Electronic Colloquium on Computational Complexity, Report No. 120 (2023)

Characterizing Direct Product Testing via Coboundary Expansion

Mitali Bafna [*] Dor Minzer[†]

Abstract

A d-dimensional simplicial complex X is said to support a direct product tester if any locally consistent function defined on its k-faces (where $k \ll d$) necessarily come from a function over its vertices. More precisely, a direct product tester has a distribution μ over pairs of k-faces (A, A'), and given query access to $F\colon X(k) \to \{0,1\}^k$ it samples $(A, A') \sim \mu$ and checks that $F[A]|_{A\cap A'} = F[A']|_{A\cap A'}$. The tester should have (1) the "completeness property", meaning that any assignment F which is a direct product assignment passes the test with probability 1, and (2) the "soundness property", meaning that if F passes the test with probability s, then F must be correlated with a direct product function.

Dinur and Kaufman showed that a sufficiently good spectral expanding complex X admits a direct product tester in the "high soundness" regime where s is close to 1. They asked whether there are high dimensional expanders that support direct product tests in the "low soundness", when s is close to 0.

We give a characterization of high-dimensional expanders that support a direct product tester in the low soundness regime. We show that spectral expansion is insufficient, and the complex must additionally satisfy a variant of coboundary expansion, which we refer to as *Unique-Games coboundary expanders*. This property can be seen as a high-dimensional generalization of the standard notion of coboundary expansion over non-Abelian groups for 2-dimensional complexes. It asserts that any locally consistent Unique-Games instance obtained using the low-level faces of the complex, must admit a good global solution.

1 Introduction

The problem of testing direct product functions lies at the intersection of many areas within theoretical computer science, such as error correcting codes, probabilistically checkable proofs (PCPs), hardness amplification and property testing. In its purest form, one wishes to encode a function $f\colon [n] \to \{0,1\}$ using local views in a way that admits local testability/ local correction. More precisely, given a parameter $1 \leqslant k < n$, the encoding of f using subsets of size k can be viewed as $F\colon \binom{[n]}{k} \to \{0,1\}^k$ that to each subset $A \subseteq [n]$ of size k assigns a vector of length k describing the restriction of f to A.[1] We refer to this encoding as the direct product encoding of f according to the Johnson graph (for reasons that will become apparent shortly). The obvious downside of this encoding scheme is, of course, that its length is much larger than the description of f (roughly n^k vs $\Theta(n)$). However, as this encoding contains many redundancies, one hopes that it more robustly stores the information in the function f, thereby being more resilient against corruptions.

[*]Carnegie Mellon University. Supported in part by the Computer Science Department, CMU and a gift from CYLAB, CMU.

[†]Department of Mathematics, Massachusetts Institute of Technology. Supported by a Sloan Research Fellowship, NSF CCF award 2227876 and NSF CAREER award 2239160.

[1]To be more specific, one identifies $A = \{a_1, \ldots, a_k\}$ with the ordered tuple $(a_1, \ldots, a_k)$ where $a_1 < a_2 < \ldots < a_k$ and defines $F[A] = (f(a_1), \ldots, f(a_k))$.

ISSN 1433-8092

1.1 Direct Product Testing with 2 Queries

Indeed, one of the primary benefits of the above direct product encoding is that it admits local testers using a few queries. These testing algorithms also go by the name "agreements testers" or "direct product testers", and are often very natural to design. A direct product tester for the above encoding, which we parameterize by a natural number $1 \leqslant t \leqslant k$ and denote by $\mathcal{T}_t$, proceeds as follows:

1. Choose two subsets $A, A' \subseteq [n]$ uniformly at random conditioned on $|A \cap A'| = t$.
2. Query $F[A]$, $F[A']$ and check that $F[A]$ and $F[A']$ agree on $A \cap A'$.

These type of testers have been first considered and used by Goldreich and Safra [GS00] in the context of the PCP theorem. They later have been identified by Dinur and Reingold [DR06] as a central component in gap amplification. To get some intuition to this test, note that a direct product function clearly passes the test with probability 1. Thus, we say that the tester has perfect completeness. The soundness of the test – namely the probability that a table F which is far from a direct product encoding passes the test – is more difficult to analyze. Intuitively, querying F at a single location gives the value of a (supposed) f on k inputs. Thus, if F is far from any direct product function, the chance this will be detected should grow with k. Formalizing this intuition is more challenging however, and works in the literature are divided into two regimes: the so-called 99% regime, and the 1% regime. To be more precise, suppose the table F passes the direct product tester $\mathcal{T}_t$ with probability at least $s > 0$; what can be said about its structure?

In the 99% regime, namely the case where $s = 1 - \varepsilon$ is thought of as close to 1, results in the literature [DR06, DS14b] show that F has to be close to a direct product function. More specifically, for $t = \Theta(k)$ the result of [DS14b] asserts that there exists $f\colon [n] \to \{0,1\}$ such that $F[A] = f|_A$ for $1 - O(\varepsilon)$ fraction of the k-sets A. A structural result of this form is a useful building block in several applications. It can be used to construct constant query PCPs with constant soundness; it also serves as a building block in other results within complexity theory; see for instance [DDG$^+$17, DFH19].

The 1% regime, namely the case where $s = \delta$ is thought of as a small constant, is more challenging. In this case, the works [DG08, IKW09] show that F has to be correlated with a direct product function. More specifically, these works show that for (say) $t = \sqrt{k}$ if $\delta \geqslant 1/k^{\Omega(1)}$, then there exists $f\colon [n] \to \{0,1\}$ such that for at least $\delta^{O(1)}$ fraction of the k-sets A, we have that

$$\Delta(F[A], f|_A) \leqslant k^{-\Omega(1)},$$

where for two strings $x, y \in \{0,1\}^k$, $\Delta(x,y) = \frac{\#\{i \in [k] \mid x_i \neq y_i\}}{k}$ denotes the fractional Hamming distance between them.[2] The motivation for studying this more challenging regime of parameters stems mainly from the perspective of hardness amplification (where one wishes to show that if a given task is somewhat hard, then repeating this task k-times in parallel gets exponentially harder) as well as from the study of PCPs with small soundness. Indeed, in [IKW09] the authors show that direct product testers similar to the above facilitate soundness amplification schemes for PCPs with similar performance to parallel repetition theorems [Raz98, Hol09, Rao11, BG15, DS14a]. Direct product testers in the low soundness regime have additional applications in property testing, as well as in the study of the complexity of satisfiable constraint satisfaction problems [BKM22, BKM23a, BKM23b, BKM23c].

[2]In contrast to the 99% regime, in this case one has to settle with agreement with f only on a small portion of the k-sets, and furthermore this agreement is not perfect; it is on $(1-o(1))$ fraction of the elements in the k-sets. As discussed in [DG08, IKW09], qualitatively speaking (namely, up to the precise parameters) this is the best type of results possible.

1.2 Size Efficient 2-Query Direct Product Testing

In the context of PCPs and hardness amplification, one typically thinks of the parameter n as very large, and k as a large constant number. With this in mind, representing an assignment $f\colon [n] \to \{0,1\}$ using its direct product encoding incurs a polynomial blow-up in size. Indeed, this type of step is often the only step in the PCP reduction that introduces a polynomial (as opposed to just linear) blow-up in the instance size. In this light, a natural question is whether it is possible to perform hardness amplification with a significantly smaller blow up in the encoding/instance size. Efficient schemes of this type are often referred to as "derandomized direct product tests", "derandomized hardness amplification" or "derandomized parallel repetition theorems".

In [IKW09] a more efficient hardness amplification procedure is proposed. Therein, instead of considering all k-sets inside $[n]$, the domain $[n]$ is thought of as a vector space $\mathbb{F}_q^d$ and one considers all subspaces of dimension $\log_q(k)$. It is easy to see that the encoding size then becomes $n^{\Theta(\log k)}$, making it more efficient. The paper [IKW09] shows that direct product testers analogous to the tester above work in this setting as well; they essentially match all of the results achieved by the Johnson scheme. Building upon [IKW09], Dinur and Meir [DM11] show how to establish parallel repetition theorems using the more efficient direct product encoding via subspaces. This parallel repetition theorem works for structured instances, which the authors show to still capture the entire class NP.

High dimensional expanders (HDX), which have recently surged in popularity, can be seen as sparse models of the Johnson graph. This leads us to the main problem considered in this paper, due to Dinur and Kaufman [DK17]:

Do high dimensional expanders facilitate direct product testers in the low soundness regime?

The main goal of this paper is to investigate the type of expansion properties that are necessary and sufficient for direct product testing with low soundness. It is known that there are HDXs of size $O_k(n)$ and $O_k(1)$ degree, and if any of these objects facilitates a direct product tester with small soundness, they would essentially be the ultimate form of derandomized direct product testers.[3] To state our results, we first define the usual notion of spectral high dimensional expansion, followed by our variant of the well-known notion of coboundary expansion.

1.2.1 High Dimensional Local Spectral Expanders

A d-dimensional complex is composed of $X(0) = \{\emptyset\}$, a set of vertices $X(1)$, which is often identified with $[n]$ and a set of i-uniform hyperedges, $X(i) \subseteq \binom{X(1)}{i}$, for each $i = 2,\ldots,d$. A d-dimensional complex $X = (X(0), X(1), \ldots, X(d))$ is called simplicial if it is downwards closed. Namely, if for every $1 \leqslant i \leqslant j \leqslant d$, and every $J \in X(j)$, if $I \subseteq J$ has size i, then $I \in X(i)$. The size of a complex is the total number of hyperedges in X. The degree of a vertex $v \in X(1)$ is the number of faces in $X(d)$ containing it, and the degree of a complex X is the maximum of the degree over all the vertices in $X(1)$.

We need a few basic notions regarding simplicial complex, and we start by presenting the notion of links and spectral expansion.

[3]We remark that to be useful, it seems that a derandomized direct product tester would need to roughly have equal degrees. This is because in applications, each one of the values of the encoded function $f\colon [n] \to \{0,1\}$ is "equally important". In that case, the requirement of being a $O(n)$ sized direct product tester is equivalent to having $O(1)$ degree.

Definition 1.1. *For a d-dimensional simplicial complex $X = (X(0), X(1), \ldots, X(d))$, $0 \leqslant i \leqslant d-2$ and $I \in X(i)$, the link of I is the $(d-i)$-dimensional complex X_I whose faces are given as*

$$X_I(j-i) = \{J \setminus I \mid J \in X(j), J \supseteq I\}.$$

For a d-dimensional complex $X = (X(0), X(1), \ldots, X(d))$ and $I \in X$ of size at most $d-2$, the graph underlying the link of I is the graph whose vertices are $X_I(1)$ and whose vertices are $X_I(2)$.

Distributions over the complex. It is convenient to equip a complex X with a measure μ_k for each one of its levels $X(k)$. For $k = d$ we consider the measure μ_d which is uniform over $X(d)$; for each $k < d$, the measure μ_k is the push down measure of μ_d: to generate a sample according to μ_k, sample $D \sim \mu_d$ and then $K \subseteq D$ of size k uniformly. Abusing notation, we will refer to all of the measures μ_k simply as μ, as the cardinality of the sets in discussion will always be clear from context. The set of measures in the link of I is the natural set of measures we get by conditioning μ on containing I.

Equipped with measures over complexes, we may now define the notion of spectral HDX.

Definition 1.2. *A d-dimensional simplicial complex X is called a γ two-sided local spectral expander if for every $I \in X$ of size at most $d-2$, the second singular value of the normalized adjacency matrix of the graph underlying the link of I is at most γ.*

In this work, we will only be concerned with simplicial complexes that are very strong spectral expanders. With this regard, following the works of [LSV05b, LSV05a, EK16] one can show that for every $\gamma > 0$ and every $d \in \mathbb{N}$ there exists an infinite family of d-dimensional complexes of linear size that are γ two-sided local expanders (see [DK17, Lemma 1.5]).

1.2.2 Results in the High Soundness Regime

Dinur and Kaufman [DK17] were the first to consider the question of direct product testing over HDX. They showed that a sufficiently good high dimensional spectral expander admits a direct product tester in the high soundness regime. The tester they consider is essentially the same as the tester in the Johnson scheme; one thinks of k which is much larger than 1 but much smaller than the dimension of the complex d. The tester has parameters $1 \leqslant s \leqslant k/2$ and is given oracle access to a table $F \colon X(k) \to \{0,1\}^k$, and proceeds as follows:

Agreement-Test 1 (F, k, s)**.**

1. Sample $D \sim \mu_d$.
2. Sample $I \subseteq D$ of size s uniformly.
3. Sample $I \subseteq A, A' \subseteq D$ of size k uniformly.
4. Accept if $F[A]|_I = F[A']|_I$.

Henceforth, we refer to this test as the (k, s) direct product tester over X. Dinur and Kaufman consider the case where $s = k/2$, and proved that for every $\varepsilon > 0$, provided that γ is sufficiently small, if $F \colon X(k) \to \{0,1\}^k$ passes the above test with probability at least $1-\varepsilon$, then there exists $f \colon X(1) \to \{0,1\}$ such

$$\Pr_{A \sim \mu_k}[F[A] \equiv f|_A] \geqslant 1 - O(\varepsilon).$$

A follow-up work by Dikstein and Dinur [DD19] further refined this result, and investigated more general structures that support direct product testing in the high soundness regime.

A problem related to direct product testing, called the list agreement testing problem, was considered in the high soundness regime by Gotlib and Kaufman [GK23]. In the list agreement testing problem, each face is assigned a list of $m = O(1)$ functions, and one performs a local test on these lists to check that they are consistent. With this in mind, the result of Gotlib and Kaufman [GK23] asserts that under certain structural assumptions on the lists, if the underlying complex has sufficiently good coboundary expansion, then one can design a 3-query list agreement tester that is sound. The list agreement problem will play an important role in the current work, and while we do not know how to use the result of Gotlib and Kaufman for our purposes, their work inspired us to look at connections between agreement testing and notions of coboundary expansion.

1.3 Main Results

Despite considerable interest, no positive nor negative results are known regarding the question of whether HDX support direct product testers in the low-soundness regime. In fact, the majority of applications of HDX are in the high soundness regime, with the first construction of c^3-locally testable codes [DEL+22] and quantum LDPC codes [EKZ20, PK22, LZ22]. At a first glance, this seems surprising: very good expander graphs give rise to objects in the low-soundness regime, and high dimensional expanders are essentially their higher order analogs.

The main contribution of this work is an explanation to this phenomenon. We show that, to facilitate direct product testers in the low-soundness regime, a high dimensional spectral expander must posses a property that may be seen as a generalization of *coboundary expansion* [LM06]. Thus, to construct constants degree, sparse complexes facilitating direct product testing, one should first come up with spectral expanders that are also coboundary expanders. Currently, no such constructions are known, even for the simplest forms of coboundary expansion.

Below, we state our main results regarding the soundness of the test, which give analysis of the (k, s) tester defined above assuming expansion properties of the complex X. In a concurrent and independent work, Dikstein and Dinur [DD23a] established related results.

1.3.1 Coboundary Expansion

For convenience, we follow the presentation of coboundary expansion from [DD23b]. Suppose we have a function $f\colon X(2) \to \mathbb{F}_2$. The function f is said to be consistent on the triangle $\{u, v, w\} \in X(3)$ if it holds that $f(\{u, v\}) + f(\{v, w\}) + f(\{u, w\}) = 0$. What can we say about the structure of functions f which are consistent with respect to $1 - \xi$ measure of the triangles? Clearly, if f is a function of the form $f(\{u, v\}) = g(u) + g(v)$ for some $g\colon X(1) \to \mathbb{F}_2$, then it is consistent with respect to all triangles. In the case that X is a coboundary expander, the converse is also true: any f which is $(1 - \xi)$ triangle consistent is $O(\xi)$-close to a function of this form.

More broadly, the notion of coboundary expansion often refers to a property of higher dimensional faces, and to more general groups beyond $\mathbb{F}_2$. We refrain from defining these notions precisely and instead turn to our variant of coboundary expansion, which we show governs the soundness of direct product testing.

1.3.2 Unique-Games Coboundary Expansion

Our notion of coboundary expansion replaces the group $\mathbb{F}_2$ with non-Abelian groups, more specifically with the permutation groups S_m; we also need to consider higher dimensional faces. Some definitions in this spirit have been made, for example in [DM19, GK23], and our notion is inspired by theirs.

Given a d-dimensional complex X and an integer $t \leqslant d/3$, we consider the graph $G_t[X] = (X(t), E_t(X))$ whose vertices are the t-faces of X, namely $X(t)$, and (u, v) is an edge if $u \cup v \in X(2t)$. We say $T = (u, v, w)$ is a triangle in $G_t[X]$ if each of $u, v, w \in X(t)$ and $u \cup v \cup w \in X(3t)$.

Definition 1.3. *Let X be a d-dimensional complex and let t be an integer such that $t \leqslant d/3$. Let $\pi\colon E_t(X) \to S_m$ be a function that satisfies $\pi(u, v) = \pi(v, u)^{-1}$ for all $(u, v) \in E_t[X]$. We say that π is consistent on the triangle (u, v, w) in $G_t[X]$ if $\pi(u, v)\pi(v, w) = \pi(u, w)$.*

We say that π is $(1 - \xi)$-consistent on triangles if sampling $T \sim \mu_{3t}$ and then splitting T as a triangle $u \cup v \cup w$ uniformly where $|u| = |v| = |w| = t$,

$$\Pr_{\substack{T \sim \mu_{3t} \\ T = u \cup v \cup w}} \left[\pi(u, v)\pi(v, w) = \pi(u, w)\right] \geqslant 1 - \xi.$$

One way to think of this definition is as a locally consistent instance of Unique-Games (see Definition 2.4). Indeed, a π as above specifies a Unique-Games (UG) instance on the graph $G_t[X]$ whose constraints are locally consistent on triangles. The goal in this UG instance may be thought of assigning elements from $[m]$ to the vertices of $G_t[X]$, namely finding an assignment $A\colon X(t) \to [m]$, so as to maximize the fraction of edges (u, v) for which $\pi(u, v)A(v) = A(u)$.

With this definition in mind, we can now present a simplified version of our notion of coboundary expansion. One way to arrive at a locally consistent UG instance as in Definition 1.3 is to first pick some function $g\colon X(t) \to S_m$ and then define $\pi(u, v) = g(u)^{-1}g(v)$. Thus, a natural question is whether there are other ways to construct locally consistent UG instances on $G_t[X]$. In simple terms, our simplified notion of UG coboundary expansion asserts that this is essentially the only way to arrive at instances of this form. More precisely:

Definition 1.4. *We say that a d-dimensional simplicial complex X is an (m, r, ξ, c) UG coboundary expander if for all $t \leqslant r$ and for all functions $f\colon E_t[X] \to S_m$ that are $(1 - \xi)$-consistent on triangles, there is $g\colon X(t) \to S_m$ such that*

$$\Pr_{u \cup v \sim \mu_{2t}} \left[\pi(u, v) = g(u)^{-1}g(v)\right] \geqslant 1 - c.$$

We remark that if a complex X is an (m, r, ξ, c) UG coboundary expander, then given a $(1 - \xi)$-locally consistent instance of Unique-Games on $G_t[X]$ for some $t \leqslant r/3$, one may find an assignment satisfying at least $1 - c$ fraction of the constraints. Indeed, by definition, given the constraint map π we may find $g\colon X(t) \to S_m$ such that $\pi(u, v) = g(u)^{-1}g(v)$ with probability at least $1 - c$ over the choice of $u \cup v \sim \mu_{2t}$. Thus, taking the labeling $A(v) = g(v)^{-1}(1)$, we see that A satisfies all edges on which $\pi(u, v) = g(u)^{-1}g(v)$.

The first result of this paper asserts that a spectral HDX which is a UG coboundary expander admits a direct product tester in the low soundness regime.

Theorem 1.5. *Suppose that a simplicial complex X is a sufficiently good spectral and UG coboundary expander. If $F\colon X(k) \to \{0, 1\}^k$ passes the $(k, \sqrt{k})$ direct product test on X with probability δ, then there is $f\colon X(1) \to \{0, 1\}$ such that*

$$\Pr_{A \sim \mu_k} \left[\Delta(f|_A, F[A]) = o(1)\right] \geqslant \Omega_\delta(1).$$

In words, being a UG coboundary expander is a sufficient condition for a spectral expander to support a low soundness direct product tester. As far as we know, however, this condition may not be necessary; below, we present a condition which is both necessary and sufficient. Nevertheless, we chose to present its simpler to state version, Definition 1.4, as we find it more appealing, intuitive and resembling non-Abelian variants of the usual notion of coboundary expansion.

