Electronic Colloquium on Computational Complexity, Report No. 205 (2016)

# Cube vs. Cube Low Degree Test

Amey Bhangale * Irit Dinur† Inbal Rachel Livni Navon‡

December 18, 2016

**Abstract**

We revisit the Raz-Safra plane-vs.-plane test and study the closely related cube vs. cube test. In this test the tester has access to a "cubes table" which assigns to every cube a low degree polynomial. The tester randomly selects two cubes (affine sub-spaces of dimension 3) that intersect on a point $x \in \mathbb{F}^m$, and checks that the assignments to the cubes agree with each other on the point $x$. Our main result is a new combinatorial proof for a low degree test that comes closer to the soundness limit, as it works for all $\epsilon \geq \text{poly}(d)/|\mathbb{F}|^{1/2}$, where $d$ is the degree. This should be compared to the previously best soundness value of $\epsilon \geq \text{poly}(m,d)/|\mathbb{F}|^{1/8}$. Our soundness limit improves upon the dependence on the field size and does not depend on the dimension of the ambient space.

Our proof is combinatorial and direct: unlike the Raz-Safra proof, it proceeds in one shot and does not require induction on the dimension of the ambient space. The ideas in our proof come from works on direct product testing which are even simpler in the current setting thanks to the low degree.

Along the way we also prove a somewhat surprising fact about connection between different agreement tests: it does not matter if the tester chooses the cubes to intersect on points or on lines: for every given table, its success probability in either test is nearly the same.

## 1 Introduction

Low degree tests are local tests for the property of being a low degree function. These were the first property testing results that were discovered, and are an important component in PCP constructions. Such tests were studied in the 1990's and their ballpark soundness behavior was more or less understood. In this work we revisit these tests and give a new and arguably simpler analysis for the cube vs. cube low degree test. Our proof method allows us to get a soundness guarantee that is much closer to the conjectured optimal value. Discovering the precise point in which soundness starts to hold is an intriguing open question that captures an interesting aspect of local-testing in the small soundness regime.

Let us begin with a short introduction to low degree tests. A low degree test can be described as a game between a prover and a verifier, in which the prover wants to convince the verifier that a function $f : \mathbb{F}^m \to \mathbb{F}$ is a low degree polynomial. The most straightforward way for the prover

*Department of Computer Science, Rutgers University, USA. amey.bhangale@rutgers.edu

†Faculty of Computer Science and Mathematics, Weizmann Institute, Rehovot, Israel. irit.dinur@weizmann.ac.il

‡Computer Science and Mathematics, Weizmann Institute, Rehovot, Israel. inbal.livni@weizmann.ac.il



ISSN 1433-8092

to specify $f$ would be to give its value on each point $x \in \mathbb{F}^m$. However, in this way, to check that $f$ has degree at most $d$ the verifier would have to read $f$ on at least $d+2$ points. If we want a verifier that makes fewer queries while keeping the error small, it is useful to move to a more redundant representation of $f$. For example, the verifier can ask the prover to specify for every cube (affine subspace of dimension 3) $C \subset \mathbb{F}^m$, a function $f_C : C \to \mathbb{F}$ that is defined on the cube and is obtained by restricting $f$ to that cube. This is called a "cubes-table", and similarly one can consider a lines table (with an entry for every line), or a planes table (with an entry for each plane).

Thus, in the cubes representation of a low degree function $f : \mathbb{F}^m \to \mathbb{F}$, we have a table entry $T(C)$ for every cube $C$ and the value of that entry is supposed to be $T(C) = f|_C$. A general cubes table is a table $T(\cdot)$ indexed by all possible cubes and the $C$-th entry is a low degree function on the cube $C$. Each $T(C)$ is viewed as a local function. Indeed the number of bits needed to specify $T(C)$ is only $O(d^3 \log |\mathbb{F}|)$ which is much smaller than $\binom{m+d}{d} \log |\mathbb{F}|$ - the number of bits needed to represent a general degree $d$ function $f$ on $\mathbb{F}^m$.

The prover may cheat, as provers do, by giving a cubes table whose entries cannot be "glued together" into any one global low degree function. This is where the *agreement* test comes in. The verifier can check the table by reading two entries corresponding to two cubes that have a non-trivial intersection, and checking that the function $T(C_1)$ and the function $T(C_2)$ agree on points in the intersection of $C_1 \cap C_2$.

**Test 1** Cube vs. Cube agreement test.

1. Select a point $x \in \mathbb{F}^m$.
2. Pick affine cubes $C_1, C_2$ randomly conditioned on $C_1, C_2 \ni x$.
3. Read $T(C_1), T(C_2)$ from the table and accept iff $T(C_1)(x) = T(C_2)(x)$.

Let $\alpha_{\mathcal{C}x\mathcal{C}}(T)$ be the *agreement* of the table $T$, i.e. the probability of acceptance of the test.

The test is local in that it accesses only two cubes. Different tests may differ in the distribution underlying the agreement test (for example, Raz and Safra look at two planes that intersect in a line, which clearly is a different distribution from choosing two planes that intersect in a point), but they all check agreement on the intersection, so we generally refer to all of these as agreement tests.

The interesting point, as proven by both Raz and Safra in [RS97], and by Arora and Sudan in [AS97], is that such tests have small soundness error. For example, the plane vs. plane theorem of Raz Safra is as follows,

**Theorem 1.1** (Raz-Safra [RS97])**.** *There is some $\delta > 0$ such that for every $d$ and prime power $q$ and every $m \geq 3$ the following holds. Let $\mathbb{F}$ be a finite field $|\mathbb{F}| = q$, and let $T(\cdot)$ be a planes table, assigning to each plane $P \subset \mathbb{F}^m$ a bivariate degree $d$ polynomial $T(P) : P \to \mathbb{F}$. Let $\alpha_{\mathcal{P}\ell\mathcal{P}}(T)$ be as defined in Test 2.*

*For every $\epsilon \geq (md/q)^\delta$, if $\alpha_{P\ell P}(T) \geq \epsilon$ then there is a degree $d$ function $g : \mathbb{F}^m \to \mathbb{F}$ such that $T(P) = g|_P$ on an $\Omega(\epsilon)$ fraction of the planes.*

A similar theorem was proven by Arora and Sudan for $T$ a lines table and for a natural test that checks if two intersecting lines agree on the point of intersection.

These results are called low degree tests although it makes sense to think of them as theorems relating local agreement to global agreement. We refer to them as low degree *agreement* test theorems.

**Test 2** The Raz-Safra Plane vs. Plane agreement test.

1. Select an affine line $\ell \subset \mathbb{F}^m$.
2. Choose affine planes $P_1, P_2$ randomly conditioned on $P_1, P_2 \supset \ell$.
3. Read $T(P_1), T(P_2)$ from the table and accept iff $T(P_1)(x) = T(P_2)(x)$ for all $x \in \ell$.

Let $\alpha_{\mathcal{P}\ell\mathcal{P}}(T)$ be the *agreement* of the table $T$, i.e. the probability of acceptance of the test.

**Towards the soundness threshold.** The most important aspect of the low degree agreement theorems of [RS97, AS97] is the fact that they have small soundness. Small soundness means that a cheating prover won't be able to fool the verifier into accepting with even a tiny $\epsilon > 0$ probability, unless the table has some non-trivial agreement with a global low degree function. Small soundness of low degree tests was used inside PCP constructions for getting PCPs with the smallest known soundness error. The fact that soundness holds for all values of $\epsilon \geq (d/q)^\delta$ was sufficient for the PCP constructions of [RS97, AS97]. It is likely that finding the minimal threshold beyond which soundness is guaranteed to hold will be important for determining the best possible PCP gaps.

Regardless of the PCP application, this encoding of a function $f$ by its restrictions to cubes (or to planes) is quite natural, and is a rare example of a property that has such strong testability. The low degree agreement test theorems guarantee that even the passing of the test with tiny $\epsilon$ probability has non-trivial structural consequences. Perhaps the best known comparable scenario is that of the long code, defined in [BGS98], that has similar properties, and for which an extensive line of work has been able to determine the precise threshold of soundness. Another setting with a similarly strong soundness is related to the inverse theorems for the Gowers uniformity norms. In that setting the function is given as a points-table, and the Gowers norm measures success in a low degree test, so it is not altogether dissimilar from the situation here.

To summarize, one of our goals is to pinpoint the absolute minimal soundness value for which a theorem as above holds. Can this threshold be, as it is in the aforementioned cases, as small as the value of a random assignment? In other words, could it be true that for every table whose agreement parameter is an additive $\epsilon > 0$ above the value that we expect from a random table, already some structure exists?

The best known value for $\delta$ for the plane vs. plane test is due to Moshkovitz and Raz who proved in [MR08] that the plane vs. plane test has soundness for all $\epsilon \geq \text{poly}(d)/q^{1/8}$. But what is the correct exponent of $q$ ?

We make progress on this question not for the plane vs. plane test but rather for the cube vs. cube test. For our test, since the intersection consists of one point, the soundness can not go below $1/q$ because the agreement of *every* table, even a random one, is always at least $1/q$.

Our main theorem is,

**Theorem 1.2.** *There exist constants $\beta_1, \beta_2 > 0$ such that for every $d$, large enough prime power $q$ and every $m \geq 3$ the following holds:*

*Let $\mathbb{F}$ be a finite field, $|\mathbb{F}| = q$. Let $T$ be a cubes table, assigning to each cube $C \subset \mathbb{F}^m$ a degree $d$ polynomial $T(C) : C \to \mathbb{F}$. Let $\alpha_{\mathcal{C}x\mathcal{C}}(T)$ be as defined in Test 1. If $\alpha_{\mathcal{C}x\mathcal{C}}(T) \geq \epsilon$ for $\epsilon \geq \beta_1 d^4/q^{1/2}$, then there is a degree $d$ function $g : \mathbb{F}^m \to \mathbb{F}$ such that $T(C) = g|_C$ on an $\beta_2\epsilon$ fraction of the cubes.*

The improvement over previous theorems is that the dependence on $q$ is $1/q^{1/2}$ compared to

$1/q^{1/8}$, It is an intriguing question whether the dependence on $q$ can be made inversely linear, i.e. $1/q$.

**Remark 1.3.** *We don't know the precise dependence of $\epsilon$ on the degree $d$. In this work we made no attempt to optimize this dependence. We would like to point out that our proof can be modified to change the dependence from $d^4$ to $d^3$. See Remark 3.14 for more details.*

**Simplified analysis.** While the line vs. line test considered by Arora and Sudan [AS97] is the most natural to come up with, it is rather difficult to analyze. In contrast, one of the captivating aspects of the Raz-Safra proof is that it is combinatorial, and the low degree aspect of the table plays a role only in that it guarantees distance between distinct polynomials on a line. Our analysis continues this combinatorial approach, and further simplifies it. Unlike the Raz-Safra proof, we do not need to use induction on the dimension of the ambient space $m$ but rather recover the global structure from $T$ "in one shot". We rely on ideas from direct product testing, [DG08, IKW12, DS14], and on some spectral properties of incidence graphs such as the cube-point graph.

**Proof Outline.** Given a table $T$, whose agreement is some small $\epsilon$, the proof must somehow come up with the global low degree function $g : \mathbb{F}^m \to \mathbb{F}$ and then argue that on many of the cubes indeed $T(C) = g|_C$. Naively, we might try to define $g$ at each point $x$ according to the most common value among all cubes containing $x$. This is a viable approach when the agreement is close to 1, as is done, e.g. in the linearity testing theorem of [BLR90]. However, when the agreement is a small $\epsilon > 0$, this will simply not work as we can see by considering the table half of whose entries are $T(C) \equiv 0$ and the other half $T(C) \equiv 1$. The agreement of this table is an impressive $\alpha_{\mathcal{C}x\mathcal{C}}(T) = 1/2$, and yet the suggested definition of $g$ according to majority will yield a random function that might be quite far from any low degree function.

We get around this problem by taking a *conditional majority*. For every point $x \in \mathbb{F}^m$ and value $\sigma \in \mathbb{F}$ we consider only cubes containing $x$ for which $T(C)(x) = \sigma$. These cubes already agree with each other on $x$ and are thus likely to agree on any other point of their intersection. Since the cubes containing $x$ cover every $y \in \mathbb{F}^m$, we can define a function $f_{x,\sigma} : \mathbb{F}^m \to \mathbb{F}$ on the entire space $\mathbb{F}^m$ by taking the most popular value among these cubes (i.e. the set of cubes whose value on $x$ is $\sigma$). We choose a best $\sigma$ for each $x$ and are left with a global function $f_x$ for each $x$.

The proof proceeds in three steps.

- Local structure: We show that this conditional majority definition is good, obtaining for each $x$ and $\sigma$ a function $f_x : \mathbb{F}^m \to \mathbb{F}$ that is "local" in that it comes from the cubes containing a point $x$. This is done in Section 3.1.
- Global Structure: We then show that there are many pairs $x, y$ for which $f_x \approx f_y$ thus finding a global $g$ that agrees with many of the cubes. This is done in Section 3.2.
- Low Degree: Finally, we show that $g$ is very close to a true low degree function. This is done by reduction to the Rubinfeld-Sudan low degree test [RS96] that works in the high-soundness regime. This is done in Section 3.3.

**Agreement tests: low degree tests and direct product tests.** The proof outline above resembles works on direct product testing, and this is no coincidence. The low degree testing setting can be generalized to a more abstract "agreement testing" in which a function $f : X \to \Sigma$ is represented

not as a truth table but as a collection of restrictions $(f|_S)_{S\in\mathcal{S}}$ where $\mathcal{S} = \{S \subset X\}$ is a collection of subsets of $X$. A natural agreement test can be defined and studied. This type of question was first suggested in work of Goldreich and Safra [GS97] in an attempt to separate the algebraic aspect of the low degree test from the combinatorial. There has been a follow-up line of work on this, [DR06, DG08, IKW12, DS14], focusing especially on the case where $X$ is a finite set, $X = [n]$, and $\mathcal{S}$ is the collection of all $k$-element subsets of $X$.

In the work here we bring some of the ideas from that line of work, most notably from [IKW12], back to the low degree testing question. The fact that our table entries have low degree gives us extra power which makes our proof simpler than that in the abstract setting, yielding a particularly direct proof of a low degree agreement test.

Our proof makes an explicit use of the expansion properties of the relevant incidence graphs (cube vs. line, cube vs. point etc.). This allows us to prove that for every table $T$, different tests have similar agreement.

