Reduction From Non-Unique Games To Boolean Unique Games

Ronen Eldan

Department of Mathematics, Weizmann Institute of Science, Israel

Dana Moshkovitz

Department of Computer Science, University of Texas at Austin, USA

Abstract

We reduce the problem of proving a "Boolean Unique Games Conjecture" (with gap $1-\delta$ vs. $1-C\delta$, for any $C > 1$, and sufficiently small $\delta > 0$) to the problem of proving a PCP Theorem for a certain non-unique game. In a previous work, Khot and Moshkovitz suggested an inefficient *candidate* reduction (i.e., without a proof of soundness). The current work is the first to provide an efficient reduction along with a proof of soundness. The non-unique game we reduce from is similar to non-unique games for which PCP theorems are known.

Our proof relies on a new concentration theorem for functions in Gaussian space that are restricted to a random hyperplane. We bound the typical Euclidean distance between the low degree part of the restriction of the function to the hyperplane and the restriction to the hyperplane of the low degree part of the function.

2012 ACM Subject Classification Theory of computation → Problems, reductions and completeness

Keywords and phrases Unique Games Conjecture, hyperplane encoding, concentration of measure, low degree testing

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

Funding *Ronen Eldan*: Supported by a European Research Council Starting Grant (ERC StG) and by an Israel Science Foundation grant no. 715/16.
Dana Moshkovitz: This material is based upon work supported by the National Science Foundation under grants number 1218547 and 1648712.

Acknowledgements We are thankful to Subhash Khot and Bo'az Klartag for discussions. We are especially grateful to Bo'az for suggesting to use Schur's Lemma which is key to the proof of Theorem 8

1 Introduction

1.1 The Unique Games Conjecture

The Unique Games Conjecture was introduced by Khot [20] (see also the survey [21]) in order to prove optimal inapproximability results that eluded existing techniques.

▶ **Definition 1** (Unique Game). *The input of a unique game consists of a regular graph $G = (V, E)$, an alphabet Σ of size k, and permutations $\pi_e : \Sigma \to \Sigma$ for the edges $e = (u, v) \in E$. The task is to label each vertex with a symbol $\sigma(v) \in \Sigma$, as to maximize the fraction of edges $e = (u, v) \in E$ that are satisfied, i.e., $\pi_e(\sigma(u)) = \sigma(v)$.*

The following two prover game describes a unique game instance: a verifier interacts with two all-powerful provers. The verifier picks uniformly an edge $e = (u, v) \in E$; sends u to one prover and sends v to the other prover. Each prover is supposed to respond with a label from Σ. The verifier accepts if the two received labels $\sigma(u), \sigma(v)$ satisfy $\pi_e(\sigma(u)) = \sigma(v)$. Note that for every response of one prover in the game, there is a *unique* response of the other prover that is acceptable to the verifier. Hence, this two prover game is called a *unique* *game*. The *value* of the game is the probability that the verifier accepts when the provers play optimally.

The Unique Games Conjecture says that it is NP-hard to distinguish unique games of value close[1] to 1 from unique games of value close to 0:

▶ **Conjecture 2** (Unique Games Conjecture). *For every $\varepsilon, \delta > 0$, there exists $k = k(\varepsilon, \delta)$, such that it is NP-hard, given a unique game instance with alphabet of size k, to distinguish between the case where at least $1 - \delta$ fraction of the edges are satisfied and the case where at most ε fraction of the edges are satisfied.*

We refer to the problem of distinguishing instances where at least $1 - \delta$ fraction of the edges can be satisfied and instances where at most ε fraction of the edges can be satisfied as $1 - \delta$ vs. ε unique games.

The Unique Games Conjecture is known to imply optimal NP-hardness of approximation for problems like Max-Cut [22] and Vertex-Cover [28] that eluded optimal inapproximability results via existing techniques [18, 9]. Moreover, under the Unique Games Conjecture one can prove inapproximability for wide families of approximation problems. Most notably, basic semidefinite programming (SDP)-based algorithms are optimal for all local constraint satisfaction problems [37].

There are efficient algorithms for unique games in four cases: (i) Sufficiently small alphabet $k \leq \exp(1/\delta)$ [20, 10]; (ii) Sufficiently small $\delta = O(1/\log n)$ where n is the size of the graph [41, 17, 10, 11]; (iii) Large run-time $2^{n^{\mathrm{poly}(\delta)}}$ [1]; (iv) Random-like structure of G [2, 30].

There is an NP-hardness result for unique games for $\delta = 1/2$ and any $\varepsilon > 0$ as follows from the recently proved 2-to-2 Theorem [24, 13, 12, 6, 23, 25]. There is also a hardness result for any $\delta > 0$ and $\varepsilon = 1 - 2\delta$ [19, 25] that holds in the Boolean case $k = 2$.

The Boolean case $k = 2$ is the first interesting case of unique games, and it captures problems like Max-Cut and 2Lin(2). The assignments to the variables are ± 1, and each edge either requires its two endpoints to have the same assignment or different assignment. It is conjectureectured (and, indeed, follows from the Unique Games Conjecture [22]) that the best algorithm for Boolean unique games is the Goemans-Williamson SDP-based algorithm [16] that can distinguish value $1 - \delta$ from value $\varepsilon = 1 - \Theta(\sqrt{\delta})$. We focus on a weaker conjectureecture:

▶ **Conjecture 3** (Boolean Unique Games Conjecture). *For every $C \geq 1$, for sufficiently small $\delta > 0$, it is NP-hard to distinguish between unique games with $k = 2$ where $1 - \delta$ fraction of the edges can be satisfied, and ones where only $1 - C\delta$ fraction of the edges can be satisfied.*

The Unique Games Conjecture can be thought of as an amplified version of Conjecture 3, with the soundness error close to 0 rather than close to 1 and the alphabet size appropriately increased. It is open whether the Unique Games Conjecture follows from Conjecture 3. There were past attempts to prove this implication via a "strong parallel repetition", but those attempts uncovered an obstacle [39, 5].

1.2 This Work

In a previous work Khot and Moshkovitz [27] suggested a *candidate* reduction for proving hardness of $1 - \delta$ vs. $1 - C\delta$ Boolean unique games, however they could not prove the

[1] For unique games there is an efficient algorithm to distinguish games of value exactly 1 from games of value smaller than 1. Hence, it is necessary to focus on games of value close to 1 rather than 1.

soundness of the reduction. In this work we define a problem, Subspaces Near-Intersection, and show a provably sound reduction from Subspaces Near-Intersection to $1-\delta$ vs. $1-C\delta$ Boolean unique games. Importantly, the NP-hardness of Subspaces Near-Intersection – which we conjectureecture but do not prove – is in the same spirit of known PCP Theorems, and resembles in many ways the 2-to-2 Theorem.

▶ **Theorem 4** (Main Theorem)**.** *Assume the Subspaces Near-Intersection Conjecture (Conjecture 7 in the sequel). For any $C \geq 1$, for any sufficiently small $\delta > 0$, distinguishing $1-\delta$ vs. $1-C\delta$ Boolean unique games is NP-hard. In fact, if the Subspaces Near-Intersection problem requires time T, then distinguishing $1-\delta$ vs. $1-C\delta$ Boolean unique games requires time $\Omega(T)$.*

Our reduction has the added benefit of being highly efficient (linear-sized). In contrast, the reduction in [27] had an exponential blowup, as it was only meant to rule out polynomial time algorithms for unique games under plausible assumptions on exponential hardness. Like for the 2-to-2 problem, one would expect a reduction from Sat to Subspaces Near-Intersection to map size-n instances of Sat to size $n^{c(\delta)}$ instances of Subspaces Near-Intersection, where δ is the completeness error in Subspaces Near-Intersection and $c(\delta) \geq 1/\delta$ is a function of δ.

Subspaces Near-Intersection is discussed in the next section. The main ideas of the proof of Theorem 4 are discussed in Section 1.4. A key lemma is a new concentration theorem for the restriction of a function in Gaussian space to a random hyperplane. The lemma bounds the Euclidean distance between the degree-d part of the restriction and the restriction of the degree-d part. The formal statement and more details appear in Section 1.5.

1.3 Subspaces Near-Intersection Conjecture

First we discuss existing PCP theorems (projection games), and a projection game based on 3Lin($\mathbb{R}$), then we define the new conjectureecture.

1.3.1 Projection Games

Existing optimal hardness of approximation results follow from the *proven* NP-hardness of approximating *projection games* [4, 3, 38, 32]. In (the symmetric version of) projection games, the verifier tests the answer of each prover *separately* in a way that depends solely on the question to the prover, and then checks *equality* between parts of the two answers (the projections). For instance, given a Sat instance the verifier may ask each prover for the assignment to a subset of the variables. Each subset spans clauses and the verifier checks that those clauses are satisfied (a separate test for each prover that depends only on the question to the prover). The two subsets intersect, and the verifier checks that the provers agree on the assignments to the variables in the intersection (a comparison on parts of the answer). Formally:

▶ **Definition 5** (Projection Game)**.** *The input of a projection game consists of a bi-regular graph $G = (X, Y, E)$ whose X-degree is denoted q, an alphabet Σ and sets $L_x \subseteq \Sigma^q$ for every vertex $x \in X$. The task is to label each vertex $x \in X$ with a symbol $\sigma(x) \in L_x$, as to maximize the probability that, when one picks $e = (x, y), (x', y) \in E$, it holds $\sigma(x)_y = \sigma(x')_y$. Sometimes one describes the game over the graph $(X, \{(x, x')\})$.*

It is known that it is NP-hard to distinguish projection games of value 1 from projection games of value close to 0 [4, 3, 38, 32], and moreover that it requires time $2^{n^{1-o(1)}}$ assuming the widely believed Exponential Time Hypothesis[2] as follows from an almost-linear sized reduction from Sat to projection games [32].

2-to-2 games are projection games where given $\sigma(x)_y \in \Sigma$ there are only *two* possibilities for $\sigma(x) \in L_x \subseteq \Sigma^q$. It is known that it is NP-hard to distinguish 2-to-2 games of value close to 1 from 2-to-2 games of value close to 0 [24, 13, 12, 6, 23, 25]. However, 2-to-2 games are easier than general projection games, since they have algorithms that run in time $2^{n^{\mathrm{poly}(\delta)}}$ [1]. Appropriately, the known NP-hardness reduction to 2-to-2 games maps size n inputs of Sat to size $n^{c(\delta)}$ 2-to-2 games for a function $c(\delta) \geq 1/\delta$.

1.3.2 3Lin($\mathbb{R}$) Projection Game

Subspaces Near-Intersection is a proxy for the following projection game based on the Khot-Moshkovitz [26] robust real 3Lin: The verifier picks uniformly at random $100k$ real 3Lin equations $c_1, \ldots, c_{100k}$ and two sets S_1, S_2 of k variables among their variables, where $|S_1 \cap S_2| = k - 1$. Note that any subset of the linear equations induced on S_1 or on S_2 forms a linear subspace of $\mathbb{R}^k$. The verifier sends S_1 to one prover, and receives a unit vector that represents an assignment to S_1's variables. The unit vector must satisfy a random linear constraint on S_1. The verifier sends S_2 to the other prover, and receives a unit vector that represents an assignment to S_2's variables. The vector must satisfy a random linear constraint on S_2. The verifier projects each of the vectors on the $k-1$ coordinates that correspond to the intersection $S_1 \cap S_2$, and measures the Euclidean distance between the projections. Suppose that there exists a prover strategy where the projections are identical with probability $1 - \delta$. The task is to efficiently compute a prover strategy that minimizes the average Euclidean distance between the projections.

Simple approximation algorithms for this problem guarantee distances $O(\sqrt{\delta/k})$ and $O(1/k)$:

- *Basic semidefinite programming* achieves *square distance* δ/k, since in the completeness case one achieves deviation 0 with probability $1 - \delta$ and deviation $1/\sqrt{k}$ with probability δ. As a result, this algorithm can efficiently guarantee distance $O(\sqrt{\delta/k})$.
- *Correlated sampling* is the strategy in which the provers guess a clause in $S_1 \cap S_2$, satisfy it (with a norm 1 assignment) and assign all other coordinates 0. It achieves distance 1 with probability[3] $1/k$, and deviation 0 with the remaining probability.

Hence, the question is whether one can efficiently compute a prover strategy where the average distance between the projections is, say, $0.0001 \cdot \min\left\{\sqrt{\delta/k}, 1/k\right\}$.

Subspaces Near-Intersection is closely related to this projection game: there one compares the vectors on their projection to a *generic* hyperplane in $\mathbb{R}^k$, as opposed to an *axis-parallel* hyperplane.

1.3.3 Subspaces Near-Intersection

The Subspaces Near-Intersection game is a projection game that is defined over the reals[4]. Each vertex is associated with a linear subspace in $\mathbb{R}^k$, and a labeling to the vertex is a unit

[2] The Exponential Time Hypothesis postulates that Sat requires time $2^{\Omega(n)}$ on inputs of size n.
[3] Note that the error probability of correlated sampling can be made C/k if one considers a projection onto a subspace of dimension $k - C$ instead of $k - 1$.
[4] The intention is to consider real numbers up to a finite precision, so the errors introduced by the finite precision are much smaller than any other quantity involved. For the sake of clarity in exposition we do not explicitly address precision errors.

vector that satisfies the constraints. Each edge is associated with a hyperplane in $\mathbb{R}^k$. The vectors on the endpoints of the edge should have the same restriction to the hyperplane of the edge.