Remark 1.6. *The usual definition of coboundary expansion in the literature refers to Abelian groups such as $\mathbb{F}_2$, see for example [KKL14, KM22a, KM22b, GK23, DD23b]. In the $\mathbb{F}_2$ setting, coboundary expansion for the base graph can be seen as a UG instance over $\mathbb{F}_2$, but it is often phrased in topological notions using the boundary and coboundary maps; these definitions extend well to higher dimensional faces. Coboundary expansion has also been defined for non-Abelian groups [DM19, KM22b, GK23], however, as far as we know, these definitions coincide with ours only for the case that $t = 1$ in Definition 1.4.*

1.3.3 A Necessary and Sufficient Condition for Low Soundness Direct Product Testing

We now move on to stating a more complex version of Definition 1.4 which is both necessary and sufficient for low soundness direct product testing. Let us again consider the graph $G_t[X]$ and a $(1-\xi)$ triangle consistent assignment of permutations on the edges $\pi\colon E_t[X] \to S_m$. However, unlike before, these permutations are guaranteed to satisfy an additional premise. Precisely, suppose that each face $u \in X(t)$ is assigned a list of m elements from $\{0,1\}^t$, say $L(u) = (L_1(u), \dots, L_m(u))$, and each face $T \in X(3t)$ is also assigned a list $L'(T) = (L'_1(T), \dots, L'_m(T))$. In words, we would like the permutations π to be consistent with the lists with respect to concatenations. Towards this end, we introduce a convenient but informal notation to compare strings. Given $u, v \in X(t)$ that are disjoint and strings $L_i(u), L_i(v) \in \{0,1\}^t$, we shall think of $L_i(u)$ as an assignment to the vertices in u and of $L_i(v)$ as an assignment to the vertices in v. Thus, the notation $L_i(u) \circ L_i(v)$ will be a string in $\{0,1\}^{2t}$ which encodes the assignment to $u \cup v$ provided by the concatenation of the two assignments. More generally, given u, v disjoint and list assignments $L(u), L(v)$ we define

$$L(u) \circ L(v) = (L_1(u) \circ L_1(v), \dots, L_m(u) \circ L_m(v)).$$

Lastly, given a list $L(u)$ as above and $\pi \in S_m$, we define $\pi L(u) = (L_{\pi(1)}(u), \dots, L_{\pi(m)}(u))$.

Definition 1.7. *Let $L\colon X(t) \to (\{0,1\}^t)^m$, $L'\colon X(3t) \to (\{0,1\}^{3t})^m$, and $\xi > 0$. We say π is $(1-\xi)$-consistent with the lists L and L' if choosing $T \sim \mu_{3t}$ and a splitting $T = u \cup v \cup w$ into a triangle, we have that*

$$\Pr_{\substack{T \sim \mu_{3t} \\ T = u \cup v \cup w}} \left[L'(T) = L(u) \circ \pi(u,v) L(v) \circ \pi(u,w) L(w) \right] \geqslant 1 - \xi.$$

We say that π is $(1-\xi)$-strongly triangle consistent if there are lists L and L' such that π is $(1-\xi)$-consistent with respect to the lists L and L'.

It is easy to see that if π is $(1-\xi)$-strongly triangle consistent, then π is $(1-O(\xi))$-triangle consistent (see Claim A.1). Thus, the class of triangle consistent functions π is more restrictive. With the notion of strong triangle consistency we are now ready to state a weaker variant of Definition 1.4; the only difference between the two definitions is that in the definition below, we only require that any strongly triangle consistent assignment admits a global structure. More precisely:

Definition 1.8. *We say that a d-dimensional simplicial complex X is a weak (m, r, ξ, c) UG coboundary expander if the following condition is satisfied for all $t \leqslant r$. Suppose $\pi\colon E_t[X] \to S_m$ is a $(1-\xi)$-strongly* *triangle consistent function. Then there exists $g\colon X(t) \to S_m$ such that*

$$\Pr_{u\cup v\sim\mu_{2t}}[\pi(u,v) = g(u)^{-1}g(v)] \geqslant 1-c.$$

The parameter r in Definition 1.8 is often referred to as the level at which UG coboundary expansion holds. With the notion of weak UG coboundary expansion, we can now state a stronger version of Theorem 1.5. Roughly speaking, the following two results asserts that for a sufficiently good spectral simplicial complex X, the direct product tester over X works in the low soundness regime if and only if X is a weak UG coboundary expander with sufficiently good parameters.

Theorem 1.9. *The following results hold for any simplicial complex X.*

1. ***Weak UG-coboundary is Necessary:*** *If a simplicial complex X is a sufficiently good spectral expander which is not a UG coboundary expander, then there is $\delta > 0$ such that for sufficiently large k, there is $F\colon X(k) \to \{0,1\}^k$ that passes the $(k, \sqrt{k})$ direct product tester with probability δ and yet for all $f\colon X(1) \to \{0,1\}$ we have that*

$$\Pr_{A\sim\mu_k}[\Delta(F[A], f|_A) = o(1)] = o(1).$$

2. ***Weak UG-coboundary is Sufficient:*** *For all $\varepsilon, \delta > 0$, if a simplicial complex X is a sufficiently good spectral expander and a weak UG coboundary expander on level $O(1)$, then the direct product test over X with respect to sufficiently large k has soundness δ. Namely, if $F\colon X(k) \to \{0,1\}^k$ passes the $(k, \sqrt{k})$ direct product tester with respect to X with probability at least δ, then there is $f\colon X(1) \to \{0,1\}$ such that*

$$\Pr_{A\sim\mu_k}[\Delta(F[A], f|_A) \leqslant \varepsilon] \geqslant \Omega(1).$$

We refer the reader to Theorems 3.1 and 4.1 for more formal statements. We would like to remark that Theorem 1.9 only uses UG coboundary expansion on a constant level r, which is potentially much smaller than k. The structure for F we get is relatively weak though, and only asserts that with significant probability over the choice of $A \sim \mu_k$, we have that $F[A]_i = f(i)$ for $(1-\varepsilon)$ fraction of $i \in A$. In the next theorem, we show that if the level r on which coboundary expansion holds is linear in k, then the conclusion of Theorem 1.9 can be strengthened to say that with significant probability over $A \sim \mu_k$, it holds that $F[A]_i = f(i)$ for all but constantly many of $i \in A$.[4]

Theorem 1.10. *If a simplicial complex X is a sufficiently good spectral expander, and for $k \in \mathbb{N}$ it holds that X is a sufficiently good weak UG coboundary expander on level $\Omega(k)$, then the direct product test over X with respect to k has soundness δ. Namely, for all $\delta > 0$ there is $\eta > 0$ such that if $F\colon X(k) \to \{0,1\}^k$ passes the $(k, \eta k)$ direct product tester with respect to X with probability at least δ, then there is $f\colon X(1) \to \{0,1\}$ such that*

$$\Pr_{A\sim\mu_k}\left[\Delta(F[A], f|_A) \leqslant O(1/k)\right] \geqslant \Omega(1).$$

[4] For that, we need to consider a direct product tester with intersection parameter s, which is significantly smaller than k but is linear in it. Indeed, it is easy to see that the conclusion of Theorem 1.10 would fail if either $s \leqslant k^{0.99}$ or $s \geqslant k/100$.

In Section 6 we examine several well known complexes. We are not aware of any constant degree spectral expander which is a UG coboundary expander. In fact, we use our result above to conclude that some of the best known spectral expanders – namely some LSV complexes – do not support direct product testers in the low soundness regime precisely because they fail to satisfy coboundary expansion. As the result of Dinur and Kaufman [DK17] asserts that LSV complexes admit direct product testers in the high soundness regime, we conclude that the low soundness regime is qualitatively different.

1.4 Proof Overview

In this section we give an overview for the proofs of Theorems 1.9 and 1.10. The proofs of these two theorems is basically the same; the only point in which they defer is what direct product theorem is used for the Johnson scheme. Thus, we will focus on the setting in Theorem 1.9.

1.4.1 The Proof that weak UG Coboundary Expansion is Necessary

Suppose that weak UG coboundary expansion for X fails. That means that we have a complex X violating the weak UG coboundary expansion, hence there is a $(1-o(1))$-strongly triangle consistent Unique-Games instance on $G_r[X]$ that has no $(1-c)$-global solution. Namely, denoting the constraints of $G_r[X]$ by $\pi\colon E_r(X) \to S_m$, for all $g\colon X(r) \to S_m$ it holds that $\pi(u,v) = g(u)^{-1}g(v)$ with probability at most $1-c$ over the choice of $(u,v) \sim \mu_{2r}$. Furthermore, we may find lists L and L' such that f is $(1-o(1))$-consistent with L, L' as in Definition 1.7. The parameters m and r should be thought of as large constants, and $c > 0$ should be thought of as a small constant bounded away from 0. The proof now proceeds in the following steps:

Preprocessing: We claim that (after appropriate pre-processing) we may assume that $G_r[X]$ doesn't have a solution that satisfies more than $(1-c/m)$ of the constraints of $G_r[X]$. Indeed, if there exists such a solution $A\colon X(r) \to [m]$, we may (1) remove $A(u)$ from the list $L[u]$, (2) For all edges (u,v) remove the pair $(A(u), A(v))$ as a satisfying assignment of $\pi(u,v)$. We argue that this process must terminate and in the end must produce a Unique-Games instance with no good global solution. For simplicity of notation, we assume henceforth that $G_r[X]$ doesn't have an assignment satisfying at least $(1-c')$ of the constraints to begin with (where $c' = c/m$).

Lifting the lists: Take $k \gg r$, and consider k-faces. Taking $K \sim \mu_k$ randomly and taking a random triangle over r-faces in K, we see that it is marginally distributed as a random triangle in X. Thus, by linearity of expectation the number of inconsistent triangles in K is at most $o(k^{3r}) = o(1)$. It follows that with probability at least $1-o(1)$ over the choice of K, all r-triangles in K are consistent. In this case, it is easily seen that there exists a unique list $L[K]$ of m strings from $\{0,1\}^k$ which is consistent with all the r-faces inside K, and we fix it. For sake of simplicity, we assume that we have constructed a list $L[K]$ for each $K \in X(k)$.

Constructing the assignment and its soundness: It is easy to see that sampling $D \sim \mu_d$, $I \subseteq D$ of size $\sqrt{k}$ and then $I \subseteq K, K' \subseteq D$ uniformly, with probability $1-o(1)$ the lists of K and K' are in 1-to-1 correspondence with respect to agreement on I. [5] Thus, choosing an assignment $F\colon X(k) \to \{0,1\}^k$ by assigning each k-face K a random element from $L[K]$ gives an assignment which in expectation satisfies at least $1/m$ fraction of the constraints, and we fix F achieving this expectation henceforth.

[5] Roughly speaking, the reason is that the argument above could be made also for $2k$-faces, and one gets a list of assignments that is in 1-to-1 correspondence to the lists of K faces with respect to containment.

No global structure: Suppose for contradiction that $h\colon X(1) \to \{0,1\}$ has that $h(K) \equiv F[K]$ for at least η fraction of the k-faces (strictly speaking, we may only assume that $h(K)$ is close to $F[K]$, but we ignore this issue for the sake of clarity), and let the set of these k-faces be denoted by $\mathcal{K}$. Consider $\mathcal{R} = \{R \in X(r) \mid \exists K \in \mathcal{K}, R \subseteq K\}$. By the sampling property of X, provided that k is sufficiently large, we get $\mu_r(\mathcal{R}) \geqslant 1 - c/10$. Note that for $R \in \mathcal{R}$, $h_R = (h|_K)|_R \in L[K]|_R = L[R]$ where $K \in \mathcal{K}$ contains R. It follows that h satisfies at least $1 - c$ fraction of the constraints of $G_r[X]$, and contradiction.

1.4.2 The Proof that weak UG Coboundary Expansion is Sufficient

Suppose we have an assignment $F\colon X(k) \to \{0,1\}^k$ that passes the (k,s) direct product test over X with probability at least δ; throughout, $s = \sqrt{k}$. We also take a parameter t where $k \ll t \ll d$. The proof now proceeds by the following steps.

We start by localizing to Johnson graphs, and show that almost all of them can be equipped with a short list of assignments that explain almost all of the agreement inside them. This naturally leads us to *list-agreement testing* as described above (and defined more formally below). This problem was addressed in the work of Gotlib and Kaufman [GK23] who analyzed a certain list agreement tester and proved it was sound in the high soundness regime. Although our result for list agreement testing is similar in spirit, our tester is different from theirs, and we do not how to know use their techniques/tester for our purposes.

1.4.3 Reduction from 1% Direct Product Testing to 99% List Agreement Testing

Localizing to a Johnson: The first part of the proof is to localize the test to Johnson schemes. For $T \in X(t)$ define $\mathsf{pass}_t(F;T)$ as the probability the following test passes: sample $I \subseteq T$ of size s and then $I \subseteq A, B \subseteq T$ independently, and test that $F[A]|_I = F[B]|_I$. [6] Similarly, we may define $\mathsf{pass}_d(F;D)$ for $D \in X(d)$. Clearly, one has that $\mathbb{E}_{T\sim\mu_t}[\mathsf{pass}_t(F;T)], \mathbb{E}_{D\sim\mu_d}[\mathsf{pass}_d(F;D)] \geqslant \delta$ and hence for a large fraction of the D's we have that $\mathsf{pass}_d(F;D) \geqslant \delta/2$. By the sampling properties of HDX (which follow as X is a sufficiently good spectral expander) we are able to derive the stronger conclusion that $\mathsf{pass}_d(F;D) \geqslant \delta^2/100$ for $1 - O_\delta(\gamma)$ for the d-faces D. We refer to such faces D as good.

Getting a list on each good Johnson: Fix a good $D \in X(d)$, consider the Johnson graph $\binom{D}{k}$ and the assignment F_D which the restriction of F to $\binom{D}{k}$. Then the fact that $\mathsf{pass}_d(F;D) \geqslant \delta^2/100$ implies that F_D passes the Johnson scheme direct product test inside T with probability at least $\delta^2/100$. Thus, by direct product theorems over the Johnson scheme – and more precisely by the result of [DG08] – we conclude that there is a function $f_D\colon D \to \{0,1\}$ such that $\Delta(f_D|_A, F_D[A]) = o(1)$ for at least $\delta' = \delta^{O(1)}$ fraction of the k-faces $A \subseteq D$. To simplify terminology, we refer to an A on which $\Delta(f_D|_A, F_D[A]) = o(1)$ as an A on which f_D and $F_D[A]$ agree.

We would like to form a list of all of the functions that achieve δ' agreement with F_D. The list of all of these functions though may be large (its size may typically depend on the parameters k and d, whereas we wish our list sizes to only be a function of δ and δ'). To remedy this situation, we create these lists in a more careful manner. One way to go about it is to construct an "ε-net" for all these functions, and the reader should have this in mind (our precise execution is a bit different but morally the same). Thus, we are able to find, for each good D, a maximal list $f_{1,D}, \dots, f_{m,D}$ of functions that have at least δ' agreement with F_D, and we have a list size bound $m \leqslant m(\delta')$. Here, maximality asserts that no function f that is somewhat

[6] The diligent reader may notice that the distribution of (A,B) in the test inside T and in our direct product tester over X is not quite the same. The probability that $A \cap B = I$ in the test inside T is $1 - \Theta(k/t)$, whereas it is $1 - \Theta(k/d)$ in the direct product tester. Conditioned on this event (which has probability close to 1) the two distributions are identical, and thus they are close in total variation distance. Therefore, we will think of the two distribution as essentially the same.

far from all $f_{i,D}$ has agreement at least δ' with F_D. We also remark that a-priori, the list size m could also depend on the identify of the face D, but we omit it from the notation for now.[7]

Generating a gap: Consider the integer valued map $m : \delta' \to m(\delta')$, mapping a soundness parameter to an upper bound on the list size corresponding to it. Considering its values in the interval $[\delta'/2, \delta']$, we see that its maximum value is at most some $M(\delta')$. Partition the interval $[\delta'/2, \delta']$ into $R \gg M$ intervals of equal length, and towards this end consider $\delta_i = \delta' - i\frac{\delta'}{2R}$ and the intervals $[\delta_{i+1}, \delta_i]$ for $i = 0, \ldots, R-1$. Among these intervals there are at most M intervals on which the value of m changes. Thus, we will choose i randomly, and in fact apply the above list-decoding procedure for δ'_i (as opposed to δ'). The benefit of this procedure will be that it generates a gap: with probability $1 - O(M/R)$ we get a list of functions that all have agreement at least δ'_i with F_T, and all functions with agreement δ'_{i+1} are quite close to at least one of the functions in the list. Indeed, the idea is that of it wasn't the case, then the list size would exceed beyond M for $i = R-1$.

Thus, after this step for $1-O(M/R)$ of the good D's we get a list of functions that have at least $\delta'' := \delta'_i$ agreement with F_D, and any function that has agreement at least $\delta''' := \delta'_{i+1} < \delta''$ with F_D is close to some function in the list. We refer to such D's as very good henceforth, and define the list $L[D]$ to be the list we created for D. With additional work, one may guarantee that any two functions in $L[D]$ are somewhat far from each other. More precisely, we are able to guarantee that they differ on at least $\Omega_\delta(1)$ fraction of the points in D; the key point here is that the distance between the functions in the list exceeds the closeness parameter any function with agreement at least δ''' has.

Consistency of the Local Lists: The steps described above for faces of size d can be applied also for faces of size $d/2$. Indeed, we define the notion of good and very good faces there as well. Thus, we now have lists $L[D]$ and $L[P]$ for each very good $D \in X(d)$ and $P \in X(d/2)$.

Consider sampling $D \sim \mu_d$ and then $P \subseteq D$ of size $d/2$ uniformly. Looking at the list of D, $L[D] = (f_{1,D}, \ldots, f_{m,D})$ naturally gives the restricted functions $f_{i,D}|_P \colon P \to \{0,1\}^{d/2}$ as candidate functions for the list of P. Indeed, we show that with probability $1 - o(1)$ it is the case that each one of these still has agreement at least $\delta'' - o(1)$ with F_P. Furthermore, we know that if a function f is far from all $f_{i,D}$, then it has agreement at most δ''' with F_D, and one can show that with probability $1 - o(1)$ it will be the case that $f|_P$ has agreement at most $\delta''' + o(1)$ with F_D. However, the number of these f's is too large for us to use the union bound.

To circumvent this issue, we formulate the problem as a Max-CSP problem and think of the CSP problem on P as a random sub-instance on the CSP problem over D. Appealing to results about random sub-instances of dense CSPs [AdlVKK02],[8] we are able to avoid the union bound and show that with probability $1-o(1)$, any f which is far from all $f_{i,D}$ has agreement at most $\delta''' + o(1)$ in P.

We conclude that with probability $1 - o(1)$, the projection of the list $L[D]$ to P constitutes a list of size m functions that each has agreement at least $\delta'' - o(1)$ with F_P and any function far from it has agreement at most $\delta''' + o(1) < \delta'' - o(1)$. This implies that we may find a natural 1-to-1 correspondence between $L[D]$ and $L[P]$: we pair functions that are closest in Hamming distance. Among other things, this asserts that the list sizes of D and P are the same with probability $1 - o(1)$.

In other words, the lists $L[D]$ pass the following list agreement tester with probability $1 - o(1)$:

1. Sample $P \sim \mu_{d/2}$.

[7]The reason is that, as we prove in subsequent steps, the list size typically does not depend on the identity of the face D.

[8]These in return, rely on combinatorial ideas revolving around (weak) regularity lemmas.

2. Sample $D, D' \supseteq P$ independently according to μ_d, and check that $L[D]|_P = L[D']|_P$.

In the next part of the argument we prove that the list agreement test above is sound.

1.4.4 List Agreement Testing using Coboundary Expansion

Designing the Unique Games instance and proving triangle consistency: With the downwards consistency step done, one is naturally led to consider Unique-Games instances over a graph on $X(d)$ similar to the one defined earlier in the introduction. Namely, take D and D' that intersect on a $d/2$ face, call it P, one has that marginally the distribution of (D, P) and (D', P) is as in the downwards consistency step, and we get a 1-to-1 correspondence between lists of D, P and then to D' again. Composing these correspondences, we get a 1-to-1 correspondence between the lists of D and D'.

The issue with the UG instance over $X(d)$ is that we only know the UG coboundary property to hold up to some level r, which is substantially smaller than k and in particular much smaller than d (e.g., it could just be some function of δ in some of our results).

To remedy this situation, we show that one may "project" this UG instance to a UG instance on $G_r[X]$, while retaining the constraint structure. To be more explicit, suppose that we sample $D \sim \mu_d$, then $R \subseteq D$ of size r uniformly and inspect the projection of the list $L[D]$ to R, i.e. $L[D]|_R = (f_{1,D}|_R, \dots, f_{m,D}|_R)$. First, as the pairwise distance between the functions $f_{i,D}$'s is sufficiently large (and in particular larger than d/r) the projected functions in $L[D]|_R$ remain distinct. Second, we show that for a typical R, there exists a list $L[R]$ such that $L[D]|_R = L[R]$ for almost all $D \supseteq R$. [9] This gives lists on the r-faces, and one now has to inherit the constraint structure from $X(d)$. To do that, note that one may also define $L[R']$ for faces $2r$ of size R' and argue in the same way. Thus, constraints over the graph $G_r[X]$ follow naturally as: sample a $2r$-face R' and partition it randomly as $R_1 \cup R_2$ where $|R_1| = r = |R_2|$. Note that with probability $1 - o(1)$ there is a 1-to-1 correspondence between $L[R_1]$ and $L[R']$ as well as $L[R_2]$ and $L[R']$, and thus one gets a 1-to-1 correspondence between $L[R_1]$ and $L[R_2]$.

Applying UG coboundary expansion: With the UG instance over $G_r[X]$ defined, strong triangle consistency is easily proved by using ideas that are similar to the construction of the constraints (looking at $3r$-faces instead of $2r$-faces). Thus, we may appeal to the UG-coboundary expansion of the complex X and conclude that the constraints $\pi(u, v)$ on $G_r[X]$ have a form as specified in Definition 1.8.

As shown earlier in the introduction, a UG instance with this type of constraints admits a solution satisfying almost of its constraints. Namely, we may find a map $\ell\colon X(r) \to [m]$ that satisfies at least $1 - o(1)$ fraction of the constraints on $G_r[X]$.