**Lemma 1.4.** *Let $T$ be a planes table, and let $\alpha_{\mathcal{P}x\mathcal{P}}(T)$ be the success probability of a test with two planes that intersects on a point. Let $\alpha_{\mathcal{P}\ell\mathcal{P}}(T)$ be the success probability of Test 2, then*

$$\alpha_{\mathcal{P}x\mathcal{P}}(T)\left(-\frac{d}{q}\right) \leq \alpha_{\mathcal{P}\ell\mathcal{P}}(T) \leq \alpha_{\mathcal{P}x\mathcal{P}}(T) + \frac{1}{q}(1+o(1)).$$

In fact, we proved a more general equivalence between tests, the general statement appears on Section 4.

## 2 Preliminaries and Notations

### 2.1 Notations

All the graphs we discuss throughout the paper are bipartite bi-regular graphs. Given such graph $G$, whose sides are $A, B$ we denote by $\mathbf{1}$ the all one vector, its size will be implied by the context. For a subset of vertices $A' \subset A$, we denote by $\mathbf{1}_{A'}$ the indicator vector for $A'$. For a vertex $a \in A$, we denote by $N(a) \subseteq B$ the neighbors of $a$ in $G$.

We use normalized inner product, such that for $x, y \in \mathbb{R}^n$, $\langle x, y\rangle = \frac{1}{n}\sum_i x_i y_i$, which means that $\langle \mathbf{1}, \mathbf{1}\rangle = 1$. The norm is defined by $\|x\| = \sqrt{\langle x, x\rangle}$.

We use the notation $x \sim S$ to denote $x$ being sampled uniformly at random (u.a.r) from the set $S$, in case this set $S$ equals the entire space, we omit this symbol and simply write $\mathbf{Pr}_a$ or $\mathbf{E}_a$ to describe choosing a uniform vertex $a \in A$. We use the notation $\mathbb{I}(E)$ to denote the indicator random variable of the event $E$.

For two vectors $u, v$, we use the notation $u \overset{\gamma}{\approx} v$ if $u$ and $v$ are equal on at least $1-\gamma$ of the coordinates.

Fix a vector space $\mathbb{F}^m$. An affine space of $S$ dimension $k$ is defined by $k+1$ vectors $x_0, x_1, \ldots, x_k$ such that $x_1, \ldots, x_k$ are linearly independent,

$$S = x_0 + \mathtt{span}(x_1, \ldots, x_k) = \{x_0 + t_1 x_1 + \ldots t_k x_k \mid t_1, \ldots, t_k \in \mathbb{F}.\}$$

A *line* is a 1-dimensional affine space, a *plane* is a 2-dimensional affine space, and a *cube* is a 3-dimensional affine space. We will denote the set of all lines and cubes by $\mathcal{L}$ and $\mathcal{C}$ be respectively.

For a point $x \in \mathbb{F}^m$ let

$$\mathcal{L}_x = \{\ell \in \mathcal{L} \mid \ell \ni x\} \qquad \mathcal{C}_x = \{C \in \mathcal{C} \mid C \ni x\}.$$

Similarly for a line $\ell \in \mathcal{L}$ let $\mathcal{C}_\ell$ be the set of all cubes that contains $\ell$.

## 2.2 Spectral Expansion Properties

In this section, we prove two properties of bi-regular bipartite graphs with good spectral parameters. In an expander, the following is well known: if we sample a random neighbor of a small, but not too small, set of vertices, we get a nearly uniform distribution over the entire set of vertices. For our purposes, we will require something more. We need to consider not only the distribution over the vertices, but also the distribution over the edges. This is done in two lemmas below.

**Definition 2.1.** *Let $G = (A \cup B, E)$ be a bi-regular bipartite graph, and let $M \in \mathbb{R}^{A \times B}$ be the adjacency matrix normalized such that $\|M\mathbf{1}\| = 1$, denote by $\lambda(G)$ the value*

$$\lambda(G) = \max_{v \perp \mathbf{1}} \left\{ \frac{\|Mv\|}{\|v\|} \right\}.$$

This is really the second largest singular value of $M$, with a different normalization (such that the maximal singular value equals 1).

**Definition 2.2.** *Let $G = (A \cup B, E)$ be a bi-regular bipartite graph and let $B' \subseteq B$ be a subset of vertices. Define the following two distributions $D_i : A \times B \cup \perp \to [0, 1]$ for $i = 1, 2$.*

- *$D_1$ : Pick $b \in B'$ u.a.r. then pick $a \in N(b)$ u.a.r.*
- *$D_2$ : Pick $a \in A$ u.a.r. If $B' \cap N(a) = \emptyset$, return $\perp$. Else, pick $b \in N(a) \cap B'$ u.a.r.*

Clearly if $B' = B$ then $D_1 = D_2$. Moreover, if $G$ is sufficiently expanding, then even for smaller $B' \subsetneq B$, the distributions are similar. Indeed, for any event defined on the edges, i.e. a subset $E' \subset E$, the following lemma shows that the probability of $E'$ is roughly the same under the two distributions.

**Lemma 2.3.** *Let $D_1, D_2$ as defined in Definition 2.2. Let $G = (A \cup B, E)$ be a bi-regular bipartite graph, then for every subset $B' \subset B$ of measure $\mu > 0$ and every $E' \subset E$*

$$\left| \Pr_{(a,b)\sim D_1}[(a, b) \in E'] - \Pr_{(a,b)\sim D_2}[(a, b) \in E'] \right| \leq \frac{\lambda(G)}{\sqrt{\mu}}.$$

*Where it is understood that if $D_2$ output $\perp$, we treat it as if $(a, b) \notin E'$.*

We now state a similar lemma, for sampling two adjacent edges instead of a single edge. We will need the graph to satisfy one more requirement.

**Definition 2.4.** *Let $G = (A \cup B, E)$ be a bi-regular bipartite graph, such that every two distinct $b_1, b_2 \in B$ have exactly the same number of common neighbors (i.e for all distinct $b_1, b_2 \in B$, $|N(b_1) \cap N(b_2)|$ is the same), and this number is non-zero. Let $B' \subseteq B$ be a subset of vertices, we define the following distributions $D_i : (A \times B \times B) \cup \perp \to [0, 1]$, for $i = 3, 4$.*

- *$D_3$ : Pick $b_1, b_2 \in B'$ u.a.r. then pick $a \in N(b_1) \cap N(b_2)$ u.a.r.*
- *$D_4$ : Pick $a \in A$ u.a.r. If $B' \cap N(a) = \emptyset$, return $\perp$. Else, pick $b_1, b_2 \in N(a) \cap B'$ u.a.r.*

**Lemma 2.5.** *Let $D_3, D_4$ be as defined in Definition 2.4. Let $G = (A \cup B, E)$ be a bi-regular bipartite graph, such that every two distinct $b_1, b_2 \in B$ have exactly the same number of common neighbors (i.e for all distinct $b_1, b_2 \in B$, $|N(b_1) \cap N(b_2)|$ is the same), and this number is non-zero. Then for every subset $B' \subset B$ of measure $\mu > 0$ and every $E' \subset E$*

$$\left| \Pr_{a,b_1,b_2 \sim D_3}[(a,b_1)(a,b_2) \in E'] - \Pr_{a,b_1,b_2 \sim D_4}[(a,b_1)(a,b_2) \in E'] \right| \leq \frac{2\lambda(G)}{\mu} + \frac{1}{\mu^2 d_A} + \frac{1}{\mu^2 |B|},$$

*where $d_A$ is the degree on $A$ side, and it is understood that if $D_4$ output $\perp$, we treat it as if $(a,b) \notin E'$.*

The proofs of these two lemmas appear in Appendix B.

### 2.3 Inclusion Graphs and Their Spectral Gap

We record here the expansion of several bi-partite *inclusion graphs* that will be relevant for our analysis. We prove the claims about these spectral gaps in Appendix A. Unless otherwise stated, $G(A, B)$ denotes a bipartite inclusion graph between $A$ and $B$ where $a \in A$ is connected to $b \in B$ if $a \subseteq b$. The relation of containment will be clear from the sets $A$ and $B$.

For example, the in the graph $G_1(\mathcal{L} \setminus \mathcal{L}_x, \mathcal{C}_x)$, the left side vertices $A$ are all the lines that do not contain $x \in \mathbb{F}^m$, and the right side vertices are all the cubes that contain $x$. There is an edge between a line $\ell$ and a cube $C$ if $\ell \subset C$.

Recall Definition 2.1 of $\lambda(G)$ for a bipartite graph $G$.

**Lemma 2.6.** *We have for every $m \geq 6$,*

(1) *For $G_1(\mathcal{L} \setminus \mathcal{L}_x, \mathcal{C}_x)$ , $\lambda(G_1) \approx \frac{1}{\sqrt{q}}$.*

(2) *For $G_2(\mathcal{L}_x, \mathcal{C}_x))$ , $\lambda(G_2) \approx \frac{1}{q}$.*

(3) *For $G_3(\mathbb{F}^m \setminus \ell, \mathcal{C}_\ell)$ , $\lambda(G_3) \approx \frac{1}{\sqrt{q}}$.*

(4) *For $G_4(\mathbb{F}^m, \mathcal{C})$ , $\lambda(G_4) \approx \frac{1}{q^{3/2}}$.*

(5) *For $G_5(\mathbb{F}^m \setminus \{x\}, \mathcal{C}_x)$ , $\lambda(G_5) \approx \frac{1}{q}$.*

*And for every $m \geq 3$*

(6) *For $G_6(\mathbb{F}^m, \mathcal{L})$, $\lambda(G_6) \approx \frac{1}{\sqrt{q}}$.*

*where $\approx$ denotes equality up to a multiplicative factor of $1 \pm o(1)$, and $o(1)$ denotes a function that approaches zero as $q \to \infty$.*

In general one can see that $\lambda \approx \frac{1}{\sqrt{q^p}}$ where $p$ is the number of degrees of freedom left after choosing a left hand vertex. We prove this lemma in Appendix A.

# 3 Proof of the Main Theorem

In this section we prove Theorem 1.2 in three steps - local structure, global structure and finally proving the agreement with a low degree polynomial. These parts are proved in the subsequent subsections.

Let $T$ be a degree $d$ cubes table, i.e. for every $C \in \mathcal{C}$, $T(C) : C \to \mathbb{F}$ is a degree $d$ polynomial. Further assume that $\alpha_{\mathcal{C}x\mathcal{C}}(T) \geq \epsilon$, where $\epsilon = \Omega(d^4/\sqrt{q})$.

## 3.1 Local Structure

In this section we show that for many points $x \in \mathbb{F}^m$, there exists a function $f_x : \mathbb{F}^m \to \mathbb{F}$ for which $f_x|_C \overset{2\gamma}{\approx} T(C)$ for a good fraction of the cubes containing $x$, for $\gamma = \Omega(1/d^3)$. Recall that $\overset{2\gamma}{\approx}$ means that the two functions agree on $1 - 2\gamma$ fraction of the points in their domain.

For each $x \in \mathbb{F}^m$ and $\sigma \in \mathbb{F}$, we define

$$\mathcal{C}_{x,\sigma} = \{C \in \mathcal{C}_x | T(C)(x) = \sigma\}.$$

Following [IKW12] we have the following important definition,

**Definition 3.1** (Excellent pair)**.** *$(x, \sigma)$ is $(\frac{\epsilon}{2}, \gamma)$-excellent if:*

1. $\mathbf{Pr}_{C \in \mathcal{C}_x}[C \in \mathcal{C}_{x,\sigma}] \geq \frac{\epsilon}{2}$.
2. *Let $C_1, \ell, C_2$ be chosen by the following probability distribution, $C_1 \in \mathcal{C}_{x,\sigma}$ u.a.r, $\ell \subset C_1$ a random line that contains $x$ and $C_2 \in \mathcal{C}_{x,\sigma} \cap \mathcal{C}_\ell$ (a random cube in $\mathcal{C}_{x,\sigma}$ that contains $\ell$).*

$$\mathbf{Pr}_{C_1,\ell,C_2}[T(C_1)_{|\ell} \neq T(C_2)_{|\ell}] \leq \gamma.$$

*A point $x \in \mathbb{F}^m$ is $(\frac{\epsilon}{2}, \gamma)$-excellent, if exists $\sigma \in \mathbb{F}$ such that $(x, \sigma)$ is $(\frac{\epsilon}{2}, \gamma)$-excellent.*

Note that in the definition of excellent, the marginal distribution of both $C_1, C_2$ is uniform in $\mathcal{C}_{x,\sigma}$.

In the sequel, we fix $\gamma = \Omega(1/d^3)$ and say that a point is excellent if it is $(\frac{\epsilon}{2}, \gamma)$-excellent. We now state the main lemma in this section.

**Lemma 3.2** (Local Structure)**.** *For $\gamma = \Omega(\frac{1}{d^3})$, let $T$ be a cubes table that passes Test 1 with probability larger than $\epsilon = \Omega(\frac{d^4}{\sqrt{q}})$, then at least $\frac{\epsilon}{3}$ of the points $x \in \mathbb{F}^m$ are excellent, and for each excellent $x$ there exist a function $f_x : \mathbb{F}^m \to \mathbb{F}$ such that*

$$\mathbf{Pr}_{C \sim \mathcal{C}_x}[T(C) \overset{2\gamma}{\approx} f_{x|_C}] \geq \frac{\epsilon}{4}.$$

We will consider the distribution $\mathcal{D}$ on $(x, \ell, C_1, C_2)$ obtained by choosing $x$ uniformly, choosing $\ell \in \mathcal{L}_x$ uniformly, and then choosing $C_1, C_2 \in \mathcal{C}_\ell$ uniformly.

This distribution induces a distribution $(x, T(C_1)(x))$ on pairs of point $x$ and value $\sigma \in \mathbb{F}$.