▶ **Definition 6** (Subspaces Near-Intersection). *The input is a regular graph $G = (V, E)$, $k \times k$ matrices A_v with entries in $[-1, 1]$ for the vertices $v \in V$, and unit vectors $\Theta_e \in \mathbb{R}^k$ for the edges. We assume that, per vertex $v \in V$, when one picks a uniform edge $e = (u, v) \in E$ that touches v, the vector Θ_e is uniform. The task is to label each vertex with a unit vector $\sigma(v) \in \mathbb{R}^k$ such that $A_v\sigma(v) = 0$, as to maximize the number of edges $e = (u, v) \in E$ with $Proj_{\Theta_e^\perp}(\sigma(u)) = Proj_{\Theta_e^\perp}(\sigma(v))$ ("satisfied edges"). We say that the edge is α-satisfied if $|Proj_{\Theta_e^\perp}(\sigma(u)) - Proj_{\Theta_e^\perp}(\sigma(v))|_2 \leq \alpha$.*

As before, in the case that there exists an assignment where the distance between the projections is 0 with probability $1-\delta$ and $1/\sqrt{k}$ with probability δ, a semidefinite programming algorithm that minimizes the square distance between the projections, would lead to distance $\sqrt{\delta/k}$ between the projections. There is a natural matching semidefinite programming integrality gap for Subspaces Near-Intersection described in Appendix A. The correlated sampling algorithm we described for the 3Lin($\mathbb{R}$) projection game in Sub-section 1.3.2 no longer applies.

There is an analogy between the games considered in the recent proof of the 2-to-2 Theorem and the Subspaces Near-Intersection game: in both games for every edge the label of one endpoint does not uniquely determine the label of the other endpoint, but rather *nearly* determines it, leaving out one "degree of freedom". In the 2-to-2 games of [24, 13, 25], labels are vectors over the binary finite field, and one degree of freedom means that there are two possibilities for the answer of the other prover. Here labels are real vectors and one of their "coordinates" remains undetermined.

For technical reasons, and similarly to the proof of the 2-to-2 Theorem, we will define a slight strengthening using *zoom-ins*. For a linear subspace $Y \subseteq \mathbb{R}^k$ we define the Y-zoom-in Subspaces Near-Intersection game as follows: Focus on edges $e \in E$ where $Y \subseteq \Theta_e^\perp$, i.e., one can write $\Theta_e^\perp = Y + S_e$, where S_e is a hyperplane in $Y^\perp$. An edge is satisfied if $Proj_{S_e}(\sigma(u)) = Proj_{S_e}(\sigma(v))$ and is α-satisfied if $|Proj_{S_e}(\sigma(u)) - Proj_{S_e}(\sigma(v))|_2 \leq \alpha$.

▶ **Conjecture 7** (Subspaces Near-Intersection Conjecture). *There exists a global constant $0 < \alpha < 1$, such that for any $\varepsilon, \delta > 0$, $r \in \mathbb{N}$, there exists $k \geq 1$ such that $\sqrt{\delta/k} \gg 1/k$, and the following is NP-hard: The input is an instance of the Subspaces Near-Intersection problem. The task is to distinguish between the cases:*

- Completeness: *There exists a labeling $\sigma : V \to \mathbb{R}^k$ that satisfies*[5] *at least $1 - \delta$ fraction of the edges $e = (u, v) \in E$. The remaining edges are $O(1/\sqrt{k})$-far from satisfied.*
- Soundness: *For any r-dimensional $Y \subseteq \mathbb{R}^k$, for any labeling $\sigma : V \to \mathbb{R}^k$, the probability over the choice of $e = (u, v)$ in the Y-zoom-in, that e is $\alpha\sqrt{\delta/k}$-satisfied is at most ε.*

1.4 Main Ideas

This work builds on an idea suggested by Khot and Moshkovitz [27] for proving hardness of unique games. Like[6] [27] we replace the commonly used long code and Hadamard code by

[5] Near satisfaction suffices; see Section 1.6.
[6] The candidate reduction in [27] had a variation on half-space encoding, namely, $\mathsf{interval}(\langle a, x\rangle)$, where $\mathsf{interval}$ changes sign as one crosses any integer point, not just 0. Crucially, we use half-spaces in the current paper.

an encoding by half-spaces. We first explain the half-space idea, and then describe our new ideas in using and analyzing half-space encodings.

The half-space defined by $a \in \mathbb{R}^k$ is $h_a : \mathbb{R}^k \to \{\pm 1\}$, where $h_a(x) = \mathsf{sign}(\langle a, x\rangle)$. The half-space encoding of a is the truth-table of h_a where we enumerate over all $x \in \mathbb{R}^k$ up to a precision that makes the rounding error sufficiently smaller than any of the other quantities involved.

Half-space encoding is similar in structure to the Hadamard encoding, where a vector $a \in \{0,1\}^k$ is encoded as the linear function $l_a(x) = \langle a, x\rangle$ for all $x \in \{0,1\}^k$, and arithmetic is done over the finite field $\{0,1\}$. This similarity gains us two benefits that the Hadamard encoding has:

1. We can test linear conditions on $a \in \mathbb{R}^k$ by testing its encoding. Specifically, $\langle a, c\rangle = 0$ for a vector $c \in \mathbb{R}^k$ iff $h_a(x + c) = h_a(x)$ for every $x \in \mathbb{R}^k$. (On the soundness side we need $|\langle a, c\rangle| \gg 0$ to detect that the inequality does not hold; this the reason we require robustness).
2. Encodings of similar strings have common parts. Suppose that the projections of $a, a' \in \mathbb{R}^k$ on a hyperplane $\Theta^\perp$ are the same. Then, when one picks $x \in \Theta^\perp$ it holds that $\langle a, x\rangle = \langle a', x\rangle$. Importantly, the union of all hyperplanes covers $\mathbb{R}^k$ uniformly.

Note that both equations $h_a(x + c) = h_a(x)$ and $h_a(x) = h_{a'}(x')$ are unique tests. We remark that a property like the first is used in any optimal inapproximability result that uses the Hadamard code, and a property like the second was used in the proof of the 2-to-2 Games Theorem (under the name "sub-code covering"). Crucially, half-space encoding has a property that the Hadamard encoding does not have, but the long code does have, namely, a unique test:

3. Noise stability test. Half-spaces optimize the success probability of the following test: pick random Gaussian $x \in \mathbb{R}^k$, perturb x to obtain $x' \in \mathbb{R}^k$ also distributed as a Gaussian. Check whether $h_a(x) = h_a(x')$.

In discrete space, the long code encoding $d_i(x) = x_i$ optimizes the analogous noise stability test, and this was used to show hardness of Boolean unique games assuming the Unique Games Conjecture [22].

In [27] it was suggested that to prove NP-hardness of Boolean unique games one needs robustness of the noise stability test:

> Suppose that a half-space passes the noise stability test with probability $1-\delta$. Assume that a balanced function $f : \mathbb{R}^k \to \{\pm 1\}$ passes the test with probability $1 - C\delta$ for $C > 1$. Does f correspond to a half-space?

Works that dealt with robustness in noise stability [34, 33, 14] proved such results for functions that pass the test with probability at least $1 - \delta - \epsilon$ for $\epsilon \ll \delta$. Such must be the same as a half-space almost everywhere. When the acceptance probability is $1 - C\delta$, the function f can have many forms, including functions of C half-spaces, low degree threshold functions, and many more. In particular, the function may have no correlation with any half-space. Mossel and Neeman [35] note that functions that pass the noise stability test with constant probability have to correlate with a half-space *after a large random shift*, but we are unable to use this fact since a shift hurts the second property above.

Our idea is not to focus on a half-space that correlates with f (which corresponds to the linear part of f), but rather consider the *low degree part* of f (where the low degree part is obtained from the Hermite expansion of f). By the noise stability of f, its low degree part must be large. We argue about consistency between low degree parts of functions that are partly similar. We also argue about the ability to extract vectors that satisfy linear tests from low degree parts that satisfy the same tests.

Crucially, all our estimates must be extremely tight, since the gap for Boolean unique games is extremely narrow to begin with, $1-\delta$ vs. $1-\Theta(\sqrt{\delta})$. We obtain the required tightness using two tools: hypercontractivity and concentration.

Hypercontractive inequalities (see, e.g., [36]) bound norms of a "smoothed" function by norms of the original function. Here we use the Gaussian hypercontractive inequality, through the implied *level-d inequalities* (see, e.g., [36]), to show that Boolean functions that are the same with probability at least $1-\delta$ over the input must have low degree parts that are $\approx \delta$-close in l_2 distance. In contrast, a less careful estimate, not using Booleanity and hypercontractivity, only gives $\sqrt{\delta}$-closeness, which is useless in our context. Note that the functions we compare are restrictions of functions f to hyperplanes (as in the second property above).

Concentration is discussed in Section 1.5. It considers functions restricted to a random hyperplane, and bounds the typical Euclidean distance of the low degree part of the restriction from the restriction of the low degree part. We use concentration to argue consistency between the low degree parts of the restrictions of a function to different hyperplanes. We note that the much easier to prove distance of $O(1/\sqrt{k})$ rather than $O(1/k)$ would have been useless for our application.

1.5 Concentration of Degree-d Part

Let $f:\mathbb{R}^n \to \mathbb{R}$, and let $f^{\leq d}$ be the degree-d part of f. Note that $f^{\leq d}$ is a *global* property of f. Let Θ be uniformly distributed in the $(n-1)$-dimensional sphere, so $\Theta^\perp$ is a random hyperplane in $\mathbb{R}^n$. Denote the restriction of f to $\Theta^\perp$ by $f_{|\Theta^\perp}$. This is a *local* part of f. We show a *local-to-global theorem*: the degree-d part of $f_{|\Theta^\perp}$ is extremely close to the restriction of $f^{\leq d}$ to $\Theta^\perp$:

▶ **Theorem 8** (Concentration of degree-d part)**.** *For any $\varepsilon > 0$, for every 0-homogeneous*[7] *$f : \mathbb{R}^n \to \mathbb{R}$ with bounded 2-norm, with probability at least $1-\varepsilon$ over Θ, $\left|(f_{|\Theta^\perp})^{\leq d} - (f^{\leq d})_{|\Theta^\perp}\right|_2 \leq O_{d,\varepsilon}(1/n)$.*

Local-to-global theorems, like linearity testing [7] and low degree testing [40] over finite fields, are key to PCP. With Theorem 8 we add a new, tight, low degree testing -type theorem, this time in the highly challenging case of real functions and approximate equality. To get intuition for why this case is so challenging, note that two different real low degree polynomials can be similar on much of the space (Carbery-Wright (Lemma 14) gives tight bounds). In contrast, two different low degree polynomials over a finite field are vastly different, and this is key to existing combinatorial and algebraic techniques, which we cannot use. Standard analytic techniques (e.g., Hermite analysis, or a sampling theorem of Klartag and Regev [29]) give an upper bound of $O(1/\sqrt{n})$ rather than $O(1/n)$ even for $d = 1$. As we remarked above, such bounds are useless for our needs.

Our proof is by a delicate second moment argument using symmetry considerations. Crucially, the second moment is a rotationally-invariant quadratic form in f, and hence we can use Schur's lemma from representation theory that classifies rotationally-invariant quadratic forms. The lemma implies that the second moment depends only on the spectrum of f, and not on its identity. Our calculations can therefore be significantly simplified by

[7] f is 0-homogeneous if $f(cx) = f(x)$ for every $x \in \mathbb{R}^n$ and $c > 0$.

focusing on f that depends only on one of its variables. Given a function that depends on one direction, the expression that we need to bound will only depend on the angle between this direction and Θ. The technical bulk of the proof then amounts to showing that this dependence is quadratic in the scalar product, meaning that it is typically of the order $1/n$.

1.6 The Road Ahead

This paper suggests two paths to NP-hardness of Boolean unique games:

1. Prove NP-hardness of Subspaces Near-Intersection as in Conjecture 7. This paper implies that NP-hardness of Boolean unique games would follow.
2. Lift the reduction in this paper to a reduction from the Khot-Moshkovitz NP-hard 3Lin($\mathbb{R}$) to Boolean unique games. The reduction was outlined in Sub-section 1.3.2.

In this sub-section we give more details about each of these paths.

One can weaken the Subspaces Near-Intersection conjectureecture substantially and the analysis in this paper would still go through (with modifications): The verifier can project onto subspaces of dimension, say, $k-100$, instead of dimension $k-1$. In the completeness case there could be approximate equality (with deviation $O(\delta/\sqrt{k})$) rather than exact equality. It is enough to have large soundness error, say $\varepsilon = 0.99$, instead of low error. The distance of the projections in the soundness case can be of the order of $\tilde{\Theta}(\delta/\sqrt{k} + 1/k)$, rather than $\Theta(\sqrt{\delta/k})$.

The reduction in this paper can be lifted to a reduction from a 3Lin($\mathbb{R}$) projection game like we described in Sub-section 1.3.2 (instead of Subspaces-Near Intersection) to Boolean unique games. In this setting, we suggest to focus on projections onto subspaces of dimension sufficiently smaller than $k-1$, as to decrease the probability that the correlated sampling algorithm achieves distance 0. To analyze such a reduction one would need to address subspaces that are axes-parallel rather than generic, and this requires ideas beyond the ones in this paper. In particular, the concentration theorem we prove is no longer directly applicable. In the authors' opinion, this path is the most promising path towards hardness of Boolean unique games.