Concluding the global structure: The next step is to consider the assignment $F'\colon X(r) \to \{0,1\}^r$ defined as $F'[R] = L[R]_{\ell(T)}$, and note that F' passes the direct product tester with respect to X with probability close to 1! Indeed, this can be shown as an easy corollary of the fact that the labeling ℓ satisfies almost all of the constraints of $G_r[X]$. This means that we are precisely in the setting of Dinur and Kaufman [DK17], and applying the result from there one concludes that F' has global structure, namely that there is a function $f\colon X(1) \to \{0,1\}$ such that

$$\Pr_{R \sim \mu_r}[\Delta(F'[R], f|_R) = o(1)] \geqslant 1 - o(1).$$

Looking at a random $D \sim \mu_d$ and a uniformly chosen $R \subseteq D$, we see that there is a 1-to-1 correspondence between the list of R and the list of D. As the labeling ℓ chooses a function from the list of R for each

[9] Intuitively, the reason for that is that the lists $L[D]|_R$ and $L[D']|_R$ agree with probability close to 1 when D and D' intersect in size $d/2$, and this transition operator has second eigenvalue bounded away from 1.

R in a way that satisfies almost all of the constraints, we conclude that for a typical D there is $\ell(D)$ such that the function $L[D]_{\ell(D)}$ agrees with almost all of the r-faces $R \subseteq T$. Using this fact, one quickly concludes that $\Delta(f|_D, L[D]_{\ell(D)}) = o(1)$ for a typical D. Thus, sampling a k-face $K \subseteq D$ one has that $L[D]_{\ell(D)}|_K \equiv F[D]$ with probability at least δ', and by Chernoff's inequality

$$|\{i \in K \mid L[D]_{\ell(D)}(i) \neq f(i)\}| = o(k)$$

with probability $1 - o(1)$, and we conclude that with probability $\Omega(\delta')$ we have that $\Delta(F[K], f|_K) = o(1)$.

2 Preliminaries

In this section we present a few standard notations as well as tools that will be used throughout.

Notation: Given a string $x \in \{0,1\}^n$ and a subset $A \subseteq [n]$, we denote by x_A the substring of x corresponding to keeping only the symbols in the coordinates of A. Given two strings $x, y \in \{0,1\}^n$, we denote by $\Delta(x, y)$ the fractional Hamming distance between x and y, and given a set $A \subseteq [n]$, we define $\Delta_A(f, g) = \Delta(x_A, y_A)$. We use the notation $x \neq_{\leqslant \eta} y$ to denote that $\Delta(x, y) \leqslant \eta$. Given a list L of strings in $\{0,1\}^n$ we say that the distance η if all distinct $x, y \in L$ we have have $\Delta(x, y) \geqslant \eta$.

We use standard big-O notations: we denote $A = O(B)$ or $A \lesssim B$ if $A \leqslant c \cdot B$ for some absolute constant $c > 0$. Similarly, we denote $A = \Omega(B)$ or $A \gtrsim B$ if $A \geqslant cB$ for some absolute constant $c > 0$. We also denote $k \ll d$ to denote the fact that d is taken to be sufficiently large compared to any function of k.

Whenever we have a d-dimensional simplicial complex X and $1 \leqslant k \leqslant d$, we denote by $A \sim X(k)$ a sample according to the distribution measure μ_k over $X(k)$ (as defined in the introduction). We use $B \subset_t A$ to denote that B is a uniform t-sized subset of A. Similarly, for B of size t, when we write $A \supset_k B$ we mean that A is distributed according to $A \sim X(k)$ conditioned on containing B.

2.1 Concentration bounds

We will need the following version of Chernoff's inequality:

Theorem 2.1. *Suppose X_i are independent random variables taking values in $\{0,1\}$ and X denotes their sum. If $\mathbb{E}[\sum X_i] = \mu$ then,*

$$\Pr[|X - \mu| > \delta\mu] \leqslant \exp(-\delta^2\mu), \quad \textit{for } \delta \in (0,1),$$

$$\Pr[X > (1+\delta)\mu] \leqslant \exp(-\delta\mu), \quad \textit{for } \delta \geqslant 1.$$

2.2 Constraint Satisfaction Problems: Value and Random Sub-Instances

Our argument will make use of instances of the *max-k-CSP* problem and properties of random sub-instances of a given instance.

Definition 2.2. *Let $k \in \mathbb{N}$. An instance Ψ of (Boolean) max-k-CSP consists of a set of variables $\{x_i\}_{i \in I}$, along with constraints, each one of the form $P(x_{i_1}, \ldots, x_{i_k}) = 1$ for some $P \colon \{0,1\}^k \to \{0,1\}$.*

Given an instance Ψ of max-k-CSP, the goal is to find an assignment to the variables of Ψ that satisfies as many of the constraints as possible. We refer to this maximum fraction as the value of Ψ, and denote it by $\mathrm{val}(\Psi)$.

Given an instance of max-k-CSP Ψ with variables V and a subset of variables $Q \subseteq V$, we define the induced instance on Q, $\Psi|_Q$, to be the instance of max-k-CSP resulting from Ψ by keeping only the variables of Q, and only the constraints of Ψ that involve only variables from Q.

Of special interest to us will be dense instances of max-k-CSP. In fact, we will be concerned with instances wherein there is a constraint for every subset of size k of the variables, and where the number of variables d is much larger than k. Given such an instance Ψ, we will want to consider random induced sub-instances of Ψ and their value. With this regard, the following result from [AdlVKK02] asserts that the value of the random sub-instance remains roughly the same.

Theorem 2.3. *For all $\gamma, \tau \in (0,1)$, $k \in \mathbb{N}$ and $d \geqslant \mathrm{poly}(k/\tau)\exp(1/\gamma^2)$,*[10] *consider a k-CSP with $\binom{n}{k}$ constraints that each depend on a unique k-set of variables. If $q \geqslant \mathrm{poly}(k/\tau\gamma)$ then:*

$$\Pr_{Q \subset_q [d]}\left[|\mathrm{val}(\Psi|_Q) - \mathrm{val}(\Psi)| \leqslant \gamma\right] \geqslant 1 - \tau.$$

We will also consider a special type of max-2-CSP, called Unique-Games. Unique-Games have already been considered in the introduction, and we will make use of them later on in our argument.

Definition 2.4. *An instance of Unique-Games $\Psi = (G, \Pi)$ consists of a graph $G = (V, E)$, a finite alphabet Σ and a collection of constraints, $\Pi = \{\pi_e\}_{e \in E}$, one for each edge in G. For all $e \in E$, the constraint π_e takes the form $\pi_e = \{(\sigma, \pi_e(\sigma)) \mid \sigma \in \Sigma\}$, where $\pi_e \colon \Sigma \to \Sigma$ is a permutation.*

The goal in the Unique-Games problem is to find an assignment $A \colon V \to \Sigma$ that satisfies the maximum possible number of constraints. That is, maximizing the number of edges $e = (u, v) \in E$ such that $(A(u), A(v)) \in \pi_e$. We define the value of the instance Ψ as:

$$\mathsf{val}(\Psi) = \max_{A \colon V \to \Sigma} \frac{\#\{e \mid A \text{ satisfies } e\}}{|E|}.$$

2.3 Properties of Expanders

We need the following well known version of the expander mixing lemma for bipartite graphs.

Lemma 2.5. *Let $G = (U, V, E)$ be a bipartite graph in which the second singular value of the normalized adjacency matrix is at most λ. Then for all $A \subset U$ and $B \subset V$ we have that*

$$\left|\Pr_{(u,v) \in E}[u \in A, v \in B] - \mu(A)\mu(B)\right| \leqslant \lambda\sqrt{\mu(A)(1-\mu(A))\mu(B)(1-\mu(B))}.$$

We also use the following standard sampling property of bipartite expanders.

Lemma 2.6. *Let $G = (U, V, E)$ be a bipartite graph with second singular value at most λ. If $B \subset U$ has $\Pr[B] = \delta$, then the set $T = \left\{v \in V \mid \Pr_{u \sim E|_v}[u \in B] > \varepsilon + \delta\right\}$ has $\Pr[T] \leqslant \lambda^2\delta/\varepsilon^2$.*

[10] For general dense k-CSPs they incur a $\exp(2^{2^k})$ dependence in d, which comes from the fact that there can be 2^{2^k} constraints in Ψ that can be satisfied by setting a particular set of variables $I \subset_k [d]$ to a fixed assignment $z \in \{0,1\}^k$. In our setting, there could only be one constraint that gets satisfied by such fixing, and therefore we do not incur this triple-exponential dependence on k (though this wouldn't matter for us in any case).

2.4 Properties of Two-Sided Local Spectral Expanders

Recall that we associated with each d-dimensional simplicial complex X a sequence of measures $\{\mu_k\}_{1\leqslant k\leqslant d}$, where μ_k is a probability measure over $X(k)$. Note that for all $0 \leqslant t \leqslant r \leqslant d$, a sample according to μ_t can be drawn by first sampling $R \sim \mu_r$, and then sampling $T \subseteq_t R$ uniformly. The converse is also true: a sample from μ_r can be drawn by first sampling $T \sim \mu_t$, and then sampling R from μ_r conditioned on containing T. These observations give rise to the standard "up" and "down" operators, which we present next. We only mention a few of their properties that are necessary for our arguments, and refer the reader to [DDFH18] for a more comprehensive exposition.

Definition 2.7. *The operator U_i^{i+1} is a map from $L_2(X(i);\mu_i)$ to $L_2(X(i+1);\mu_{i+1})$ defined as*

$$U_i^{i+1}f(u) = \mathop{\mathbb{E}}_{v\subset_i u}\left[f(v)\right]$$

for all $u \in X(i+1)$. For $j \geqslant k+1$, we define U_k^j via composition of up operators: $U_k^j = U_{j-1}^j \circ \ldots \circ U_k^{k+1}$.

Definition 2.8. *The operator D_i^{i+1} is a map from $L_2(X(i+1);\mu_{i+1})$ to $L_2(X(i);\mu_i)$ defined as*

$$D_i^{i+1}f(u) = \mathop{\mathbb{E}}_{v\supseteq_{i+1} u}\left[f(v)\right]$$

for all $u \in X(i)$. For $j \geqslant k+1$, we define D_k^j via composition of down operators: $D_k^j = D_k^{k+1} \circ \ldots \circ D_{j-1}^j$.

Abusing notations, we use the notations U_k^j, D_k^j to denote the operators, as well as the real valued matrices associated with them. A key property of the down and up operators is that they are adjoint:

Claim 2.9. *For all $k \leqslant j \leqslant d$, U_k^j and D_k^j are adjoint operators: for all functions $f\colon X(k) \to \mathbb{R}$ and $g\colon X(j) \to \mathbb{R}$ it holds that $\langle U_k^j f, g\rangle = \langle f, D_k^j g\rangle$.*

We need the following lemma regarding the second eigenvalue of the down-up walks $U_k^j D_k^j$ on $X(j)$ ($j \geqslant k$), that can be found in [AL20].

Lemma 2.10. *Let (X,μ) be a d-dimensional γ two-sided local spectral expander. For all $i \leqslant d$ and $\alpha \in (1/i, 1)$, the largest singular value of $U_{\alpha i}^i$ and $D_{\alpha i}^i$ is at most $\sqrt{\alpha} + \mathrm{poly}(i)\gamma$. Thus the down-up random walk $U_{\alpha i}^i D_{\alpha i}^i$ on $X(i)$ has second largest singular value at most $\alpha + \mathrm{poly}(i)\gamma$.*

We will use the following theorem from [DK17] that shows that spectral HDXs support an agreement test in the 99% regime.[11]

Theorem 2.11 ([DK17])**.** *Let X be a d-dimensional λ two-sided local spectral expander and let $t^2 < d$, $\lambda < 1/d$ and $\varepsilon > \varepsilon_0(t,\lambda)$. Let $F : X(t) \to \{0,1\}^t$ such that:*

$$\Pr_{\substack{D\sim\mu_d \\ Q\subseteq_{t/2} D \\ Q\subseteq B, B'\subseteq_t D}}\left[F[B]|_Q = F[B']|_Q\right] \geqslant 1-\varepsilon.$$

Then, there exists a function $G : X(1) \to \{0,1\}$ such that,

$$\Pr_{B\sim X(t)}\left[F[B] = G[B]\right] \geqslant 1 - O(\varepsilon).$$

[11] Strictly speaking, the direct product tester that Dinur and Kaufman analyze a bit different. The formulation we give is a bit more convenient for us to apply, and the proof in [DK17] applies to that setting in exactly the same way to give the soundness guarantee as stated in Theorem 2.11.

3 UG Coboundary is Necessary

In this section we prove the "necessary" part of Theorem 1.9, stated formally below.

Theorem 3.1. *For all $c, \eta > 0$, $m, r \in \mathbb{N}$, there exist $d, k \in \mathbb{N}$, $\xi, \gamma > 0$ such that the following holds. If a simplicial complex X is a d-dimensional γ-spectral expander and is not (m, r, ξ, c) coboundary expander, then there exists $F\colon X(k) \to \{0,1\}^k$ that passes the $(k, \sqrt{k})$ direct product tester with probability at least $\frac{1}{m} - O(\sqrt{\xi})$, and yet for all $f\colon X(1) \to \{0,1\}$ we have that*

$$\Pr_{A \sim \mu_k}\left[\Delta(F[A], f|_A) \geqslant \Omega_{m,r,c}(1)\right] \leqslant \eta.$$

Henceforth in this section, we assume that X is **not** a weak (m, r, ξ, c) UG coboundary expander. That is, for some $t \leqslant r$ there exists a collection of lists $\{L'(R)\}_{R \in X(t)}$, $\{L'(T)\}_{T \in X(3t)}$ and a set of permutations $\{\pi'(S)\}_{S \in X(2t)}$ such that,

$$\Pr_{\substack{T \sim \mu_{3t} \\ A \cup B \cup C = T}}[L'(T) = L'(A) \circ \pi'(A, B) L'(B) \circ \pi'(B, C) L'(C)] \geqslant 1 - \xi,$$

and yet for all $P\colon X(t) \to S_m$ it holds that $\pi'(A, B) = P(A)P(B)^{-1}$ with probability at most $1 - c$ over the choice of $A \cup B \sim \mu_{2t}$. We refer to such P as an S_m-solution to the Unique-Games instance $\Psi' = (G_t[X], \Pi')$. This step is summarized in the following lemma.

3.1 Preprocessing

The first step of the proof is to show that we can convert Ψ' as above to a (possibly different) UG instance Ψ over $G_t[X]$ that doesn't have any $[m]$-valued solution satisfying more than $(1 - c/m)$ weight of the constraints in $G_t[X]$.

Lemma 3.2. *There exists a collection of lists $\mathcal{L} = \{L(R)\}_{R \in X(t)} \cup \{L(T)\}_{T \in X(3t)}$ such that:*

$$\Pr_{\substack{T \sim X(3t) \\ A \cup B \cup C \sim T}}[L(T) = L(A) \circ \pi(A, B) L(B) \circ \pi(B, C) L(C)] \geqslant 1 - \xi,$$

and a set of permutations $\Pi = \{\pi(S)\}_{S \in X(2t)}$ such that the corresponding UG instance $\Psi = (G_t[X], \Pi)$ has no solution $I : G_t[X] \to [m]$ with $\mathrm{val}(I) \geqslant 1 - c/m$.

Proof. Start with the collection of lists $\mathcal{L}' = \{L'(R)\}_{R \in X(t)} \cup \{L'(T)\}_{T \in X(3t)}$ and permutations $\Pi' = \{\pi'(S)\}_{S \in X(2t)}$. If for all $A : X(t) \to [m]$, $\mathrm{val}(A) \leqslant 1 - c/m$ we are done, therefore let us assume that there exists such a solution $A\colon X(t) \to [m]$ satisfying $1 - c/m$ fraction of the edges. In this case we will do the following:

1. Remove $A(u)$ from the list $L[u]$ to get a new list of size $m - 1$.

2. For all edges (u, v), change $\pi'(u, v)$ to a permutation $\pi''(u, v) \in S_{m-1}$, by "removing" $(A(u), A(v))$, i.e. $\pi''(u, v)(i) = \pi'(u, v)(i)$ for all $i \neq A(u)$.

Continue this procedure until there is no such assignment, and let $(\mathcal{L}, \Pi)$ be the set of lists and permutations when this ends. We argue that these lists are non-empty; in fact they will be of size at least 2. Given that, the fact that $(\mathcal{L}, \Pi)$ is strongly triangle consistent is obvious.

To see that the lists are non-empty, first notice that the process must terminate within $m-2$ steps. Indeed, otherwise for each step $1 \leqslant i \leqslant m-1$ define $A_i(R)$ to be the assignment given to vertex R in the ith iteration, and define $A_m(R)$ to be the last assignment left in the list of R after all iterations are done. Thus, defining the permutation valued assignment $P\colon X(t) \to S_m$ as $P(R)$ being the unique permutation for which $P(u)L'(R) = (g_1(R), \ldots, g_m(R))$, we see that $\pi'(U,V) = P(U)P(V)^{-1}$ unless the edge (U,V) was violated by at least one of the assignments $A_1, \ldots, A_m$. By the union bound, the total weight of edges violated by at least one of $A_1, \ldots, A_m$ is at most $m \cdot \frac{c}{m} = c$, hence $\pi'(U,V) = P(U)P(V)^{-1}$ for at least weight $1-c$ of the edges, in contradiction.

□

Henceforth, we fix a Unique-Games instance $(\mathcal{L}, \Pi)$ as in Lemma 3.2, and denote $c' = c/m$.

3.2 Lifting the Lists

The next step in the proof is to lift the lists $\mathcal{L}$ to lists on $X(k)$, and to do so we use Kneser graphs. For every $A \in X(k)$, denote by $K(A,t)$ the Kneser graph on A, whose vertex set is $\binom{A}{t}$, and two t-sets T, T' are adjacent if $T \cap T' = \emptyset$. Note that this is exactly the subgraph of $G_t[X]$ induced by the t-faces contained inside A. We say a k-face A triangle-consistent if all the triangles in $K(A,t)$ are consistent with respect to $(\mathcal{L}, \pi)$, and denote by $\mathcal{K}(k) \subseteq X(k)$ the collection of triangle-consistent k-faces.

Claim 3.3. *If $\xi \leqslant \exp(-t \log k)$, then $\mu_k(\mathcal{K}(k)) \geqslant 1 - \sqrt{\xi}$.*

Proof. We know that a triangle picked as $T \sim X(3t)$ and $a \cup b \cup c = T$ is inconsistent with probability at most ξ. This is the same distribution as picking a k-face from $K \sim X(k)$, $T \subset_{3t} K$ and $a \cup b \cup c \sim T$. Thus by linearity of expectation the number of inconsistent triangles in K is at most $k^{3t}\xi$, which implies that the probability that $K \sim \mu_k$ contains at least one inconsistent triangle is at most $k^{3t}\xi \leqslant \sqrt{\xi}$. □

We now show how to lift the lists to k-faces in $\mathcal{K}(k)$. To do so we will need the following simple claim about Kneser graphs. We will show that triangle consistent UG instances on Kneser graphs are satisfiable, i.e. any set of permutations on the edges of $K([k],t)$ that is consistent on triangles has value 1.

Claim 3.4. *For all $r \geqslant 1, k \geqslant 5t$ the following holds. Let $\Phi = (K([k],t), \Pi)$ be a UG instance on $K([k],t)$ with alphabet $[m]$, in which all triangles are consistent. Then $\mathrm{val}(\Phi) = 1$ and furthermore there exist m distinct satisfiable assignments.*

Proof. Let K denote $K([k],t)$. We will show that all cycles in K are consistent, i.e. for any ℓ-cycle ($\ell \geqslant 3$) $C = (v_1, \ldots, v_\ell) \in K$ it holds that $\pi(v_1,v_2) \cdot \ldots \cdot \pi(v_\ell, v_1) = \mathrm{id}$. We can show this by induction on the length of the cycle. By assumption 3-cycles are consistent. Given that all $\ell-1$-cycles are consistent we can prove that all ℓ-cycles in K are consistent. Given cycle $C = (v_1, \ldots, v_\ell)$, consider any vertex u such that u has an edge to v_1, v_2, v_3 and v_4. Such a vertex exists because $|\cup_{i=1}^4 v_i| \leqslant 4t$ and $k \geqslant 5t$. Since u is connected to $v_1, \ldots, v_4$ we get that:

$$\begin{aligned}\pi(v_1,v_2)\pi(v_2,v_3)\pi(v_3,v_4) &= \pi(v_1,u)\pi(u,v_2)\pi(v_2,u)\pi(u,v_3)\pi(v_3,u)\pi(u,v_4)\\ &= \pi(v_1,u)\pi(u,v_4)\end{aligned}$$

where we used the fact that the triangles $(v_1,u,v_2), \ldots, (u,v_3,v_4)$ are consistent. Therefore we have reduced the task of showing that the cycle C is consistent, to the task of showing that the cycle $C' = (v_1, u, v_4, \ldots, v_\ell)$ of length $\ell - 1$ is consistent; this last assertion is true by the induction hypothesis.

Given that all cycles in this graph are consistent, we can assign the first vertex in the graph to be some element from $[m]$, and find an assignment to the rest of the vertices via propagation. Note that we do not run into contradictions because all cycles are consistent. It is easy to see that such an assignment will satisfy all the edges of the graph, and we will get a set of m distinct satisfiable assignments in this way. $\square$

For convenience of notation, we will say that an assignment $(F[U], F[V])$ (for $U, V \in X(t), U \cap V = \emptyset$) satisfies the edge (U, V) if $F[U] = L_i(U)$ and $F[V] = L_j[V]$ for some $i, j \in [m]$ and $j = \pi(i, j)i$.