**Claim 3.3.** *For every $\gamma = \Omega(\frac{1}{d^3})$,*

$$\mathbf{Pr}_{(x,\sigma)}[(x, \sigma) \text{ is } (\tfrac{\epsilon}{2}, \gamma) \text{ - excellent}] \geq \frac{\epsilon}{3}.$$

**Proof:** We consider $(x, \ell, C_1, C_2)$ chosen according to $\mathcal{D}$, and we note that the marginal distribution over all elements is uniform. We also write $\sigma = T(C_1)(x)$. We define the following events on $(x, \ell, C_1, C_2)$:

1. $E$ : "$\ell$ is confusing for $x$": $T(C_1)(x) = T(C_2)(x)$, $T(C_1)_{|\ell} \neq T(C_2)_{|\ell}$.
2. $H$ : "$x, C_1$ is heavy": $\mathbf{Pr}_{C \sim \mathcal{C}_x}[T(C)(x) = T(C_1)(x)] \geq \frac{\epsilon}{2}$

Since $T(C_1)_{|\ell}, T(C_2)_{|\ell}$ are two degree $d$ polynomials, and $x$ is a random point in $\ell$,

$$\mathop{\mathbf{Pr}}_{(x,\ell,C_1,C_2)}[E] \leq \frac{d}{q}.$$

Using the fact that $\alpha_{\mathcal{C}x\mathcal{C}}(T) \geq \epsilon$, and averaging, we get

$$\mathop{\mathbf{Pr}}_{(x,\ell,C_1,C_2)}[H] \geq \frac{\epsilon}{2}. \tag{1}$$

Instead of picking $C_1$ as a uniform cube containing $x$, we can choose it by the following process, pick $\sigma$ proportional to its weight in $\mathcal{C}_x$, then pick $C_1 \sim \mathcal{C}_{x,\sigma}$. This process describes the same distribution.

Note that after deciding $x, \sigma$, the event $H$ is already determined, so (1) becomes $\mathbf{Pr}_{x,\sigma}[H] \geq \epsilon/2$. Also, notice that conditioned on $x, \sigma$, the distribution $\mathcal{D}$ is choosing $C_1$ uniformly from $\mathcal{C}_{x,\sigma}$ and then $\ell \subset C_1$ a random line containing $x$ and then $C_2$ a random cube containing $\ell$ (and we do not require that $T(C_2)(x) = \sigma$). The event $H$ is already fixed by $x, \sigma$, but the event $E$ will occur only if $C_2 \in \mathcal{C}_{x,\sigma}$ and also $T(C_1)|_\ell \neq T(C_2)|_\ell$.

We want to bound the probability of $x, \sigma$ such that $H = 1$, but $\mathbf{E}_{C_1,\ell,C_2}[E|x, \sigma] \leq \gamma \cdot \frac{\epsilon}{2}$. We know that

$$\mathop{\mathbf{E}}_{x,\sigma}[\mathbf{Pr}[H \wedge E \mid x, \sigma]] = \mathbf{Pr}[H \wedge E] \leq \mathbf{Pr}[E] \leq \frac{d}{q}.$$

Therefore, by averaging, the probability over $x, \sigma$ that we have $\mathbf{Pr}[H \wedge E|x, \sigma] > \epsilon\gamma/2$ is at most $\frac{d/q}{\epsilon\gamma/2}$. So for at least $\epsilon/2 - \frac{d/q}{\epsilon\gamma/2} \geq \epsilon/3$ of the pairs $x, \sigma$, we have that both $H$ occurs, and that $\mathbf{E}_{C_1,\ell,C_2}[E|x, \sigma] \leq \epsilon\gamma/2$.

We end by showing that such $x, \sigma$ are excellent. The first requirement follows by the fact that $H$ occurs, for the second we need to show that for $C_1 \in \mathcal{C}_{x,\sigma}$, a uniform $\ell \in \mathcal{C}_1$ and a uniform $C_2 \in \mathcal{C}_{x,\sigma} \cap C_\ell$ the probability of $T(C_1)_{|\ell} \neq T(C_2)_{|\ell}$ is lower than $\gamma$.

We notice that after fixing $(x, \sigma)$, the distribution $\mathcal{D}$ chooses $C_1 \in \mathcal{C}_{x,\sigma}$, a uniform $\ell \in \mathcal{C}_1$, but then a uniform $C_2 \in \mathcal{C}_\ell$.

The event $E$ can be written as $E = E_1 \wedge E_2$ where $E_1$ is the event "$T(C_1)(x) = T(C_2)(x)$" and $E_2$ is the event "$T(C_1)_{|\ell} \neq T(C_2)_{|\ell}$". In this notation

$$\begin{aligned}
\mathop{\mathbf{E}}_{C_1,\ell,C_2}[E|x, \sigma] &= \mathop{\mathbf{E}}_{C_1,\ell,C_2}[E_1 \wedge E_2|x, \sigma] \\
&= \mathop{\mathbf{E}}_{C_1,\ell,C_2}[E_1|x, \sigma] \mathop{\mathbf{E}}_{C_1,\ell,C_2}[E_2|E_1, x, \sigma] \\
&\geq \frac{\epsilon}{2} \cdot \mathop{\mathbf{E}}_{C_1,\ell,C_2}[E_2|E_1, x, \sigma]. && \text{(since } H \text{ occurs)}
\end{aligned}$$

We notice that if $E_1$ occurs, then $C_2 \in \mathcal{C}_{x,\sigma}$, therefore

$$\mathop{\mathbf{E}}_{C_1,\ell,C_2}[T(C_1)_{|\ell} \neq T(C_2)_{|\ell}|C_2 \in \mathcal{C}_{x,\sigma}, x, \sigma] \leq \frac{2}{\epsilon} \cdot \mathop{\mathbf{E}}_{C_1,\ell,C_2}[E|x, \sigma] \leq \frac{2}{\epsilon}\frac{\epsilon}{2}\gamma \leq \gamma,$$

which means that $(x,\sigma)$ is $(\frac{\epsilon}{2},\gamma)$ - excellent. ■

For each $(x,\sigma)$ we define $f_{x,\sigma}$ by plurality over all cubes $C \in \mathcal{C}_{x,\sigma}$.

**Definition 3.4.** *For a pair $(x,\sigma)$ define a function $f_{x,\sigma}:\mathbb{F}^m \to \mathbb{F}$ as follows:*

$$f_{x,\sigma}(y) = \underset{C\sim\mathcal{C}_y\cap\mathcal{C}_{x,\sigma}}{\texttt{argmax}} \{T(C)(y)\}.$$

*If $\mathcal{C}_y \cap \mathcal{C}_{x,\sigma} = \emptyset$, define $f_{x,\sigma}(y)$ arbitrarily.*

**Claim 3.5.** *For an $(\frac{\epsilon}{2},\gamma)$ excellent pair $(x,\sigma)$,*

$$\underset{C\sim\mathcal{C}_{x,\sigma},y\sim C}{\mathbf{Pr}}[f_{x,\sigma}(y) = T(C)(y)] \geq 1-\gamma.$$

**Proof:** Fix an $(\frac{\epsilon}{2},\gamma)$ excellent pair $(x,\sigma)$, and denote $f = f_{x,\sigma}$. If we pick a uniform $C_1 \in \mathcal{C}_{x,\sigma}$, then $y \in C_1$ such that $y \neq x$, and a uniform $C_2 \in \mathcal{C}_{x,\sigma} \cap \mathcal{C}_y$, then

$$\underset{C_1,y,C_2}{\mathbf{Pr}}[T(C_1)(y) \neq T(C_2)(y)] \leq \underset{C_1,y,C_2}{\mathbf{Pr}}[T(C_1)_{|\ell(x,y)} \neq T(C_2)_{|\ell(x,y)}] \leq \gamma,$$

since $(x,\sigma)$ is $(\frac{\epsilon}{2},\gamma)$ excellent.
For each $y$, denote $\gamma_y = \mathbf{Pr}_{C_1,C_2\sim\mathcal{C}_{x,\sigma}\cap\mathcal{C}_y}[T(C_1)(y) \neq T(C_2)(y)]$. From the above we get that $\mathbb{E}_y[\gamma_y] \leq \gamma$, where $y$ is distributed according to it's weight in $\mathcal{C}_{x,\sigma}$. For each $y$,

$$\begin{aligned}
1-\gamma_y &= \sum_{\theta\in\mathbb{F}} \underset{C\sim\mathcal{C}_{x,\sigma}\cap\mathcal{C}_y}{\mathbf{Pr}}[T(C)(y)=\theta]^2 \\
&\leq \underset{C\sim\mathcal{C}_{x,\sigma}\cap\mathcal{C}_y}{\mathbf{Pr}}[T(C)(y)=f(y)] \sum_{\theta\in\mathbb{F}} \underset{C\sim\mathcal{C}_{x,\sigma}\cap\mathcal{C}_y}{\mathbf{Pr}}[T(C)(y)=\theta] \quad (f(y) \text{ is the most frequent value}) \\
&\leq \underset{C\sim\mathcal{C}_{x,\sigma}\cap\mathcal{C}_y}{\mathbf{Pr}}[T(C)(y)=f(y)].
\end{aligned}$$

Since it is true for each $y$, it is also true when taking expectation over $y$, for any distribution:

$$\underset{C\sim\mathcal{C}_{x,\sigma},y\sim C}{\mathbf{Pr}}[f(y)=T(C)(y)] = \underset{y}{\mathbf{E}}\left[\underset{C\sim\mathcal{C}_{x,\sigma}\cap\mathcal{C}_y}{\mathbf{E}}[\mathbb{I}(T(C)(y)=f(y))]\right] \geq \underset{y}{\mathbf{E}}[1-\gamma_y] \geq 1-\gamma.$$

In expectation, each $y$ is chosen with probability proportional to it's weight in $\mathcal{C}_{x,\sigma}$, as before. ■

**Proof of Lemma 3.2:** From Claim 3.3 we know that the probability of $(x,\sigma)$ to be $(\frac{\epsilon}{2},\gamma)$-excellent is at least $\frac{\epsilon}{3}$. Since $x$ is chosen uniformly, it means that for at least $\frac{\epsilon}{3}$ of the inputs $x \in \mathbb{F}^m$ there exists some $\sigma \in \mathbb{F}$ such that $(x,\sigma)$ is excellent. If there is more than one such $\sigma$ choose one arbitrarily.
Fixing an excellent $x$, let $\sigma$ be the value such that $(x,\sigma)$ is excellent. For this $\sigma$, $\mathbf{Pr}_{C\in\mathcal{C}_x}[C \in \mathcal{C}_{x,\sigma}] \geq \frac{\epsilon}{2}$. From Claim 3.5, $\mathbf{Pr}_{C\sim\mathcal{C}_{x,\sigma},y\sim C}[f_{x,\sigma}(y) = T(C)(y)] \geq 1-\gamma$. By averaging, at least half of the cubes $C \in \mathcal{C}_{x,\sigma}$ satisfy $\mathbf{Pr}_{y\sim C}[f_{x,\sigma}(y) = T(C)(y)] \geq 1-2\gamma$. For all these cubes $T(C) \overset{2\gamma}{\approx} f_{x,\sigma}$, and they are at least $\frac{\epsilon}{4}$ fraction of the cubes in $\mathcal{C}_x$. ■

### 3.2 Global Structure

In this section, we prove the following lemma:

**Lemma 3.6** (Global Structure). *Let $T$ be a cubes table that passes Test 1 with probability at least $\epsilon = \Omega(\frac{d^4}{\sqrt{q}})$, then for every $\gamma = \Omega(\frac{1}{d^3})$, there exists an $(\frac{\epsilon}{2}, \gamma)$-excellent $x$ such that $f = f_x : \mathbb{F}^m \to \mathbb{F}$ satisfies*

$$\Pr_C[T(C) \overset{32\gamma}{\approx} f_{|C}] \geq \frac{\epsilon}{16}.$$

Let $X^\star \subseteq \mathbb{F}^m$ the set of $(\frac{\epsilon}{2}, \gamma)$ excellent points.

The main idea in the proof of the global structure, is showing that there exist many pairs of excellent points $x, y \in X^\star$, such that for many cubes $C$, the $T(C)$ is similar both to $f_x$ and to $f_y$ (Claim 3.8). If this is the case, then the functions $f_x, f_y$ must be very similar (Claim 3.9). Finally, the lemma is proven by averaging and finding a single $x$ such that $f_x$ agrees simultaneously with many of the $f_y$'s and their supporting cubes.

**Definition 3.7** (Supporting cubes). *For any excellent $x \in X^\star$, we denote by $F_x$ the set of cubes "supporting" $f_x$,*

$$F_x = \left\{ C \in \mathcal{C}_x \,\middle|\, T(C) \overset{2\gamma}{\approx} f_{x|_C} \right\}.$$

**Claim 3.8.** *Let $\mathcal{D}$ be the following process: choose $x, y \in X^\star$ independently and uniformly at random, let $C$ be a random cube containing both $x$ and $y$. Then*

$$\Pr_{x,y,C\sim D}[C \in F_x \cap F_y] \geq \frac{\epsilon^2}{26}.$$

**Proof:** Since each $x \in X^\star$ is excellent, we know from the local structure lemma, Lemma 3.2, that $\Pr_{C\sim\mathcal{C}_x}[C \in F_x] \geq \frac{\epsilon}{4}$. This is of course also true when taking a uniform $x \in X^\star$, thus, $\Pr_{x\sim X^\star, C\sim\mathcal{C}_x}[C \in F_x] \geq \frac{\epsilon}{4}$.