2 Preliminaries

2.1 Hermite Polynomials

Let $\mathcal{G}^n$ denote the n-dimensional Gaussian distribution with n independent mean-0 and variance-1 coordinates. The space of all real functions $f : \mathbb{R}^n \to \mathbb{R}$ with $\mathbf{E}_{x\sim\mathcal{G}^n}\left[f(x)^2\right] < \infty$ is denoted $L^2(\mathbb{R}^n, \mathcal{G}^n)$. This is an inner product space with inner product $\langle f, g\rangle \doteq \mathbf{E}_{x\sim\mathcal{G}^n}\left[f(x)g(x)\right]$. For a natural number j, the j'th *Hermite polynomial* $H_j : \mathbb{R} \to \mathbb{R}$ is $H_j(x) = \frac{1}{\sqrt{j!}} \cdot (-1)^j e^{x^2/2} \frac{d^j}{dx^j} e^{-x^2/2}$. The first few Hermite polynomials are $H_0 \equiv 1$, $H_1(x) = x$, $H_2(x) = \frac{1}{\sqrt{2}} \cdot (x^2 - 1)$, $H_3(x) = \frac{1}{\sqrt{6}} \cdot (x^3 - 3x)$, $H_4(x) = \frac{1}{2\sqrt{6}} \cdot (x^4 - 6x^2 + 3)$. The Hermite polynomials satisfy:

▶ **Proposition 9** (Orthonormality). *For every j, $\langle H_j, H_j\rangle = 1$. For every $i \neq j$, $\langle H_i, H_j\rangle = 0$. In particular, for every $j \geq 1$, $\mathbf{E}_{x\in\mathcal{G}}\left[H_j(x)\right] = 0$.*

The multi-dimensional Hermite polynomials are: $H_{j_1,\ldots,j_n}(x_1, \ldots, x_n) = \prod_{i=1}^n H_{j_i}(x_i)$. For multi-indices $L = (l_1, \ldots, l_n)$ and $T = (t_1, \ldots, t_n)$ we denote $L \leq T$ if $l_i \leq t_i$ for every i. We write $T - L$ to denote $(t_1 - l_1, \ldots, t_n - l_n)$. We write C^T to denote $C^{\sum_i t_i}$ and $\binom{T}{L}$ to denote $\binom{t_1}{l_1} \cdots \binom{t_n}{l_n}$. The Hermite polynomials form an orthonormal basis for the space $L^2(\mathbb{R}^n, \mathcal{G}^n)$.

Hence, every function $f \in L^2(\mathbb{R}^n, \mathcal{G}^n)$ can be written as $f(x) = \sum_{S \in \mathbb{N}^n} \hat{f}(S)\ H_S(x)$, where S is multi-index, i.e. an n-tuple of natural numbers, and $\hat{f}(S) \in \mathbb{R}$ (Hermite expansion). The size of a multi-index $S = (S_1, \ldots, S_n)$ is defined as $|S| = \sum_{i=1}^{n} S_i$. The degree-d part of f is $f^{=d} = \sum_{|S|=d} \hat{f}(S) H_S(x)$. The part of degree at most d is $f^{\leq d} = \sum_{i=0}^{d} f^{=i}$. When f is anti-symmetric, i.e. $\forall x \in \mathbb{R}^n, f(-x) = -f(x)$, we have $\widehat{f}(\vec{0}) = \mathbf{E}[f] = 0$ and $f^{\leq 0} \equiv 0$.

The noise operator (more commonly known as the Ornstein-Uhlenbeck operator) T_ρ takes a function $f \in L^2(\mathbb{R}^n, \mathcal{G}^n)$ and produces a function $T_\rho f \in L^2(\mathbb{R}^n, \mathcal{G}^n)$ that averages the value of f over local neighborhoods: $T_\rho f(x) = \mathbf{E}_{y \in \mathcal{G}^n}\left[f(\rho x + \sqrt{1-\rho^2} y)\right]$. The Hermite expansion of $T_\rho f$ can be obtained from the Hermite expansion of f as follows:

▶ **Proposition 10.** $T_\rho f = \sum_S \rho^{|S|} \hat{f}(S) H_S$.

2.2 Some classical inequalities

The hypercontractive inequality is given in the next lemma.

▶ **Lemma 11** (Hypercontractive inequality). *Let $f, g : \mathbb{R}^k \to \mathbb{R}$. For $0 \leq \rho \leq \sqrt{rs} \leq 1$, $\langle f, T_\rho g \rangle \leq |f|_{1+r}\, |g|_{1+s}$.*

The inequality is often used to show the small sets cannot have much weight on low degree parts. Similarly, we will use a corollary of it to show that Boolean functions that are almost always the same must have low degree parts that are similar. The corollary is known as *level-k inequality*:

▶ **Lemma 12** (Level-k inequality). *Let $f : \mathbb{R}^k \to \{0, 1\}$ have mean $\mathbf{E}[f] = \alpha$ and let $k \leq 2\ln(1/\alpha)$. Then, $\left|f^{\leq k}\right|_2^2 \leq \left(\frac{2e}{k} \ln(1/\alpha)\right)^k \alpha^2$.*

A convenient re-formulation is

▶ **Lemma 13.** *Let $A \subseteq \mathbb{R}^k$ be a set of probability α. Let $p : \mathbb{R}^k \to \mathbb{R}$ be a polynomial of degree at most $k \leq 2\ln(1/\alpha)$ with $|p|_2 = 1$. Then, for χ_A, the indicator function of A, $|\mathbf{E}_x [p(x)\chi_A(x)]| \leq \left(\frac{2e}{k} \ln(1/\alpha)\right)^{k/2} \alpha$.*

Proof. Since p is of degree at most k, we have $\langle \chi_A, p \rangle = \langle \chi_A^{\leq k}, p \rangle$. By Cauchy-Schwarz inequality, $\langle \chi_A^{\leq k}, p \rangle \leq \left|\chi_A^{\leq k}\right|_2 |p|_2 \leq \left|\chi_A^{\leq k}\right|_2$. The lemma follows from a level-k inequality (Lemma 12) invoked on χ_A. ◀

The Carbery-Wright anti-concentration inequality shows that a low degree polynomial cannot be concentrated around any point:

▶ **Lemma 14** (Carbery-Wright Anti-concentration [8]). *For $t \in \mathbb{R}$ and $\varepsilon > 0$, for a polynomial p of degree d, $|p|_2 = 1$, $\Pr_{x \sim \mathcal{G}^n}[|p(x) - t| \leq \varepsilon] \leq O(d)\varepsilon^{1/d}$.*

The Gaussian Poincaré inequality upper bounds the variance of a function in terms of its derivative:

▶ **Lemma 15** (Gaussian Poincaré inequality). *Let $f : \mathbb{R}^k \to \mathbb{R}$ have continuous derivatives. Then, $\mathrm{Var} f \leq \mathbf{E}\left[|\nabla f|^2\right]$.*

Klartag and Regev showed that a random subspace samples well any function:

▶ **Lemma 16** (Sampling [29]). *Let $f : \mathbb{R}^k \to \mathbb{R}$ with $|f|_2 < \infty$. Let $0 < \varepsilon < 1$. Let S be a uniform subspace of dimension $k-1$. Then, $\Pr_S[|\mathbf{E}_S[f] - \mathbf{E}[f]| \geq \varepsilon |f|_2] \leq O\left(\exp\left(-\Omega\left(\frac{\varepsilon k}{\log(2/\varepsilon)}\right)\right)\right)$.*

Their formulation referred to functions on spheres, but immediately implies the same for functions in Gaussian space by averaging over all possible radii. Their formulation referred to non-negative functions and multiplicative approximation, but immediately extends to general functions and additive approximation by separately considering the negative and positive parts of the function.

The next lemma follows from Lemma 16 (in fact, one needs a much weaker version of Lemma 16):

▶ **Lemma 17.** *For any constants $0 < \delta < 1$ and $d \geq 1$, For any subset $\mathcal{H}$ of fraction δ of $(k-1)$-dimensional subspaces in $\mathbb{R}^k$, the distribution induced on d-dimensional subspaces by picking $H \in \mathcal{H}$ and $S \subseteq H$, $dim(S) = d$, is $\tilde{O}_{d,\delta}(1/k)$-close in statistical distance to the uniform distribution over d-dimensional subspaces.*

3 Boolean Unique Game Construction

Let $C \geq 1$. Fix an instance of the Subspaces Near-Intersection Problem, given by $G = (V, E)$, k, $\{A_v\}_v$, $\{\Theta_e\}_e$. Let δ and ε be the completeness and soundness errors, respectively, where $\delta > 0$ is sufficiently small and ε is a constant, say $1/10$. We will construct a Boolean unique games instance with completeness error $O(\delta/\sqrt{k})$ (where the $O(\cdot)$ hides a small absolute constant, independent of C) and soundness error $1 - C\delta/\sqrt{k}$.

The unique game we construct consists of encodings of the labeling for the $v \in V$ via half-spaces.

▶ **Definition 18** (half-space encoding)**.** *The half-space encoding of $\sigma \in \mathbb{R}^k$ is the Boolean function $\mathbb{R}^k \to \{\pm 1\}$ defined as $\mathsf{HS}_\sigma(x) = \mathsf{sign}(\langle \sigma, x\rangle)$.*

For every $v \in V$ and $x \in \mathbb{R}^k$ we have a unique game variable corresponding to v, x that is supposed to be assigned $\mathsf{HS}_{\sigma(v)}(x)$ (The actual construction involves a discretization of $\mathbb{R}^k$ up to a very high precision in each coordinate. The precision depends on k and $1/\delta$). We denote by $f_v : \mathbb{R}^k \to \{\pm 1\}$ the actual assignment to the variables that correspond to v.

Next we group together variables in order to enforce certain basic structural properties on the f_v's in a technique called *folding*. The properties we consider are ones that half-spaces have.

Half-spaces are anti-symmetric, i.e., for every $x \in \mathbb{R}^k$, $\mathsf{HS}_\sigma(-x) = -\mathsf{HS}_\sigma(x)$. While f_v may not necessarily be $\mathsf{HS}_{\sigma(v)}$, we will enforce anti-symmetry by having only one variable for every pair of $x, -x$ where $x \in \mathbb{R}^k$.

▶ **Definition 19** (anti-symmetry folding)**.** *In the unique games construction the functions f_v satisfy $f_v(-x) = -f_v(x)$ for every $x \in \mathbb{R}^k$.*

Half-spaces are 0-homogeneous, i.e., for every $x \in \mathbb{R}^k$ and $c > 0$ it holds $\mathsf{HS}_\sigma(c \cdot x) = \mathsf{HS}_\sigma(x)$. We enforce 0-homogeneity as follows:

▶ **Definition 20** (0-homogeneity folding)**.** *In the unique games construction the functions f_v satisfy $f_v(cx) = f_v(x)$ for every $x \in \mathbb{R}^k$ and $c > 0$.*

For every A such that $A\sigma = 0$, for every $x, y \in \mathbb{R}^k$, $\alpha, \beta \in \mathbb{R}$, we have:

$$
\begin{aligned}
\mathsf{HS}_\sigma(\alpha x A + \beta y) &= \mathsf{sign}(\langle \sigma, \alpha x A + \beta y\rangle) \\
&= \mathsf{sign}(\alpha \cdot \langle \sigma, xA\rangle + \langle \sigma, \beta y\rangle) \\
&= \mathsf{sign}(\alpha \cdot \langle A\sigma, x\rangle + \langle \sigma, \beta y\rangle) \\
&= \mathsf{sign}(\langle \sigma, \beta y\rangle)
\end{aligned}
$$

Therefore we enforce:

▶ **Definition 21** (constraints folding). *In the unique games construction the functions* f_v *satisfy* $f_v(\alpha x A_v + \beta y) = f_v(\alpha z A_v + \beta y)$ *for every* $x, y, z \in \mathbb{R}^k$, $\alpha, \beta \in \mathbb{R}$.

To complete the definition of the unique games instance, we define the equations over the variables. The equations correspond to two local tests: (1) Noise test on f_v for $v \in V$; (2) Consistency test on f_u, f_v for $(u, v) \in E$. The equations are specified in Figure 1.

Verifier$\{f_v\}$
Folding: We assume that the f_v's are folded as in Definitions 19, 20 and 21.
Set $\beta = 1/(10^{10}C^2)$, $p = \delta/\sqrt{\beta k}$. The verifier performs the noise test with probability p; the consistency test with probability $1 - p$:

- *Noise Test:* Pick at random $v \in V$. Pick $y, x, z \sim \mathcal{G}^k$ and set $\tilde{x}, \tilde{z} \in \mathbb{R}^k$ as follows: $\tilde{x} = (1-\beta)y + \sqrt{2\beta - \beta^2}x$, $\tilde{z} = (1-\beta)y + \sqrt{2\beta - \beta^2}z$. Check $f_v(\tilde{x}) = f_v(\tilde{z})$.
- *Consistency Test:* Pick at random $e = (u, v) \in E$. Pick a random Gaussian $x \in \Theta_e^{\perp}$. Check $f_u(x) = f_v(x)$.

Figure 1 Unique game

The size of the construction is linear in the size of the Subspaces Near-Intersection instance and a function of (the constants) k and $1/\delta$.

3.1 Completeness

Suppose that there is an assignment $\sigma : V \to \mathbb{R}^k$ as in the completeness case of Subspaces Near-Intersection. Further, assume that each f_v corresponds to a half-space encoding of $\sigma(v)$. The probability that the noise test rejects is $O(\sqrt{\beta})$ and it is performed with probability p, so its total contribution is $O(\delta/\sqrt{k})$. By the completeness of Subspaces Near-Intersection, with probability $1 - \delta$ the consistency test always passes, and with probability δ it passes except with probability $1/\sqrt{k}$. Overall, the probability of rejection is $O(\delta/\sqrt{k})$.

4 Soundness

Assume that $\{f_v\}_{v \in V}$ pass the unique tests with probability at least $1 - C\delta/\sqrt{k}$. We will construct a constant-dimensional $Y \subset \mathbb{R}^k$ and an assignment $\sigma : V \to \mathbb{R}^k$. Each $\sigma(v)$ is a unit vector such that $A_v\sigma(v) = 0$, and with constant probability over $e = (u, v) \in E_Y$, when one writes $\Theta_e^{\perp} = Y + S_e$ for S_e orthogonal to Y, it holds that

$$|Proj_{S_e}(\sigma(u)) - Proj_{S_e}(\sigma(v))|_2 \le \tilde{O}_C(\delta/\sqrt{k} + 1/k),$$

where the $\tilde{O}_C(\cdot)$ hides logarithmic factors in $\sqrt{k}/\delta$, k, as well as factors that depend on C, and the deviation is therefore $\ll \sqrt{\delta/k}$.

The plan for the analysis is as follows: Use the noise stability to decode a large low degree part for almost every vertex $v \in V$. Use concentration to argue consistency between the restriction of the low degree part to an edge hyperplane and the low degree part of the restriction to the hyperplane, for most edges. The low degree parts of the restrictions to the edge hyperplane are close in l_2 distance for most edges thanks to the consistency test and hypercontractivity. Obtain from each low degree polynomial a vector by repeatedly differentiating the polynomial. The differentiation will be in random directions we pick, and we focus on zoom-in's so we can restrict to hyperplanes that contain the random directions. We use consistency along edges to argue about consistency of the derivatives and of the number of differentiations.