Lemma 3.5. *If $k \geqslant 5t$, then for every $A \in \mathcal{K}(k)$, there exists a unique list $L(A)$ such that $L(A)|_R = L(R)$ for all $R \subset_t A$. Furthermore, for every assignment $F \in L(A)$ and edge $(U, V) \in K(A, t)$, $(F|_U, F|_V)$ satisfies the edge (U, V).*

Proof. Fix an $A \in \mathcal{K}(k)$. Since A is triangle-consistent we know that the permutations on its edges are triangle consistent, i.e. for all triangles (a, b, c) in $K(A, r)$, $\pi(a, b)\pi(b, c)\pi(c, a) = \mathrm{id}$. In Claim 3.4 we showed that all triangle consistent UG instances on $K(A, t)$ (for $|A| \geqslant 5t$) have m global solutions that satisfy all the edges. Let these assignments by $S_1, \ldots, S_m$ that map the vertices of $K(A, t)$ to $[m]$.

For each S_i construct the following assignment: $B_i : K(A, t) \to \{0, 1\}^t$, defined as $B_i(T) = L_{S_i(T)}(T)$. Since S_i satisfies the permutations on every edge, we know that, for all triangles (T_1, T_2, T_3) in $K(A, t)$, $B_i(T_1) \circ B_i(T_2) \circ B_i(T_3) \in L(T_1 \cup T_2 \cup T_3)$. In particular by the connectivity of the graph we can check that in fact $B_i(T_1) \circ B_i(T_2) \circ B_i(T_3) = B_i(T_1') \circ B_i(T_2') \circ B_i(T_3')$ for two different splittings of the $3t$-sized set $T_1 \cup T_2 \cup T_3$. This immediately implies that there is a unique assignment $C_i \in \{0, 1\}^k$ to A such that $C_i|_T = B_i(T)$ for all $T \subset_t A$. By definition $(B_i(U), B_i(V))$ satisfies the edge $(U, V) \in K(A, t)$, therefore so does the assignment $(C_i|_U, C_i|_V)$.

Putting all the assignments $C_1, \ldots, C_m$ in a list we get $L(A)$ that satisfies $L(A)|_T = L(T)$ for all $T \in K(A, t)$. $\square$

3.3 Constructing the Assignment and Its Soundness

Fix the lists $\{L(A)\}_{A \in X(k)}$ as in Lemma 3.5. Below, we show how to construct an assignment F that passes the direct product test with probability $\Omega(1/m)$. In fact, it passes the related list-agreement test with probability close to 1.

Lemma 3.6. *For all $(5t)^2 \leqslant k \leqslant d$, there exists a function $F : X(k) \to \{0, 1\}^k$ that satisfies $F[A] \in L(A)$, for all $A \in X(k)$, such that:*

$$\Pr_{\substack{D \sim \mu_d \\ B \subseteq_{\sqrt{k}} D \\ B \subset A, A' \subset_k D}} \left[F(A)|_B = F(A')|_B \right] \geqslant \frac{1}{m} \left(1 - O(\sqrt{\xi}) \right).$$

Proof. We will first show that the list-agreement test passes with probability $1 - O(\sqrt{\xi})$. Let D, B, A, A' be generated as in the statement of the lemma, and denote by $\mathcal{D}$ the distribution of (B, A, A'). Note that the marginal distribution on $B \sim \mathcal{D}$ is $\mu_{\sqrt{k}}$ and on A and $A' \sim \mathcal{D}$ is μ_k. Therefore, by the union bound we get that with probability $1 - 3\sqrt{\xi}$ it holds that $B \in \mathcal{K}(\sqrt{k})$ and $A, A' \in \mathcal{K}(k)$ (Claim 3.3). Call such a triple (B, A, A') as good.

Fix a good triple (B, A, A'). By virtue of being in $\mathcal{K}(\sqrt{k})$, $L(B)|_T = L(T)$ for all $T \subset_t B$ and the same holds for $L(A), L(A')$, which in particular implies that $L(A)|_T = L(B)|_T = L(A')|_T$ for all $T \subset_t B$.

There can only be one list on B that satisfies $L(B)|_T = L(T)$ for all $T \subset_t B$. Therefore we get that $L(A)|_B = L(A')|_B$ for all good triples (B, A, A'), hence,

$$\Pr_{(B,A,A')\sim\mathcal{D}}[L(A)|_B = L(A')|_B] \geqslant 1 - O(\sqrt{\xi}).$$

Thus, choosing $F[A]$ to be a random element from the list $L(A)$ we get that

$$\mathop{\mathbb{E}}_{F}\left[\Pr_{(B,A,A')\sim\mathcal{D}}[F(A)|_B = F(A')|_B]\right] \geqslant \frac{1}{m}\left(1 - O(\sqrt{\xi})\right).$$

Therefore we can pick an assignment F such that the above holds. $\square$

3.4 No Global Structure

We finish by showing that for F as constructed in Lemma 3.6, there is no global function $f\colon X(1) \to \{0,1\}$ that has significant agreement with it. In fact, we show for any F on $X(k)$ such that $F[A] \in L(A)$, there is no global function on $X(1)$ that agrees with F on a large fraction of A's.

Lemma 3.7. *Suppose that the UG instance $\Psi = (G_t[X], \Pi)$ has no solution of value $\geqslant 1 - c'$. Then, for all functions $F : X(k) \to \{0,1\}^k$ where $F[A] \in L(A)$ for all A, and for any $G : X(1) \to \{0,1\}$*

$$\Pr_{A\sim\mu_k}\left[\Delta(G[A], F[A]) \leqslant O\left(\frac{c'}{t}\right)\right] \leqslant O\left(\frac{t}{kc'^2}\right) + \sqrt{\xi}.$$

Proof. Fix such a function F and suppose for contradiction that there is $G : X(1) \to \{0,1\}$ such that

$$\Pr_{A\sim\mu}[\Delta(G[A], F[A]) \leqslant \varepsilon] \geqslant \alpha + \sqrt{\xi} \tag{1}$$

for ε, α to be determined later. In this case we will construct an assignment to the UG instance Ψ with large value, which will be a contradiction.

Consider the following assignment $I : G_t[X] \to [m]$. For each $T \in X(t)$ let $I[T] = i$ if $G[T] = L_i(T)$, else assign $I[T]$ arbitrarily. Thus

$$\mathrm{val}(I) \geqslant \Pr_{\substack{T\sim\mu_{2t}\\ U\cup V\sim T}}\big[(G[U], G[V]) \text{ satisfies the edge } (U,V)\big],$$

where $A \cup B$ is a random split of T into two sets of size t each. We show that $\mathrm{val}(I)$ is close to 1.

Let $\mathcal{K}'$ be the set of k-faces A where $\Delta(G[A], F[A]) \leqslant \varepsilon$ and $A \in \mathcal{K}(k)$. By Claim 3.3 it holds that $\mu_k(\mathcal{K}) \geqslant 1 - \sqrt{\xi}$, and combining with (1) yields that $\mu_k(\mathcal{K}') \geqslant \alpha$.

Fix $A \in \mathcal{K}'$. For $T \subseteq_{2t} A$, we get that $\Delta_T(G[A], F[A]) = 0$ with probability at least $1 - 2\varepsilon t$, so

$$\Pr_{T\subset_{2t}A}[F[A]|_T = G[T]] \geqslant 1 - O(\varepsilon t). \tag{2}$$

Thus, for $A \in \mathcal{K}'$ define $\mathrm{Good}(A) = \{T \in X(2t) \mid F[A]|_T = G[T]\}$, and define $\mathcal{T} = \bigcup_{A\in\mathcal{K}'} \mathrm{Good}(A)$. First note that for all $T \in \mathcal{T}$ and every splitting (U,V) of T, $(G[U], G[V])$ satisfies the edge (U,V). To see this, consider some $T \in \mathcal{T}$, where $T \in \mathrm{Good}(A)$ for $A \in \mathcal{K}'$. Since $A \in \mathcal{K}$ and $F[A] \in L(A)$, by Lemma 3.5 $(F[A]|_U, F[A]|_V)$ satisfies the edge (U,V). Since F and G are equal on T this immediately implies that G also satisfies the edges (U,V) for $U \cup V = T$. Therefore to get a bound on the value of I it suffices to lower bound the measure of $\mathcal{T}$.

Lower bounding the measure of $\mathcal{T}$: Consider the bipartite graph $G_{k,2t} = (X(k), X(2t), D^k_{2t})$ where the edges are weighted according to the down walk from $X(k)$ to $X(t)$. Namely, an edge in $G_{k,2t}$ is sampled by picking $K \sim \mu_k$, and then taking $T \subset_{2t} K$ uniformly. We will be interested in counting the number of edges between $\mathcal{K}'$ and $\mathcal{T}$.

For $\mathcal{A} \subseteq X(k)$ and $\mathcal{B} \subseteq X(2t)$, we denote by $E(\mathcal{A}, \mathcal{B})$ the set of edges between $\mathcal{A}$ and $\mathcal{B}$, and we denote by $\mu(E(\mathcal{A}, \mathcal{B}))$ the total weight of edges in $E(\mathcal{A}, \mathcal{B})$. Using these notations, we have that

$$\mu(E(\mathcal{K}', \mathcal{T})) \geqslant \mu_k(\mathcal{K}') \Pr_{T \subset_{2t} K}[T \in \mathcal{T} \mid K \in \mathcal{K}'] \geqslant \alpha(1 - O(\varepsilon t)), \tag{3}$$

where in the last inequality we used (2). By Lemma 2.5 we have

$$|\mu(E(\mathcal{K}', \mathcal{T})) - \mu(\mathcal{K}')\mu(\mathcal{T})| \leqslant \lambda(D^k_{2t})\sqrt{\mu(\mathcal{K}')} \leqslant O(\sqrt{t/k}), \tag{4}$$

where the last inequality is by Lemma 2.10 and the fact that $\gamma < 1/\mathrm{poly}(k)$. Combining (3) and (4) and simplifying gives that $\mu(\mathcal{T}) \geqslant 1 - O(t\varepsilon) - \sqrt{2t/k\alpha}$ which is at least $1 - O(t\varepsilon)$ if $\alpha \geqslant \frac{1}{tk\varepsilon^2}$. In that case, we conclude that:

$$\mathrm{val}(I) \geqslant \Pr_{\substack{T \sim X(2t) \\ U \cup V \sim T}}[(G[U], G[V]) \text{ satisfies the edge } (U, V)] \geqslant \mu_{2t}(\mathcal{T}) \geqslant 1 - O(t\varepsilon),$$

which is a contradiction to Ψ having value at most $1 - c'$ if $\varepsilon < O(c'/t)$. It follows that $\alpha \leqslant \frac{1}{tk\varepsilon^2}$, and the proof is concluded by choosing $\varepsilon = \frac{c''c'}{t}$ for sufficiently small $c'' > 0$. □

Proof of Theorem 3.1. The result follows immediately by combining Lemmas 3.6 and 3.7. □

4 Proof of Theorem 1.9: UG Coboundary is Sufficient

In this section, we prove the "sufficient" part of Theorem 1.9, formally stated below.

Theorem 4.1. *There is $c > 0$ such that for all $\varepsilon, \delta > 0$ there is $\xi, \eta > 0$ and $m, r \in \mathbb{N}$ such that for sufficiently large k, sufficiently large d and γ small enough function of d, the following holds. If a d-dimensional simplicial complex X is a γ-spectral expander and (m, r, ξ, c) weak UG coboundary expander, then the direct product test over X with respect to sufficiently large k has soundness δ. Namely, if $F\colon X(k) \to \{0,1\}^k$ passes the $(k, \sqrt{k})$ direct product tester with respect to X with probability at least δ, then there is $f\colon X(1) \to \{0,1\}$ such that*

$$\Pr_{A \sim \mu_k}[\Delta(F[A], f|_A) \leqslant \varepsilon] \geqslant \eta.$$

We begin by setting up some notations that will be helpful throughout the proof. Given a global function $f : [d] \to \{0,1\}$ and a set $B \subseteq [d]$ we let $f(B)$ denote the assignment to B using f. For a function $f : [d] \to \{0,1\}$ and an assignment $F : X(k) \to \{0,1\}^k$ we let $\mathrm{Agr}(f, F)$ denote the subset of $X(k)$ where $f(s) = F(s)$ and $\mathrm{agr}(f, F)$ denote the probability of this event under the measure μ_k. Furthermore for $\nu \in (0,1)$ let $\mathrm{Agr}_\nu(f, F)$ denote the subset of $X(k)$ where $f(B)$ and $F(B)$ agree on $(1-\nu)$-fraction of the elements in B and $\mathrm{agr}_\nu(f, F)$ denotes the probability of this event under μ_k.

4.1 High Level Structure of the Proof

The proof of Theorem 4.1 follows the outline given in the introduction. For convenience we break it into two parts, encapsulated in the following two lemmas. In the first lemma we implement the first four steps in the plan and reduce the problem of direct product testing to the problem of "list agreement" testing. In this problem, for each d-face D in a complex X we have a list $L[D]$ of $O(1)$ functions, and we test whether these lists are in 1-to-1 correspondence according to the up-down-up walk on the complex. More precisely, the problem is defined as follows:

List-Agreement-Test 1.
Input: a list $L(D)$ for each $D \in X(d)$ and a parameter $\eta \in (0,1)$.

1. Choose random $B \sim X(d/2)$.
2. Choose independently random $A, A' \supseteq_d B$ from $X(d)$.
3. Accept iff both lists are non-empty and $L[A]|_B \neq_{<\eta} L[A']|_B$.

With the list agreement problem formally defined, we can now state the lemma encapsulating the first few steps in the argument, saying that an assignment that passes the direct product test with probability bounded away from 1 implies a natural list assignment passing the list agreement test with probability close to 1.

Lemma 4.2. *For all $\delta > 0$, for sufficiently large $k \in \mathbb{N}$, $d \geqslant \mathrm{poly}(k)2^{\mathrm{poly}(1/\delta)}$, sufficiently small γ compared to d and $\tau = O(\delta^{68})$, the following holds. Suppose that X is a d-dimensional simplicial complex which is a γ-spectral expander, and $F : X(k) \to \{0,1\}^k$ passes the $(k, \sqrt{k})$-agreement-test 1 with probability δ. Then, there exists $2^{-1/\delta^{1200}} \leqslant \eta \leqslant \delta^{101}$ and lists $(L[D])_{D \in X(d)}$ satisfying:*

1. ***Short, non-empty lists:*** *With probability $1 - O(\tau)$ over the choice of $D \sim X(d)$, the list $L[D]$ is non-empty and has size at most $O(1/\delta^{12})$.*
2. ***Good agreement:*** *For all $D \in X(d)$ and every $f \in L[D]$, we have that $agr_\nu(f, F|_D) \geqslant \Omega(\delta^{12})$ for $\nu = 1/k^{\Omega(1)}$.*
3. ***Distance in the lists:*** *With probability at least $1 - O(\tau)$ over the choice of $D \sim X(d)$, the list $L[D]$ has distance at least $\delta^{-100}\eta$.*

Furthermore the lists above pass the List-Agreement-Test 1 with probability $1 - \tau$.

Armed with the conversion of our assignment F to lists that pass the list agreement test with probability close to 1, we implement the next three steps in the introduction. Namely, we show that if X is a sufficiently good UG coboundary expander, then we can use the lists above to define a locally consistent instance of Unique-Games on low levels of the complex and apply UG coboundary expansion to deduce the existence of a global solution.

Lemma 4.3. *Assume there exists a collection of lists $\{L[D]\}_{D \in X(d)}$ that satisfy the premise of Lemma 4.2, and assume that X is a γ-spectral expander for $\gamma < 1/\mathrm{poly}(d)$ and a weak $(O(1/\delta^{12}), t, O(\sqrt{\tau}), c)$ UG coboundary expander for $t = \Theta\left(\frac{\tau\delta^{12}}{\eta}\right)$. Then there exists $G : X(1) \to \{0,1\}$ such that*

$$\Pr_{D \sim X(d)}[\Delta(G(D), L[D]) \leqslant \delta] \geqslant 1 - O(c^{1/2} + \tau^{1/4} + \gamma).$$

Here, the distance between a function $G(D)$ and a list of functions $L[D]$ is the minimal distance between $G(D)$ and any function in the list.

The proof of Theorem 4.1 now readily follows from the above two lemmas.

Proof of Theorem 4.1. In the setting of Theorem 4.1, first assume that $\varepsilon = \delta$ (otherwise we lower both of them to be the minimum of ε and δ)). Apply Lemma 4.2 and then Lemma 4.3 to conclude that there is a function $G\colon X(1) \to \{0,1\}$ such that

$$\Pr_{D\sim\mu_d}[\Delta(G(D), L[D]) \leqslant \varepsilon] \geqslant \frac{1}{2}.$$

Fix $D \in X(d)$ such that $\Delta(G(D), L[D]) \leqslant \varepsilon$, and let $f \in L[D]$ be such that $\Delta(G(D), f) \leqslant \varepsilon$. Sampling $A \subseteq_k D$, we have by the "good agreement" property of the list that $F[A] \neq_{<\nu} f|_A$ with probability at least $\Omega(\delta^{12})$. By Chernoff's bound we have that $G(D)|_A \neq_{<2\varepsilon} f|_A$ with probability $1 - o(1)$. It follows that with probability at least $\Omega(\delta^{12})$ over $A \subset_k D$, $F[A]$ and $G(A)$ differ on at most $2\varepsilon + \nu \leqslant 3\varepsilon$ fraction of the coordinates of A. Since the fraction of good Ds is $\geqslant 1/2$, $\Delta(F[A], G[A]) \leqslant 3\varepsilon$ on at least $\Omega(\delta^{12})$ fraction of $X(k)$ as required. □

4.2 Auxiliary Claims

Our proof requires a few basic auxiliary probabilistic claims, which we record here. The first claim asserts that if the distance between two functions $f, g\colon [d] \to \{0,1\}$, then choosing a random subset $A \subseteq_k [d]$, we have that the distance between $f|_A$ is also very close to R. More precisely:

Claim 4.4. *Suppose $R \in (0,1)$, and let $f, g\colon [d] \to \{0,1\}$ be functions such that $\Delta(f,g) = R$. Then, for $\frac{1}{R^2} \leqslant k \leqslant d$ we have that:*

1. $\Pr_{A\subseteq_k[d]}\left[\Delta_A(f,g) > 2R\right] \leqslant 2^{-\Omega(Rk)}$.
2. $\Pr_{A\subseteq_k[d]}\left[\Delta_A(f,g) < R/2\right] \leqslant 2^{-\Omega(Rk)}$.

Proof. Both of the items are immediate consequences of Chernoff's inequality. The arguments are essentially identical, and we give a proof of the first item only.

To see this, sample $A \subseteq [d]$ by including each element in A with probability k/d. Let $I \subseteq [d]$ be the set of $i \in [d]$ such that $f(i) \neq g(i)$, and for each $i \in I$ define the random variable Z_i to be the indicator of $i \in A$. Define $Z = \sum\limits_{i\in I} Z_i$, and note that $\Delta_A(f,g) = \frac{1}{|A|}Z$. Noting that $\mathbb{E}[Z] = \frac{k|I|}{d} = Rk$, by Theorem 2.1 we get that $\Pr\left[Z \geqslant 1.1Rk\right] \leqslant 2^{-\Omega(Rk)}$; also, by another application of Theorem 2.1 we get that $|A| \geqslant 0.9k$ with probability $1 - 2^{-\Omega(k)}$. It follows that except with probability $2^{-\Omega(Rk)}$ we have that $\Delta_A(f,g) \leqslant \frac{1.1Rk}{0.9k} \leqslant 2R$. The probability that $|A| = k$ is $\Omega(1/\sqrt{k})$, and conditioned on that A is distributed as $A \subseteq_k [d]$, hence we get that the probability in the first item is at most $O\left(\sqrt{k}2^{-\Omega(Rk)}\right) = 2^{-\Omega(Rk)}$. □

The second claim asserts that if two functions f and g are relatively far, then there are not many k-sets A on which they roughly agree. More precisely:

Claim 4.5. *Suppose that $F\colon X(k) \to \{0,1\}^k$ is an assignment, that $D \in X(d)$ is a face and that $f, g\colon D \to \{0,1\}$ are functions such that $\Delta(f,g) > C\frac{t}{k}$, where $C \geqslant 6$. Then*

$$\Pr_{A\subseteq_k D}\left[A \in Agr_t(f,F) \cap Agr_t(g,F)\right] \leqslant 2^{-\Omega(Ct)}.$$

Proof. By Claim 4.4, sampling $A \subseteq_k D$ we get that $\Delta_A(f,g) \geqslant \frac{Ct}{2k}$ with probability $1 - 2^{-\Omega(Ck)}$; we claim that such A cannot both be in $\text{Agr}_t(f,F)$ and in $\text{Agr}_t(g,F)$. Indeed, otherwise we would get that

$$\frac{Ct}{2k} \leqslant \Delta(f|_A, g|_A) \leqslant \Delta(F[A], f|_A) + \Delta(F[A], g|_A) \leqslant 2\frac{t}{k},$$

and contradiction. □

4.3 Proof of Lemma 4.2: Reduction from agreement to list agreement testing

4.3.1 Localizing to a Johnson

The first step of the proof is to localize to a random d-face $D \sim \mu_d$, and show that with probability close to 1, the assignment F passes the direct product test inside T with noticeable probability. More precisely:

Lemma 4.6. *If (k,s)-Agreement-Test 1 on F passes with probability δ, then*

$$\Pr_{D \sim X(d)}\left[\textit{the } (k,s) - \textit{direct product test passes with probability } \geqslant \delta^2/16 \textit{ inside } D\right] \geqslant 1 - o(1).$$

Proof. Let $\mathcal{D}_1$ be the distribution on (A, A', I) induced by Agreement-Test 1, and consider the following distribution $\mathcal{D}_2$ over (A, A', I):

1. Sample $B \sim \mu_{\sqrt{d}}$.
2. Sample $I \subseteq D$ of size s uniformly.
3. Sample $I \subseteq A, A' \subseteq B$ of size k uniformly.

