Study, Microsoft Research and Weizmann Institute of Science. He is interested broadly in theoretical computer science, analysis, probability and combinatorics.

Dor Minzer is an Associate Professor in the Mathematics Department at Massachusetts Institute of Technology. He has a Ph. D. from Tel Aviv University, completed in 2018 under the guidance of Muli Safra, after which he spent 2 years as a postdoctoral fellow at the Institute for Advanced Study, Princeton. Dor is interested in Hardness of Approximation, Analysis of Boolean functions and relations to adjacent areas such as extremal combinatorics and error correcting codes. He started his academic studies at the Open University of Israel, where he met Professor Vadim Greenstein who has greatly influenced Dor and his passion for math.

Muli Safra is a Professor of Computer Science at Tel Aviv University. He has a Ph. D. from the Weizmann Institute, completed in 1990 under the supervision of Amir Pnueli. He is one of the recipients of the 2001 Gödel Prize.