For all $v \in V$ we have $|f_v|_2 = 1$. By the success of the functions f_v in the unique game, the noise test must pass except with probability $C\delta/(\sqrt{k}p) \leq C\sqrt{\beta}$ and the consistency test must pass except with probability $C\delta/(\sqrt{k}(1-p)) \leq 2C\delta/\sqrt{k}$. We say that $v \in V$ is *typical* if the noise test rejects with probability at most $100C\sqrt{\beta}$ when v is chosen. In other words, for a typical $v \in V$, $\langle f_v, T_{1-\beta} f_v \rangle \geq 1 - 200C\sqrt{\beta}$. Note that all $v \in V$ are typical except for at most 0.1 fraction. We say that an edge $e = (u, v) \in E$ is *typical* if both u and v are typical and the consistency test rejects with probability at most $20C\delta/\sqrt{k}$ when e is chosen. At least 0.7 fraction of the edges are typical.

4.1 Approximation By Low Degree

Our first lemma shows that the low degree part of a noise stable function approximates it:

▶ **Lemma 22** (Noise stable functions have large low degree part). *Let $f : \mathbb{R}^k \to \mathbb{R}$, $|f|_2 < \infty$. Let $0 \leq \rho \leq 1$ and $d \geq 0$. Then, $f^{\leq d}|_2^2 \geq \langle f, T_\rho f \rangle - \rho^d \, |f|_2^2$.*

Proof. We can decompose f to its low degree part and its high degree part, $f = f^{\leq d} + f^{>d}$, and then $\langle f, T_\rho f \rangle = \langle f^{\leq d}, T_\rho f^{\leq d} \rangle + \langle f^{>d}, T_\rho f^{>d} \rangle$. By Cauchy-Schwarz inequality, $\langle f^{\leq d}, T_\rho f^{\leq d} \rangle \leq |f^{\leq d}|_2 |T_\rho f^{\leq d}|_2 \leq |f^{\leq d}|_2^2$. Therefore, by Proposition 10 and Parseval identity, $|f^{\leq d}|_2^2 \geq \langle f^{\leq d}, T_\rho f^{\leq d} \rangle \geq \langle f, T_\rho f \rangle - \langle f^{>d}, T_\rho f^{>d} \rangle > \langle f, T_\rho f \rangle - \rho^d \, |f|_2^2$. ◀

Lemma 22 implies that the low degree part of f_v approximates f_v for a typical $v \in V$: $f_v^{\leq d}|_2^2 \geq 1 - 200C\sqrt{\beta} - (1-\beta)^d$. In the above we used that $|f_v|_2 = 1$. We set $d = \Theta(1/\beta)$, so $|f_v^{\leq d}|_2^2 \geq 0.99$.

4.2 Consistency of Degree-d Parts

In this section we use the high acceptance probability of the consistency test in order to show that for most edges $(u, v) \in E$ the projections of the barycenters of f_u, f_v onto $\Theta_e^\perp$ are extremely close to each other. The proof uses the main technical tools we discussed in the introduction, namely hypercontractivity and concentration.

By hypercontractivity, Boolean functions that are the same except with probability $O(\delta)$ have low degree parts that are $\tilde{O}(\delta)$ apart in 2-norm (note that there is a simple upper bound relying on Parseval identity alone, but it gives the worse upper bound $O(\sqrt{\delta})$), as proven in the following lemma:

▶ **Lemma 23** (Low degree consistency). *Let $f, g : \mathbb{R}^k \to \{\pm 1\}$ be anti-symmetric functions. Let $0 \leq \rho \leq 1$ and $d \leq 2\ln(1/\delta)$. Let $\delta > 0$ be sufficiently small. If $f(x) = g(x)$ with probability $1 - \delta$ over Gaussian $x \in \mathbb{R}^k$, then $|f^{\leq d} - g^{\leq d}|_2 \leq 2\left(\frac{2e}{d}\ln(2/\delta)\right)^{d/2}\delta$.*

Proof. We have $\left|f^{\leq d} - g^{\leq d}\right|_2 = \left|(f-g)^{\leq d}\right|_2$. Let p be a polynomial of degree at most d and 2-norm 1 that maximizes the correlation with $f - g$. Then, $\left|(f-g)^{\leq d}\right|_2 = \langle f - g, p \rangle$. Since f and g are anti-symmetric, so is $f - g$. Hence, p is anti-symmetric. Let $A \subseteq \mathbb{R}^k$ be the set of x with $f(x) > g(x)$. Since $f(x) > g(x)$ iff $g(-x) > f(-x)$, the probability of A is $\delta/2$, and $\langle f - g, p \rangle = \mathbf{E}_x\left[(2p(x) - 2p(-x))\chi_A(x)\right] = 4\mathbf{E}\left[p(x)\chi_A(x)\right]$. By Lemma 13, since $d \leq 2\ln(1/\delta)$ for sufficiently small $\delta > 0$, $4\mathbf{E}_x\left[p(x)\chi_A(x)\right] \leq 4\left(\frac{2e}{d}\ln(2/\delta)\right)^{d/2}(\delta/2)$. The lemma follows by collecting all of the above. ◀

Let $(u,v) \in E$ be a typical edge. By the consistency test, it holds that $f_{u|\Theta_e^\perp}(x) = f_{v|\Theta_e^\perp}(x)$ for random $x \in \Theta_e^\perp$ except with probability $O(\delta/\sqrt{k})$. Thus, by Lemma 23, $\left|(f_{u|\Theta_e^\perp})^{\leq d} - (f_{v|\Theta_e^\perp})^{\leq d}\right|_2 \leq \tilde{O}(\delta/\sqrt{k})$. By Theorem 8, for each $v \in V$, for at least 0.99 fraction of edges $e=(u,v) \in E$, $\left|(f_{v|\Theta_e^\perp})^{\leq d} - (f_v^{\leq d})_{|\Theta_e^\perp}\right|_2 \leq O(1/k)$. By the regularity of the graph, the triangle inequality and a union bound, with probability at least 0.6 over $(u,v) \in E$, the edge is typical, and

$$\begin{aligned}\left|(f_u^{\leq d})_{|\Theta_e^\perp} - (f_v^{\leq d})_{|\Theta_e^\perp}\right|_2 &\leq \left|(f_{u|\Theta_e^\perp})^{\leq d} - (f_u^{\leq d})_{|\Theta_e^\perp}\right|_2 + \left|(f_{v|\Theta_e^\perp})^{\leq d} - (f_v^{\leq d})_{|\Theta_e^\perp}\right|_2 \\ &\leq \tilde{O}(\delta/\sqrt{k} + 1/k). \qquad (1)\end{aligned}$$

4.3 Defining The Assignment

In Section 4.2 we showed that for most edges $e=(u,v) \in E$ the degree-d polynomials $f_u^{\leq d}$ and $f_v^{\leq d}$ are close over $\Theta_e^\perp$. In this section we show how to extract from the degree-d polynomials *unit vectors* that satisfy the constraints and their projections onto $\Theta_e^\perp$ are close.

We next describe the main ideas behind the construction of unit vectors. Close degree-d polynomials, like $f_u^{\leq d}$ and $f_v^{\leq d}$ over $\Theta_e^\perp$, imply close degree-1 parts, and the degree-1 parts correspond to vectors in the linear subspaces associated with u and v. Hence, if the degree-1 parts of the polynomials were known to be of large 2-norm, then one could have assigned each vertex its normalized linear part. Unfortunately, the degree-1 part of the polynomials can be $\vec{0}$. The idea is to differentiate the degree-d polynomials sufficiently many times until the degree-1 part is of sufficiently large 2-norm. The consistency deteriorates with the number of differentiations, but since the degree d is constant, the number of differentiations is constant and the deterioration is limited.

To carry through the above plan we differentiate along random directions $y_1,\ldots,y_{d-1}$, and focus only on hyperplanes $\Theta_e^\perp$ that contain $Y = span\{y_1,\ldots,y_{d-1}\}$, since for those hyperplanes differentiation and restriction to $\Theta_e^\perp$ commute. This is the reason we focus on a zoom-in of the Subspaces Near-Intersection game. This also introduces a certain asymmetry in favor of the directions in Y. To eliminate this asymmetry, we focus on random affine shifts of the space $Y^\perp$. The random choices of Y and the shift would be useful in the analysis, but eventually we will fix them so they satisfy desired properties.

The assignment $\sigma : V \to \mathbb{R}^k$ for the Subspaces Near-Intersection instance is defined by the algorithm in Figure 2. Our analysis closely follows the algorithm.

The first lemma upper bounds the degree and lower bounds the norm on $D_v^{(i)}$ from the algorithm in Figure 2 for $0 \leq i \leq d-1$:

▶ **Lemma 24** (Norm lemma). *For every typical $v \in V$, during the execution of the algorithm in Figure 2, for every $0 \leq i \leq d-1$,*

1. *For all $y_1,\ldots,y_i$, the function $D_v^{(i)}$ is a polynomial of degree at most $d-i$.*
2. $\mathbf{E}_{y_1,\ldots,y_i}\left[\left|\mathbf{E}\left[D_v^{(i)}\right]\right|^2\right] < \eta$.
3. $\mathbf{E}_{y_1,\ldots,y_i}\left[\left|D_v^{(i)}\right|_2^2\right] \geq 0.99 - \eta i$.

Proof. We prove that the three items of the lemma hold by induction on $0 \leq i \leq d-1$. First consider the case of $i=0$ where $D_v^{(0)} = f_v^{\leq d}$.

1. $f_v^{\leq d}$ is a polynomial of degree at most d.
2. By the anti-symmetry folding, $\mathbf{E}\left[f_v^{\leq d}\right] = 0$.
3. For a typical v we have $\left|f_v^{\leq d}\right|_2^2 \geq 0.99$.

Assume that the statement holds for $i-1$ and let us prove it for i.

Global parameters:
- For sufficiently small constants $0 < c_0 < c_1 < 1$ (depending on the constant in Lemma 14), pick uniformly at random
$$\eta \in \left[c_0 \cdot 2^{-2d\log d}, c_1 \cdot 2^{-2d\log d}\right].$$
- Pick Gaussian vectors $y_1, \ldots, y_{d-1} \in \mathbb{R}^k$. Let $Y = span\{y_1, \ldots, y_{d-1}\}$.
- Pick Gaussian vector $y \in Y$.

For every typical $v \in V$ we define the assignment $\sigma(v)$ as follows (for other v's leave $\sigma(v)$ undefined):
1. Let $D_v^{(0)} = f_v^{\leq d}$ and $i = 0$.
2. Let $D_{v,y}^{(0)} : Y^{\perp} \to \mathbb{R}$ be the affine shift $D_{v,y}^{(0)}(x) = D_v^{(0)}(y + x)$
3. While $\left|(D_{v,y}^{(i)})^{=1}\right|_2^2 < \eta$,
   a. $i \leftarrow i + 1$.
   b. Let $D_v^{(i)} = \frac{\partial}{\partial y_i} D_v^{(i-1)}$.
   c. Let $D_{v,y}^{(i)} : Y^{\perp} \to \mathbb{R}$ be the affine shift $D_{v,y}^{(i)}(x) = D_v^{(i)}(y + x)$.
4. $i_v \leftarrow i$.
5. Let $vec_v \in Y^{\perp}$ be $(D_{v,y}^{(i_v)})^{=1}$.
6. $\sigma(v) \leftarrow \frac{vec_v}{|vec_v|_2}$.

Figure 2 The assignment $\sigma : V \to \mathbb{R}^k$ for the Y-zoom-in of Subspaces Near-Intersection

1. The function $D_v^{(i)}$ is a polynomial of degree at most $deg(D_v^{(i-1)}) - 1$. The degree bound therefore follows from the inductive hypothesis.

2. $\mathbf{E}\left[D_v^{(i)}\right]$ is the constant part of $D_v^{(i)} = \langle \nabla D_v^{(i-1)}, y_i \rangle$. Moreover, $\nabla D_v^{(i-1)}$ depends on $y_1, \ldots, y_{i-1}$ and is independent of y_i. Thus, $\mathbf{E}\left[D_v^{(i)}\right] = \langle (D_v^{(i-1)})^{=1}, y_i \rangle$ is a normal variable with standard deviation $\left|(D_v^{(i-1)})^{=1}\right|_2$. By the design of the algorithm, $\left|(D_v^{(i-1)})^{=1}\right|_2^2 < \eta$ and hence $\mathbf{E}_{y_1,\ldots,y_{d-1}}\left[\left|\mathbf{E}\left[D_v^{(i)}\right]\right|^2\right] < \eta$.

3. We have $D_v^{(i)} = \langle \nabla D_v^{(i-1)}, y_i \rangle$, where $\nabla D_v^{(i-1)}$ depends on $y_1, \ldots, y_{i-1}$ and is independent of y_i. Thus, for every $x \in \mathbb{R}^k$, it holds that $D_v^{(i)}(x)$ is a normal variable with standard deviation $\left|\nabla D_v^{(i-1)}(x)\right|_2$. Hence, $\mathbf{E}_{y_1,\ldots,y_{d-1},x}\left[(D_v^{(i)})(x)^2\right] = \mathbf{E}\left[\left|\nabla D_v^{(i-1)}(x)\right|_2^2\right]$. By the Gaussian Poincaré inequality (Lemma 15), for any $y_1, \ldots, y_i$,

$$\mathop{\mathbf{E}}_x\left[\nabla D_v^{(i-1)}(x)^2\right] \geq \mathrm{Var} D_v^{(i-1)} = \left|D_v^{(i-1)}\right|_2^2 - \mathbf{E}\left[D_v^{(i-1)}\right]^2.$$

By the inductive hypothesis, $\mathbf{E}\left[\left|D_v^{(i-1)}\right|_2^2\right] \geq 0.99 - \eta(i-1)$ and $\mathbf{E}\left[D_v^{(i-1)}\right]^2 < \eta$. Hence, $\mathbf{E}\left[(D_v^{(i)}(x))^2\right] \geq 0.99 - \eta(i-1) - \eta = 0.99 - \eta i$. ◀

By the following proposition and the constraints folding (see Definition 21), whenever $\sigma(v)$ is defined it satisfies $A_v \sigma(v) = \vec{0}$.