From Lemma 2.6(4) , the inclusion graph $G = G(\mathbb{F}^m, \mathcal{C})$ has $\lambda(G) = \lambda \leq (1 + o(1))\frac{1}{q^{3/2}}$. Denote the measure of $X^\star$ by $\mu$, from Lemma 3.2, $\mu \geq \frac{\epsilon}{3}$. Hence, by the application of Lemma 2.3 on the graph $G$ with $A = \mathcal{C}, B = \mathbb{F}^m$ and $B' = X^\star$, we get

$$\left| \Pr_{x\sim X^\star, C\sim\mathcal{C}_x}[C \in F_x] - \Pr_{C\sim\mathcal{C}, x\sim C\cap X^\star}[C \in F_x] \right| \leq \frac{\lambda}{\sqrt{\mu}} \leq \frac{2\lambda}{\sqrt{\epsilon}}. \tag{2}$$

For each $C \in \mathcal{C}$, let $p_C = \Pr_{x\sim C\cap X^\star}[C \in F_x]$, this measures for every cube $C$ how many points $x \in C$ are such that $f_{x|_C} \overset{2\gamma}{\approx} T(C)$. In this notation, (2) implies $\mathbf{E}_C[p_C] \geq \frac{\epsilon}{4} - \frac{2\lambda}{\sqrt{\epsilon}} \geq \frac{\epsilon}{5}$. We can use this to bound the probability of the event $C \in F_x \cap F_y$ by first choosing $C$, then two independent points in $C \cap X^\star$,

$$\Pr_{\substack{C\sim\mathcal{C} \\ x,y\sim C\cap X^\star}}[C \in F_x \cap F_y] = \mathop{\mathbf{E}}_C[p_C^2] \geq \left(\mathop{\mathbf{E}}_C[p_C]\right)^2 \geq \frac{\epsilon^2}{25}.$$

We observe that this distribution is very similar to the required distribution $D$. The only difference is that here we first pick $C \in \mathcal{C}$ and then two excellent points in $C$, whereas in $D$ we first pick two points in $X^\star$ and then a common neighbor $C$. The graph $G$ satisfies that every two distinct points

$x, y \in \mathbb{F}^m$ have exactly the same number of common neighbors. Therefore, we can use Lemma 2.5 on the graph $G$ with $A = \mathcal{C}, B = \mathbb{F}^m$ and $B' = X^\star$ to get

$$\left| \mathop{\mathbf{Pr}}_{\substack{C\sim\mathcal{C} \\ x,y\sim C\cap X^\star}}[C \in F_x \cap F_y] - \mathop{\mathbf{Pr}}_{x,y,C\sim D}[C \in F_x \cap F_y] \right| \leq \frac{2\lambda}{\mu} + \frac{1}{\mu^2 d_A} + \frac{1}{\mu^2 |B|} \leq \frac{6\lambda}{\epsilon} + \frac{9}{q^m\epsilon^2} + \frac{9}{q^3\epsilon^2}.$$

Recall that $\lambda \leq (1+o(1))\frac{1}{q^{3/2}}$ and since $\epsilon = \Omega(\frac{d^4}{\sqrt{q}})$, we conclude that $\mathbf{Pr}_{x,y,C\sim D}[C \in F_x \cap F_y] \geq \frac{\epsilon^2}{25} - \frac{6\lambda}{\epsilon} - \frac{9}{q^m\epsilon^2} - \frac{9}{q^3\epsilon^2} \geq \frac{\epsilon^2}{26}$. ■

**Claim 3.9.** *Let $x \neq y \in X^\star$, and let $\ell$ be the line containing $x$ and $y$, if $\mathbf{Pr}_{C\sim\mathcal{C}_\ell}[C \in F_x \cap F_y] \geq \frac{\epsilon^2}{100}$ then $f_x \overset{5\gamma}{\approx} f_y$.*

**Proof:** Consider the graph $G = G(\mathbb{F}^m \setminus \ell, \mathcal{C}_\ell)$. This is a bi-regular bipartite graph, and by Lemma 2.6(3) it has $\lambda = \lambda(G) \leq (1+o(1))\frac{1}{\sqrt{q}}$. Let $F = F_x \cap F_y$. By assumption, $F$ has measure at least $\frac{\epsilon^2}{100}$ inside $\mathcal{C}_\ell$.
We denote by $E' \subset E$ the edges of $G$ that indicate agreement with both $f_x$ and $f_y$,

$$E' = \{(z, C) \mid T(C)(z) = f_x(z) = f_y(z)\}.$$

Every cube $C \in F$ has $1-2\gamma$ of the points $z \in C$ satisfying $T(C)(z) = f_x(z)$ and $1-2\gamma$ of the points satisfying $T(C)(z) = f_y(z)$. By a union bound we get $\mathbf{Pr}_{C\in F, z\in N(C)}[(z, C) \in E'] \geq 1 - 4\gamma$. By Lemma 2.3 on $G$ when $A = \mathbb{F}^m \setminus \ell, B = \mathcal{C}_\ell, B' = F$,

$$\left| \mathop{\mathbf{Pr}}_{C\sim F, z\sim N(C)}[(z, C) \in E'] - \mathop{\mathbf{Pr}}_{z,C\sim N(z)\cap F}[(z, C) \in E'] \right| \leq \frac{20\lambda}{\epsilon},$$

which means that $\mathbf{Pr}_{z\sim\mathbb{F}^m, C\sim N(z)\cap F}[(z, C) \in E'] \geq 1 - 4\gamma - \frac{20\lambda}{\epsilon} \geq 1 - 5\gamma$. By the definition of $E'$, for each point $z \in \mathbb{F}^m$ that has an adjacent edge in $E'$, $f_x(z) = f_y(z)$. This means that

$$\mathop{\mathbf{Pr}}_{z}[f_x(z) = f_y(z)] \geq \mathop{\mathbf{Pr}}_{z}[\exists C \textbf{ s.t. } (z, C) \in E'] \geq \mathop{\mathbf{Pr}}_{z,C\sim N(z)\cap F}[(z, C) \in E'] \geq 1 - 5\gamma.$$

■

The above claim showed that if two functions have a large set of cubes on which they almost agree then these functions are similar. In order to prove the global structure, we also need to show that in this case, most of $C \in F_y$ will also be close to $f_x$.

**Claim 3.10.** *Let $x, y \in X^\star$ such that $f_x \overset{5\gamma}{\approx} f_y$, then*

$$\mathop{\mathbf{Pr}}_{C\sim F_y}[T(C) \overset{32\gamma}{\approx} f_{x|_C}] \geq \frac{1}{2}.$$

Note that the function $f_x$ may not be a low degree polynomial, so $T(C) \overset{32\gamma}{\approx} f_{x|_C}$ doesn't imply equality.

**Proof:** Let $G = G(\mathbb{F}^m \setminus \{y\}, \mathcal{C}_y)$, by Claim 2.6(5) it has $\lambda = \lambda(G) \approx \frac{1}{q}$. First, we denote by $E'_y$ the following set of edges,

$$E'_y = \{(z, C) \mid T(C)(z) = f_y(z)\}.$$

For each $C \in F_y$, we know that $\mathbf{Pr}_{z \in N(C)}[(z, C) \in E'_y] \geq 1 - 2\gamma$. From Lemma 2.3 on $G$ when $A = \mathbb{F}^m \setminus y, B = \mathcal{C}_y, B' = F_y$, we know that

$$\left| \mathop{\mathbf{Pr}}_{C \sim F_y, z \sim N(C)}[(z, C) \in E'_y] - \mathop{\mathbf{Pr}}_{z, C \in N(z) \cap F_y}[(z, C) \in E'_y] \right| \leq \frac{4\lambda}{\epsilon},$$

since the measure of $F_y$ is at least $\frac{\epsilon}{4}$. This implies that $\mathbf{Pr}_{z, C \in N(z) \cap F_y}[(z, C) \in E'_y] \geq 1 - 3\gamma$.
We define a second set of edges, $E'_x$ to be the same only for $f_x$,

$$E'_x = \{(z, C) \mid T(C)(z) = f_x(z)\}.$$

We notice that if $z$ is a point such that $f_x(z) = f_y(z)$, then $(z, C) \in E'_y \Rightarrow (z, C) \in E'_x$.

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{z, C \sim N(z) \cap F_y}[(z, C) \in E'_x] \geq & \mathop{\mathbf{Pr}}_{z}[f_x(z) = f_y(z)] \cdot \mathop{\mathbf{Pr}}_{z, C \sim N(z) \cap F_y}[(z, C) \in E'_y \mid f_x(z) = f_y(z)] \\
\geq & (1 - 5\gamma) \cdot \mathop{\mathbf{Pr}}_{z, C \sim N(z) \cap F_y}[(z, C) \in E'_y \mid f_x(z) = f_y(z)] \qquad (\text{since } f_x \overset{5\gamma}{\approx} f_y) \\
\geq & (1 - 5\gamma) \cdot \left( \mathop{\mathbf{Pr}}_{z, C \sim N(z) \cap F_y}[(z, C) \in E'_y] - 5\gamma \right) \\
\geq & 1 - 15\gamma.
\end{aligned}$$

Therefore, we can use Lemma 2.3 again on the same graph $G$ and set $F_y$, now with the edge set $E'_x$, to conclude that

$$\mathop{\mathbf{Pr}}_{C \sim F_y, z \sim N(C)}[(z, C) \in E'_x] \geq \mathop{\mathbf{Pr}}_{z, C \sim N(z) \cap F_y}[(z, C) \in E'_x] - \frac{4\lambda}{\epsilon} \geq 1 - 16\gamma,$$

By averaging, at least half of $C \in F_y$ satisfies $T(C) \overset{32\gamma}{\approx} f_{x|_C}$. ∎

We are now ready to prove the global structure.

**Proof of Lemma 3.6:** Let $T$ be the cubes table that passes Test 1 with probability at least $\epsilon = \Omega(\frac{d^4}{\sqrt{q}})$. From the local structure, Lemma 3.2, we know that there exists a set $X^\star$ of excellent points, such that each $x \in X^\star$ has a function $f_x$, and $|F_x| \geq \frac{\epsilon}{4} |\mathcal{C}_x|$.

From Claim 3.8, we know that $\mathbf{Pr}_{x, y, C \sim D}[C \in F_x \cap F_y] \geq \frac{\epsilon^2}{26}$, when $x, y$ are chosen uniformly from $X^\star$ and $C$ is a common neighbor. Therefore, there must be $x \in X^\star$ such that $\mathbf{Pr}_{y \sim X^\star, C \sim N(x) \cap N(y)}[C \in F_x \cap F_y] \geq \frac{\epsilon^2}{26}$.

Fix such $x \in X^\star$, and let $X'$ be the set of $y \in X^\star$ such that $|F_x \cap F_y| \geq \frac{\epsilon^2}{100} |\mathcal{C}_\ell|$. By averaging, $|X'| \geq \frac{\epsilon^2}{100} |X^\star| \geq \frac{\epsilon^3}{400} |\mathbb{F}|^m$.

By Claim 3.9, for all $y \in X'$, $f_y \overset{5\gamma}{\approx} f_x$. For each $y \in X'$, let

$$F'_y = \{C \in F_y \mid T(C) \overset{32\gamma}{\approx} f_{x|_C}\}.$$

At this point we have a large collection of $y$'s and for each one a large collection of cubes $F'_y$ such that all of these support the same function $f_x$. It is immediate that $f_x$ is supported by some $\text{poly}(\epsilon)$ fraction of all of the cubes. Since we are aiming for a better quantitative bound of $\Omega(\epsilon)$ fraction of $\mathcal{C}$, we will rely on the expansion once more.

In order to finish the proof, we need to show that $\left|\cup_{y\in X'} F'_y\right| \geq \frac{\epsilon}{16}|\mathcal{C}|$.

Let $G = G(\mathbb{F}^m, \mathcal{C})$, by Lemma 2.6(4) $\lambda(G) \leq q^{-\frac{3}{2}}$. We use $X'$ as the set of vertices, and define

$$E' = \{(y, C) \mid T(C) \overset{32\gamma}{\approx} f_{x|_C}\}.$$

By Lemma 2.3 on $G$ with $A = \mathcal{C}, B = \mathbb{F}^m, B' = X'$,

$$\left|\Pr_{y\sim X', C\sim N(y)}[(y, C) \in E'] - \Pr_{C\sim\mathcal{C}, y\sim N(C)\cap X'}[(y, C) \in E']\right| \leq \frac{20\lambda}{\sqrt{\epsilon^3}} \leq \frac{20q^{-\frac{3}{2}}}{q^{-\frac{3}{4}}} \leq 20q^{-\frac{3}{4}} \leq \frac{\epsilon}{16},$$

where we used the fact that $\epsilon \geq \frac{1}{\sqrt{q}}$.

Claim 3.10 lets us bound the first term on the left, since for each $y \in X'$, $\Pr_{C\sim N(y)}[C \in F'_y] \geq \frac{1}{2}\Pr_{C\sim N(y)}[C \in F_y] \geq \frac{\epsilon}{8}$. Thus,

$$\Pr_{C\sim\mathcal{C}, y\sim N(C)\cap X'}[(y, C) \in E'] \geq \frac{\epsilon}{8} - \frac{\epsilon}{16} = \frac{\epsilon}{16}.$$

We notice that a cube with even a single adjacent edge in $E'$ satisfies $T(C) \overset{32\gamma}{\approx} f_{x|_C}$, so we are done. ■

### 3.3 Low Degree

The last step is to prove that the global function discovered in the previous section can be modified to make it a low degree function, while still maintaining large support for it among the cubes.

**Theorem 3.11** (Theorem 1.2 restated)**.** *For every $d$ and large enough prime power $q$ and every $m \geq 3$ the following holds. Let $T$ be a cubes table that passes Test 1 with probability at least $\epsilon = \Omega(\frac{d^4}{\sqrt{q}})$, then there exist a degree $d$ polynomial $g : \mathbb{F}^m \to \mathbb{F}$ such that $T(C) = g|_C$ on an $\Omega(\epsilon)$ fraction of the cubes.*

From Lemma 3.6, we get a function $f$ such that $\Omega(\epsilon)$ of the cubes have $T(C) \approx f_{|C}$. In this section, we will show that this function $f$ is close to a degree $d$ polynomial $g$. Afterwards, we also need to show that $\Omega(\epsilon)$ of the cubes satisfies $T(C) = g_{|C}$

To show the first part, we will use a robust characterization of low degree polynomials given by Rubinfeld and Sudan.

**Theorem 3.12** ([RS96, Theorem 4.1])**.** *Let $f : \mathbb{F}^m \to \mathbb{F}$ be a function, and let $N_{y,h} = \{y + i(h - y) \mid i \in \{0, \ldots, d + 1\}\}$, if $f$ satisfies*

$$\Pr_{y,h\in\mathbb{F}^m}[\exists \deg d \text{ polynomial } p \textbf{ s.t. } p_{|N_{y,h}} = f_{|N_{y,h}}] \geq 1 - \delta,$$

*for $\delta \leq \frac{1}{2(d+2)^2}$, then there exists a degree $d$ polynomial $g$ such that $f \overset{2\delta}{\approx} g$.*

For completeness, we present proof of the above theorem in Appendix C.

**Claim 3.13.** *Fix any $\gamma \leq \frac{1}{100(d+2)^3}$, let $f : \mathbb{F}^m \to \mathbb{F}$ and $x \in \mathbb{F}^m$ such that $\mathbf{Pr}_{C \in \mathcal{C}_x}[T(C) \overset{32\gamma}{\approx} f_{|C}] \geq \frac{\epsilon}{4}$, then exists a degree $d$ polynomial $g$ such that $f \overset{84d\gamma}{\approx} g$.*

**Proof:** Denote by $F \subseteq \mathcal{C}_x$ the following set

$$F = \{C \in \mathcal{C}_x \mid T(C) \overset{32\gamma}{\approx} f_{|C}\}.$$

Our first goal is to show that for nearly all lines, $f$ agrees with a low degree function on almost all of the points of the line.