Note that conditioned on $|A \cap A'| = s$, the distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ are identical. Thus, as the probability of this event is $1 - O(k^2/\sqrt{d}) = 1 - o(1)$ in both distributions, it follows that the statistical distance between $\mathcal{D}_1$ and $\mathcal{D}_2$ is $o(1)$. Therefore,

$$\Pr_{(A,A',I) \sim D_2}\left[F[A]|_I = F[A']|_I\right] \geqslant \delta - o(1).$$

Denote by $\mathcal{D}_2(B)$ the distribution on (A, A', I) conditioned on sampling B, and by p_B the probability that $F[A]|_I = F[A']|_I$ if B was chosen. By an averaging argument, with probability at least $\frac{\delta}{4}$ over the choice $B \sim \mu_{\sqrt{d}}$ we have that $p_B \geqslant \frac{\delta}{2}$; we call such B good, and denote the set of good B's by $\mathcal{B}$.

By Lemma 2.6 we get that

$$\Pr_{D \sim \mu_d}\left[\Pr_{B \subseteq_{\sqrt{d}} D}[B \in \mathcal{B}] \geqslant \frac{\delta}{8}\right] \geqslant 1 - O\left(\frac{1}{\sqrt{d}} + \gamma\right) = 1 - o(1).$$

Fix a d-face D satisfying the above event. Thus, picking $B \subset_{\sqrt{d}} D$ and $(A, A', I) \sim \mathcal{D}_2(B)$ passes the direct product test with probability at least $\frac{\delta^2}{8}$. Let this distribution be $\mathcal{D}_2(D)$. As before, letting the distribution $\mathcal{D}_1(A)$ be the distribution over $(A, A', I) \sim D_1$ conditioned on sampling D, the statistical distance between $\mathcal{D}_1(D)$ and $\mathcal{D}_2(D)$ is $o(1)$. Therefore we get that,

$$\Pr_{D \sim \mu_d}\left[\text{the } (k,s) - \text{direct product test passes w.p. } \geqslant \delta^2/8 - o(1) \text{ inside } D\right] \geqslant 1 - o(1),$$

which completes the proof. □

We refer to a d-face $D \in X(d)$ for which the event in Lemma 4.6 holds as good, and thus conclude that $1 - o(1)$ fraction of the d-faces are good. Note that the above argument would also work for $d/2$-faces, and thus we similarly define the notion of good $d/2$-faces.

4.3.2 Getting a list on each good Johnson and generating a gap

Fix a good d-face D, and consider the assignment F when restricted to k-sets inside k. For notational convenience, we denote this restricted assignment by F_D. Thus, the event in Lemma 4.6 translates to saying that the direct product tester over the Johnson scheme passes inside D with noticeable probability. Thus, using direct product testing results over the Johnson scheme, we may "explain" this consistency via correlations of F_D with true direct product functions. Towards this end, we use a result due to [DG08] (see also [IKW09], who state a version that is more convenient for our purposes).

Theorem 4.7. *Suppose that F_D passes the $(k, \sqrt{k})$ direct product test in D with probability ε. Then there is a function $g : [d] \to \{0, 1\}$ such that*

$$\Pr_{A \subseteq_k D}\left[\Delta(g|_A, F_D[A]) \leqslant 1/k^{\Omega(1)}\right] \geqslant \Omega(\varepsilon^6).$$

Theorem 4.7 by itself is not enough for us, and we need an idea that is often useful in conjunction with such results: list decoding. We wish to consider all direct product functions that are correlated with F_D and have these as the lists. Alas, there is a technical issue: the number of direct product functions that are correlated with F_D need not be bounded in terms of ε, the probability that the test passes. To remedy this issue we require the notion of η-covers, defined below.

Definition 4.8. *Let $\mathcal{F} \subseteq \mathcal{G}$ be two families of functions from $[d]$ to $\{0, 1\}$. We say that $\mathcal{F}$ is an η-cover for $\mathcal{G}$ if for any $g \in \mathcal{G}$ there exists $f \in \mathcal{F}$ such that $\Delta(f, g) \leqslant \eta$.*

We are now ready to present a procedure that, given a good d-face D, generates a short list of functions that "explain" most of the probability that F_D passes the direct product test inside D, and which is also short. The procedure takes as input a restriction of the assignment F to a face D, which below we denote by G, and finds one by one direct product functions that are correlated with G, following by randomizing G at appropriate places.

Algorithm 1. the short list algorithm.

Input: $G : \binom{[d]}{k} \to \{0, 1\}^k$, $\delta > 0$, $r \in \mathbb{N}$, $\eta \in (0, 1)$.

Output: List of functions $\{f_1, \ldots, f_m\}$ from $[d] \to \{0, 1\}$.

Operation:

1. Set $t = k^{-\Omega(1)}$, $\delta_0 = \Theta(\delta^6)$, $\widetilde{G}_0 = G$, and initialize $L_1, I_1 = \emptyset$.
2. For $i \in \{0, \ldots, \lfloor 1/\delta^{80} \rfloor\}$:
    - If there exists f with $\text{agr}_t(f, \widetilde{G}_i) > \delta_i$, add i to I_1 and f_i to L_1.
    - Obtain $\widetilde{G}_{i+1}$ by randomizing $\widetilde{G}_i$ on k-sets $A \in \text{Agr}_t(f, \widetilde{F}_i)$.
    - $\delta_{i+1} = \delta_i - \delta^{100}$.
3. Create lists I_2, L_2 as follows: for all $i \in I_1$, add i to I_2 and f_i to L_2 iff $i \geqslant r$.
4. Construct a graph G whose vertices are L_2, and $f, g \in L_2$ are adjacent if $\Delta(f, g) < \eta$. Take a maximal independent set in G and add the corresponding functions to L_3.

5. Output L_3.

The following lemma summarizes some of the basic properties of the short list algorithm.

Lemma 4.9. *When ran on $G = F_D$ for a good d-face D with parameter $\Theta(\delta^2)$ in place of δ, setting $\delta' = \Theta(\delta^{12})$, with probability $1 - o(1)$ Algorithm 1 outputs a list $L = \{(i, f_i)\}_{i \in I}$ with $I \subset \{0, \dots, 1/\delta'^{80}\}$ such that,*

1. *$0 \neq |I_1| \leqslant \frac{2}{\delta'}$.*
2. *For all $i \in I_1$, $agr_t(f_i, G) > \delta' - i\delta'^{100} - o(1)$.*
3. *If $i \notin I_1$ then for all g, $agr_t(g, \widetilde{G}_i) < \delta_i$.*
4. *For all $i \in \lfloor 1/\delta'^{80} \rfloor$ and $B \subseteq_{d/2} A$, if $g \colon B \to \{0,1\}$ is a function such that $\min_{j \in I_1, j \geqslant r} \Delta(g, f_j|_B) > 202 \log(1/\delta') t/k$ and $agr_t(g, \widetilde{F}_{i+1}|_B) < \theta$, then $agr_t(g, F|_B) < \theta + \delta'^{200}$.*

Proof. First note that by Theorem 4.7 we get that there is at least one function with $\mathrm{agr}_t(f) \geqslant \delta'$, therefore the list is non-empty. Let us start by proving the upper bound on the size.

Proof of (1): At the i^{th} iteration we add a function to the list only if $\mathrm{agr}_t(f_i, \tilde{G}_i) > \delta_i$ which is always at least $\delta' - \delta'^{20}$. Let $\mathcal{R} \subseteq \binom{D}{k}$ be the k-sets that have been randomized in the algorithm so far, so $|\mathcal{R}| \geqslant (\delta - \delta^{20})\binom{d}{k}$. Using the Chernoff bound we get that every function $g \colon D \to \{0,1\}$ satisfies:

$$\Pr\left[\frac{|\mathrm{Agr}_t(g) \cap \mathcal{R}|}{|\mathcal{R}|} > \frac{2k^t}{2^k}\right] \leqslant \exp\left(-\frac{k^t}{2^k}\delta\binom{d}{k}\right) \leqslant \exp(-(d/4)^k).$$

Therefore by a union bound we get that with probability $1 - o(1)$, for all functions on D the above holds, and we condition on this event. Hence, the contribution of $\mathcal{R}$ to the agreement of function found in later steps in the procedure is always at most $o(1)$. Thus, each newly found function in the process increases the measure of $\mathcal{R}$ by at least $\delta' - \delta'^{20} - o(1) \geqslant \delta'/2$. Therefore, with probability $1 - o(1)$ the process terminates after at most $2/\delta'$ steps, which is thus also an upper bound on the list size I_1.

Proof of (2): If we inserted f into the list at step i, then $\mathrm{agr}_t(f, \tilde{G}_i) \geqslant \delta' - i\delta'^{100}$. As we have already argued, with probability $1 - o(1)$ at most $o(1)$ of this agreement comes from k-sets in which $\tilde{G}_i$ was randomized, and it follows that $\mathrm{agr}_t(f, F_D) \geqslant \delta' - i\delta'^{100} - o(1)$.

Proof of (3): If $i \notin I_1$ then the process terminated before step i, meaning that the assignment at that time no longer was δ_i-correlated with any direct product function.

Proof of (4): Denote by $\mathcal{R}_i$ the collection of all k-sets in which the assignment has been randomized in steps prior to the $i + 1$th iteration, and consider $\widetilde{G}_{i+1}$. By Claim 4.5

$$\Pr_{A \subseteq_k B}\left[A \in \mathrm{Agr}_t(g, F|_B) \cap \mathrm{Agr}_t(f_j|_B, F|_s)\right] \leqslant \delta'^{202},$$

and so

$$\Pr_{A \subseteq_k B}\left[A \in \mathrm{Agr}_t(g, F|_B) \cap \mathcal{R}_i\right] \leqslant \delta'^{202} + o(1) \leqslant \delta'^{200}.$$

It follows from the that

$$\begin{aligned}\Pr_{A\subseteq_k B}\left[A \in \mathrm{Agr}_t(g, F|_B)\right] &= \Pr_{A\subseteq_k D}\left[A \in \mathrm{Agr}_t(g, F|_B) \cap \mathcal{R}_i\right] + \Pr_{A\subseteq_k D}\left[A \in \mathrm{Agr}_t(g, F|_B) \cap \overline{\mathcal{R}_i}\right]\\ &\leqslant \delta'^{200} + \Pr_{A\subseteq_k D}\left[A \in \mathrm{Agr}_t(g, \tilde{F}_{i+1}|_B) \cap \overline{\mathcal{R}_i}\right]\\ &\leqslant \delta'^{200} + \Pr_{A\subseteq_k D}\left[A \in \mathrm{Agr}_t(g, \tilde{F}_{i+1}|_B)\right],\end{aligned}$$

which is at most $\delta'^{200} + \theta$. □

We will now consider the run of the short list algorithm on a d-face with various options for parameters, and its relationship with direct product functions on $d/2$ sub-faces. We will especially care about the relationship between the functions in the list of the d-face $D \in X(d)$, and direct product functions on its $d/2$-faces that have large correlation with the assignment F. In a sense, we will want to show that these are "the same functions"; ultimately, this is where the local consistency of the lists comes from.

Towards this end, we will run the algorithm above for D faces, and denote the outputted list by $L[D]$, For $d/2$ sub-faces of D, we will let $L[B]$ be an η-cover for functions that have sufficient agreement with $F|_B$. The following lemma summarizes the properties of such runs of the short list algorithm:

Lemma 4.10. *Let $\varepsilon, \delta > 0$, let k be sufficiently large and let $d \geqslant \mathrm{poly}(k)\exp(\mathrm{poly}(1/\delta))$. Suppose that F_D passes the $(k, \sqrt{k})$ direct product tester inside D with probability at least δ. Then choosing $r, i \sim \lfloor 1/\delta^{80} \rfloor$ uniformly and running Algorithm 1 with parameters r and $\eta' = \delta^{-100i}\eta$ on D and on all $d/2$ sub-faces, with probability $1 - O(\delta^{68})$ the algorithm outputs a list $L[D]$ such that:*

1. ***Non-empty, short list:*** *$0 \neq |L[D]| \leqslant 1/\delta'$, where $\delta' = \Theta(\delta^6)$.*
2. ***Significant correlation:*** *For all $f \in L$, $agr_t(f, F_D) \geqslant \delta_r := \delta' - r\delta'^{100}$, where $t = k^{-\Omega(1)}$.*
3. ***Large distance in the list:*** *$\Delta(L[D]) > \delta^{-100}\eta'$.*
4. ***Downwards consistent:*** *$\Pr_{B\subseteq_{d/2} D}[\forall f \in L[D], \exists g \in L[B] \text{ with } \Delta(f|_B, g) \leqslant \eta'] \geqslant 1 - o(1)$. In words, for each function in the list of D, projecting it onto a random $B \subseteq_{d/2} D$ yields a function which is very close to a function in the list of B.*
5. ***Upwards consistent:*** *$\Pr_{B\subseteq_{d/2} D}[\forall g \in L[B], \exists f \in L[D] \text{ with } \Delta(g, f|_B) \leqslant 2\eta'] \geqslant 1 - o(1)$. In words, choosing a random $B \subseteq D$, every function in the list $L[B]$ is close to a projection of some function from the list $L[D]$.*

For each $B \subseteq_{d/2} D$, $L[B]$ is an η'-cover for functions on B with $agr_t(g, F|_B) > \delta_r - \delta^{200}$.

The first four items in Lemma 4.10 are not too hard to establish; the fifth item however requires more care, and this is where we are going to utilize results from random sub-instances of max-k-CSPs. In particular, we require the following lemma which follows from results in [AdlVKK02] (and more precisely, from Theorem 2.3).

Lemma 4.11. *For all $\tau, \zeta \in (0, 1)$, $d \geqslant \mathrm{poly}(k/\tau)\exp(1/\zeta^2)$, and all functions $G : \binom{[d]}{k} \to \{0, 1\}^k$ that satisfy $agr_t(g, G) \leqslant \alpha$ for all $g \colon [d] \to \{0, 1\}$, the following holds:*

$$\Pr_{B\subseteq_{d/2}[d]}[\max_g agr_t(g|_B, G|_B) < \alpha + \gamma] \geqslant 1 - \tau.$$

Proof. Consider the following Max-k-CSP $\Psi = ([d], \mathcal{F})$. The constraints in $\mathcal{F}$ are as follows: for every k-subset I we have the constraint $f_I : \{0,1\}^k \to \{0,1\}$ defined as,

$$f_I(x) = \begin{cases} 1, & \text{if } \Delta(x, G[A]) \leqslant t, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, the value of Ψ is $\mathrm{val}(\Psi) = \max_g \mathrm{agr}_t(g, G)$. Applying Theorem 2.3 we get that with probability $1-\tau$ over the choice of $B \subseteq_{d/2} [d]$, $\mathrm{val}(\Psi|_B) \leqslant \mathrm{val}(\Psi) + \zeta$, which is at most $\alpha + \zeta$. Noting that $\mathrm{val}(\Psi|_B) = \max_g \mathrm{agr}_t(g|_B, G|_B)$ finishes the proof. $\square$

We are now ready to prove Lemma 4.10.

Proof of Lemma 4.10. The proofs of (1) and (2) are immediate from point (1) and (2) of Lemma 4.9.

Proof of (3): Consider the lists produced by the algorithm and consider the pairwise distances $\Delta(f_i, f_j)$ for $f_i, f_j \in L[D]$. Since $|L_2| \leqslant 1/\delta'$ there are at most $1/\delta'^2$ different pairwise distances, therefore with probability $1-O(\delta^{68})$ over $i \in \{0, \ldots, 1/\delta^{80}\}$ we have that for all $i \neq j$ either $\Delta(f_i, f_j) < \eta'$ or $> \delta^{-100}\eta'$. In that case, a maximal independent set L_3 obtained in the foruth step of the short list algorithm satisfies that for all $f_i, f_j \in L_3$, $\Delta(f_i, f_j) > \delta^{-100}\eta'$.

Proof of (4): By Lemma 2.6, we get that for each $f \in L[D]$, with probability $1 - o(1)$ over the choice of $B \subseteq_{d/2} D$ we have that $\mathrm{agr}_t(f|_B, F|_B) \geqslant \delta_r - o(1)$. Thus, by the upper bound on the size of $L[D]$ and the union bound we get

$$\Pr_{B \subseteq_{d/2} D} \left[\forall f \in L[D], \mathrm{agr}_t(f|_a) \geqslant \delta_r - o(1)\right] \geqslant 1 - o(1).$$

By the property of η'-covers we conclude that

$$\Pr_{B \subseteq_{d/2} D}[\forall f \in L[D], \exists g \in L[B] \text{ with } \Delta(f|_B, g) \leqslant \eta'] \geqslant 1 - o(1).$$

Proof of (5): Note that the list L_2 has size at most $1/\delta'$, hence with probability at least $1-O(\delta^{74})$ over the choice of r, we get that $r+1 \notin I_1$. This means that we have a gap: $\forall h, \mathrm{agr}_t(h, \widetilde{F}_{r+1}) < \delta_{r+1}$. Condition on r being chosen so that this holds; by Lemma 4.11 we get that

$$\Pr_{B \subseteq_{d/2} D} \left[\max_h \mathrm{agr}_t(h|_B, \widetilde{F}_{r+1}|_B) < \delta_{r+1} + \delta^{200}\right] \geqslant 1 - o(1).$$

Fix a B where the above holds, let $L[B]$ be an η'-cover as in the statement of the lemma, and take $g \in L[B]$. Assume for contradiction that $\Delta(f|_B, g) > \eta' + \Omega(t/k)$ for all $f \in L_3$. By the maximality of the independent set L_3, we get that for all $f \in L_2 \setminus L_3$, there exists $f' \in L_3$ such that $\Delta(f, f') < \eta'$. Therefore if g is $\Omega(t/k) + \eta$-far from all $f \in L_3$, then it is $\Omega(t/k)$-far from all $f' \in L_2$ and in particular from all $f_j \in L_1$ for $j \geqslant r, j \in I_1$. Since $\mathrm{agr}_t(g, \widetilde{F}_{r+1}|_B) < \delta_{r+1} + \delta^{200}$, we may apply the fourth item in Lemma 4.9 to get that $\mathrm{agr}_t(g, F|_B) < \delta_{r+1} + 2\delta^{200} < \delta_r - \delta^{200}$, which is a contradiction to g being in $L[B]$. $\square$

4.3.3 Consistency of the local lists

In this section, we finish the proof of Lemma 4.2. Fix parameters as therein, let $\mathcal{D}$ be the set of good faces (namely, faces in which the $(k, \sqrt{k})$ agreement test passes with probability at least $\delta' = \delta^2/16$), and recall that by Lemma 4.6 we have that $\mu_d(\mathcal{D}) \geqslant 1 - o(1)$.

We sample r and i integers between 1 and $\lceil 1/\delta^{80} \rceil$ uniformly, set $\eta' = \delta^{-100i}\eta$ and run the short list algorithm on each $D \in \mathcal{D}$ with the parameters r and i. For each D, with probability $1 - O(\delta^{68})$ we get a list $L[D]$ as in Lemma 4.10. It follows by linearity of expectation and an averaging argument that we may choose r and i such that we get lists $L[D]$ for at least $1 - O(\delta^{68})$ of $D \in \mathcal{D}$ such that $L[D]$ satisfies the conditions of Lemma 4.10, and we fix such r and i henceforth. Below, we refer to a good D that additionally has a list $L[D]$ satisfying the conditions of Lemma 4.10 as very good, and we note that the probability that D is very good is at least $1 - O(\delta^{68}) - o(1) = 1 - O(\delta^{68})$. For each $B \in X(d/2)$, we fix $L[B]$ to be an η' cover of the collection of functions $g \colon B \to \{0, 1\}$ such that $\mathrm{agr}_t(g, F|_B) \geqslant \delta_r = \delta' - r\delta'^{100}$.

The first and the second item in the statement of Lemma 4.2 clearly hold by Lemma 4.10, and in the rest of the argument we argue about the third item. Towards this end, consider a generation of queries for the list agreement problem. Namely, sample $B \sim \mu_{d/2}$ and independently sample $D, D' \supset_d B$ from. We say a triple (D, B, D') is good if:

1. The d-faces D and D' are very good.
2. It holds that $\Delta(L[D]|_B), \Delta(L[D']|_B) > \frac{1}{2}\delta^{-100}\eta'$.
3. For all $f \in L[D]$, there exists $g \in L[B]$ with $\Delta(f|_B, g) < \eta$, and for all $g \in L[B]$ there exists $f \in L[D]D$ with $\Delta(g, f|_B) < 2\eta$. The same holds when D is replaced by D'.

Note that since marginally, each one of D and D' is distributed according to μ_d, we get that the first item holds with probability $1 - O(\delta^{68})$. Note that the marginal distribution of (B, D) is the same as sampling $D \sim \mu_d$, and then $B \subseteq_{d/2} D$. Thus, if the first item holds, then $\Delta(L[D]) \geqslant \eta'$, hence by Claim 4.4 we get that the second item holds with probability $1 - o(1)$. Lastly, if the first item holds, then by Lemma 4.10 we get that the third item holds with probability $1 - o(1)$. Overall by the union bound, we get that all of the events above holds together with probability at least $1 - O(\delta^{68})$.