▶ **Proposition 25.** *Let $f : \mathbb{R}^k \to \mathbb{R}$. If f satisfies a constraints folding, then so do $f^{=i}$ for any i, any derivative of f, and any scalar multiplication of f.*

The next lemma uses Lemma 24 to argue that $\sigma(v)$ is well-defined for most vertices $v \in V$.

▶ **Lemma 26** (Assignment lemma). *Let $v \in V$ be typical. With probability at least 0.99 over $y_1, \ldots, y_{d-1}$ and y, the algorithm in Figure 2 terminates, i_v is well-defined, and $\left|(D_{v,y}^{(i_v)})^{=1}\right|_2^2 \geq \eta$.*

Proof. The algorithm terminates and i_v is well-defined iff there exists $0 \leq i \leq d-1$ such that $\left|(D_{v,y}^{(i)})^{=1}\right|_2^2 \geq \eta$. Assume on way of contradiction that there is no such i. By Lemma 24, when the algorithm reaches $i = d-1$, the polynomial $D_v^{(d-1)}$ is of degree 1 and $\mathbf{E}_{Y,y}\left[\left|D_{v,y}^{(d-1)}\right|_2^2\right] = \mathbf{E}_{y_1,\ldots,y_{d-1}}\left[\left|D_v^{(d-1)}\right|_2^2\right] \geq 0.9$. Since each coordinate of the coefficients vector $\nabla D_{v,y}^{(d-1)}$ is a polynomial of degree at most d in $y_1, \ldots, y_{d-1}$ and y, the norm $\left|D_{v,y}^{(d-1)}\right|_2^2$ is a polynomial of degree at most $2d$ in $y_1, \ldots, y_{d-1}, y$. By convexity, $\mathbf{E}_{y_1,\ldots,y_{d-1},y}\left[\left|D_{v,y}^{(d-1)}\right|_2^4\right] \geq \left(\mathbf{E}\left[\left|D_{v,y}^{(d-1)}\right|_2^2\right]\right)^2 \geq 0.81$. By Carbery-Wright anti-concentration (Lemma 14), $\left|D_{v,y}^{(d-1)}\right|_2^2 \geq \eta$ with probability at least 0.99 over $y_1, \ldots, y_{d-1}$ and y. In this case, the loop in the algorithm in Figure 2 terminates and $i_v = d-1$. ◀

The next lemma argues consistency between $D_u^{(i)}$ and $D_v^{(i)}$ across most edges $e = (u,v) \in E$, provided that $y_1, \ldots, y_{d-1} \in \Theta_e^\perp$ (note that the degree d is constant so the large dependence in d – which we state here explicitly, and later omit in the $O(\cdot)$ notation – is permissible).

▶ **Lemma 27** (Consistency lemma). *With probability at least 0.6 over $e = (u,v) \in E$, for every $0 \leq i \leq d-1$,*

$$\mathop{\mathbf{E}}_{y_1,\ldots,y_i \in \Theta_e^\perp}\left[\left|(D_u^{(i)})_{|\Theta_e^\perp} - (D_v^{(i)})_{|\Theta_e^\perp}\right|_2\right] \quad \leq \quad (O(d))^i \cdot \tilde{O}(\delta/\sqrt{k} + 1/k).$$

Proof. By induction over i. For $i = 0$, the inequality follows from inequality (1): for at least 0.6 of the edges $e = (u,v) \in E$ we have $\left|(f_u^{\leq d} - f_v^{\leq d})_{|\Theta_e^\perp}\right|_2 \leq \tilde{O}(\delta/\sqrt{k} + 1/k)$. Assume that the claim holds for $i-1$, and let us prove it for i. Let $(u,v) \in E$. We have $D_u^{(i)} - D_v^{(i)} = \langle \nabla(D_u^{(i-1)} - D_v^{(i-1)}), y_i \rangle$, where $\nabla(D_u^{(i-1)} - D_v^{(i-1)})$ depends on $y_1, \ldots, y_{i-1}$ and is independent of y_i. Thus, for every $y_1, \ldots, y_{i-1}$ and $x \in \mathbb{R}^k$, it holds that $(D_u^{(i)} - D_v^{(i)})(x)$ is a normal variable with standard deviation $\left|\nabla(D_u^{(i-1)} - D_v^{(i-1)})(x)\right|_2$. Thus, by concavity and the inductive hypothesis,

$$\begin{aligned}
\mathop{\mathbf{E}}_{y_1,\ldots,y_i \in \Theta_e^\perp}\left[\sqrt{\mathop{\mathbf{E}}_{x \in \Theta_e^\perp}\left[(D_u^{(i)} - D_v^{(i)})(x)^2\right]}\right] \quad &\leq \quad \mathop{\mathbf{E}}_{y_1,\ldots,y_{i-1}}\left[\sqrt{\mathop{\mathbf{E}}_{x,y_i}\left[\left|\nabla(D_u^{(i-1)} - D_v^{(i-1)})(x)\right|_2^2\right]}\right] \\
&\leq \quad O(d) \cdot \mathop{\mathbf{E}}_{y_1,\ldots,y_{i-1}}\left[\sqrt{\mathop{\mathbf{E}}_{x}\left[(D_u^{(i-1)} - D_v^{(i-1)})(x)^2\right]}\right] \\
&\leq \quad (O(d))^i \tilde{O}(\delta/\sqrt{k} + 1/k).
\end{aligned}$$

◀

The next lemma is similar to Lemma 27, but applies to the shifted $D_{u,y}^{(i)}$ and $D_{v,y}^{(i)}$ rather than to $D_u^{(i)}$ and $D_v^{(i)}$. Recall that $Y = span\{y_1, \ldots, y_{d-1}\}$ and $E_Y = \left\{e \in E \mid Y \subseteq \Theta_e^\perp\right\}$. For each $e \in E_Y$ we write $\Theta_e^\perp = Y + S_e$. The subspace S_e is a uniform hyperplane in $Y^\perp$.

▶ **Lemma 28.** *With probability at least* 0.99 *over* Y *and* y*, with probability at least* 0.6 *over* $e = (u, v) \in E_Y$*, for every* $0 \le i \le d-1$*,* $\left|(D^{(i)}_{u,y})_{|S_e^\perp} - (D^{(i)}_{v,y})_{|S_e^\perp}\right|_2 \le \tilde{O}(\delta/\sqrt{k} + 1/k)$.

Proof. By Lemma 27, with probability at least 0.6 over $e = (u, v) \in E$, for every $0 \le i \le d-1$,

$$\mathop{\mathbf{E}}_{Y \subseteq \Theta_e^\perp}\left[\sqrt{\mathop{\mathbf{E}}_{y \in Y, x \in S_e}\left[(D^{(i)}_{u,y} - D^{(i)}_{v,y})(x)^2\right]}\right] \quad \le \quad \tilde{O}(\delta/\sqrt{k} + 1/k). \tag{2}$$

By concavity, with probability at least 0.6 over $e = (u, v) \in E$, for every $0 \le i \le d-1$,

$$\mathop{\mathbf{E}}_{Y \subseteq \Theta_e^\perp}\left[\mathop{\mathbf{E}}_{y \in Y}\left[\sqrt{\mathop{\mathbf{E}}_{x \in S_e}\left[(D^{(i)}_{u,y} - D^{(i)}_{v,y})(x)^2\right]}\right]\right] \quad \le \quad \tilde{O}(\delta/\sqrt{k} + 1/k). \tag{3}$$

By Markov's inequality, with probability at least 0.6 over $e = (u, v) \in E$, with probability at least 0.99 over $Y \subseteq \Theta_e^\perp$ and $y \in Y$, we have

$$\left|(D^{(i)}_{u,y})_{|S_e} - (D^{(i)}_{v,y})_{|S_e}\right|_2 \quad \le \quad \tilde{O}(\delta/\sqrt{k} + 1/k). \tag{4}$$

By Lemma 17, the distribution induced on e and Y by first picking $e \in E$ out of the set of fraction 0.6, and then picking $Y \subseteq \Theta_e^\perp$, is close to the distribution that picks Y by picking Gaussian $y_1, \ldots, y_{d-1}$, $Y = span\{y_1, \ldots, y_{d-1}\}$, and then picks $e \in E_Y$ that belongs to the set of fraction 0.6. Therefore, with probability 0.99 over Y, y, the above event also holds with probability 0.6 over $e \in E_Y$. ◀

By Lemmas 26 and 28, there exist $y_1, \ldots, y_{d-1}$ and y, such that with probability at least 0.5 over $e = (u, v) \in E_Y$, the following two conditions holds (recall that when one picks $e = (u, v) \in E_Y$ uniformly, the distribution over v is uniform over V, and that 0.9 fraction of the vertices $v \in V$ are typical):

1. $\left|(D^{(i_v)}_{v,y})^{=1}\right|_2^2 \ge \eta$.
2. For every $0 \le i \le d-1$, $\left|(D^{(i)}_{u,y})_{|S_e^\perp} - (D^{(i)}_{v,y})_{|S_e^\perp}\right|_2 \le \tilde{O}(\delta/\sqrt{k} + 1/k)$.

The second item implies that for every $0 \le i \le d-1$, $\left|((D^{(i)}_{u,y})_{|S_e})^{=1} - ((D^{(i)}_{v,y})_{|S_e})^{=1}\right|_2 \le \tilde{O}(\delta/\sqrt{k} + 1/k)$. The case $d = 1$ of Theorem 8 implies that for every $u \in V$ with probability at least 0.999 over the edge $e = (u, v) \in E_Y$, for every i,

$$\left|((D^{(i)}_{u,y})_{|S_e})^{=1} - (D^{(i)}_{u,y})^{=1})_{|S_e}\right|_2 \quad \le \quad \tilde{O}(\delta/\sqrt{k} + 1/k). \tag{5}$$

Applying the same to $v \in V$ and taking a union bound and a triangle inequality, with probability at least 0.49 over $(u, v) \in E_Y$, for every i,

$$\left|((D^{(i)}_{u,y})^{=1})_{|S_e} - ((D^{(i)}_{v,y})^{=1})_{|S_e}\right|_2 \quad \le \quad \tilde{O}(\delta/\sqrt{k} + 1/k). \tag{6}$$

Note that inequality (6) implies consistency between vectors corresponding to u and to v restricted to the hyperplane of interest. It remains to argue that $i_u = i_v$ with high probability. As a consequence of inequality (6), with probability at least 0.49 over $e = (u, v) \in E_Y$, for every i,

$$\left|\left|((D^{(i)}_{u,y})^{=1})_{|S_e}\right|_2^2 - \left|((D^{(i)}_{v,y})^{=1})_{|S_e}\right|_2^2\right| \quad \le \quad \tilde{O}(\delta/\sqrt{k} + 1/k). \tag{7}$$

By sampling (Lemma 16) and union bound, for every $u \in V$, with probability at least 0.999 over $e = (u, v) \in E_Y$, for every i,

$$\left|\left|((D^{(i)}_{u,y})^{=1})_{|S_e}\right|_2^2 - \left|(D^{(i)}_{u,y})^{=1}\right|_2^2\right| \quad \le \quad \tilde{O}(1/k). \tag{8}$$

A similar bound holds for v. Hence, from inequalities (7) and (8) via a union bound and a triangle inequality, with probability at least 0.47 over $e = (u,v) \in E_Y$, for every i,

$$\left| \left|(D_{u,y}^{(i)})^{=1}\right|_2^2 - \left|(D_{v,y}^{(i)})^{=1}\right|_2^2 \right| \quad \leq \quad \tilde{O}(\delta/\sqrt{k} + 1/k). \tag{9}$$

By the design of the algorithm in Figure 2, inequality (9) guarantees that $i_u = i_v$ except with probability $\tilde{O}(\delta + 1/k)$. In this case, by inequality (5),

$$|Proj_{S_e}(vec_u) - Proj_{S_e}(vec_v)| \quad \leq \quad \tilde{O}(\delta/\sqrt{k} + 1/k). \tag{10}$$

The vectors vec_u and vec_v are normalized to obtain $\sigma(u)$ and $\sigma(v)$, respectively. Hence, by inequalities (10) and (9), and since $\left|(D_{u,y}^{(i_u)})^{=1}\right|_2 \geq \Omega(1)$, with probability at least 0.47 over $e = (u,v) \in E_Y$, $|Proj_{S_e}(\sigma(u)) - Proj_{S_e}(\sigma(v))| \leq \tilde{O}(\delta/\sqrt{k} + 1/k)$.

5 Concentration of the restricted Hermite tensors

In this section we prove Theorem 8. **Note:** In this section we use n to denote the dimension.

5.1 Overview of the proof

We first sketch some of the main steps of the proof. Consider the functional $Q(f) = \mathbb{E}_\Theta \left|(f_{|\Theta^\perp})^{\leq d} - (f^{\leq d})_{|\Theta^\perp}\right|_2^2$ where Θ is uniformly distributed in the sphere. It is not hard to check that $Q(f)$ is a quadratic form in f, which is invariant under compositions of f with orthogonal transformations.

Here we allude to Schur's lemma, which states that rotational invariant quadratic forms on functions *on the sphere* can be expressed as linear combinations of the L_2 norms of the projections onto eigenspaces of the Laplacian. This means that the maximum of the quadratic form among functions with a perscribed L_2 norm must be attained on a function which only depends on the first coordinate x_1.

Our quadratic form, however, is a functional of functions on $\mathbb{R}^n$ rather than the sphere; this issue can be bypassed by considering homogeneous functions and using the concentration of the Gaussian in a thin spherical shell. Thus the first step of the proof roughly implies that it is sufficient to consider functions of the form $f(x_1, ..., x_n) = g(x_1)$.