Fix $C \in F$, if we pick a uniform $\ell \subset C$ we expect that $T(C)_\ell \overset{O(\gamma)}{\approx} f_{|\ell}$. Using the spectral properties we show that almost all lines satisfy this property. Let $G_C = G(A \cup B, E)$ be the following bipartite inclusion graph where $A$ is all the points in $C$, and $B$ is all the affine lines in $C$. Let $A' \subset A$ be $A' = \{y \in A \mid T(C)(y) \neq f(y)\}$, and $B' \subset B$ be $B' = \{\ell \in B \mid |N(\ell) \cap A'| \geq 40\gamma\,|N(\ell)|\}$. From Lemma 2.6(6) with $m = 3$ (we apply the lemma where "$\mathbb{F}^m$" is the cube $C$), $\lambda_C = \lambda(G_C) \leq \frac{2}{\sqrt{q}}$. We apply Lemma 2.3 on $G_C$ and the set $B'$, where the set of edges is all the edges adjacent to $A'$:

$$\left| \mathop{\mathbf{Pr}}_{y \in A, \ell \in N(y) \cap B'}[y \in A'] - \mathop{\mathbf{Pr}}_{\ell \in B', y \in N(\ell)}[y \in A'] \right| \leq \frac{\lambda_C}{\sqrt{\frac{|B'|}{|B|}}}.$$

We notice that $\mathbf{Pr}_{y \in A}[y \in A'] \leq 32\gamma$. By the definition of $B'$, $\mathbf{Pr}_{\ell \in B', y \in N(\ell)}[y \in A'] \geq 40\gamma$. Therefore $|B'| \leq \left(\frac{\lambda_C}{8\gamma}\right)^2 |B| < \gamma\,|B|$.

We have shown that for every cube $C \in F$, almost all lines in it satisfy $T(C)_\ell \overset{40\gamma}{\approx} f_{|\ell}$. Now we need to show that the set $F$ is large enough to cover $(1 - O(\gamma))$ of all the lines in $\mathcal{L}$. The inclusion graph $G = G(\mathcal{L} \setminus \mathcal{L}_x, \mathcal{C}_x)$ has $\lambda = \lambda(G) \leq \frac{1}{\sqrt{q}}$, by Lemma 2.6(1). We denote by $E'$ the set of edges $(\ell, C)$ such that $T(C)_{|\ell} \overset{40\gamma}{\approx} f_{|\ell}$. As we've seen above, for every $C \in F$, $\mathbf{Pr}_{\ell \in N(C)}[(\ell, C) \in E'] \geq 1 - \gamma$.
By Lemma 2.3 on $G$, with $A = \mathcal{L} \setminus \mathcal{L}_x$, $B = \mathcal{C}_x$, $B' = F$,

$$\left| \mathop{\mathbf{Pr}}_{\ell, C \sim N(\ell) \cap F}[(\ell, C) \in E'] - \mathop{\mathbf{Pr}}_{C \sim F, \ell \sim C}[(\ell, C) \in E'] \right| \leq \frac{\lambda}{\sqrt{\epsilon}} \leq \gamma,$$

which means that

$$\mathop{\mathbf{Pr}}_{\ell}[\exists C \textbf{ s.t. } (\ell, C) \in E'] \geq \mathop{\mathbf{Pr}}_{\ell, C \sim N(\ell) \cap F}[(\ell, C) \in E'] \geq 1 - 2\gamma.$$

This means that for $1 - 2\gamma$ of the lines in $\mathcal{L}$, $f$ agrees with a degree $d$ function on $1 - 40\gamma$ fraction of the points of each line.
We are very close to being able to apply the low degree test of Rubinfeld and Sudan [RS96], that works in the high soundness regime. For this, we need to move to neighborhoods. For $y, h \in \mathbb{F}^m$, we define the neighborhood of $y, h$,

$$N_{y,h} = \{y + i(h - y) \mid 0 \leq i \leq d + 1\}.$$

Notice that $N_{y,h} \subset \ell(y, h)$. We show that on almost all of the neighborhoods $N_{y,h}$, the function $f_{|N_{y,h}}$ equals a degree $d$ polynomial, by showing that for almost all $N_{y,h}$, there exists some cube $C$ such that $f_{|N_{y,h}} = T(C)_{|N_{y,h}}$ ($T(C)$ is a degree $d$ polynomial).

Picking a random neighborhood $N_{y,h}$ is equivalent to picking a random line $\ell \in \mathcal{L}$ and then uniform $y, h \in \ell$. We have already showed that almost all lines $\ell \in \mathcal{L}$, there exists a cube $C$ such that $T(C)_\ell \overset{\Omega(\gamma)}{\approx} f_{|\ell}$.

Now we can bound the same probability over neighborhoods

$$\begin{aligned}
\Pr_{y,h\sim\mathbb{F}^m}[\exists C \text{ s.t. } f(N_{y,h}) = T(C)(N_{y,h})] \geq & \Pr_{\ell}[\exists C \text{ s.t. } (\ell, C) \in E'] \cdot \\
& \Pr_{\ell,y,h\sim\ell}[f(N_{y,h}) = T(C)(N_{y,h}) \mid \exists C \text{ s.t. } (\ell, C) \in E'] \\
\geq & (1-2\gamma) \Pr_{\ell,y,h\sim\ell}[f(N_{y,h}) = T(C)(N_{y,h}) \mid \exists C \text{ s.t. } (\ell, C) \in E'] \\
\geq & (1-2\gamma)(1-(d+2)\cdot 40\gamma), \qquad (3) \\
\geq & 1 - 42d\gamma,
\end{aligned}$$

where (3) is due to union bound on the neighborhoods inside $\ell$. Therefore, the function $f$ equals a degree $d$ polynomial on $(1-42d\gamma)$ of the neighborhoods. Since $\gamma \leq 100(d+2)^{-3}$, by Theorem 3.12, we get that there exists a degree $d$ polynomial $g$, such that $f \overset{84d\gamma}{\approx} g$. ∎

**Proof of Theorem 3.11:** Fix the cubes table $T$, and let $f : \mathbb{F}^m \to \mathbb{F}$ be the function promised from Lemma 3.6. This function satisfies the conditions of Claim 3.13, so there exists a degree $d$ polynomial $g$ such that $f \overset{84d\gamma}{\approx} g$.

Since $g$ is a degree $d$ polynomial, for every cube $C$ either $T(C) = g_{|C}$, or else they are very different. Let $G$ be the inclusion graph $G = G(\mathbb{F}^m, \mathcal{C})$, and let

$$F = \{C \in \mathcal{C} \mid T(C) \overset{32\gamma}{\approx} f_{|C}\}$$

From Lemma 3.6, the measure of $F$ is at least $\frac{\epsilon}{16}$, let $A'$ be the set of points on which $f \neq g$. By Lemma 2.6(4), $\lambda(G) \leq q^{-\frac{3}{2}}$. We use Lemma 2.3 on $G$ with $A = \mathbb{F}^m, B = \mathcal{C}, B' = F$,

$$\left| \Pr_{C\in F, y\in N(C)}[y \in A'] - \Pr_{y, C\in N(y)\cap F}[y \in A'] \right| \leq \frac{q^{-\frac{3}{2}}}{\epsilon} \leq \gamma$$

We know that $\Pr_{y,C\in N(y)\cap F}[y \in A'] \leq \Pr_y[y \in A'] \leq 84d\gamma$, which implies that $\Pr_{C\in F, y\in N(C)}[y \in A'] \leq 85d\gamma$.

By averaging, for at least half of the cubes $C \in F$, $\Pr_{y\in C}[y \in A'] \leq 200d\gamma \leq \frac{1}{2}$. For all these cubes $T(C) = g_{|C}$, because $\Pr_{y\in C}[T(C)(y) = g(y)] \geq \Pr_{y\in C}[T(C)(y) = f(y), y \notin A'] \geq 1 - 32\gamma - \frac{1}{2} > d/q$, and since $g_{|C}, T(C)$ are both degree $d$ polynomials, they must be equal. ∎

**Remark 3.14.** *Instead of Theorem 3.12, we can use another similar characterization from [RS96], where the neighborhood is defined as $N_{y,h} = \{y + i(h-y) \mid i \in \{0, \ldots, 10d\}\}$. The advantage of using this new neighborhood is that we can conclude $f \overset{(1+o(1))\delta}{\approx} g$ as long as $\delta = O(1/d)$. This will help in reducing the exponent of $d$ by 1 in our main theorem. We chose to use Theorem 3.12 for a self contained proof.*

## 4 Comparing between different tests and their agreement parameter

There are many variants for the low degree test, in this section we look into equivalences between similar low degree agreement tests. We first prove the equivalence in a more general setting and as a corollary we get some interesting results.

Throughout this section, we will work over $\mathbb{F}^m$ where $\mathbb{F}$ is a field of size $q$ and let $s \leq m/2$ be fixed. Also, let $T$ denotes a table which maps every $s$ dimensional affine subspace in $\mathbb{F}^m$ to a degree $d$ polynomial. Let $\mathcal{A}^s$ denote the set of all $s$ dimensional affine subspaces in $\mathbb{F}^m$. For $r < s$ and for $R \in \mathcal{A}^r$ let $\mathcal{A}^s_R \subseteq \mathcal{A}^s$ denote all subspaces in $\mathcal{A}^s$ which contain a particular subspace $R$,

$$\mathcal{A}^s_R = \{S \subset \mathbb{F}^m \mid \dim(S) = s, R \subseteq S\}.$$

For parameters $s > k \geq r$ consider the following test:

---

**Test 3** Subspace agreement test : $\alpha_{sks(r)}$

---

1. Select $K \in \mathcal{A}^k$ u.a.r.
2. Pick $S_1, S_2 \in \mathcal{A}^s_K$ u.a.r.
3. Pick a $r$ dimensional subspace $R \subseteq K$ u.a.r.
4. Accept iff $T(S_1)_{|R} = T(S_2)_{|R}$.

Let $\alpha_{sks(r)}(T)$ be the *agreement* of the table $T = (f_S)_{S \in \mathcal{A}^s}$, i.e. the probability of acceptance of the test.

---

When $r = k$ we simply denote the agreement as $\alpha_{sks}(T)$. With these notations, the success probability of Test 1 is denoted by $\alpha_{3,0,3}(T)$, and of Test 2 by $\alpha_{2,1,2}(T)$.

In this section, we prove the following main lemma.

**Lemma 4.1.** *Let $0 \leq r < k < s \leq \frac{m}{2}$, we have*

$$\alpha_{srs}(T)\left(1 - \left(\frac{d}{q}\right)^{r+1}\right) \leq \alpha_{sks}(T) \leq \alpha_{srs}(T) + (1 + o(1))q^{-(s-2k+r+1)},$$

From Lemma 4.1, we can deduce the following corollary,

**Corollary 4.2.** *Let $\alpha_{\mathcal{C}\ell\mathcal{C}}(T) = \alpha_{3,1,3}(T)$ be the success probability of Test 3 with $s = 3, k = r = 1$, i.e checking consistency of two cubes that intersect on a line. Then for every cubes table $T$,*

$$\alpha_{\mathcal{C}x\mathcal{C}}(T)\left(1 - \frac{d}{q}\right) \leq \alpha_{\mathcal{C}\ell\mathcal{C}}(T) \leq \alpha_{\mathcal{C}x\mathcal{C}}(T) + \frac{1}{q^2}(1 + o(1)).$$

The corollary implies that Theorem 1.2 holds if we modify the test as selecting two cubes u.a.r from a pair of cubes intersecting in a line and checking consistency on the whole line.

Using Lemma 4.1, we can also compare the Raz-Safra Plane vs. Plane agreement tests where planes intersect at a point and on a line. Recall that $\alpha_{\mathcal{P}\ell\mathcal{P}}(T)$ is the acceptance probability of Test 2. Invoking Lemma 4.1 with $s = 2$, $k = 1$ and $r = 0$, we get the following corollary.

**Corollary 4.3** (Lemma 1.4 restated)**.** *Let $T$ be a planes table, and let $\alpha_{\mathcal{P}x\mathcal{P}}(T)$ be the success probability of Test 3 with $s = 2, k = r = 0$, i.e two planes that intersects on a point. Let $\alpha_{\mathcal{P}\ell\mathcal{P}}(T)$ be the success probability of Test 2 from the introduction (two planes that intersects on a line), then*

$$\alpha_{\mathcal{P}x\mathcal{P}}(T)\left(1 - \frac{d}{q}\right) \leq \alpha_{\mathcal{P}\ell\mathcal{P}}(T) \leq \alpha_{\mathcal{P}x\mathcal{P}}(T) + \frac{1}{q}(1 + o(1)).$$

### 4.1 Proof of Lemma 4.1

We prove a few claims that together with the observation $\alpha_{sks(r)}(T) \geq \alpha_{sks}(T)$, prove the lemma.

The following claim shows that two distinct low degree polynomials agree on a random subspace of fixed dimension with very small probability.