To finish the proof, we argue that if (D, B, D') is good, then the list agreement test passes on it. For that, we show that for each $f \in L[D]$ there exists a unique $f' \in L[D']$ s.t. $\Delta_B(f, f') \leqslant 3\eta'$ and vice versa. We show the argument only in one of the directions, and the other direction is identical. Take $f \in L[D]$ and consider $f|_B$; by the η'-cover property we can find a $g \in L[B]$ with $\Delta(f|_B, g) \leqslant \eta'$. By the third property above, for g we may find $f' \in L[D']$ with $\Delta(g, f'|_B) \leqslant 2\eta'$, so by the triangle inequality $\Delta_B(f, f') \leqslant 3\eta$. Next, we show that uniqueness of f'. For any $f'' \in L[D'] \setminus \{f'\}$, by the second property above $\Delta(f''|_B, f'|_B) \geqslant \frac{1}{2}\delta^{-100}\eta'$, so

$$\Delta(f''|_B, f|_B) \geqslant \Delta(f''|_B, f'|_B) - \Delta(f'_B, f|_B) \geqslant \frac{1}{2}\delta^{-100}\eta' - 3\eta' \geqslant 100\eta'.$$

4.4 List agreement testing using UG coboundary expansion: proof of Lemma 4.3

The goal of this section is to prove Lemma 4.3. Throughout this section, we fix lists $\{L[D]\}_{D \in X(d)}$ satisfying the premise of Lemma 4.2. We refer to a d-face $D \in X(d)$ for which the properties in Lemma 4.2 are satisfied as good, and note that the measure of the set of good d-faces under μ_d is at least $1 - O(\tau)$. Our first goal is to define a locally consistent instance of Unique-Games on which we can apply coboundary expansion. At the moment though we have assignments only to the d-faces, and our UG coboundary expansion only holds for much lower levels. Thus, we will first show how to project our list assignments to lower levels.

4.4.1 Global consistency of the list sizes

We begin with establishing several basic claims that will be useful in the projection process. The following claim asserts that almost all of the lists $L[D]$ have the same size. More precisely,

Claim 4.12. *There exists $\ell \leqslant \mathsf{poly}(1/\delta)$ such that $\Pr_{D\sim\mu_d}\left[|L(D)| \neq \ell\right] \leqslant 10\tau$.*

Proof. Suppose this is not the case. Then the set of d-faces $X(d)$ can be broken into two disjoint parts, $P_1 \cup P_2$ such that $\mu_d(P_1), \mu_d(P_2) \geqslant 10\tau$ and for every $D \in P_1$, $D' \in P_2$ we have that $|L[D]| \neq |L[D']|$. Note that in that case, there can never be a 1-to-1 correspondence between $L[D]$ and $L[D']$, and hence we conclude that the list agreement test fails whenever it picks $D \in P_1$ and $D' \in P_2$. Suppose without loss of generality that $\mu_d(P_1) \leqslant 1/2$

On the other hand, considering the graph G on $X(d)$ generated by the list agreement test, which equivalently can be stated as pick $D \sim \mu_d$, $B \subseteq_{d/2} D$ and then $D' \supseteq_d B$ according to $D' \sim \mu_d$. By Lemma 2.10 second eigenvalue of G is at most $\frac{1}{2} + O(d^2\gamma)$. It follows from Cheeger's inequality that the edge expansion of P_1 is at least $\frac{1}{4} - O(d^2\gamma) \geqslant 1/8$. It follows that a randomly sampled edge goes from P_1 to P_2 with probability at least $\mu_d(P_1)/8 \geqslant 10\tau/8$. This contradicts the fact that the probability that the list agreement test fails is at most τ. ◻

We pick ℓ to be the list size parameter from Claim 4.12. In the next claim we prove that the fact that the list $L[D]$ typically has a large distance implies that its projection onto a sub-face has the same size.

Claim 4.13. *For $t \geqslant 102\frac{\log(\ell/\tau)}{\delta^{-100}\eta}$ it holds that $\Pr_{\substack{D\sim\mu_d\\ B\subset_t D}}\left[|L(D)|_B| \neq \ell\right] \leqslant O(\tau)$.*

Proof. We prove that for each good D, conditioning on D, the above probability is at most $O(\tau)$, and the claim trivially follows. Fix a good D and consider distinct $f, g \in L[D]$. Then $\Delta(f, g) \geqslant \delta^{-100}\eta$, and the probability over the choice of B that $f|_B = g|_B$ is at most $(1-\delta^{-100}\eta)^t \leqslant e^{-\delta^{-100}\eta t} \leqslant \frac{\tau^{50}}{\ell^{50}}$. By a union bound we have that it follows that the probability there are distinct $f, g \in L[D]$ such that $f|_B = g|_B$ is at most $\ell^2\frac{\tau^{50}}{\ell^{50}} \leqslant \tau$, completing the proof. ◻

4.4.2 Majority decoding

Next, we show that for t that is not too large, for a typical t-face B, almost all of the d-faces D have the same projection of $L[D]$ onto B. More precisely:

Claim 4.14. *For $t \leqslant \frac{\tau}{\eta}$, with probability at least $1 - O(\sqrt{\tau})$ over the choice of $B \sim \mu_t$ it holds that*

$$\Pr_{D,D'\supseteq_d B}\left[L[D]|_B = L[D']|_B\right] \geqslant 1 - O(\sqrt{\tau}).$$

Proof. Consider the following sampling procedure: sample $B \sim \mu_t$, then $C \supseteq_{d/2} B$ and then sampling $D, D' \supseteq_d C$ independently. We first claim that with probability $1-\tau$ it holds that $L[D]|_C \neq_{<\eta} L[D']|_B$. Indeed, first note that the marginal distribution of D and D' is μ_d, and the marginal distribution of (D, C, D') is according to the list agreement test. Hence, with probability $1 - O(\tau)$ it holds that D, D' are good and the list agreement test passes on (D, C, D'). In that case, we may find a matching $\pi\colon L[D] \to L[D']$ such that $\Delta(\pi(f)|_C, f|_C) \leqslant \eta$ for all $f \in L[D]$. Noting that B is a random subset of C of size t, we get that $\Delta(\pi(f)|_B, f_B) = 0$ with probability at least $1 - t\eta \geqslant 1 - \tau$.

For each $B \in X(t)$, let

$$p_B = \Pr_{\substack{C \supseteq_{d/2} B \\ D, D' \supseteq C}} [L[D]|_B = L[D']|_B].$$

Rephrasing the conclusion of the previous discussion, we have that $\mathbb{E}_{B \sim \mu_t}\left[p_B\right] \geqslant 1 - \tau$. By an averaging argument, defining $\mathcal{B} = \{B \in X(t) \mid p_B \geqslant 1 - O(\sqrt{\tau})\}$ we have that $\mu_t(\mathcal{B}) \geqslant 1 - O(\sqrt{\tau})$. Fix $B \in \mathcal{B}$; we argue that there exists a list of functions $L[B]$ on B such that

$$\Pr_{D \supseteq_d B}\left[L[D]|_B = L[B]\right] \geqslant 1 - O(\sqrt{\tau}).$$

Indeed, otherwise we may partition the set of D's containing B into P_1 and P_2 of relative measure at least $c'\sqrt{\tau}$ so that $L[D]|_B \neq L[D']|_B$ for all $D \in P_1$, $D' \in P_2$, where c' is an absolute constant to be determined. Consider the G on d-faces containing B, whose edges are sampled by first picking $C \supseteq_{d/2} B$ and then $D, D' \supseteq_d C$ independently; by Lemma 2.10 this graph has second eigenvalue at most $1/2 + \mathrm{poly}(d)\gamma$, and thus the fraction of edges inside the graph that go from P_1 to P_2 is at least $(1/4 - O(d^2\gamma))c'\sqrt{\tau} \geqslant c'\sqrt{\tau}/8$. On any such edge we have that $L[D]|_B \neq L[D']|_B$, and it follows that $p_B \leqslant 1 - c'\sqrt{\tau}/8$, and contradiction provided that c' is sufficiently large. $\square$

With Claim 4.14 in hand, one may naturally project the lists that we have on d faces to t-faces in a way that "preserves their essence". More precisely, take a parameter t in the range

$$102\frac{\delta^{100}\log(\ell/\tau)}{\eta} \leqslant t \leqslant \frac{\tau}{\eta}. \tag{5}$$

For each $B \in X(t)$ define a list for B using weighted majority

$$L[B] := \mathrm{Maj}_{D \supset_d B}\left(L[D]|_B\right),$$

where the weight of D is $\Pr_{D' \supseteq_d B}\left[D' = D\right]$

Claim 4.15. *For t in the range as in (5), we have that:*

1. $\Pr_{B \sim \mu_t}\left[|L[B]| = \ell\right] \geqslant 1 - O(\sqrt{\tau})$.
2. *Choosing $B \sim \mu_t$, with probability at least $1 - O(\sqrt{\tau})$ it holds that $\Pr_{D \supseteq_d B}\left[L[D]|_B = L[B]\right] \geqslant 1 - O(\sqrt{\tau})$.*

Proof. The second item holds for every B satisfying the conclusion of Claim 4.14, and hence it follows. The first item follows from the second item when it is combined with Claim 4.13 using the union bound. $\square$

4.4.3 Designing the Unique Games instance and proving triangle consistency

Fix a t as in (5). Our next goal is to define a Unique-Games instance on the weighted graph G whose vertices are $X(t)$ and whose edge correspond to $2t$-faces: the edges are (u, v) where $u \cup v \in X(2t)$, and its weight is proportional to $\mu_{2t}(u \cup v)$. We remark that strictly speaking, we only define a partial Unique-Games instance on the subset of t-faces B where $|L[B]| = \ell$. By Claim 4.15 these t-faces constitute almost all of $X(t)$, and we encourage the reader to ignore this point.[12]

[12] Alternatively, one may think of picking an arbitrary list of size ℓ for every $B \in X(t)$ where $|L[B]| \neq \ell$.

List ordering, permutations and concatenation. Towards this end, we fix an ordering for each one of the lists constructed thus far (both for d-faces as well as for t-faces). Thus, we will think of the list of $B \in X(t)$ as $L[B] = (L_1[B], \ldots, L_\ell[B])$. For a permutation $\pi \in S_\ell$, we define $\pi(L[B]) = (L_{\pi(1)}[B], \ldots, L_{\pi(\ell)}[B])$. For $u, v \in X(t)$ such that $u \cup v \in X(2t)$ and $\pi \in S_\ell$, we denote

$$L[u] \circ \pi(L[v]) = \left(L_1[u] \circ L_{\pi(1)}[v], \ldots, L_\ell[u] \circ L_{\pi(\ell)}[v]\right),$$

and think of it as a list of assignments to $u \cup v$.

Defining the constraints of the Unique Games instance Consider the set of $2t$-faces $W \in X(2t)$, and note that one has the analog of Claim 4.15 for these as well, and thus we fix lists $L[W]$ satisfying Claim 4.15 for $2t$-faces as well. Let $\mathcal{W} \subseteq X(2t)$ be the collection of all $2t$-faces for the items in Claim 4.15 hold.

We now define a Unique-Games instance Ψ over G by describing the constraints on the graph G. For each edge (u, v), we put a constraint as follows. If $u \cup v \notin \mathcal{W}$, we put an arbitrary constraint. Else, we put a constraint between u and v if $L[u \cup v]|_u = L[u]$ and $L[u \cup v]|_v = L[v]$. Note that in that case, there is a natural 1-to-1 correspondence between $L[u]$, $L[u \cup v]$ and $L[v]$, and we fix it to be the constraint between $L[u]$ and $L[v]$. Stated otherwise, the constraint on (u, v) is the unique permutation $\pi = \pi(u, v) \in S_\ell$ such that $L[u \cup v] = L[u] \circ \pi(L[v])$ (when both sides are thought of as assignments to $u \cup v$). We think of edges as being directed, and note that then $\pi(u, v) = \pi(v, u)^{-1}$.

The following claim asserts that Ψ is $(1 - O(\sqrt{\tau}))$ strongly triangle consistent.

Claim 4.16. $\Pr_{\substack{Z \sim \mu_{3t} \\ Z = u \cup v \cup w}} [(u, v, w) \textit{ is strongly consistent in } \Psi] \geqslant 1 - O(\sqrt{\tau})$.

Proof. We use Claim 4.15 for $3t$-faces, and denote the set of $3t$-faces for which the items there hold by $\mathcal{Z}$. Thus, $\mu_{3t}(\mathcal{Z}) \geqslant 1 - O(\sqrt{\tau})$. Note that sampling $D \sim \mu_d$, then $Z \subseteq_{3t} D$ and then writing $Z = u \cup v \cup w$, with probability $1 - O(\sqrt{\tau})$ we have that there is a 1-to-1 correspondence between the list of each one of u, v, w, the lists of $u \cup v$, $v \cup w$, $u \cup w$, the list of Z and the list of D. We get a 1-to-1 correspondence between the list of u and the list of $v \cup w$, and we denote it by $\pi(u, v \cup w)$, and all of these correspondences are consistent. In particular, we get that $\pi(u, w) = \pi(u, v) \circ \pi(v, w)$ (as both can be thought of as re-alignments of the list of w to concatenate with the list of u so that they agree with $L[u \cup w]$), and hence $\pi(w, u)\pi(u, v)\pi(v, w) = \text{id}$. This proves triangle consistency, and strong triangle consistency readily follows. $\square$

4.4.4 Applying UG coboundary expansion

By Claim 4.16 we get that Ψ is $(1 - O(\sqrt{\tau}))$ strongly triangle consistent, and applying the Unique-Games coboundary expansion we get that there is $g\colon X(t) \to S_m$ such that

$$\Pr_{u \cup v \in X(2t)} \left[\pi(u, v) = g(u) g(v)^{-1}\right] \geqslant 1 - c. \tag{6}$$

We now pick an element from the list of each u. More precisely, define $h\colon X(t) \to [\ell]$ defined by $h(v) = g(v)(1)$. Note that if (u, v) is an edge such that the event in (6) holds, then

$$\pi(u, v)(h(v)) = \pi(u, v) g(v)(1) = g(u)(1) = h(u).$$

In other words, for each vertex u we picked an assignment from the list of u in a locally consistent way. We may thus define the assignment $R(u) = L[u]_{h(u)}$; our goal is to show that there is a global function on $X(1)$ that agrees with many of these selections. Towards this end, we first show that R passes the standard direct product test with probability close to 1.

Lemma 4.17. *We have that*

$$\Pr_{\substack{D\sim\mu_d\\ Q\subseteq_{t/2}D\\ Q\subseteq B,B'\subseteq_t D}}\left[R(B)|_Q = R(B')_Q\right] \geqslant 1 - O(\tau^{1/4}+c^{1/2}).$$

Proof. Sample $Z\sim\mu_{3t}$, and write $Z = u\cup v\cup w$, $Z = u\cup v'\cup w'$ independently. Note that by the strong triangle consistency, get that with probability at least $1-O(\sqrt{\tau}+c)$ we have that

$$L[Z] = L[u]\circ\pi(u,v)L[v]\circ\pi(u,w)L[w] = L[u]\circ\pi(u,v')L[v']\circ\pi(u,w')L[w']$$

and the edges $(u,v),(u,w),(u,v'),(u,w')$ are satisfied. In that case, we conclude that $R(u)\circ R(v)\circ R(w)$ and $R(u)\circ R(v')\circ R(w')$ correspond to the same function in the list of Z, and so we get that

$$\Pr_{\substack{Z\sim\mu_{3t}\\ Z=u\cup v\cup w=u\cup v'\cup w'}}\left[R(u)\circ R(v)\circ R(w) = R(u)\circ R(v')\circ R(w')\right] \geqslant 1-O(\sqrt{\tau}+c). \tag{7}$$

For $Z\in X(3t)$, we associate splittings as $Z = u\cup v\cup w$ points in the multi-slice

$$\binom{[3t]}{t,t,t} = \left\{x\in\{0,1,2\}^{3t} \mid \forall j\in\{0,1,2\}, \left|\{i\mid x_i = j\}\right| = t\right\}$$

by identifying u with the set of coordinates equal to 0, v with the set of coordinates equal to 1 and w with the set of coordinates equal to 2. We define $\tilde{R}_Z(x) = R(u)\circ R(v)\circ R(w)$. For each $j\in\{0,1,2\}$, consider the Markov chain T_j on $\binom{[3t]}{t,t,t}$ that from x moves to y where the set of coordinates that are 0 are kept, and the rest are randomized. Then (7) implies that

$$\Pr_{\substack{Z\sim\mu_{3t}\\ x\in\binom{[3t]}{t,t,t}, y\sim \mathrm{T}_0 x}}\left[\tilde{R}_Z(x) = \tilde{R}_Z(y)\right] \geqslant 1-O(\sqrt{\tau}+c)$$

and analogously we have the same statement for T_1 and T_2, hence by the union bound

$$\Pr_{\substack{Z\sim\mu_{3t}\\ x\in\binom{[3t]}{t,t,t}, y\sim \mathrm{T}_2\mathrm{T}_1\mathrm{T}_0 x}}\left[\tilde{R}_Z(x) = \tilde{R}_Z(y)\right] \geqslant 1-O(\sqrt{\tau}+c).$$

Therefore, for at least $1-O(\tau^{1/4}+\sqrt{c})$ of Z, we have that

$$\Pr_{x\in\binom{[3t]}{t,t,t}, y\sim \mathrm{T}_2\mathrm{T}_1\mathrm{T}_0 x}\left[\tilde{R}_Z(x) = \tilde{R}_Z(y)\right] \geqslant 1-O(\tau^{1/4}+\sqrt{c}),$$

and we call such Z decisive. Fix a decisive Z; the Markov chain $\mathrm{T}_2\mathrm{T}_1\mathrm{T}_0$ has second eigenvalue at most $1-\Omega(1)$, and thus from the above it follows that

$$\Pr_{x,y\in\binom{[3t]}{t,t,t}}\left[\tilde{R}_Z(x) = \tilde{R}_Z(y)\right] \geqslant 1-O(\tau^{1/4}+\sqrt{c}),$$

and we define $R(Z)$ to be the most popular value of $\tilde{R}_Z(x)$. Concluding, for decisive Z we get

$$\Pr_{Z=u\cup v\cup w}\left[R(z) = R(u)\circ R(v)\circ R(w)\right] \geqslant 1-O(\tau^{1/4}+c^{1/2}).$$

Fix a decisive Z, and consider the following direct product tester over Z: choose $Q\subseteq_{t/2} Z$, and then $Q\subseteq B,B'\subseteq_t Z$ such that $B\cap B' = Q$. With probability at least $1-O(\tau^{1/4}+c^{1/2})$ we get that $R(B)_Q = R(Z)|_Q = R(B')|_Q$. Noting that sampling $Z\sim\mu_{3t}$ and then generating Q,B,B' yields a distribution of (B,Q,B') that is $O(t^2/d) = o(1)$ close to the distribution of Q,B,B' in the direct product tester in the lemma, so the conclusion follows. □

4.4.5 Concluding the global structure

With Lemma 4.17 in hand, we apply Theorem 2.11 to get that there exists a global function $G : X(1) \to \{0,1\}$ such that

$$\Pr_{B \sim \mu_t}[G|_B = R(B)] \geqslant 1 - O(\tau^{1/4} + c^{1/2} + \gamma).$$

In the next lemma we analyze the agreement of G with our lists $L[D]$ for $D \in X(d)$, thereby completing the proof of Lemma 4.3.

Claim 4.18. $\Pr_{D \sim \mu_d}\left[\Delta(G|_D, L[D]) \leqslant \frac{100\log(2\ell)}{t}\right] \geqslant 1 - O(\tau^{1/4} + c^{1/2} + \gamma)$

Proof. Sample $D \sim \mu_d$ and then $B \subseteq_t D$. Then $L[B] = L[D]|_B$ with probability at least $1 - O(\tau)$, and $G|_B \in L[B]$ with probability $1 - O(\tau^{1/4} + c^{1/2} + \gamma)$, hence

$$\Pr_{B \subseteq_t D \sim \mu_d}\left[\Delta(G|_B, L[D]|_B) = 0\right] \geqslant 1 - O(\tau^{1/4} + c^{1/2} + \gamma).$$

We get that with probability at least $1-O(\tau^{1/4}+c^{1/2}+\gamma)$ over the choice of D, it holds that $\Delta(G|_B, L[D]|_B) = 0$ with probability at least $1/2$ over the choice of B, and we argue that event in question holds for each such D. To see that, first note that fixing $f \colon D \to \{0,1\}$ such that $\Delta(G|_D, f) \geqslant 100\log(\ell)/t$, it holds that $G|_B = f|_B$ with probability at most

$$\left(1 - \frac{100\log(2\ell)}{t}\right)^t \leqslant (2\ell)^{-100}.$$

Thus, if $\Delta(G|_D, L[D]) \geqslant 100\log(\ell)/t$, then by the union bound $\Delta(G|_B, L[D]|_B) = 0$ with probability at most $(2\ell)^{-99} < 1/2$. □

5 Proof of Theorem 1.10

In this section we outline the proof of Theorem 1.10, restated formally below:

Theorem 5.1. *There is $c > 0$ such that for all $\delta > 0$, there are $\xi, \eta, \gamma > 0$, $m \in \mathbb{N}$, $C > 1$ and $H \in \mathbb{N}$ such that the following holds. Suppose $k \in \mathbb{N}$ is such that X is an $(m, k/C, \xi, c)$ weak UG coboundary expander, and $F \colon X(k) \to \{0,1\}^k$ passes the (k, s) direct product tester over X for $s = \eta k$ with probability at least δ. Then there exists $f \colon X(1) \to \{0,1\}$ such that*

$$\Pr_{A \sim \mu_k}[\Delta(F[A], f|_A) \leqslant H/k] \geqslant \eta.$$

The proof of Theorem 5.1 proceeds in exactly the same way as the proof of Theorem 1.9, except that we use a different agreement theorem over the Johnson scheme: we seek a global function f that has stronger agreement with the assignment F. Namely, we want $\Delta(f|_A, F[A]) = O(1)$ for a large fraction of the k-faces A. Below is a formal statement of the Johnson agreement theorem we need.