By applying rotations around the first vector of the standard basis, e_1, it is not hard to see that when f is of the above form, the quantity $\theta \to \left|(f_{|\theta^\perp})^{\leq d} - (f^{\leq d})_{|\theta^\perp}\right|_2^2$ only depends on $\theta_1 := \langle \theta, e_1 \rangle$. By concentration of measure, this angle is of the order $1/\sqrt{n}$. The technical bulk of the proof is to show that the above expression behaves like θ_1^4 for small θ_1.

5.2 Preliminaries

Identify a tensor T of degree ℓ with a multilinear polynomial $T[x^1, ..., x^\ell] = \sum_{i_1,...,i_\ell \in [n]^\ell} T_{i_1,...,i_\ell} x^1_{i_1} \cdot ... \cdot x^\ell_{i_\ell}$. For any $x \in \mathbb{R}^n$, denote by $H^{(k)}(x)$, the k-th Hermite tensor associated with x, defined by $H^{(k)}(x) := (-1)^k \phi(x)^{-1}(\nabla^k \phi(x))$, where $\phi(x) = \exp(-|x|^2/2)$. For example, we have $H^{(1)}(x) = x$, $H^{(1)}(x)[y] = \langle x, y\rangle$, $H^{(2)}(x) = x^{\otimes 2} - \mathrm{I}_n$, $H^{(1)}(x)[y,z] = \langle x,y\rangle\langle x,z\rangle - \langle y,z\rangle$. and $H^{(3)}(x)[y,z,w] = \langle x,y\rangle\langle x,z\rangle\langle x,w\rangle - \langle x,y\rangle\langle z,w\rangle - \langle x,z\rangle\langle y,w\rangle - \langle x,w\rangle\langle y,z\rangle$, (see [31, p. 157]). For two tensors T, U of degree ℓ, define the Hilbert-Schmidt inner product by $\langle T, U\rangle_{HS} = \sum_{(i_1,...,i_\ell)\in[n]^\ell} T_{i_1,...,i_\ell} U_{i_1,...,i_\ell}$ and the corresponding norm $\|T\|_{HS}^2 = \langle T, T\rangle_{HS}$. We will allow ourselves to abbreviate the notation and write $\|T\|$ and $\langle T, U\rangle$ whenever this causes no confusion. For a function f, we define its k-barycenter by $b_k(f) := \int H^{(k)}(x)f(x)d\gamma(x)$ and also denote $\alpha_k(f)^2 := \|b_k(f)\|_{HS}^2$. For a tensor T of degree ℓ and an orthogonal projection P, define $PT[x_1,...,x_\ell] := T[Px_1,...,Px_\ell]$. It is not hard to verify that for $f : \mathbb{R}^n \to \mathbb{R}$ and for $\theta \in \mathbb{S}^{n-1}$, one has

$$P_{\theta^\perp} \int H^{(k)}(x)f(x)d\gamma(x) = \int H^{(k)}(x)f_\theta(x)d\gamma(x) \tag{11}$$

where $f_\theta(x) = \int_{-\infty}^{\infty} f(P_{\theta^\perp}x + t\theta)d\gamma(t)$ is the marginal of f on $\theta^\perp$.

For a unit vector $\theta \in \mathbb{S}^{n-1}$, let γ_θ be the Gaussian measure restricted to $\{\langle x, \theta\rangle = 0\}$, in other words, $d\gamma_\theta(x) = \frac{1}{(2\pi)^{(n-1)/2}}e^{-|x|^2/2}\mathbf{1}_{\langle x,\theta\rangle=0}d\mathcal{H}_{n-1}(x)$. For a function $f : \mathbb{R}^n \to \mathbb{R}$ define, by slight abuse of notation, $b_k(f;\theta) = \int H^{(k)}(x)f(x)d\gamma_\theta$. By the orthogonality of Hermite polynomials, we have for all $f^{=k}(x) = \frac{1}{k!}\langle H^{(k)}(x), b_k(f)\rangle, \quad \forall x \in \mathbb{R}^n$. Likewise, for all $\theta \in \mathbb{S}^{n-1}$ $(f_{|\theta^\perp})^{=k} = \frac{1}{k!}\langle H^{(k)}(x), b_k(f;\theta)\rangle, \quad \forall x \in \theta^\perp$. Therefore, by Parseval's identity, we have $\left|(f_{|\theta^\perp})^{=k} - (f^{=k})_{|\Theta^\perp}\right|_2 = \frac{1}{k!}\|P_{\theta^\perp}(b_k(f;\theta) - b_k(f))\|_{HS}$. Thus, for a function $f : \mathbb{R}^n \to \mathbb{R}$, we define $Q(f) = Q_k(f) := \mathbb{E}_{\theta\sim\sigma}\|P_{\theta^\perp}(b_k(f;\theta) - b_k(f))\|_{HS}^2$. Theorem 8 will follow immediately from the next result.

▶ **Theorem 29.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be 0-homogeneous.* $\mathbb{E}_{\theta\sim\sigma}\|b_k(f;\theta) - b_k(f))\|_{HS}^2 = O_k(1/n^2)$.

Proof of Theorem 8. Apply Theorem 29 for and $k \le d$ and use Chebyshev's inequality and a union bound. ◀

5.3 A reduction to functions depending on one variable

The proof of the above theorem relies on the following lemma, which essentially reduces the problem to the case that f is a low-degree polynomial which only depends on one variable.

▶ **Lemma 30.** *For any 0-homogeneous function f with $\|f\|_{L_2(\gamma)} = 1$, there is a polynomial $h : \mathbb{R} \to \mathbb{R}$ of degree at most $8k$ such that, defining $\tilde{f}(x) = h\left(\frac{x_1}{|x|/\sqrt{n}}\right)$, we have $\|\tilde{f}\|_{L_2(\gamma)} = 1$ and $\left|Q_k(f) - Q_k(\tilde{f})\right| = O(1/n^2)$.*

The main step towards the lemma is the following proposition:

▶ **Proposition 31.** *Assuming that f is 0-homogeneous, There exists a polynomial q on $\mathbb{R}$, of degree at most $8k$, such that $Q_k(f) = \int_{\mathbb{S}^{n-1}}\int_{\mathbb{S}^{n-1}} f(x)f(y)q(\langle x, y\rangle)d\gamma(x)d\gamma(y) + O(1/n^2)$.*

Before we prove Proposition 31, we need two additional propositions, whose proofs are deferred to the end of this section.

▶ **Proposition 32.** *There exist constants C_n,C'_n such that $C_n, C'_n < C$ for some universal constant $C > 0$, and such that the following holds. Let $x, y \in \mathbb{R}^n$ and let θ be uniformly distributed in $\mathbb{S}^{n-1}$. Then, for every continuous $g : \mathbb{S}^{n-1} \to \mathbb{R}$, $\lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2}\mathbb{E}\Big[\mathbf{1}\{|\langle x, \theta\rangle| \le \varepsilon, |\langle y, \theta\rangle| \le \varepsilon\}g(\theta)\Big] = C_n \frac{1}{|x||y|\sqrt{1-\left\langle \frac{x}{|x|}, \frac{y}{|y|}\right\rangle^2}}\mathbb{E}g(\theta_1)$, where θ_1 is uniform in $\mathbb{S}^{n-1} \cap x^\perp \cap y^\perp$. Furthermore, $\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathbb{E}\left[\mathbf{1}\{|\langle x, \theta\rangle| \le \varepsilon\}g(\theta)\right] = \frac{C'_n}{|x|}g(\theta_2)$, where θ_2 is uniform in $\mathbb{S}^{n-1} \cap x^\perp$.*

▶ **Proposition 33.** *For every $k, n \in \mathbb{N}$ there exist polynomials p_1, p_2, p_3, p_4 in 3 variables, of degree at most $3k$, with coefficients bounded by $O_k(n^k)$, such that the following holds. For each $x, y \in \mathbb{R}^n$, let θ_1 be uniform in $\mathbb{S}^{n-1} \cap x^\perp \cap y^\perp$ and let θ_2 be uniform in $\mathbb{S}^{n-1} \cap x^\perp$. Then we have the representations $\mathbb{E}\langle P_{\theta_1^\perp}H^{(k)}(x), P_{\theta_1^\perp}H^{(k)}(y)\rangle = p_1(|x|, |y|, \rho(x, y)) + \sqrt{1-\rho(x, y)^2} \cdot p_2(|x|, |y|, \rho(x, y))$ and $\mathbb{E}\langle P_{\theta_2^\perp}H^{(k)}(x), P_{\theta_2^\perp}H^{(k)}(y)\rangle = p_3(|x|, |y|, \rho(x, y)) + \sqrt{1-\rho(x, y)^2} \cdot p_4(|x|, |y|, \rho(x, y))$ where $\rho(x, y) := \left\langle \frac{x}{|x|}, \frac{y}{|y|}\right\rangle$.*

Proof of Proposition 31. By an approximation argument, we may assume that f is continuous. We then have, $\beta(f;\theta) = \lim_{\varepsilon\to 0} \frac{\sqrt{2\pi}}{2\varepsilon} \int \mathbf{1}\{|\langle x,\theta\rangle| \le \varepsilon\} H^{(k)}(x) f(x) d\gamma$. Therefore, we have the following expression for $Q(f)$:

$$\mathbb{E}_{\theta\sim\sigma} \left\| \lim_{\varepsilon\to 0} P_{\theta^\perp} \frac{\sqrt{2\pi}}{2\varepsilon} \int \mathbf{1}\{|\langle x,\theta\rangle| \le \varepsilon\} H^{(k)}(x) f(x) d\gamma(x) - P_{\theta^\perp} \int H^{(k)}(x) f(x) d\gamma(x) \right\|_{HS}^2$$
$$= \lim_{\varepsilon\to 0} \left(\mathbb{E}_{\theta\sim\sigma} \left[\frac{\pi}{2\varepsilon^2} \int \mathbf{1} \left\{ \begin{array}{c} |\langle x,\theta\rangle| \le \varepsilon \\ |\langle y,\theta\rangle| \le \varepsilon \end{array} \right\} \langle P_{\theta^\perp} H^{(k)}(x), P_{\theta^\perp} H^{(k)}(y)\rangle f(x) f(y) d\gamma(x,y) \right.\right.$$
$$- \frac{\sqrt{2\pi}}{\varepsilon} \int \mathbf{1}\{|\langle x,\theta\rangle| \le \varepsilon\} \langle P_{\theta^\perp} H^{(k)}(x), P_{\theta^\perp} H^{(k)}(y)\rangle f(x) f(y) d\gamma(x,y)$$
$$\left.\left. + \int \langle P_{\theta^\perp} H^{(k)}(x), P_{\theta^\perp} H^{(k)}(y)\rangle f(x) f(y) d\gamma(x,y) \right] \right)$$
$$= \int \left(h_1(x,y) - 2h_2(x,y) + h_3(x,y)\right) f(x) f(y) d\gamma(x,y),$$

where

- $h_1(x,y) = \lim_{\varepsilon\to 0} \mathbb{E}_{\theta\sim\sigma} \frac{\pi}{2\varepsilon^2} \mathbf{1} \left\{ \begin{smallmatrix} |\langle x,\theta\rangle| \le \varepsilon \\ |\langle y,\theta\rangle| \le \varepsilon \end{smallmatrix} \right\} \langle P_{\theta^\perp} H^{(k)}(x), P_{\theta^\perp} H^{(k)}(y)\rangle$,
- $h_2(x,y) = \lim_{\varepsilon\to 0} \mathbb{E}_{\theta\sim\sigma} \frac{\sqrt{2\pi}}{\varepsilon} \mathbf{1}\left\{|\langle x,\theta\rangle| \le \varepsilon\right\} \langle P_{\theta^\perp} H^{(k)}(x), P_{\theta^\perp} H^{(k)}(y)\rangle$,
- $h_3(x,y) = \mathbb{E}_{\theta\sim\sigma} \langle P_{\theta^\perp} H^{(k)}(x), P_{\theta^\perp} H^{(k)}(y)\rangle$.

By Proposition 32, we have

$$h_1(x,y) = \frac{C_n}{|x||y|\sqrt{1 - \left\langle \frac{x}{|x|}, \frac{y}{|y|} \right\rangle^2}} \mathbb{E}_{\theta_1 \sim U(\mathbb{S}^{n-1}\cap x^\perp \cap y^\perp)} \langle P_{\theta_1^\perp} H^{(k)}(x), P_{\theta_1^\perp} H^{(k)}(y)\rangle$$

for some constant C_n depending only on the dimension, and which is smaller than a universal constant $C > 0$. From this point on, the expression C_k will denote a constant that depends only on k, whose value may vary between different instances.

By Proposition 33 there are polynomials p_1, p_2 of degree at most $3k$, with coefficients bounded by $C_k n^k$, such that $h_1(x,y) = \frac{1}{|x||y|} \left(\frac{p_1(\rho(x,y),|x|,|y|)}{\sqrt{1-\rho(x,y)^2}} + p_2(\rho(x,y),|x|,|y|) \right)$ where $\rho(x,y) = \left\langle \frac{x}{|x|}, \frac{y}{|y|} \right\rangle$.