**Claim 4.4.** *Let $P_1, P_2 : \mathbb{F}^t \to \mathbb{F}$ be two distinct degree $d$ polynomials. For $r \leq t$*

$$\Pr_{R \in \mathcal{A}^r}\left[(P_1)_{|R} \equiv (P_2)_{|R}\right] \leq \left(\frac{d}{q}\right)^{r+1}.$$

**Proof:** Consider the following way of choosing an $r$ dimensional *affine* subspace from $\mathcal{A}^r$ uniformly at random: Pick $x_0, x_1, x_2, \ldots, x_r$ from $\mathbb{F}_q^t$ independently and u.a.r. Then pick a $r$ dimensional affine subspace $R$ containing $\{x_0 + \mathtt{span}(x_1, x_2, \ldots, x_r)\}$ u.a.r ($R$ is determined by $x_0, x_1, x_2, \ldots, x_r$, unless $\dim \mathtt{span}(x_1, x_2, \ldots, x_r) < r$). It is easy to see that $R$ is distributed uniformly in $\mathcal{A}^r$. Now, $P_1$ and $P_2$ agreeing on the whole subspace $R$ implies that they agree on the points $\{x_0, x_0 + x_1, x_0 + x_2, \ldots, x_0 + x_r\}$ as all these points are contained in $R$. Therefore,

$$\begin{aligned}\Pr_{R \in \mathcal{A}^r}[(P_1)_{|R} \equiv (P_2)_{|R}] &\leq \Pr_{x_0, x_1, x_2, \ldots, x_r \sim \mathbb{F}^t}\left[P_1(x_0) = P_2(x_0) \wedge_{i=1}^r P_1(x_0 + x_i) = P_2(x_0 + x_i)\right] \\ &= \left(\Pr_{x \in \mathbb{F}_q^t}\left[P_1(x) = P_2(x)\right]\right)^{r+1} \leq \left(\frac{d}{q}\right)^{r+1},\end{aligned}$$

where the last inequality is because two different degree $d$ polynomial agree on at most $\frac{d}{q}$ fraction of the points (Schwartz-Zippel lemma). ■

**Claim 4.5.** *Let $M_{m \times n}$ be the adjacency matrix of a bi regular bipartite graph $G$, and let $f$ be a $n$-dimensional $\{0,1\}$ vector such that $\mathbf{E}[f] = \mu$. Then*

$$\langle Mf, Mf \rangle \leq \mu^2 + \lambda(G)^2 \mu.$$

**Proof:** Let $\mathbf{1}$ be the unit vector. We write $f$ as $f = f_1 + f_1^{\perp}$ where $f_1$ is in the direction of $\mathbf{1}$, the singular vector with the maximal singular value, and $f_1^{\perp}$ is its orthogonal component. We note that $f_1 = \mu \mathbf{1}$, and hence $\langle f_1, f_1 \rangle = \mu^2$. Also,

$$\mu = \langle f, f \rangle = \langle f_1 + f_1^{\perp}, f_1 + f_1^{\perp} \rangle = \langle f_1, f_1 \rangle + \langle f_1^{\perp}, f_1^{\perp} \rangle \geq \langle f_1^{\perp}, f_1^{\perp} \rangle.$$

Using this we can bound:

$$\begin{aligned}\langle Mf, Mf \rangle =& \langle Mf_1 + Mf_1^{\perp}, Mf_1 + Mf_1^{\perp}) \rangle \\ =& \langle f_1, f_1 \rangle + \langle Mf_1^{\perp}, Mf_1^{\perp} \rangle \\ \leq& \mu^2 + \lambda(G)^2 \langle f_1^{\perp}, f_1^{\perp} \rangle \\ \leq& \mu^2 + \lambda(G)^2 \mu.\end{aligned}$$

■

**Claim 4.6.** $\alpha_{sks(r)}(T) \geq \alpha_{srs}(T)$.

**Proof:** We start by fixing $R \in \mathcal{A}^r, \sigma \in \mathbb{F}^{q^r}$. For each $k$ dimensional subspace $K \in \mathcal{A}^k_R$, denote by $p_K$ the following probability $p_K = \mathbf{Pr}_{S \sim \mathcal{A}^s_K}[T(S)_{|R} \equiv \sigma]$. In this notation

$$\mathop{\mathbf{Pr}}_{\substack{K \sim \mathcal{A}^k_R \\ S_1, S_2 \sim \mathcal{A}^s_K}} [T(S_1)_{|R} \equiv T(S_2)_{|R} \equiv \sigma] = \mathop{\mathbf{E}}_{K}[p_K^2] \geq \left( \mathop{\mathbf{E}}_{K}[p_K] \right)^2 = \mathop{\mathbf{Pr}}_{S_1, S_2 \sim \mathcal{A}^s_R} [T(S_1)_{|R} \equiv T(S_2)_{|R} \equiv \sigma]. \quad (4)$$

Now, we average over $R, \sigma$ to get $\alpha_{srs}(T)$ and $\alpha_{sks(r)}(T)$:

$$\alpha_{srs}(T) = \mathop{\mathbf{Pr}}_{\substack{R \sim \mathcal{A}^r \\ S_1, S_2 \sim \mathcal{A}^s_R}} [T(S_1)_{|R} \equiv T(S_2)_{|R}] = \mathop{\mathbf{E}}_{R \sim \mathcal{A}^r} \left[ \sum_{\sigma \in \mathbb{F}^{q^r}} \mathop{\mathbf{Pr}}_{S_1, S_2 \sim \mathcal{A}^s_R} [T(S_1)_{|R} \equiv T(S_2)_{|R} \equiv \sigma] \right]. \quad (5)$$

Picking a uniform $R \in \mathcal{A}^r$ then $K \in \mathcal{A}^k_R$ is the same as picking $K \in \mathcal{A}^k$ and then a random $r$ dimensional subspace $R$ in $K$, so by definition

$$\alpha_{sks(r)}(T) = \mathop{\mathbf{Pr}}_{\substack{R \sim \mathcal{A}^r, K \sim \mathcal{A}^k_R \\ S_1, S_2 \sim \mathcal{A}^s_K}} [T(S_1)_{|R} \equiv T(S_2)_{|R}] = \mathop{\mathbf{E}}_{R \sim \mathcal{A}^r} \left[ \sum_{\sigma \in \mathbb{F}^{q^r}} \mathop{\mathbf{Pr}}_{\substack{K \sim \mathcal{A}^k_R \\ S_1, S_2 \sim \mathcal{A}^s_K}} [T(S_1)_{|R} \equiv T(S_2)_{|R} \equiv \sigma] \right]. \quad (6)$$

Using (4), (5) and (6), we get $\alpha_{sks(r)}(T) \geq \alpha_{srs}(T)$. ■

**Claim 4.7.** $\alpha_{sks}(T) \geq \alpha_{sks(r)}(T) \left(1 - \left(\frac{d}{q}\right)^{r+1}\right)$.

**Proof:** By the definition of the agreement,

$$\alpha_{sks}(T) = 1 - \mathop{\mathbf{E}}_{K \sim \mathcal{A}^k} \left[ \mathop{\mathbf{Pr}}_{S_1, S_2 \sim \mathcal{A}^s_K} [T(S_1)_{|K} \neq T(S_2)_{|K}] \right],$$

and

$$\alpha_{sks(r)}(T) = 1 - \mathop{\mathbf{E}}_{K \sim \mathcal{A}^k} \left[ \mathop{\mathbf{Pr}}_{\substack{R \sim K, \\ S_1, S_2 \sim \mathcal{A}^s_K}} [T(S_1)_{|R} \neq T(S_2)_{|R}] \right],$$

where we use $R \sim K$ to denote a random $r$ dimensional subspace in $K$. For every subspace $K \in \mathcal{A}^k$, $R \subseteq K$ is uniform and is independent of $S_1, S_2$.

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{\substack{R \sim K, \\ S_1, S_2 \sim \mathcal{A}^s_K}} [T(S_1)_{|R} \neq T(S_2)_{|R}] &= \mathop{\mathbf{Pr}}_{\substack{R \sim K, \\ S_1, S_2 \sim \mathcal{A}^s_K}} [T(S_1)_{|K} \neq T(S_2)_{|K}, T(S_1)_{|R} \neq T(S_2)_{|R}] \\
&= \mathop{\mathbf{Pr}}_{S_1, S_2 \sim \mathcal{A}^s_K} [T(S_1)_{|K} \neq T(S_2)_{|K}] \cdot \\
&\quad \mathop{\mathbf{Pr}}_{\substack{R \sim K, \\ S_1, S_2 \sim \mathcal{A}^s_K}} [T(S_1)_{|R} \neq T(S_2)_{|R} \mid T(S_1)_{|K} \neq T(S_2)_{|K}] \\
&\geq \mathop{\mathbf{Pr}}_{S_1, S_2 \sim \mathcal{A}^s_K} [T(S_1)_{|K} \neq T(S_2)_{|K}] \cdot \left(1 - \left(\frac{d}{q}\right)^{r+1}\right).
\end{aligned}$$

The lower bound on the probability in the last inequality is as follows: the event $T(S_1)_{|K} \neq T(S_2)_{|K}$ implies that the degree $d$ polynomials corresponding to $T(S_1)_{|K}$ and $T(S_2)_{|K}$ are distinct. Thus, using Claim 4.4 $\mathbf{Pr}_{R\sim K}[T(S_1)_{|R} \equiv T(S_2)_{|R}] \leq (d/q)^{r+1}$. Therefore, for a $k$ dimensional subspace $K \in \mathcal{A}^k$,

$$\Pr_{\substack{R\sim K,\\ S_1,S_2\sim\mathcal{A}^s_K}}[T(S_1)_{|R} \neq T(S_2)_{|R}] \geq \Pr_{S_1,S_2\sim\mathcal{A}^s_K}[T(S_1)_{|K} \neq T(S_2)_{|K}]\left(1-\left(\frac{d}{q}\right)^{r+1}\right).$$

Finally, taking the expectation of the inequality over $K$ finishes the proof. ∎

We first state a lemma about an expansion of the kind of inclusion graphs which we will be dealing with in analyzing the Test 3, the proof of which appears in Appendix A.

**Lemma 4.8.** *Let $r \leq k < s \leq \frac{m}{2}$ be integers, and let $G$ be the inclusion graph $G = G(\mathcal{A}^k_R, \mathcal{A}^s_R)$ for a $r$ dimensional subspace $R$, where $R \neq \emptyset$. Then,*

$$\lambda(G)^2 \leq (1+o(1)) \cdot q^{-(s-2k+r+1)}.$$

**Claim 4.9.** *$\alpha_{sks(r)}(T) \leq \alpha_{srs(r)}(T) + \lambda(G)^2$ where $G$ is the inclusion graph $G = G(\mathcal{A}^k_R, \mathcal{A}^s_R)$ for an $r$ dimensional subspace $R$.*

**Proof:** Fix an $r$ dimensional affine subspace $R \in \mathcal{A}^r$. We prove the following inequality:

$$\Pr_{\substack{K\sim\mathcal{A}^k_R,\\ S_1,S_2\sim\mathcal{A}^s_K}}[T(S_1)_{|R} \equiv T(S_2)_{|R}] \leq \Pr_{S_1,S_2\sim\mathcal{A}^s_R}[T(S_1)_{|R} \equiv T(S_2)_{|R}] + \lambda(G)^2, \tag{7}$$

Note that this implies the claim if we take expectation over $R \in \mathcal{A}^r$. Towards proving (7), for each value $\sigma \in \mathbb{F}^{q^k}$, denote by $A_\sigma \subseteq \mathcal{A}^s_R$ the following set

$$A_\sigma = \{S \in \mathcal{A}^s_R \mid T(S)_{|R} \equiv \sigma\},$$

and $\mu_\sigma = \frac{|A_\sigma|}{|\mathcal{A}^s_R|}$. Let $f_\sigma$ be the indicator function for $A_\sigma$, for $S \in A_\sigma$, $f_\sigma(S) = 1$. By definition

$$\Pr_{S_1,S_2\sim\mathcal{A}^s_R}[T(S_1)_{|R} \equiv T(S_2)_{|R}] = \sum_\sigma \mu_\sigma^2. \tag{8}$$

Let $G = G(\mathcal{A}^k_R, \mathcal{A}^s_R)$ be the inclusion graph, and denote by $M \in \mathbb{R}^{|\mathcal{A}^k_R|\times|\mathcal{A}^s_R|}$ the normalized adjacency matrix, such that each entry is either 0 or $\frac{1}{\deg(K)}$ where $K \in \mathcal{A}^k_R$.

For each $k$ dimensional subspace $K \in \mathcal{A}^k_R$, the value $(Mf_\sigma)_K$ is the fraction of $K$'s neighbors in $A_\sigma$, $(Mf_\sigma)_K = \mathbf{Pr}_{S\sim\mathcal{A}^s_K}[S \in A_\sigma]$. Therefore, the inner product gives us the expected value:

$$\langle Mf_\sigma, Mf_\sigma\rangle = \mathop{\mathbf{E}}_{K\in\mathcal{A}^k_R}\left[\mathop{\mathbf{E}}_{S\in\mathcal{A}^s_K}[S \in A_\sigma]^2\right] = \mathop{\mathbf{E}}_{K\in\mathcal{A}^k_R}\left[\mathop{\mathbf{E}}_{S_1,S_2\in\mathcal{A}^s_K}[S_1,S_2 \in A_\sigma]\right].$$

Therefore

$$\begin{aligned}
\Pr_{\substack{K\sim\mathcal{A}^k_R,\\ S_1,S_2\sim\mathcal{A}^s_K}}[T(S_1)_{|R} \equiv T(S_2)_{|R}] &= \sum_\sigma \langle Mf_\sigma, Mf_\sigma\rangle \\
&\leq \sum_\sigma \mu_\sigma^2 + \lambda(G)^2\mu_\sigma && \text{(using Claim 4.5)} \\
&= \Pr_{S_1,S_2\sim\mathcal{A}^s_R}[T(S_1)_{|R} \equiv T(S_2)_{|R}] + \lambda(G)^2. && \text{(from (8) )}
\end{aligned}$$

which proves (7). ∎

Claim 4.9 together with Lemma 4.8 gives us $\alpha_{sks}(T) \leq \alpha_{srs}(T) + (1+o(1))q^{-2(s-2k+r+1)}$. Claim 4.6 and Claim 4.7 prove the other inequality, $\alpha_{srs}(T)\left(1 - \left(\frac{d}{q}\right)^{r+1}\right) \leq \alpha_{sks}(T)$.

## References


[AS97] Sanjeev Arora and Madhu Sudan. Improved low-degree testing and its applications. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 485–495. ACM, 1997.

[BCN89] A.E. Brouwer, A.M. Cohen, and A. Neumaier. *Distance-regular graphs*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, 1989.

[BGS98] Mihir Bellare, Oded Goldreich, and Madhu Sudan. Free bits, PCPs, and nonapproximability—towards tight results. *SIAM Journal on Computing*, 27(3):804–915, June 1998.

[BLR90] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. pages 73–83, 1990.

[DG08] Irit Dinur and Elazar Goldenberg. Locally testing direct product in the low error range. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 613–622. IEEE, 2008.

[DR06] Irit Dinur and Omer Reingold. Assignment testers: Towards combinatorial proofs of the PCP theorem. *SIAM Journal on Computing*, 36(4):975–1024, 2006. Special issue on Randomness and Computation.

[DS14] Irit Dinur and David Steurer. Direct product testing. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 188–196. IEEE, 2014.

[GS97] Oded Goldreich and Shmuel Safra. A combinatorial consistency lemma with application to proving the PCP theorem. In *RANDOM: International Workshop on Randomization and Approximation Techniques in Computer Science*. LNCS, 1997.