Theorem 5.2. *For every $\varepsilon > 0$ there is $\alpha \in (0,1)$, $t \in \mathbb{N}$ such that for the following holds for sufficiently large k, and for n sufficiently large compared to k. If $F \colon \binom{[n]}{k} \to \{0,1\}^k$ passes the $(k, \alpha k)$ direct product test with probability ε, then there exists g such that,*

$$\Pr_{A \subseteq_k [n]}[F[A] \neq_{< \frac{t}{k}} g(A)] \geqslant \Omega(\varepsilon^{12}).$$

Proof. The proof is deferred to Section B. □

5.1 Deriving Theorem 5.1 from Theorem 5.2

Following the argument in Section 4 with slight modifications and using Theorem 5.2 instead of Theorem 4.7, one gets the"sufficient" part of Theorem 5.1. Below, we elaborate on the slight modifications that are necessary.

1. Following the argument in Section 4.3.1, we localize the tester to d faces again, and the test we consider is the $(k, \alpha k)$ test (just like in the overall complex).

2. We run a procedure which is the same as the short list algorithm in Section 4.3.2, however the parameters are a bit different (as the soundness of Theorem 5.2 is a 12th power of ε, as opposed to a 3rd power of ε as in Theorem 4.7. The effect of that is that all the powers in the description of the short list algorithm grow by a factor of 4.

3. When we run the analog of Lemma 4.10 in our context, the distance parameter changes. This amounts to changing the η in Lemma 4.10 (which is inherited from the distance in Theorem 4.7) to be t/k (which is the distance in Theorem 5.2). This change propogates throughout the argument.

4. In Section 4.4, when we project the lists onto level r, it is important for us that the list sizes do not collapse. As seen in the proof of Claim 4.13, this ultimately boils down to the fact that for almost all of the d-faces D, the functions in the list of D are pairwise far from each other. In the context of Section 4 this distance is a constant fractional distance (hence we can reduce to a constant level r). In the context of this section though, our distance promise is milder and stands at C/k for a large constant C, and therefore we are able to only reduce to level k/C' and retain the non-collapse property of the list. This is ultimately the reason that in Theorem 5.1 we require UG coboundary expansion for a level which is comparable to k.

This summarizes the modifications required in Section 4 to make the proof of Theorem 5.1 go through.

6 UG Coboundary Expansion for Known Complexes

6.1 LSV Complexes

In this section, we apply Theorem 3.1 to get that there exist LSV complexes [LSV05b, LSV05a] that do not support direct product tests with soundness $1/2$. We will make use of the following result due to [KKL14, EK16]. In topological language, the result asserts that there are LSV complexes with non-vanishing cohomology, and in fact elements in the cohomology must have large weight. We refrain from defining these notions and instead state their result in our language. We will need a quantitatively stronger version though, which was recently established by Dikstein and Dinur [DD23b].

Theorem 6.1. *There exists $\mu > 0$, such that for all large enough $d \in \mathbb{N}$ and $\gamma > 0$, there is an infinite sequence of d-dimensional LSV complexes $\{X_n\}$ such that the following holds. For all n, the complex X_n is a γ two-sided local spectral expander and there exists a UG instance $\Psi_n = (G_1[X_n], \Pi_n)$ over $\mathbb{F}_2$ that is 1-triangle consistent but* $\mathrm{val}(\Psi_n) \leqslant 1 - \mu$.

Using Theorem 6.1 in conjunction with Theorem 3.1 we get the following corollary.

Corollary 6.2. *There exists $\varepsilon > 0$, such that for all $\eta > 0$, there is large enough d and $k \leqslant d$ and an infinite sequence of d-dimensional LSV complexes $\{X_n\}_{n\in\mathbb{N}}$, such that the following holds for each X_n. There is* *an assignment $F : X_n(k) \to \{0,1\}^k$ that passes the $(k, \sqrt{k})$ direct product tester with probability at least $1/2$, but for all $f : X(1) \to \{0,1\}$ we have that,*

$$\Pr_{A\sim\mu_k}\left[\Delta(F[A], f|_A) \geqslant \varepsilon\right] \leqslant \eta.$$

Proof. Using Theorem 6.1, for every d, γ we can get a sequence of γ-spectral d-dimensional LSV complexes, and for each of them a locally consistent UG instance Ψ_n on $G_1[X_n]$ with value at most $1-\mu$. We can now check that in fact for each n, we can generate a collection of lists on $X(1), X(3)$ such that Π_n is strongly consistent with respect to the lists. To see this, associate each $u \in X(1)$ by the list $L(u) = \{0,1\}$ and each face $(u,v,w) \in X(3)$ by the list $\{0 \cdot \pi(u,v)(0) \cdot \pi(u,w)(0), 1 \cdot \pi(u,v)(1) \cdot \pi(u,w)(1)\}$. One can check that the latter list is well-defined ($L(u,v,w) = L(u,w,v)$ for example) because $\pi(u,w) = \pi(u,v)\pi(v,w)$. This shows that π is strongly consistent with respect to all triangles. Therefore X_n is *not* a $(m = 2, r = 1, \xi = 0, c = \mu)$ weak UG coboundary expander. Applying Theorem 3.1 we immediately get that there is a large enough k such that there exists a function $F_n : X(k) \to \{0,1\}^k$ that passes the $(k, \sqrt{k})$ agreement test with probability $1/2$ (since $\xi = 0$) but for all global functions, $\Delta(F[A], f|_A) \geqslant \varepsilon$ holds with probability at most η over $A \sim \mu_k$. □

6.2 Johnson and Grassmann Complexes

In this section we verify that the Johnson and Grassmann complexes are UG coboundary expanders as defined in Definition 1.4. They are known to be standard coboundary expanders, but we give the proof that they satisfy UG coboundary expansion for completeness.

Let $\mathcal{C}_n$ denote the complete complex with vertices being $[n]$. Let $\mathcal{C}(i)$ denote the faces at the i^{th} level.

Lemma 6.3. *For all $n \in N$, the complete complex $\mathcal{C}_n$ is a $(m, r, \xi, \xi + o(1))$ UG coboundary expander for all $m \in \mathbb{N}$, $r = o(\sqrt{n})$ and $\xi \in (0,1)$.*

Proof. We will drop the subscript n in $\mathcal{C}_n$ henceforth. Consider any UG instance $\Psi = (G_r[\mathcal{C}], \Pi)$ with $1-\xi$-fraction of locally consistent triangles. It suffices to show that Ψ has an assignment that satisfies a $1 - \xi - o(1)$ fraction of the edges.

Let $\mathcal{T}$ denote the uniform distribution over all triangles in $G_r[\mathcal{C}]$ and for a face $U \in \mathcal{C}(r)$, let $\mathcal{T}_U$ denote the distribution over triangles that contain U. By averaging we know there exists $U \in \mathcal{C}(r)$ such that,

$$\Pr_{(U,V,W)\sim\mathcal{T}_U}[(U,V,W) \text{ is consistent}] \geqslant 1 - \xi.$$

Let N_U denote the neighborhood of U in $G_r[\mathcal{C}]$, i.e. the set of faces $V \in \mathcal{C}(r)$ that are disjoint to U. One can check that N_U has fractional size at least $1 - r^2/n = 1 - o(1)$ as $r = o(\sqrt{n})$. Now define an assignment of permutations $P : \mathcal{C}(r) \to S_m$ as follows:

$$P(V) = \begin{cases} \text{id} & \text{if } V = U, \\ \pi(u,v) & \text{if } V \in N_U \\ \text{id} & \text{otherwise.} \end{cases} \tag{8}$$

Note that by definition P satisfies all the edges incident on U and in fact it is easy to check that P also satisfies all the edges (V,W) where (U,V,W) forms a consistent triangle. Therefore it suffices to lower bound the measure of such edges:

$$\text{val}(P) \geqslant \Pr_{(V,W)\sim G_r[\mathcal{C}]}[V, W \in N_U, (U,V,W) \text{ is consistent}]$$

$$\geqslant \left(1 - 2\frac{r^2}{n}\right) \Pr_{(V,W)\sim G_r[\mathcal{C}]}[(U,V,W) \text{ is consistent } \mid V,W \in N_U]$$
$$\geqslant (1-o(1))(1-\xi),$$

where in the first inequality we used that the fractional size of N_U is at least $1 - r^2/n$, and in the last inequality we used the fact that the distribution over triangles obtained by sampling a random edge in the neighborhood of U is equal to the distribution $\mathcal{T}_U$. This shows the existence of an assignment with value $1 - \xi - o(1)$ as required. □

Let $\mathrm{Gr}_{n,q}$ denote Grassmann complex over the ambient space $\mathbb{F}_q^n$ where the i-level faces in the complex for $i \leqslant n$ are the i-dimensional linear subspaces of $\mathbb{F}_q^n$. Note that this is not a simplicial complex but we can still consider the graph $G_r[\mathrm{Gr}_{n,q}]$ – the vertices are the r-level faces of the complex and we have edges between two subspaces that are disjoint.

Lemma 6.4. *For all finite fields $\mathbb{F}_q$ and $n \in N$, the Grassmann complex $Gr_{n,q}$ is a $(m, r, \xi, \xi + o(1))$ UG coboundary expander for all $m \in \mathbb{N}$, $r = o(\log_q n)$ and $\xi \in (0, 1)$.*

Proof. The proof of this lemma is exactly the same as that of Lemma 6.3, except that we note that the neighborhood of every vertex in $G_r[\mathrm{Gr}_{q,n}]$ has size at least $1 - q^r/q^n = 1 - o(1)$ if $r = o(\log_q n)$. Therefore by following the rest of the proof we get that for any UG instance that is $1 - \xi$ triangle consistent, there exists an assignment with value at least $1 - \xi - o(1)$, as required. □

7 Acknowledgements

We thank Mehtaab Sawhney for suggesting the use of the weak regularity lemma and Nikhil Vyas for bringing [AdlVKK02] to our attention, which led to a considerable simplification on our earlier attempts at proving Lemma 4.11. We thank Tali Kaufman for helpful comments on an earlier version of this paper. We also thank the Simons institute for hosting us during the “Analysis and TCS” program, where part of the work was carried out.

References

[AdlVKK02] Noga Alon, Wenceslas Fernandez de la Vega, Ravi Kannan, and Marek Karpinski. Random sampling and approximation of MAX-CSP problems. In John H. Reif, editor, *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 232–239. ACM, 2002.

[AL20] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 1198–1211. ACM, 2020.

[BG15] Mark Braverman and Ankit Garg. Small value parallel repetition for general games. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 335–340, 2015.

[BKM22] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable *k*-csps: I. In *STOC ’22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 976–988, 2022.

[BKM23a] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable k-csps: II. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 632–642, 2023.

[BKM23b] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable k-csps: III. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 643–655, 2023.

[BKM23c] Amey Bhangale, Subhash Khot, and Dor Minzer. On approximability of satisfiable k-csps: IV. *CoRR*, abs/2307.16248, 2023.

[DD19] Yotam Dikstein and Irit Dinur. Agreement testing theorems on layered set systems. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 1495–1524, 2019.

[DD23a] Yotam Dikstein and Irit Dinur. Agreement theorems for high dimensional expanders in the small soundness regime: the role of covers. 2023.

[DD23b] Yotam Dikstein and Irit Dinur. Coboundary and cosystolic expansion without dependence on dimension or degree, 2023.

[DDFH18] Yotam Dikstein, Irit Dinur, Yuval Filmus, and Prahladh Harsha. Boolean function analysis on high-dimensional expanders. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

[DDG$^+$17] Roee David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. Direct sum testing. *SIAM J. Comput.*, 46(4):1336–1369, 2017.

[DEL$^+$22] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 357–374. ACM, 2022.

[DFH19] Irit Dinur, Yuval Filmus, and Prahladh Harsha. Analyzing boolean functions on the biased hypercube via higher-dimensional agreement tests: [extended abstract]. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2124–2133, 2019.

[DG08] Irit Dinur and Elazar Goldenberg. Locally testing direct product in the low error range. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 613–622. IEEE Computer Society, 2008.

[DK17] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 974–985. IEEE Computer Society, 2017.

[DM11] Irit Dinur and Or Meir. Derandomized parallel repetition via structured pcps. *Comput. Complex.*, 20(2):207–327, 2011.

[DM19] Irit Dinur and Roy Meshulam. Near coverings and cosystolic expansion - an example of topological property testing. *Electron. Colloquium Comput. Complex.*, TR19-126, 2019.

[DR06] Irit Dinur and Omer Reingold. Assignment testers: Towards a combinatorial proof of the pcp theorem. *SIAM Journal on Computing*, 36(4):975–1024, 2006.

[DS14a] Irit Dinur and David Steurer. Analytical approach to parallel repetition. In David B. Shmoys, editor, *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 624–633. ACM, 2014.

[DS14b] Irit Dinur and David Steurer. Direct product testing. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 188–196, 2014.

[EK16] Shai Evra and Tali Kaufman. Bounded degree cosystolic expanders of every dimension. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 36–48, 2016.

[EKZ20] Shai Evra, Tali Kaufman, and Gilles Zémor. Decodable quantum LDPC codes beyond the square root distance barrier using high dimensional expanders. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 218–227. IEEE, 2020.

[FOW22] Yuval Filmus, Ryan O'Donnell, and Xinyu Wu. Log-sobolev inequality for the multislice, with applications. *Electronic Journal of Probability*, 27:1–30, 2022.

[GK23] Roy Gotlib and Tali Kaufman. List agreement expansion from coboundary expansion. In *14th Innovations in Theoretical Computer Science Conference, ITCS 2023, January 10-13, 2023, MIT, Cambridge, Massachusetts, USA*, volume 251 of *LIPIcs*, pages 61:1–61:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023.

[GS00] Oded Goldreich and Shmuel Safra. A combinatorial consistency lemma with application to proving the pcp theorem. *SIAM Journal on Computing*, 29(4):1132–1154, 2000.

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

[IKW09] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New direct-product testers and 2-query pcps. In Michael Mitzenmacher, editor, *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 131–140. ACM, 2009.

[KKL14] Tali Kaufman, David Kazhdan, and Alexander Lubotzky. Ramanujan complexes and bounded degree topological expanders. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 484–493, 2014.

[KM22a] Tali Kaufman and David Mass. Double balanced sets in high dimensional expanders. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2022, September 19-21, 2022, University of Illinois, Urbana-Champaign, USA (Virtual Conference)*, pages 3:1–3:17, 2022.

[KM22b] Tali Kaufman and David Mass. Unique-neighbor-like expansion and group-independent cosystolic expansion. *CoRR*, abs/2211.09482, 2022.

[LM06] Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. *Combinatorica*, 26(4):475–487, 2006.

[LSV05a] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. Explicit constructions of ramanujan complexes of type $\tilde{a}_d$. *Eur. J. Comb.*, 26(6):965–993, 2005.

[LSV05b] Alexander Lubotzky, Beth Samuels, and Uzi Vishne. Ramanujan complexes of type $\tilde{a}_d$. *Israel journal of Mathematics*, 149:267–299, 2005.

[LZ22] Anthony Leverrier and Gilles Zémor. Quantum tanner codes. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 872–883. IEEE, 2022.

[PK22] Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 375–388. ACM, 2022.

[Rao11] Anup Rao. Parallel repetition in projection games and a concentration bound. *SIAM J. Comput.*, 40(6):1871–1891, 2011.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.

A Proofs of Omitted Claims

Claim A.1. *If a set of lists of size m and permutations in S_m, $(\mathcal{L}, \Pi)$ on $X(t), X(2t), X(3t)$ are $1-\xi$ consistent according to Definition 1.7, then Π is $1-3\xi$ consistent according to Definition 1.3.*

Proof. Let $\mathcal{T}$ be the distribution over triangles generated by sampling $T \sim X(3t)$ and then randomly splitting it into (u, v, w). We can use a union bound to get that with probability $1-3\xi$, $(u, v, w) \sim \mathcal{T}$ satisfies,

$$\begin{aligned} L'(T) &= L(u) \circ \pi(u,v)L(v) \circ \pi(u,w)L(w) \\ &= L(w) \circ \pi(w,u)L(u) \circ \pi(w,v)L(v) \\ &= L(v) \circ \pi(v,w)L(w) \circ \pi(v,u)L(u). \end{aligned}$$

Note that for each $i \in [m]$, the first line asserts that the assignment $\pi(u,w)(i)$ in the list of $L(u)$ is consistent with the assignment i of w with respect to $L'(T)$. By the third line, this assignment is consistent with the assignment $\pi(v,u) \circ \pi(u,w)(i)$ for $L(v)$, and by the second line this is consistent with the assignment $\pi(w,v) \circ \pi(v,u) \circ \pi(u,w)(i)$ for w. As $L'(T)$ contains m distinct assignments, we conclude that it must be the case that $\pi(w,v) \circ \pi(v,u) \circ \pi(u,w)(i) = i$ for all $i \in [m]$, and so $\pi(w,u) = \pi(u,w)^{-1} = \pi(w,v)\pi(v,u)$ for such a triangle (u, v, w). $\square$

B The Johnson Agreement Theorem: Proof of Theorem 5.2

In this section we give the proof of Theorem 5.2, and throughout we use the parameters:

$$0 \ll R^{-1} \ll h^{-1} \ll \nu \ll \alpha \ll \varepsilon < 1.$$

The proof follows the lines of the argument in [IKW09], except that in the end we use small-set expansion type arguments to glue together the local decoded functions similarly to the way it is done in [BKM23b].

B.1 Auxiliary Claims

We need a few standard auxiliary claim. The first claim asserts that if two functions are far, then they disagree on many k-sets.

Claim B.1. *If $g, h\colon [n] \to \{0,1\}$ are functions such that $\Delta(g,h) \geqslant R/k$, then*

$$\Pr_{A \subseteq_k [n]}\left[g|_A = h|_A\right] \leqslant e^{-R}.$$

Proof. The probability is at most

$$\left(\frac{n - n\Delta(g,h)}{n}\right) \cdot \left(\frac{n-1-n\Delta(g,h)}{n-1}\right) \cdots \left(\frac{n-(k-1)-n\Delta(g,h)}{n-(k-1)}\right) \leqslant (1-\Delta(g,h))^k \leqslant e^{-R}.$$

□

The second claim asserts that if we have a graph in which the second singular value is small, then any set of vertices which is not-too-small contains a sizable number of edges.

Claim B.2. *Suppose that M is the normalized adjacency matrix of a graph G, and the second singular value of M is at most δ. Then for any set of vertices S of relative size ε we have*

$$\Pr_{x, y \sim Mx}\left[x \in S, y \in S\right] \geqslant \varepsilon^2 - \delta\varepsilon.$$

Proof. The probability can be written as

$$\varepsilon\langle 1_S, M1_S\rangle = \varepsilon^2 + \langle 1_S - \varepsilon, M(1_S - \varepsilon)\rangle \geqslant \varepsilon^2 - \|1_S\|\|M(1_S-\varepsilon)\|_2 \geqslant \varepsilon^2 - \sqrt{\varepsilon}\delta\sqrt{\varepsilon} = \varepsilon^2 - \varepsilon\delta.$$

□

B.2 Local Structure

We will think of a k-set as partition as $A_0 \cup B_0$ where A_0 has size αk and B_0 has size $(1-\alpha)k$. For a partition of k-set (A_0, B_0), we define

$$\mathsf{Cons}(A_0, B_0) = \left\{B \subseteq [n] \setminus A_0 \;\middle|\; F[A_0 \cup B_0]|_{A_0} = F[A_0 \cup B]|_{A_0}\right\}.$$

We use a few definitions and results from [IKW09].

B.2.1 Good and Excellent

Definition B.3. *We say* (A_0, B_0) *is good if* $\Pr_{B\subseteq_{(1-\alpha)k}[n]\setminus A_0}\left[B \in \mathsf{Cons}(A_0, B_0)\right] \geqslant \frac{\varepsilon}{2}$.

We will think of B as being split into $D \cup E$ where $|E| = \alpha k$ and $|D| = (1-2\alpha)k$.

Definition B.4. *We say* (A_0, B_0) *is excellent if it is good, and furthermore*

$$\Pr_{\substack{B_1=(D_1\cup E)\subseteq[n]\setminus A_0\\ B_2=(D_2\cup E)\subseteq[n]\setminus A_0}}\left[(D_1, E), (D_2, E) \in \mathsf{Cons}(A_0, B_0) \wedge F[A_0 \cup D_1 \cup E]|_E \neq_{>h} F[A_0 \cup D_2 \cup E]|_E\right] \leqslant \nu.$$

Lemma B.5. *The following holds:*

1. *A randomly chosen* (A_0, B_0) *is good with probability at least* $\varepsilon/2$.
2. *A randomly chosen good* (A_0, B_0) *is excellent with probability at least* $1 - \frac{2^{-\Omega(h)}}{\nu}$.