Since the coefficients of p_1 are bounded by $C_k n^k$, we have $p_1(\rho(x,y),|x|,|y|) \le C_k n^k (|x| + 1)^k (|y|+1)^k$. By taking the Taylor expansion of the function $s \to \frac{1}{\sqrt{1-s^2}}$ of order $2k+4$, we conclude that there exists a polynomial $q(\cdot)$ of degree $4k+4$ such that $h_1(x,y) = \frac{q(\rho(x,y)) p_1(\rho(x,y),|x|,|y|) + p_2(\rho(x,y),|x|,|y|)}{|x||y|} + O_k\left(n^k (1+|x|)^k (1+|y|)^k \rho(x,y)^{4k+4} \right)$. By Cauchy-Schwartz and since $\mathbb{E}_{x\sim\gamma} |x|^{2k} \le C_k n^k$ and $\mathbb{E}_{x,y\sim\gamma} \left[|\rho(x,y)|^\ell \right] \le C_\ell n^{-\ell/2}$, we have $n^k \int (|x| + 1)^k (|y|+1)^k \rho(x,y)^{4k+4} f(x) f(y) d\gamma(x,y) \le \frac{1}{n^2} C_k \|f\|_2^2$. The last two displays imply that there exists a polynomial q_1 of degree at most $8k$ so that $\int h_1(x,y) f(x) f(y) d\gamma(x,y) = \int q_1(\rho(x,y),|x|,|y|) d\gamma(x,y) + O_k\left(\frac{\|f\|_2^2}{n^2} \right)$. Following a similar argument with the terms h_2 and h_3, we conclude that there exists a polynomial p of degree at most $8k$ such that $Q(f) = \int p\left(\rho(x,y),|x|,|y|\right) f(x) f(y) d\gamma(x,y) + O_k\left(\frac{\|f\|_2^2}{n^2} \right)$. Since f is 0-homogeneous, by polar integration one learns that for all k_1,k_2,k_3, there exist constants $C_{k_1,k_2,k_3}, C'_{k_1,k_2,k_3}$ such that $\int\int \langle x,y\rangle^{k_1} |x|^{k_2} |y|^{k_3} f(x) f(y) d\gamma(x) d\gamma(y)$ can be written as

$$= C_{k_1,k_2,k_3} \int\int \left\langle \frac{x}{|x|}, \frac{y}{|y|} \right\rangle^{k_1} f(x) f(y) d\gamma(x) d\gamma(y)$$
$$= C'_{k_1,k_2,k_3} \int_{\mathbb{S}^{n-1}} \int_{\mathbb{S}^{n-1}} \left\langle \frac{x}{|x|}, \frac{y}{|y|} \right\rangle^{k_1} f(x) f(y) d\sigma(x) d\sigma(y).$$

We conclude that there exists a polynomial $q(\cdot)$ of degree at most $8k$ such that $Q_k(f) = \int_{\mathbb{S}^{n-1}\times\mathbb{S}^{n-1}} q\left(\left\langle \frac{x}{|x|}, \frac{y}{|y|}\right\rangle\right) f(x)f(y)d\sigma(x)d\sigma(y) + O_k(1/n^2)$. ◀

Proof of Lemma 30. For a function $h \in L_2(\mathbb{S}^{n-1})$, define by $\mathrm{Proj}_{\mathcal{S}_k} h$ the orthogonal projection of h into the subspace spanned by spherical harmonics of degree k. An application of Schur's lemma (or the Funk-Hecke formula) ensures that for every polynomial g degree ℓ there exist constant $\alpha_1, ..., \alpha_\ell$ such that $\int_{\mathbb{S}^{n-1}}\int_{\mathbb{S}^{n-1}} f(x)f(y)g(\langle x,y\rangle)d\gamma(x)d\gamma(y) = \sum_{i\le\ell}\alpha_i\|\mathrm{Proj}_{\mathcal{S}_i} f\|^2_{L_2(\mathbb{S}^{n-1})}$ Thus, by Proposition 31 we learn that there are some $(\alpha_i)_{i=0}^{8k}$ such that

$$Q(f) = \sum_{0\le i\le 8k} \alpha_i \|\mathrm{Proj}_{\mathcal{S}_i} f\|^2_{L_2(\mathbb{S}^{n-1})} + O_k(1/n^2). \tag{12}$$

(in the last formula, by slight abuse of notation, on the right hand side the function f should be understood as its restriction to the sphere). Now, for any $j \in \mathbb{N}$ there exists a function h_j depending only on x_1 such that $\|\mathrm{Proj}_{\mathcal{S}_i} h_j\|^2_{L_2(\mathbb{S}^{n-1})} = \mathbf{1}_{\{i=j\}}$. Therefore, defining $\tilde{f}(x) = \sum_j h_j\left(\frac{x_1}{|x|}\right)\|\mathrm{Proj}_{\mathcal{S}_i} f\|_{L_2(\mathbb{S}^{n-1})}$, we have $\|\mathrm{Proj}_{\mathcal{S}_i} f\|_{L_2(\mathbb{S}^{n-1})} = \|\mathrm{Proj}_{\mathcal{S}_i}\tilde{f}\|_{L_2(\mathbb{S}^{n-1})}$ for all i, and therefore by (12), we have $|Q(f) - Q(\tilde{f})| = O(1/n^2)$. Moreover, $\|f\|_{L_2(\gamma)} = \|f\|_{L_2(\mathbb{S}^{n-1})} = \|\tilde{f}\|_{L_2(\mathbb{S}^{n-1})}$. This completes the proof. ◀

5.4 Finishing the proof

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function which has the form $f(x_1, ..., x_n) = h\left(x_1\frac{\sqrt{n}}{|x|}\right)$ for some polynomial $h : \mathbb{R} \to \mathbb{R}$ of degree at most $8k$ and with $\|f\|_{L_2(\gamma)} = 1$. In light of Lemma 30, Theorem 29 will be concluded by showing that

$$Q(f) = O_k(1/n^2). \tag{13}$$

Let θ be uniform in $\mathbb{S}^{n-1}$. We first show that, by symmetry, we can essentially assume in our calculations that $\theta \in \mathrm{span}\{e_1, e_2\}$. Let us write $\theta_1 = \langle\theta, e_1\rangle$ and define $\tilde{\theta} := e_1\theta_1 + e_2\sqrt{1-\theta_1^2}$. By symmetry of the function f to orthogonal transformations which keep e_1 fixed, we have $Q(f) = \mathbb{E}_{\theta_1}\|P_{\tilde{\theta}^\perp}(b_k(f;\tilde{\theta}) - b_k(f))\|^2_{HS}$. In order to understand the role of the projection onto the subspace $\tilde{\theta}^\perp$, define an orthonormal basis to $\tilde{\theta}^\perp$ as follows: Set $e_1' = \sqrt{1-\theta_1^2}e_1 - \theta_1 e_2$ and $e_i' = e_{i+1}$ for $i = 2, ..., n-1$, so that $(e_i')_{i=1}^{n-1}$ form an orthonormal basis for $\tilde{\theta}^\perp$. We have,

$$\|P_{\tilde{\theta}^\perp}(b_k(f;\tilde{\theta}) - b_k(f))\|^2_{HS} = \sum_{(i_1,\dots,i_k)\in[n-1]^k} \left(b_k(f;\tilde{\theta})[e'_{i_1}, ..., e'_{i_k}] - b_k(f)[e'_{i_1}, ..., e'_{i_\ell}]\right)^2. \tag{14}$$

Fix $I = (i_1, ..., i_\ell) \in [n-1]^\ell$. There exists a function J_I and $\alpha(I) \in [k]$ such that

$$H^{(k)}(x)[e'_{i_1}, ..., e'_{i_\ell}] = H_{\alpha(I)}(\langle x, e'_1\rangle)J_I(\mathrm{Proj}_L(x)), \tag{15}$$

where $L = \mathrm{span}(e'_2, ..., e'_{n-1})$. Let $\Gamma_1 \sim \mathcal{N}(0,1), \Gamma_2 \sim \mathcal{N}(0,1), \Gamma_3 \sim N(0, \mathrm{Proj}_L)$ be independent. In this case, note that $e'_1\Gamma_1 + \tilde{\theta}\Gamma_2 + \Gamma_3 \overset{(d)}{=} \mathcal{N}(0, \mathrm{I}_n)$. We therefore have by equation (15) and by the definition of $b_k(f;\tilde{\theta})$,

$$b_k(f;\tilde{\theta})[e'_{i_1}, ..., e'_{i_\ell}] = \mathbb{E}\left[H_{\alpha(I)}(\Gamma_1)J_I(\Gamma_3)h\left(\frac{\sqrt{1-\theta_1^2}\Gamma_1}{\sqrt{(\Gamma_1^2 + |\Gamma_3|^2)/n}}\right)\right], \tag{16}$$

and on the other hand,

$$b_k(f)[e'_{i_1},...,e'_{i_\ell}] = \mathbb{E}\left[H_{\alpha(I)}(\Gamma_1)J_I(\Gamma_3)h\left(\frac{\sqrt{1-\theta_1^2}\Gamma_1+\theta_1\Gamma_2}{\sqrt{(\Gamma_1^2+\Gamma_2^2+|\Gamma_3|^2)/n}}\right)\right]. \tag{17}$$

The assumption $\|f\|_2 = 1$ amounts to

$$\mathbb{E}\left[h\left(\frac{\Gamma_1}{\sqrt{(|\Gamma_3|^2+\Gamma_1^2+\Gamma_2^2)/n}}\right)^2\right] = 1. \tag{18}$$

The next lemma follows from a direct calculation.

▶ **Lemma 34.** *Assume that n is large enough. Let $\Gamma_1, \Gamma_2 \sim \mathcal{N}(0,1)$ and $\Gamma_3 \sim \mathcal{N}(0, \mathrm{I}_{n-2})$ be independent. Let $\tilde{\gamma}$ be the density of the random varianble $\frac{\Gamma_1}{\sqrt{(|\Gamma_3|^2+\Gamma_1^2+\Gamma_2^2)/n}}$ and let γ be the standard Gaussian density. Then $\frac{1}{2} \le \frac{\tilde{\gamma}(s)}{\gamma(s)} \le 2, \quad \forall s \in [-n^{0.1}, n^{0,1}]$.*

Equation (18) and Lemma 34 imply that $\|h\|_{L_2(\gamma)} \le 2$ and

$$\mathbb{E}\left[h\left(\frac{\Gamma_1}{\sqrt{|\Gamma_3|^2/n}}\right)^2\right] \le 2. \tag{19}$$

In what follows, we denote by C_k a constant depending only on k whose value may change between different appearances. Since $H'_\ell(x) = \ell H_{\ell-1}(x)$, for every ℓ there exists a constant C_ℓ such that any Hermite polynomial H_ℓ with $\ell \le k$ satisfies $|H_\ell(x(1-s)) - H_\ell(x)| \le s|x|\ell \max_{|y|\le|x|}|H_{\ell-1}(y)| \le C_k s(2+|x|)^k, \quad \forall s \in (0,1)$. Moreover since h is a polynomial of degree at most $8k$ with $\|h\|_{L_2(\gamma)} \le 2$, we conclude that

$$|h(x(1-s)) - h(x)| \le C_k s(2+|x|)^{8k}, \quad \forall s \in (0,1). \tag{20}$$

So we can write $b_k(f;\tilde{\theta})[e'_{i_1},...,e'_{i_k}] = \mathbb{E}\left[H_\alpha(\Gamma_1)J_I(\Gamma_3)h\left(\frac{\Gamma_1}{\sqrt{|\Gamma_3|^2/n}}\right)\right] + T_{res}[e'_{i_1},...,e'_{i_k}]$ where, relying on (15) and on (16),

$$T_{res} = \mathbb{E}\left[H^{(k)}(\Gamma_2\tilde{\theta}+\Gamma_1 e'_1+\Gamma_3)\left(h\left(\frac{\Gamma_1}{\sqrt{|\Gamma_3|^2/n}}\right) - h\left(\frac{\sqrt{1-\theta_1^2}\Gamma_1}{\sqrt{(\Gamma_1^2+|\Gamma_3|^2)/n}}\right)\right)\right]$$

By Parseval's inequality, we have

$$\begin{aligned}
\|T_{res}\|_2^2 &= \mathbb{E}\left[\left(h\left(\frac{\Gamma_1}{\sqrt{|\Gamma_3|^2/n}}\right) - h\left(\frac{\sqrt{1-\theta_1^2}\Gamma_1}{\sqrt{(\Gamma_1^2+|\Gamma_3|^2)/n}}\right)\right)^2\right] \\
&\overset{(20)}{\le} C_k\mathbb{E}\left[\left(\left|\frac{\frac{\sqrt{1-\theta_1^2}\Gamma_1}{\sqrt{\Gamma_1^2+|\Gamma_3|^2}} - \frac{\Gamma_1}{\sqrt{|\Gamma_3|^2}}}{\frac{\Gamma_1}{\sqrt{|\Gamma_3|^2}}}\right|(2+|\Gamma_1|)^{8k}\right)^2\right] \\
&= C_k\mathbb{E}\left[\left(\left|\frac{\sqrt{1-\theta_1^2}}{\sqrt{\frac{\Gamma_1^2}{|\Gamma_3|^2}+1}} - 1\right|(2+|\Gamma_1|)^{8k}\right)^2\right] \\
&\le C_k\mathbb{E}\left[\left(\left(\theta_1^2+\frac{\Gamma_1^2}{|\Gamma_3|^2}\right)(2+|\Gamma_1|)^{8k}\right)^2\right] \le C_k\left(\theta_1^4+\frac{1}{n^2}\right).
\end{aligned}$$

In a similar manner, (20) and (17) imply that $b_k(f)[e'_{i_1},...,e'_{i_k}] = \mathbb{E}\left[H_{\alpha(I)}(\Gamma_1)J_I(\Gamma_3)h\left(\frac{\sqrt{1-\theta_1^2}\Gamma_1+\theta_1\Gamma_2}{\sqrt{|\Gamma_3|^2/n}}\right)\right] +$ $T'_{res}[e'_{i_1},...,e'_{i_k}]$ with $\|T'_{res}\|_2^2 \le C_k\left(\theta_1^4+\frac{1}{n^2}\right)$. Note, however, that since $H_{\alpha(I)}$ is an eigenvector of the heat operator, we have

$$\mathbb{E}\left[H_{\alpha(I)}(\Gamma_1)J_I(\Gamma_3)h\left(\frac{\sqrt{1-\theta_1^2}\Gamma_1+\theta_1\Gamma_2}{\sqrt{|\Gamma_3|^2/n}}\right)\right] = \mathbb{E}\left[J_I(\Gamma_3)\mathbb{E}\left[H_{\alpha(I)}(\Gamma_1)h\left(\frac{\sqrt{1-\theta_1^2}\Gamma_1+\theta_1\Gamma_2}{\sqrt{|\Gamma_3|^2/n}}\right)\middle|\Gamma_3\right]\right]$$
$$= (1-\theta_1^2)^{\alpha(I)/2}\mathbb{E}\left[H_{\alpha(I)}(\Gamma_1)J_I(\Gamma_3)h\left(\frac{\Gamma_1}{\sqrt{|\Gamma_3|^2/n}}\right)\right].$$