[IKW12] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New direct-product testers and 2-query PCPs. *SIAM Journal on Computing*, 41(6):1722–1768, 2012.

[MR08] Dana Moshkovitz and Ran Raz. Sub-constant error low degree test of almost-linear size. *SIAM J. Computing*, 38(1):140–180, 2008.

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[RS97] Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability pcp characterization of np. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 475–484. ACM, 1997.

[vdWANB49] van der Waerden, E. Artin, E. Noether, and F. Blum. *Modern Algebra*. Number v. 1. Frederick Ungar Publishing, 1949.

# A Spectral properties of Certain Inclusion Graphs

Let $G_{s,k}$ be the intersection graph where the vertex set is all *linear subspaces* of dimension $s$ in $\mathbb{F}_q^m$ and $U \sim U'$ iff $\dim(U \cap U') = k$. We will use the $T_{s,k}$ to denote the *Markov operator* associated with a random walk on this graph. We will need following fact about eigenvalues of $T_{k,k-1}$.

**Definition A.1.** *$k$-th $q$-ary Gaussian binomial coefficient $\begin{bmatrix} m \\ k \end{bmatrix}_q$ is given by*

$$\begin{bmatrix} m \\ k \end{bmatrix}_q := \prod_{i=0}^{k-1} \frac{q^m - q^i}{q^k - q^i}.$$

As $q$ is fixed throughout the article, we will omit the subscript from now on.

**Fact A.2.** *( [BCN89, Theorem 9.3.3]) Suppose $1 \leq k \leq \frac{m}{2}$,*

1. *The number of $k$ dimensional linear subspaces in $\mathbb{F}_q^m$ is exactly $\begin{bmatrix} m \\ k \end{bmatrix}$.*
2. *The degree of $G_{k,k-1}$ is $q\begin{bmatrix} k \\ 1 \end{bmatrix}\begin{bmatrix} m-k \\ 1 \end{bmatrix}$.*
3. *The eigen values of $T_{k,k-1}$ are*

$$\lambda_j(T_{k,k-1}) = \frac{q^{j+1}\begin{bmatrix} k-j \\ 1 \end{bmatrix}\begin{bmatrix} m-k-j \\ 1 \end{bmatrix} - \begin{bmatrix} j \\ 1 \end{bmatrix}}{q\begin{bmatrix} k \\ 1 \end{bmatrix}\begin{bmatrix} m-k \\ 1 \end{bmatrix}},$$

*with multiplicities $\begin{bmatrix} m \\ j \end{bmatrix} - \begin{bmatrix} m \\ j-1 \end{bmatrix}$ for $j = 0, 1, \ldots, k$. Asymptotically, $\lambda_j(T_{k,k-1}) = \Theta(q^{-j})$.*

**Claim A.3.** *For any $1 \leq k \leq \frac{m}{2}$ and , we have $\left|\lambda_1(T_{k,k-2}) - \lambda_1(T_{k,k-1})^2\right| = (1 + o(1))\frac{1}{q^k}$.*

**Proof:** Consider a two-step random walk on the graph $G_{k,k-1}$. We will show that with very high probability, a two-step random walk on $G_{k,k-1}$ corresponds to a single step random walk on $G_{k,k-2}$. Let $U_1, U_2, U_3$ be the vertices from a two-step random walk on $G_{k,k-1}$. Note that conditioned on the event $\dim(U_1 \cap U_3) = k-2$, the distribution of $(U_1, U_3)$ is exactly same as a single step random walk on $G_{k,k-2}$. We will upper bound the probability of the event $\dim(U_1 \cap U_3) \neq k-2$. Let $w_1 = U_1 \cap U_2$ and $w_2 = U_2 \cap U_3$, we can describe the distribution of the two-step random walk as follows:

1. Choose a uniform $k$ dimensional subspace $U_2$.
2. Choose two random $k-1$ dimensional subspaces, $w_1, w_2 \subset U_2$.
3. Choose a point $x_1 \in \mathbb{F}^m \setminus U_2$, and set $U_1 = \text{span}(w_1, x_1)$.
4. Choose a point $x_2 \in \mathbb{F}^m \setminus U_2$, and set $U_3 = \text{span}(w_2, x_2)$.

By definition, $U_2$ has $\begin{bmatrix} k \\ k-1 \end{bmatrix}$ subspaces of size $k-1$, therefore $\mathbf{Pr}_{w_1,w_2}[w_1 = w_2] = \frac{1}{\begin{bmatrix} k \\ k-1 \end{bmatrix}}$. In order to satisfy $\dim(U_1 \cap U_3) \neq k-2$ given that $w_1 \neq w_2$, the point $x_2$ should be in $U_1$. There are $q^k - q^{k-1}$

points in $U_1 \setminus U_2$, and therefore this probability equals $\frac{|U_1 \setminus U_2|}{|\mathbb{F}^m \setminus U_2|} = \frac{q^k - q^{k-1}}{q^m - q^k}$.

$$\begin{aligned}
\mathbf{Pr}[\dim(U_1 \cap U_3) \neq k-2] &= \mathbf{Pr}[w_1 = w_2] + \mathbf{Pr}[\dim(U_1 \cap U_3) \neq k-2 \wedge w_1 \neq w_2] \\
&= \frac{1}{\begin{bmatrix} k \\ k-1 \end{bmatrix}} + \left(1 - \frac{1}{\begin{bmatrix} k \\ k-1 \end{bmatrix}}\right) \mathbf{Pr}[\dim(U_1 \cap U_3) \neq k-2 \mid w_1 \neq w_2] \\
&= \frac{1}{\begin{bmatrix} k \\ k-1 \end{bmatrix}} + \left(1 - \frac{1}{\begin{bmatrix} k \\ k-1 \end{bmatrix}}\right) \cdot \frac{q^k - q^{k-1}}{q^m - q^k} =: \beta.
\end{aligned}$$

Thus, we have

$$T^2_{k,k-1} = \beta \mathcal{N} + (1-\beta) T_{k,k-2},$$

where $\mathcal{N}$ is a Markov operator corresponding to the two-step random walk on $G_{k,k-1}$, conditioning on $\dim(U_1 \cap U_3) \neq k-2$. The claim follows as $\beta = (1+o(1))1/q^k$. ∎

Following fact follows from the definition of $\lambda(G)$.

**Fact A.4.** *For a bi-regular bipartite graph $G(A,B)$, if $T$ is a Markov operator associated with a random walk of length two starting from $A$ (or $B$) then $\lambda(G)^2 = \lambda(T)$.*

We now prove Lemma 2.6.

**Lemma A.5** (Restatement of Lemma 2.6)**.** *We have for every $m \geq 6$,*

1. *For $G_1(\mathcal{L} \setminus \mathcal{L}_x, \mathcal{C}_x)$ , $\lambda(G_1) \approx \frac{1}{\sqrt{q}}$.*
2. *For $G_2(\mathcal{L}_x, \mathcal{C}_x))$ , $\lambda(G_2) \approx \frac{1}{q}$.*
3. *For $G_3(\mathbb{F}^m \setminus \ell, \mathcal{C}_\ell)$ , $\lambda(G_3) \approx \frac{1}{\sqrt{q}}$.*
4. *For $G_4(\mathbb{F}^m, \mathcal{C})$ , $\lambda(G_4) \approx \frac{1}{q^{3/2}}$.*
5. *For $G_5(\mathbb{F}^m \setminus \{x\}, \mathcal{C}_x)$ , $\lambda(G_5) \approx \frac{1}{q}$.*

*And for every $m \geq 3$*

6. *For $G_6(\mathbb{F}^m, \mathcal{L})$, $\lambda(G_6) \approx \frac{1}{\sqrt{q}}$.*

*where $\approx$ denotes equality up to a multiplicative factor of $1 \pm o(1)$.*

**Proof:** Suppose $T$ is an $n \times n$ Markov operator which is a convex combination of a bunch of other Markov operators: $T = \sum_{i=1}^k \alpha_i T_i$ where $\alpha_i \geq 0$ and $\sum_{i=1}^k \alpha_i = 1$, and that both $T$ and $T_i$'s are regular. As the row sum of each Markov operator is 1, the largest eigenvalue is 1, since both $T$ and $T_i$'s are regular, the eigenvector of the largest eigenvalue is the all 1 vector. The second largest eigenvalue of $T$ can be upper bounded by

$$\begin{aligned}
\lambda(T) &:= \max_{\substack{v \in \mathbb{R}^n, \|v\|=1, \\ v \perp \mathbf{1}}} \|Tv\| \\
&= \max_{\substack{v \in \mathbb{R}^n, \|v\|=1, \\ v \perp \mathbf{1}}} \left\| \sum_{i=1}^k \alpha_i T_i \right\| \\
&\leq \sum_{i=1}^k \max_{\substack{v \in \mathbb{R}^n, \|v\|=1, \\ v \perp \mathbf{1}}} \|\alpha_i T_i\| = \sum_{i=1}^k \alpha_i \lambda(T_i).
\end{aligned}$$

In proving the lemma, we repeatedly use the above simple fact to upper bound the eigenvalue.

1. Without loss of generality, we can assume $x = \mathbf{0}$. Let $d_L$ and $d_R$ denote the left and right degree of $G_1$ respectively. Fix a line $\ell$, $d_L$ is the number of cubes containing $\ell$ and not passing through $\mathbf{0}$. Every point $x \notin \text{span}(\ell, \mathbf{0})$ defines a cube $C = \text{span}(x, \mathbf{0}, \ell)$. Thus, the number of linear cubes containing $\ell$ equals $d_L = \frac{q^m - q^2}{q^3 - q^2}$, where the denominator is the overcounting factor, the number of points that give the same cube.

   Fix a linear cube $C$. The right degree is the number of lines in $C$ not passing through the origin which is $\frac{\binom{q^3}{2}}{\binom{q}{2}} - \frac{q^3-1}{q-1}$, where the first term counts all possible lines in $C$ (each two different points define a line, we divide by the double counting) and the second term counts all the lines in $C$ that pass through the origin.

   Let $T_1$ be the Markov operator associated with a two-step random walk in $G_1$ starting from $\mathcal{C}_x$. Using Fact A.4, in order to bound $\lambda(G_1)$ it is enough to bound the second largest eigenvalue of $T_1$. Since $G_1$ is bi-regular, the first eigenvector of $T_1$ is the all ones vector. For every cube $C$, the number of two-step walks starting from $C$ is $d_L \cdot d_R$.

   If $\dim\{C_1 \cap C_2\} = 1$, then the two cubes intersection is only on a line. Since both cubes are linear, it means that this line goes through the origin, therefore it doesn't correspond to a vertex on the left side, and there is no walk $C_1 \to \ell \to C_2$, so $(T_1)_{C_1,C_2} = 0$. Of course, the same holds if $\dim\{C_1 \cap C_2\} = 0$.

   If $\dim\{C_1 \cap C_2\} = 2$, there there is a plane going through the origin in both $C_1, C_2$. The number of walks $C_1 \to \ell \to C_2$ equals the number of lines in this plane that don't contain the origin, $\mathbf{0}$. Each pair of distinct points on the plane correspond to a line, and we divide by the double counting. Therefore the number of lines in a plane equals $\frac{\binom{q^2}{2}}{\binom{q}{2}}$. We subtract from it the number of lines in a plane that contains $\mathbf{0}$, resulting in $\frac{\binom{q^2}{2}}{\binom{q}{2}} - \frac{q^2-1}{q-1} =: \beta$.

   If $C_1 = C_2$, then exists a path $C_1 \to \ell \to C_2$ for every line $\ell$ adjacent to $C_1$, and there are $d_R$ such lines.

   Since $T_1$ is a Markov operator, we need to normalize the number of paths between $C_1, C_2$ by dividing in the total number of outgoing paths from $C_1$, which equals $d_R \cdot d_L$. Therefore,

$$(T_1)_{C_i,C_j} = \begin{cases} \frac{d_R}{d_R \cdot d_L}, & \text{if } C_i = C_j \\ \frac{\beta}{d_R \cdot d_L}, & \text{if } \dim\{C_1 \cap C_2\} = 2 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

   Thus, we can write $T_1$ as:

$$T_1 = \frac{1}{d_L} I + \frac{\beta}{d_R d_L} \cdot G_{3,2} = \frac{1}{d_L} I + \frac{\beta d'}{d_R d_L} \cdot T_{3,2},$$

   where $d'$ is the degree of a vertex in $G_{3,2}$. One can verify that $T_1$ is indeed a convex combination of two Markov operators $I$ and $T_{3,2}$. Since $G_{3,2}$ is a regular graph, the second eigenvector

of $T_{3,2}$ is also orthogonal to $\mathbf{1}$. Hence,

$$\lambda(G_1)^2 = \lambda(T_1) = \max_{\substack{v\in\mathbb{R}^{|\mathcal{C}_x|}, v\perp\mathbf{1}\\ \|v\|=1}} \|T_1 v\| = \max_{\substack{v\in\mathbb{R}^{|\mathcal{C}_x|}, v\perp\mathbf{1}\\ \|v\|=1}} \left\| \left(\frac{1}{d_L} I + \frac{\beta d'}{d_R d_L}\cdot T_{3,2}\right) v \right\|$$
$$= \frac{1}{d_L} + \frac{\beta d'}{d_R d_L}\cdot \lambda_1(T_{3,2}). \tag{10}$$

We now just need to plug in the values of $\beta, d'$ and $\lambda_1(T_{3,2})$. Using Fact A.2, $\lambda_1(T_{3,2})$ is given by the following expression,

$$\lambda_1(T_{3,2}) = \frac{q^2\begin{bmatrix}2\\1\end{bmatrix}\begin{bmatrix}m-4\\1\end{bmatrix} - \begin{bmatrix}1\\1\end{bmatrix}}{q\begin{bmatrix}3\\1\end{bmatrix}\begin{bmatrix}m-3\\1\end{bmatrix}} = (1+o(1))\frac{1}{q}.$$

As we have seen before, $d_R = \frac{\binom{q^3}{2}}{\binom{q}{2}} - \frac{q^3-1}{q-1} = (1+o(1))q^4$, $d_L = \frac{q^m-q^2}{q^3-q^2} = (1+o(1))q^{m-3}$ and $\beta = \frac{\binom{q^2}{2}}{\binom{q}{2}} - \frac{q^2-1}{q-1} = (1+o(1))q^2$. From Fact A.2, $d' = (1+o(1))q^{m-1}$. Thus,

$$\frac{1}{d_L} = (1+o(1))\frac{1}{q^{m-3}}, \qquad \frac{\beta d'}{d_R d_L}\lambda_1(T_{3,2}) = (1+o(1))\frac{1}{q}$$

Plugging these values in (10) gives $\lambda(G_1) = (1+o(1))\frac{1}{\sqrt{q}}$ as required.