Proof. These are [IKW09, Lemma 3.5], [IKW09, Lemma 3.6]. □

B.2.2 Decoding Local Structure

Lemma B.6. *Suppose* (A_0, B_0) *is excellent. Then there exists* $g_{A_0,B_0} \colon [n] \to \{0,1\}$ *such that*

$$\Pr_{B\in\mathsf{Cons}(A_0,B_0)}\left[g_{A_0,B_0}(B) \neq_{>2R/k} F[A_0 \cup B]|_B\right] \leqslant \sqrt{\nu}.$$

Proof. This is [IKW09, Lemma 3.8]. □

B.3 Gluing Local Structure

B.3.1 Getting correlation over the global functions

Consider the following joint distribution over pairs of (A_0, B_0) and (A_0', B_0'), which we call $\mathcal{D}$:

1. Sample $\tilde{A} \subseteq_{\sqrt{\alpha}k} [n]$.
2. Sample ℓ from a binomial distribution $\mathsf{Bin}(\alpha k, \sqrt{\alpha})$ and $A_0, A_0' \subseteq_{\alpha k} \tilde{A}$ conditioned on $|A_0 \cap A_0'| = \ell$.
3. Sample $B_0, B_0' \subseteq_{(1-\alpha)k} [n] \setminus \tilde{A}$.

The following lemma shows that the functions g_{A_0,B_0} and $g_{A_0',B_0'}$ for pairs generated in this way are very close to each other with noticeable probability.

Claim B.7. *It holds that*

$$\Pr_{((A_0,B_0),(A_0',B_0'))\sim\mathcal{D}}\left[\Delta(g_{A_0,B_0}, g_{A_0',B_0'}) \lesssim \frac{R\log(1/\varepsilon)}{k}\right] \gtrsim \varepsilon^6.$$

Proof. Choose $\tilde{A}$ of size $\sqrt{\alpha}k$, choose $B_0 \subseteq_{(1-\alpha)k} [n] \setminus \tilde{A}$ and $B' \subseteq_{(1-\sqrt{\alpha})k} [n] \setminus \tilde{A}$ independently, and partition $\tilde{A}$ randomly as $A_0 \cup A_1$ where the size of A_0 is αk. Then

$$\mathop{\mathbb{E}}_{\substack{\tilde{A}\\ B'\subseteq_{(1-\sqrt{\alpha})k}[n]\setminus\tilde{A}}}\left[\mathop{\mathbb{E}}_{\tilde{A}=A_0\cup A_1, B_0}\left[1_{(A_0,B_0)\text{ is excellent}} \cdot 1_{g_{A_0,B_0}(A_1\cup B')\neq_{\leqslant 2R/k} F[A_0\cup A_1\cup B']|_{A_1\cup B'}}\right]\right] \gtrsim \varepsilon^2,$$

where we used the fact that the distribution of (A_0, B_0) is uniform, hence by Lemma B.5 it is excellent with probability at least $\varepsilon/4$. Also, conditioned on A_0, B_0, the distribution of $A_1 \cup B'$ is $O(k^2/n) = o(1)$ close to uniform; thus, it is in $\mathsf{Cons}(A_0, B')$ with probability at least $\varepsilon/2 - o(1)$. Conditioned on that, the distribution of $A_1 \cup B'$ is $o(1)$ close to uniform in $\mathsf{Cons}(A_0, \tilde{B})$, so by Lemma B.6 we have that $g_{A_0,B_0}(A_1 \cup B') \neq_{\leqslant R/k} F[A_0 \cup A_1 \cup B']|_{A_1 \cup B'}$ with probability $1 - o(1)$.

It follows by an averaging argument that with probability at least $\Omega(\varepsilon^2)$ over the choice of $\tilde{A}, B'$, we have that

$$\Pr_{A_0 \subseteq \tilde{A}, B_0}\left[g_{A_0,B_0}(B') \neq_{\leqslant 2R/k} F[\tilde{A} \cup B']|_{B'}\right] \gtrsim \varepsilon^2, \tag{9}$$

and we fix such $\tilde{A}, B'$. Consider the graph induced by the sampling procedure $\mathcal{D}$, i.e. its vertices are (A_0, B_0) and the weight of an edge (A_0, B_0) and (A_0', B_0') is proportional to the probability this is sampled from $\mathcal{D}$ conditioned on $\tilde{A}$. Note that this is a product graph, where one graph is the A_0 vs A_0' graph, and the other graph is the B_0 vs B_0' graph. We argue that the absolute value of the largest non-trivial eigenvalue in each one of these graphs is at most $\sqrt{\alpha}$, and hence the same holds for the product graph. As both arguments are essentially the same, we explain the argument for the A_0 vs A_0' graph. The adjacency matrix of the (A_0, A_0') graph is $\mathrm{M} = \sum_{\ell} p(\ell) M_\ell$ where $p(\ell) = \Pr\left[\mathsf{Bin}(\alpha k, \sqrt{\alpha}) = \ell\right]$ and M_ℓ is the adjacency matrix of the Johnson graph with intersection parameter ℓ. The matrices M_ℓ have the same eigenvectors, and the largest non-trivial eigenvalue of M_ℓ in absolute value is $\frac{\ell}{\alpha k}$, hence the largest non-trivial eigenvalue of M is $\frac{1}{\alpha k}\sum_{\ell} p(\ell)\ell = \sqrt{\alpha}$. Combining this observation with (9) via Claim B.2 yields that

$$\Pr_{(A_0,A_0'),(B_0,B_0')}\left[g_{A_0,B_0}(B') \neq_{\leqslant 2R/k} F[\tilde{A} \cup B']|_{B'}, g_{A_0',B_0'}(B') \neq_{\leqslant 2R/k} F[\tilde{A} \cup B']|_{B'}\right] \gtrsim \varepsilon^4,$$

so by the triangle inequality

$$\Pr_{(A_0,A_0'),(B_0,B_0')}\left[g_{A_0,B_0}(B') \neq_{\leqslant 4R/k} g_{A_0',B_0'}(B')\right] \gtrsim \varepsilon^4$$

Taking expectation over $\tilde{A}$ and B' gives (noting that the distribution over B' is a uniform subset of $[n] \setminus \tilde{A}$ of size $(1 - \sqrt{\alpha})k$, hence it is $O(k^2/n) = o(1)$ close to uniform)

$$\mathop{\mathbb{E}}_{((A_0,B_0),(A_0',B_0'))\sim\mathcal{D}}\left[\mathop{\mathbb{E}}_{B'}\left[1_{g_{A_0,B_0}(B') \neq_{\leqslant 4R/k} g_{A_0',B_0'}(B')}\right]\right] \gtrsim \varepsilon^6.$$

By an averaging argument, with probability at least $\Omega(\varepsilon^6)$ over the choice of $((A_0, B_0), (A_0', B_0')) \sim \mathcal{D}$, we get that $\mathbb{E}_B\left[1_{g_{A_0,B_0}(B) \neq_{\leqslant 4R/k} g_{A_0',B_0'}(B)}\right] \gtrsim \varepsilon^6$. In that case, sampling $C \subseteq [n]$ of size $k/10R$ we get that $g_{A_0,B_0}|_C = g_{A_0',B_0'}|_C$ with probability $\Omega(\varepsilon^6)$, and by Claim B.1 it follows that $\Delta(g_{A_0,B_0}, g_{A_0',B_0'}) \lesssim R\log(1/\varepsilon)/k$. □

B.3.2 Applying Expansion

We identify pairs (A_0, B_0) with points in the multi-slice $\binom{[n]}{\alpha k, (1-\alpha)k, n-k}$ in the obvious way: a pair (A_0, B_0) is identified with a point $x \in \binom{[n]}{\alpha k, (1-\alpha)k, n-k}$ whose set of coordinates i such that $x_i = 0$ is A_0, whose set of coordinates i such that $x_i = 1$ is B_0, and whose set of coordinates i such that $x_i = 2$ is $[n] \setminus (A_0 \cup B_0)$. Thus, the distribution $\mathcal{D}$ defines a Markov chain T over $\binom{[n]}{\alpha k, (1-\alpha)k, n-k}$. To simplify notation, we denote the function g_{A_0,B_0} by g_x, where x is the point in the multi-slice associated with (A_0, B_0). Using these notations, Claim B.7 implies that sampling x uniformly and then $y \sim \mathrm{T}x$ we have that $\Delta(g_x, g_y) \lesssim R\log(1/\varepsilon)/k$ with probability $\Omega(\varepsilon^6)$. Looking at the definition of $\mathcal{D}$, it is clear that T is symmetric.

Claim B.8. $\lambda_2(\mathrm{T}) \lesssim \alpha^{\Omega(1)}$.

Proof. The proof is deferred to Section B.4. □

Claim B.8 gives us the following small set expansion conclusion:

Claim B.9. *Suppose* $f\colon \binom{[n]}{\alpha k,(1-\alpha)k,n-k} \to \{0,1\}$ *has expectation* p*. Then*

$$\langle f, \mathrm{T}f\rangle \leqslant p^{1.5} + \alpha^{\Omega(1)}p.$$

Proof. By Cauchy-Schwarz and Parseval

$$\langle f, \mathrm{T}f\rangle^2 \leqslant p\|\mathrm{T}f\|_2^2 = p\left(\|\mathrm{T}(f-p)\|_2^2 + p^2\right) \leqslant p\left(\alpha^{\Omega(1)}\|f-p\|_2^2 + p^2\right) \leqslant p^2\alpha^{\Omega(1)} + p^3,$$

and the result follows. □

The following claim asserts that the functions g_x are close to each other with noticeable probability.

Claim B.10. *It holds that* $\Pr_{x,y}\left[\Delta(g_x, g_y) \lesssim R^2\log(1/\varepsilon)^2/k\right] \gtrsim \varepsilon^{12}$.

Proof. From Claim B.7 we get that

$$\mathop{\mathbb{E}}_{x,y\sim \mathrm{T}x}\left[\mathop{\mathbb{E}}_{S\subseteq_{k/R^2\log(1/\varepsilon)}[n]}\left[1_{g_x|_S \equiv g_y|_S}\right]\right] \gtrsim \varepsilon^6.$$

For each $S \subseteq [n]$ of size $k/R^2\log(1/\varepsilon)$ and $v \in \{0,1\}^S$, let $f_v\colon \binom{[n]}{\alpha k,(1-\alpha)k,n-k} \to \{0,1\}$ be defined by $f_{S,v}(x) = 1$ if and only if $g_x|_S = v$. Rewriting the above by flipping the order of expectations we get that

$$\mathop{\mathbb{E}}_{S\subseteq_{k/R^2\log(1/\varepsilon)}[n]}\left[\sum_{v\in\{0,1\}^S}\langle f_{S,v}, \mathrm{T}f_{S,v}\rangle\right] \gtrsim \varepsilon^6.$$

By Claim B.9 we get that

$$\sum_{v\in\{0,1\}^S}\langle f_{S,v}, \mathrm{T}f_{S,v}\rangle \leqslant \sum_{v\in\{0,1\}^S}\|f_{S,v}\|_2^3 + \alpha^{\Omega(1)}\|f_{S,v}\|_2^2 \leqslant \alpha^{\Omega(1)} + \sqrt{\sum_v \|f_{S,v}\|_2^4} \leqslant \varepsilon^9 + \sqrt{\sum_v \|f_{S,v}\|_2^4},$$

where we used Cauchy-Schwarz and $\sum_v \|f_{S,v}\|_2^2 \leqslant 1$. We get that $\mathbb{E}_{S\subseteq_{k/R^2\log(1/\varepsilon)}[n]}\left[\sqrt{\sum_v \|f_{S,v}\|_2^4}\right] \gtrsim \varepsilon^6$, so by Cauchy-Schwarz

$$\mathop{\mathbb{E}}_{S\subseteq_{k/R^2\log(1/\varepsilon)}[n]}\left[\sum_v \|f_{S,v}\|_2^4\right] \gtrsim \varepsilon^{12}.$$

The last inequality implies that

$$\mathop{\mathbb{E}}_{S\subseteq_{k/R^2\log(1/\varepsilon)}[n]}\left[\mathop{\mathbb{E}}_{x,y}\left[1_{g_x|_S \equiv g_y|_S}\right]\right] \gtrsim \varepsilon^{12},$$

where now the distribution over x and y is independent. Flipping the order of expectations again, we get that

$$\mathop{\mathbb{E}}_{x,y}\left[\mathop{\mathbb{E}}_{S\subseteq_{k/R^2\log(1/\varepsilon)}[n]}\left[1_{g_x|_S\equiv g_y|_S}\right]\right]\gtrsim\varepsilon^{12}.$$

By an averaging argument with probability at least $\Omega(\varepsilon^{12})$ over x,y we have $\mathbb{E}_{S\subseteq_{k/R^2\log(1/\varepsilon)}[n]}\left[1_{g_x|_S\equiv g_y|_S}\right]\gtrsim\varepsilon^{12}$. The proof is concluded by Claim B.1. □

B.3.3 Concluding Theorem 5.2

We now finish off the proof of Theorem 5.2.

By Claim B.10 we may find x such that $\Delta(g_x,g_y)\lesssim R^2\log(1/\varepsilon)^2/k$ for at least $\Omega(\varepsilon^{12})$ fraction of the y's. We fix such x henceforth, and sample y independently of x. Identify y back with A_0,B_0 and then sample $B\subseteq[n]\setminus A_0$ of size $(1-\alpha)k$. Then $\Delta(g_x,g_y)\lesssim R^2\log(1/\varepsilon)^2/k$ with probability $\Omega(\varepsilon^{12})$, and we condition on that. By Chernoff's inequality we have that $\Delta_B(g_x,g_y)\lesssim R^2\log(1/\varepsilon)^3/k$ except with probability $1-\varepsilon^{100}$. By Lemma B.6, with probability at least $\Omega(\varepsilon^2)$ we get that $\Delta(g_y(B),F[A_0\cup B])\lesssim R/k$. By a union bound and triangle inequality it follows that with probability $\Omega(\varepsilon^2)$ we have $\Delta_B(g_x,F[A_0\cup B])\lesssim R^2\log(1/\varepsilon)^3/k$. In conclusion,

$$\Pr_{x,A_0,B}\left[\Delta_B(g_x,F[A_0\cup B])\lesssim R^2\log(1/\varepsilon)^3/k\right]\gtrsim\varepsilon^{12}.$$

As A_0,B are independent of x, we get that

$$\Pr_{x,A}\left[\Delta_A(g_x,F[A])\lesssim R^2\log(1/\varepsilon)^3/k\right]\gtrsim\varepsilon^{12},$$

concluding the proof. □

B.4 Proof of Claim B.8

Throughout, we denote $\mathcal{U}=\binom{[n]}{\alpha k,(1-\alpha)k,n-k}$. Consider the symmetric group S_n, and let $I,J\subseteq[n]$ be disjoint subsets of sizes αk and $(1-\alpha)k$ respectively. Let $\rho>0$ to be determined, and let $\mathrm{T}''_\rho=e^{-\rho L}$ where $L\colon L_2(S_n)\to L_2(S_n)$ is the Laplacian $Lf(\pi)=f(\pi)-\mathbb{E}_{i\neq j\in[n]}\left[f(\pi_{i,j}\pi)\right]$ where $\pi_{i,j}$ is the transposition permutation between i and j. Let $\mathrm{S}\colon L_2(S_n)\to L_2(S_n)$ be the operator defined as $\mathrm{S}f(\pi)=\mathbb{E}_{\pi':\pi'(I)=\pi(I)}\left[f(\pi')\right]$. Define $\mathrm{T}'_\rho=\mathrm{S}\circ\mathrm{T}''_\rho$.

We will now explain how T'_ρ can be thought of as an operator on $L_2(\mathcal{U})$. To do that, we associate with each permutation $\pi\in S_n$ a point $x=x(\pi)\in\mathcal{U}$, where $x_{\pi(I)}=0$, $x_{\pi(J)}=1$ and $x_{[n]\setminus\pi(I)\cup\pi(J)}=2$, where $I=\{1,\ldots,\alpha k\}$, J is the next $(1-\alpha)k$ elements. Thus, we may define an operator $\mathrm{H}\colon L_2(\mathcal{U})\to L_2(S_n)$ by $\mathrm{H}f(\pi)=f(x(\pi))$. Thus, we define the operator $\mathrm{T}_\rho=\mathrm{H}^*\mathrm{T}'_\rho\mathrm{H}\colon L_2(\mathcal{U})\to L_2(\mathcal{U})$. Our plan is to show the following two claims:

Claim B.11. *For $\rho=n\ln(1/\alpha)/2$ we have that $\|\mathrm{T}_\rho-\mathrm{T}\|_2\leqslant\alpha^{\Omega(1)}$.*

Claim B.12. *For $\rho=n\ln(1/\alpha)/2$, $\lambda_2(\mathrm{T}_\rho)\leqslant\alpha^{\Omega(1)}$.*

Together, the two claims finish the proof of Claim B.8 immediately.

B.4.1 Proof of Claim B.11

Consider a sampling of $x \sim \mathcal{U}$, $y \sim \mathrm{T}x$ and $y' \sim \mathrm{T}_\rho x$. For each $\ell \in \mathbb{N}$, let E^ℓ be the event that $|x^{-1}(0) \cap y^{-1}(0)| = \ell$, let E^ℓ_ρ be the event that $|x^{-1}(0) \cap y'^{-1}(0)| = \ell$, and note that the distributions $y \mid E^\ell$ and $y' \mid E^\ell_\rho$ are identical. Thus, the statistical distance between y and y' is at most

$$\Delta := \sum_\ell \left| \Pr[E^\ell] - \Pr[E^\ell_\rho] \right|.$$

It follows that there is a coupling between y and y' such that $\Pr[y \neq y']$. Fix f with 2-norm equal to 1 for which $\|\mathrm{T} - \mathrm{T}_\rho\|_2 = \|(\mathrm{T} - \mathrm{T}_\rho)f\|_2$. Then

$$\|\mathrm{T} - \mathrm{T}_\rho\|_2^2 = \mathop{\mathbb{E}}_{x \sim \mathcal{U}}\left[\left| \mathop{\mathbb{E}}_{y,y'}\left[f(y) - f(y') \right] \right|^2 \right] = \mathop{\mathbb{E}}_{x \sim \mathcal{U}}\left[\left| \mathop{\mathbb{E}}_{y,y'}\left[(f(y) - f(y'))1_{y \neq y'} \right] \right|^2 \right],$$

which by Cauchy-Schwarz is at most

$$\mathop{\mathbb{E}}_{x \sim \mathcal{U}}\left[\mathop{\Pr}_{y,y'}\left[y \neq y' \right] \mathop{\mathbb{E}}_{y,y'}\left[\left| f(y) - f(y') \right|^2 \right] \right] \leqslant \Delta 4 \|f\|_2^4 = 4\Delta.$$

The following claim finishes the proof of Claim B.11.

Claim B.13. $\Delta \leqslant \alpha^{\Omega(1)}$.

Proof. Consider the distribution of $|x^{-1}(0) \cap y'^{-1}(0)|$. An equivalent way to think about the Poisson sampling, is that for each $i \neq j$ we have an independent poisson random variable $Z_{i,j} \sim \mathsf{Poisson}(\rho/n(n-1))$, and then we apply the transpositions corresponding to $Z_{i,j}$. Define $Z_i = 1_{\sum_j Z_{i,j} > 0}$, $Z = \sum_{i \in I} Z_i$, and note that with probability $1 - \alpha^{\Omega(1)}$ it holds that $|x^{-1}(0) \cap y'^{-1}(0)| = \alpha k - Z$; the reason is that Z counts the number of $i \in I$ such that we picked a transposition of the form π_{i,j_i} in the process for some j_i, and with probability $1 - \alpha^{\Omega(1)}$ all of these j_i's are outside I and are all distinct. Thus, the distribution $q_\ell = \Pr[E^\ell]$ is $\alpha^{\Omega(1)}$ close in statistical distance to the distribution $p_\ell = \Pr\left[Z = \alpha k - \ell \right] = \Pr\left[\sum_{i \in I} 1 - Z_i = \ell \right]$. Note $1 - Z_i$ are independent Bernouli random variables with

$$\mathbb{E}[1 - Z_i] = \prod_{j \neq i} \Pr\left[Z_{i,j} = 0 \right] = e^{-\rho/n} = \alpha^{1/2},$$

so p_ℓ is exactly the law of $|x^{-1}(0) \cap y^{-1}(0)|$. $\square$

B.4.2 Proof of Claim B.12

Fix $f \colon \mathcal{U} \to \mathbb{R}$ with 2-norm equal to 1 and expectation 0 to be an eigenvector of T_ρ with eigenvalue $\lambda_2(\mathrm{T}_\rho)$, and define $g = \mathrm{H}f$. Note that $\mathbb{E}[g] = 0$ and

$$\lambda_2(\mathrm{T}_\rho) = \|\mathrm{H}^* \mathrm{T}'_\rho \mathrm{H} f\|_2 \leqslant \|\mathrm{T}'_\rho g\|_2.$$

Thus, it suffices to upper bound the second eigenvalue of T'_ρ. To study these eigenvalues, we use formulas (31), (32) and (33) from [FOW22, Section 4.2]. These formulas assert that for each non-trivial partition $\lambda \vdash n$, the corresponding eigenvalue is $e^{-\rho(1-c_\lambda)}$ where

$$c_\lambda = \frac{1}{n(n-1)} \sum_i \lambda_i^2 - (2i-1)\lambda_i \leqslant \frac{1}{n(n-1)}(n-1)^2 \leqslant \frac{n-1}{n}.$$

Thus, for the eigenvalue we have that $e^{-\rho(1-c_\lambda)} \leqslant e^{-\rho\frac{1}{n}} \leqslant e^{-\ln(1/\alpha)/2} = \sqrt{\alpha}$.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il