We conclude that $b_k(f;\tilde{\theta})[e'_{i_1},...,e'_{i_k}] - b_k(f)[e'_{i_1},...,e'_{i_k}]$ equals:

$$T_{res}[e'_{i_1},...,e'_{i_k}] - T'_{res}[e'_{i_1},...,e'_{i_k}] + \left(1-(1-\theta_1^2)^{\alpha(I)/2}\right)\mathbb{E}\left[H_{\alpha(I)}(\Gamma_1)J_I(\Gamma_3)h\left(\frac{\Gamma_1}{\sqrt{|\Gamma_3|^2/n}}\right)\right],$$

Now, by Parseval,

$$\sum_{I=(i_1,...,i_k)\in[n-1]^k}\left(1-(1-\theta_1^2)^{\alpha(I)}\right)^2\mathbb{E}\left[H_{\alpha(I)}(\Gamma_1)J_I(\Gamma_3)h\left(\frac{\Gamma_1}{\sqrt{|\Gamma_3|^2/n}}\right)\right]^2$$
$$\le k^2\theta_1^4\mathbb{E}\left[h\left(\frac{\Gamma_1}{\sqrt{|\Gamma_3|^2/n}}\right)^2\right] \overset{(19)}{\le} C_k\theta_1^4,$$

Combining the last two displays with equation (14), we finally attain

$$\|\mathrm{P}_{\tilde{\theta}^\perp}(b_k(f;\tilde{\theta})-b_k(f))\|_{HS}^2 \le C\theta_1^4 + 4\|T'_{res}\|_2^2 + 4\|T_{res}\|_2^2 \le C_k\left(\theta_1^4+\frac{1}{n^2}\right).$$

Since $\mathbb{E}\theta_1^4 = O(1/n^2)$, taking expectation over θ establishes (13), and completes the proof of Theorem 29.

5.5 Loose ends

Proof of Proposition 32. Denote by σ_n the unique rotationally-invariant measure on the unit sphere in $\mathbb{R}^n$. A standard calculation (see [15, Equation (24)]) shows that the density of an ℓ-dimensional marginal of σ_n has the form $\psi_{n,\ell}(x) = \psi_{n,\ell}(|x|) = \Gamma_{n,\ell}\left(1-|x|^2\right)^{\frac{n-\ell-2}{2}}$, $|x|\le 1$ for a constant $\Gamma_{n,\ell}$. By continuity, $\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}\mathbf{1}\{|\langle x,\theta\rangle|\le\varepsilon\} = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}\mathbf{1}\left\{|\langle x/|x|,\theta\rangle|\le\frac{\varepsilon}{|x|}\right\} = \frac{2}{|x|}\Gamma_{n,1}$. By the continuity of g,

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}\left[\mathbf{1}\{|\langle x,\theta\rangle|\le\varepsilon\}g(\theta)\right] = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\mathbb{E}\left[\mathbf{1}\{|\langle x,\theta\rangle|\le\varepsilon\}g\left(\frac{\mathrm{Proj}_{x^\perp}\theta}{|\mathrm{Proj}_{x^\perp}\theta|}\right)\right],$$

and the first part of the proposition follows by symmetry to revolution about x. Now, for the second part, for $\rho\in[0,1]$ denote $V(\rho) = \mathrm{Vol}\left(\left\{(x,y): \quad |x|<1, |\rho x+\sqrt{1-\rho^2}y|<1\right\}\right)$, the volume of the rhombus with angle $\arcsin(\rho)$ and height 2. A calculation shows that for all $\rho < 1/2$, $V(\rho) = \frac{4}{\sqrt{1-\rho^2}}$. So we have by continuity the following expression for $\lim_{\varepsilon\to 0}\frac{1}{\varepsilon^2}\mathbb{E}\left[g(\theta)\mathbf{1}\{|\langle x,\theta\rangle|\le\varepsilon, |\langle y,\theta\rangle|\le\varepsilon\}\right]$:

$$= \lim_{\varepsilon\to 0}\frac{1}{|x||y|\varepsilon^2}\mathbb{E}\left[g\left(\frac{\mathrm{Proj}_{x^\perp\cap y^\perp}\theta}{|\mathrm{Proj}_{x^\perp\cap y^\perp}\theta|}\right)\mathbf{1}\{|\langle\hat{x},\theta\rangle|\le\varepsilon, |\langle\hat{y},\theta\rangle|\le\varepsilon\}\right]$$
$$= \frac{\Gamma_{n,2}V(\langle\hat{x},\hat{y}\rangle)}{|x||y|}\mathbb{E}\left[g\left(\frac{\mathrm{Proj}_{x^\perp\cap y^\perp}\theta}{|\mathrm{Proj}_{x^\perp\cap y^\perp}\theta|}\right)\right].$$

The proposition follows. ◀

Proof of Proposition 33. Both expressions are invariant to orthogonal transformations applied to both x, y, and are therefore functions of $\langle x, y\rangle$, $|x|$ and $|y|$. By applying a rotation, assume that

$$x \in \text{span}(e_1), \quad y \in \text{span}(e_1, e_2), \quad x_1 \geq 0, \quad y_2 \geq 0. \tag{21}$$

Evidently, for any fixed θ and indices $i_1, ..., i_k \in [n]^k$, the expression

$$\mathrm{P}_{\theta^\perp} H^{(k)}(x)[e_{i_1}, ..., e_{i_k}]\mathrm{P}_{\theta^\perp} H^{(k)}(y)[e_{i_1}, ...e_{i_k}]$$

is a polynomial of degree at most k in x_1, y_1, y_2 with coefficients depending only on k. Since the distribution of θ_1, θ_2 does not depend on x, y given the above assumption, we have that restricted to (21), the two expressions $\mathbb{E}\langle \mathrm{P}_{\theta_{1,2}^\perp} H^{(k)}(x), \mathrm{P}_{\theta_{1,2}^\perp} H^{(k)}(y)\rangle_{HS}$, are polynomials of degree at most k in x_1, y_1, y_2 with coefficients bounded by $O_k(n^k)$. Note that under (21), we have $x_1 = |x|, \quad y_1 = \rho(x, y)|y|, \quad y_2 = \sqrt{1 - \rho(x, y)^2}|y|$. Thus, we can express the above expressions as polynomials of degree at most $2k$ in $|x|$, $|y|$, $\rho(x, y)$ and $\sqrt{1 - \rho(x, y)^2}$ as long as (21) holds. Since the above expressions are invariant under rotations, these forms will hold true in general. This completes the proof. ◀

References

1. S. Arora, B. Barak, and D. Steurer. Subexponential algorithms for unique games and related problems. In *Proc. 51st IEEE Symp. on Foundations of Computer Science*, 2010.
2. S. Arora, S. A. Khot, A. Kolla, D. Steurer, M. Tulsiani, and N. Vishnoi. Unique games on expanding constraint graphs are easy: extended abstract. In *Proc. 40th ACM Symp. on Theory of Computing*, pages 21–28, 2008.
3. S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.
4. S. Arora and S. Safra. Probabilistic checking of proofs: a new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998.
5. B. Barak, M. Hardt, I. Haviv, A. Rao, O. Regev, and D. Steurer. Rounding parallel repetitions of unique games. In *Proc. 49th IEEE Symp. on Foundations of Computer Science*, pages 374–383, 2008.
6. B. Barak, P. Kothari, and D. Steurer. Small-set expansion in shortcode graph and the 2-to-2 conjecture. In *ITCS'19*, 2019.
7. M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993.
8. A. Carbery and J. Wright. Distributional and l_q norm inequalities for polynomials over convex bodies in R^n. *Math. Res. Lett.*, 8(3):233–248, 2001.
9. S. O. Chan. Approximation resistance from pairwise independent subgroups. In *Proc. 45th ACM Symp. on Theory of Computing*, pages 447–456, 2013.
10. M. Charikar, K. Makarychev, and Y. Makarychev. Near-optimal algorithms for maximum constraint satisfaction problems. *ACM Transactions on Algorithms*, 5(3), 2009.
11. E. Chlamtac, K. Makarychev, and Y. Makarychev. How to play unique games using embeddings. In *Proc. 47th IEEE Symp. on Foundations of Computer Science*, pages 687–696, 2006.
12. I. Dinur, S. Khot, G. Kindler, D. Minzer, and S. Safra. On non-optimally expanding sets in grassmann graphs. In *Proc. 50th ACM Symp. on Theory of Computing*, 2018.
13. I. Dinur, S. Khot, G. Kindler, D. Minzer, and S. Safra. Towards a proof of the 2-to-1 games conjecture? In *Proc. 50th ACM Symp. on Theory of Computing*, pages 376–389, 2018.
14. R. Eldan. A two-sided estimate for the gaussian noise stability deficit. *Invent. Math.*, 2014.

15 R. Eldan and B. Klartag. Pointwise estimates for marginals of convex bodies. *J. Functional Analysis*, 254(8):2275–2293, 2008.
16 M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995.
17 A. Gupta and K. Talwar. Approximating unique games. In *SODA*, pages 99–106, 2006.
18 J. Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48(4):798–859, 2001.
19 J. Håstad, S. Huang, R. Manokaran, R. O'Donnell, and J. Wright. Improved NP-inapproximability for 2-variable linear equations. *Theory of Computing*, 13(19):1–51, 2017.
20 S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th ACM Symp. on Theory of Computing*, pages 767–775, 2002.
21 S. Khot. On the unique games conjecture (invited survey). In *IEEE Conference on Computational Complexity*, pages 99–121, 2010.
22 S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other two-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007.
23 S. Khot, D. Minzer, D. Moshkovitz, and S. Safra. Small set expansion in the johnson graph. Technical Report TR18-078, ECCC, 2018.
24 S. Khot, D. Minzer, and S. Safra. On independent sets, 2-to-2 games, and grassmann graphs. In *Proc. 49th ACM Symp. on Theory of Computing*, pages 576–589, 2017.
25 S. Khot, D. Minzer, and S. Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In *Proc. 59th IEEE Symp. on Foundations of Computer Science*, 2018.
26 S. Khot and D. Moshkovitz. NP-hardness of approximately solving linear equations over reals. In *Proc. 43rd ACM Symp. on Theory of Computing*, pages 413–420, 2011.
27 S. Khot and D. Moshkovitz. Candidate hard unique game. In *Proc. 48th ACM Symp. on Theory of Computing*, pages 63–76, 2016.
28 S. Khot and O. Regev. Vertex cover might be hard to approximate to within 2-epsilon. *Journal of Computer and System Sciences*, 74(3):335–349, 2008.
29 B. Klartag and O. Regev. Quantum one-way communication can be exponentially stronger than classical communication. In *Proc. 43rd ACM Symp. on Theory of Computing*, pages 31–40, 2011.
30 A. Kolla, K. Makarychev, and Y. Makarychev. How to play unique games against a semi-random adversary: Study of semi-random models of unique games. In *Proc. 52nd IEEE Symp. on Foundations of Computer Science*, pages 443–452, 2011.
31 P. McCullagh. *Tensor methods in statistics.* Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1987.
32 D. Moshkovitz and R. Raz. Two query PCP with sub-constant error. *Journal of the ACM*, 57(5), 2010.
33 E. Mossel and J. Neeman. Robust dimension free isoperimetry in gaussian space. *Annals of Probability*, 43(3):971–991, 2015.
34 E. Mossel and J. Neeman. Robust optimality of gaussian noise stability. *Journal of the European Math Society (JEMS)*, 17(2):433–482, 2015.
35 E. Mossel and J. Neeman. Noise stability and correlation with half spaces. *Electron. J. Probab.*, 23(16), 2018.
36 R. O'Donnell. *Analysis of Boolean Functions.* Cambridge University Press, 2014.
37 P. Raghavendra. Optimal algorithms and inapproximability results for every csp? In *Proc. 40th ACM Symp. on Theory of Computing*, pages 245–254, 2008.
38 R. Raz. A parallel repetition theorem. In *SIAM Journal on Computing*, volume 27, pages 763–803, 1998.
39 R. Raz. A counterexample to strong parallel repetition. *SIAM Journal on Computing*, 40(3):771–777, 2011.
40 R. Rubinfeld and M. Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

41 L. Trevisan. Approximation algorithms for unique games. *Theory of Computing*, 4(1):111–128, 2008.

A Integrality Gap for Subspaces Near-Intersection

In this section we sketch an integrality gap instance for the semidefinite program that minimizes $\mathbf{E}_{(u,v)\in E}\left[|Proj_{\Theta_e^\perp}(\sigma(u)) - Proj_{\Theta_e^\perp}(\sigma(v))|_2^2\right]$. Consider the following graph $G = (V, E)$: its vertices correspond to all unit vectors $v \in \mathbb{R}^k$ where coordinates are taken up to sufficiently large precision with respect to $\delta > 0$. The subspace associated with the vertex is the one that is spanned by v. For the vertex corresponding to vector v there is an edge that touches it for every unit vector $\Theta \in \mathbb{R}^k$ (up to the aforementioned precision) and it connects it to a vertex associated with a random vector $u \in \mathbb{R}^k$ such that $\left|v_{|\Theta^\perp} - u_{|\Theta^\perp}\right|_2 \approx \sqrt{\delta}$ (the approximation reflects the precision error). Note that this instance of Subspaces Near-Intersection has a vector solution given by the unit vector associate with every vertex, and it achieves value approximately δ by construction. Nevertheless, there is no feasible assignment $\sigma : V \to \mathbb{R}^k$ where $|Proj_{\Theta_e^\perp}(\sigma(u)) - Proj_{\Theta_e^\perp}(\sigma(v))|_2$ is typically $0.001\sqrt{\delta}$, simply because only the prescribed unit vector is in the subspace of each vertex.