2. This bound is implied from a more general Lemma 4.8 we prove below with $s = 3, k = 1$ and $r = 0$.

3. In this case, it will be easier to bound the eigenvalue of the Markov operator associated with a random walk of length two starting from $\mathbb{F}^m \setminus \ell$. Let $T_3$ be the Markov operator. Now, the path of length two starting from $x$ looks like $x \to C \to y$. Thus, the cube $C$ contains all points from the affine plane spanned by $x$ and $\ell$. Let $p(x,\ell)$ be the affine plane spanned by $x$ and $\ell$. We have $\mathbf{Pr}[y \in p(x,\ell)] = \frac{q^2-q}{q^3-q} \approx \frac{1}{q}$. If $y \notin p(x,\ell)$ then the distribution of $y$ is uniform in $\mathbb{F}^m \setminus p(x,\ell)$. Thus, we have

$$T_3 = (1-o(1))\left(1-\frac{1}{q}\right)J + (1+o(1))\frac{1}{q}\mathcal{N},$$

where $J$ is a Markov operator associated with a complete graph on $\mathbb{F}^m \setminus \ell$, with self loops and $\mathcal{N}$ is an appropriate Markov operator. Thus, we have bound $\lambda(T_3) = (1+o(1))\frac{1}{q}$. Since $\lambda(G_3)^2 = \lambda(T_3)$, the bound follows.

4. Proof of this is along the same lines as (3). The Markov operator here (starting a walk from the left side) can be written as

$$T_4 = (1\pm o(1))\frac{1}{q^3}I + \left((1\pm o(1))(1-\frac{1}{q^3})\right)J,$$

where $I$ is an identity matrix. Thus $\lambda(T_4) = (1\pm o(1))\frac{1}{q^3} = \lambda(G_4)^2$.

5. The proof of this item is also similar to (3), we look on the path of length 2 starting from the left side, i.e $y \to C \to z$, and let $T_5$ be the Markov operator. Let $\ell(x,y)$ be the line spanned by $x, y$ (where $x$ is the fixed point, $G_5(\mathbb{F}^m \setminus \{x\}, \mathcal{C}_x)$), then $\mathbf{Pr}[z \in \ell(x,y)] = \frac{|\ell(x,y)\setminus\{x\}|}{|C\setminus\{x\}|} = \frac{q-1}{q^3-1} \approx \frac{1}{q^2}$, let $\mathcal{N}$ be the appropriate Markov operator of the event that $x, y, z$ are colinear, then

$$T_5 = (1-o(1))\left(1-\frac{1}{q^2}\right)J + (1+o(1))\frac{1}{q^2}\mathcal{N}.$$

Here $J$ is the Markov operator of the complete graph on $\mathbb{F}^m \setminus \{x\}$. Thus $\lambda(G_5)^2 \approx \frac{1}{q^2}$.

6. Consider a two-step random walk in $G_6$, $x \to \ell \to y$. If we sample a random line through $x$ then conditioned on $y \neq x$, $y$ is uniformly distributed in $\mathbb{F}^m$. Thus, we can write the Markov operator $T$ associated with this process as:

$$T = \frac{1}{q}I + \left(1-\frac{1}{q}\right)T',$$

where $T'$ is a Markov operator associated with a random walk on a complete graph on $A$, without self loops and $I$ is an identity matrix. As $T' = \frac{1}{|A|-1}J - \frac{1}{|A|-1}I$, $\lambda(T') = \frac{1}{q^3-1}$. Thus, $\left|\lambda(T) - \frac{1}{q}\right| \leq \frac{1}{q^3-1}$. The claim follows as $\lambda(G_6)^2 = \lambda(T)$.

■

Next, we prove Lemma 4.8. Recall that $\mathcal{A}^s$ denotes set of all $s$ dimensional affine subspaces in $\mathbb{F}^m$. Also, for $r < s$ and for $R \in \mathcal{A}^r$, $\mathcal{A}^s_R \subseteq \mathcal{A}^s$ denotes all those subspaces in $\mathcal{A}^s$ which contains a particular subspace $R$.

**Lemma A.6** (Restatement of Lemma 4.8)**.** *Let $r \leq k < s \leq \frac{m}{2}$ be integers, and let $G$ be the inclusion graph $G = G(\mathcal{A}^k_R, \mathcal{A}^s_R)$ for an $r$ dimensional subspace $R$, where $R \neq \emptyset$. Then,*

$$\lambda(G)^2 \leq (1+o(1)) \cdot q^{-(s-2k+r+1)}.$$

**Proof:** Fix an $r$ dimensional subspace $R \subseteq \mathbb{F}^m, R \neq \emptyset$ and recall that

$$\mathcal{A}^k_R = \{K \subset \mathbb{F}^m | \dim(K) = k, R \subset K\}.$$

Let $G = G(\mathcal{A}^k_R, \mathcal{A}^s_R)$ be the biregular bipartite inclusion graph and let $d_k$ (resp. $d_s$) denote the degree of vertex in $\mathcal{A}^k_R$ (resp. $\mathcal{A}^s_R$).

For every $n, t, j \in \mathbb{N}$, let $h(n,t,j)$ be the number of $t$ dimensional subspaces in $\mathbb{F}^n$ that contain a specific dimention $j$ subspace,

$$h(n,t,j) = \frac{(q^n-q^j)\cdots(q^n-q^{t-1})}{(q^t-q^j)\cdots(q^t-q^{t-1})} \approx q^{(n-t)(t-j)}, \tag{11}$$

where $\approx$ denotes equality up to a multiplicative factor $(1 \pm o(1))$, as before. For any fixed $j$ dimensional subspace $X$, the numerator equals the number of $t-j$ linearly independent points $y_1, y_2, \ldots, y_{t-j}$ in $\mathbb{F}^n$ such that $\dim(\mathtt{span}(X, y_1, y_2, \ldots, y_{t-j})) = t$, whereas for every $t$ dimensional subspace $Z$, the denominator equals the double counting of $Z$, i.e the number of $t-j$ linearly independent points $y_1, y_2, \ldots, y_{t-j}$ such that $\mathtt{span}(X, y_1, y_2, \ldots, y_{t-j}) = Z$. We can now bound the number of vertices and the left and right degree in $G$.

$$\left|\mathcal{A}^k_R\right| = h(m,k,r), \qquad |\mathcal{A}^s_R| = h(m,s,r),$$
$$d_k = h(m,s,k), \qquad d_s = h(s,k,r).$$

Let $T$ be the two-step Markov operator on the bipartite graph $G$, starting from $\mathcal{A}_R^k$, we want to calculate the entries of $T$. Let $K_1, K_2 \in \mathcal{A}_R^k$, by definition $(T)_{K_1,K_2}$ is the probability that a two-step random walk will end at $K_2$, conditioned on it starting from $K_1$.

Let $r' = \dim(K_1 \cap K_2) \geq r$, in this notation $\dim(K_1 \cup K_2) = 2k - r'$. Any 2 step random walk from $K_1$ to $K_2$ looks like $K_1 \to S' \to K_2$ where $S'$ is an $s$ dimentional subspace containing both $K_1$ and $K_2$. The number of such $S'$ is exactly $h(m, s, 2k - r')$. Thus, $(T)_{K_1,K_2}$ equals

$$(T)_{K_1,K_2} = \mathbf{Pr}[\text{R.W ends at } K_2 | \text{ R.W starts at } K_1] = \frac{h(m, s, 2k - r')}{d_k \cdot d_s} = \frac{h(m, s, 2k - r')}{h(m, s, k) \cdot h(s, k, r)}. \quad (12)$$

This probability is the same for every $K_1, K_2 \in \mathcal{A}_R^k$ such that $\dim(K_1 \cap K_2) = r'$, so we can denote this value by $p_{r'} = (T)_{K_1,K_2}$. Notice that $p_{r'} \geq p_r$ for every $r' \geq r$.

Let $G_{r'}$ be the graph with vertex set $\mathcal{A}_R^k$, where $K_1, K_2$ are connected by an edge if $\dim(K_1 \cap K_2) = r'$. We also denote the $0/1$ adjacency matrix of graph $G_{r'}$ by $G_{r'}$. With these notations, the 2 step Markov operator $T$ equals

$$T = \sum_{r'=r}^{k} p_{r'} G_{r'}.$$

Notice that this is not a convex combination, $\sum_{r'} p_{r'} \neq 1$, but rather $p_{r'}$ are the entries of $T$, and $G_{r'}$ are $0/1$ matrices.

Let $J$ be the all 1 matrix, we know that $J = \sum_{r'=r}^{k} G_{r'}$. The first matrix in the sum $G_r$ is the only non sparse matrix, since for every subspace $K_1 \in \mathcal{A}_r^k$, almost all other subspaces intersects with $K_1$ only in $R$. Therefore we can write $G_r = J - \sum_{r'=r+1}^{k} G_{r'}$, and get

$$T = p_r J + \sum_{r'=r+1}^{k} (p_{r'} - p_r) G_{r'}.$$

Since $T$ is a Markov operator of a regular graph, the all $\mathbf{1}$ vector is the vector with the maximal eigenvalue, which equals 1. Since $G_{r'}$ are also regular graphs, $\mathbf{1}$ is the vector with the maximal eigenvalue, which equals $\deg(G_{r'})$, which is the number of $K' \in \mathcal{A}_R^k$ such that $\dim(K \cap K') = r'$ (as the adjacency matrices are not normalized).

$$\begin{aligned}\deg(G_{r'}) =& h(k, r', r) \cdot \frac{(q^m - q^k) \cdots (q^m - q^{2k-r'-1})}{(q^k - q^{r'}) \cdots (q^k - q^{k-1})} \\ \approx& q^{(k-r')(r'-r)} \cdot q^{(m-k)(k-r')} = q^{(k-r')(m-k+r'-r)}\end{aligned}$$

For every $K \in \mathcal{A}_R^k$, the factor $h(k, r', r)$ is the number of $r'$ dimensional subspace in $K$ that contain $R$, the second factor is the number of $k$ dimensional subspaces that intersect with $K$ only in a specific $r'$ dimensional subspace.

Let $v$ be the normalized eigenvector of the second eigenvalue of $T$, this means that $v \perp \mathbf{1}$ and $\|v\| = 1$. Since $J$ is the all 1 matrix, $Jv = 0$. We also know that for every $r' > r$, $\|G_{r'} v\| \leq \deg(G_{r'})$,

as it is true for every vector $v$.

$$\begin{aligned}\|Tv\| &= \left\| \sum_{r'=r+1}^{k} (p_{r'} - p_r) G_{r'} v \right\| \\ &\leq \sum_{r'=r+1}^{k} (p_{r'} - p_r) \|G_{r'} v\| && \text{(triangle inequality)} \\ &\leq \sum_{r'=r+1}^{k} p_{r'} \deg(G_{r'})\end{aligned}$$

For every $r'$, by using the expression for $p_{r'}$ from (12) and bounds on $h$ from (11) we get that

$$p_{r'} \deg(G_{r'}) \approx p_{r'} q^{(k-r')(m-s+r'-r)} \approx q^{-(r'-r)(s-2k+r')}.$$

Since $r' > r$, $(r'-r)(s-2k+r')$ is minimized when $r' = r+1$ and hence

$$\lambda(T) = \|Tv\| \leq (1+o(1)) \sum_{r'=r+1}^{k} \frac{1}{q^{s-2k+r'}} \leq (1+o(1)) \cdot \frac{1}{q^{s-2k+r+1}}.$$

The lemma statement now follows from the Fact A.4. ∎

# B Spectral Expansion Properties Proofs

**Lemma B.1** (Restatement of Lemma 2.3)**.** *Let $D_1, D_2$ as defined in Definition 2.2. Let $G = (A \cup B, E)$ be a bi-regular bipartite graph, then for every subset $B' \subset B$ of measure $\mu > 0$ and every $E' \subset E$*

$$\left| \Pr_{(a,b)\sim D_1}[(a,b) \in E'] - \Pr_{(a,b)\sim D_2}[(a,b) \in E'] \right| \leq \frac{\lambda(G)}{\sqrt{\mu}}.$$

*Where is $D_2$ returned $\perp$, we treat is as it is not in $E'$.*

**Proof:** In the proof we represent both probabilities as an inner product, and then use $\lambda(G)$ to bound the difference. Let $M \in \mathbb{R}^{A \times B}$ the adjacency matrix of the graph $G$, normalized such that $M\mathbf{1} = \mathbf{1}$ (where the first $\mathbf{1}$ is of dimension $|B|$ and the second of dimension $|A|$). We define the matrix $M'$ representing the subset of edges $E'$, $M'_{a,b} = M_{a,b} \cdot (\mathbf{1}_{E'})_{a,b}$.

Starting with the probability of $(a,b) \sim D_1$, the vector $M'\mathbf{1}_{B'}$ satisfies that for every $a \in A$, $(M'\mathbf{1}_{B'})_a = \mathbf{Pr}_{b\in N(a)}[(a,b) \in E', b \in B']$.

$$\begin{aligned}\langle \mathbf{1}, M'\mathbf{1}_{B'} \rangle &= \mathop{\mathbf{E}}_{a\sim A}\left[E_{b\sim N(a)}[\mathbb{I}((a,b) \in E', b \in B')]\right] \\ &= \Pr_{a\sim A, b\sim N(a)}[(a,b) \in E', b \in B'] && \text{(using bi-regularity of } G) \\ &= \Pr_{b\sim B, a\sim N(b)}[(a,b) \in E', b \in B'] \\ &= \Pr_{b\sim B}[b \in B'] \cdot \Pr_{b\sim B, a\sim N(b)}[(a,b) \in E' \mid b \in B'] \\ &= \mu \cdot \Pr_{(a,b)\sim D_1}[(a,b) \in E'].\end{aligned}$$

We now want to represent the second probability as an inner product. We define the vector $P \in [0,1]^A$ as follows, for each $a \in A$:

1. If $N(a) \cap B' = \emptyset$, then $P_a = 0$.
2. Else, $P_a = \mathbf{Pr}_{b \in N(a)}[(a,b) \in E' \mid b \in B']$.

In this notation $\mathbf{Pr}_{(a,b)\sim D_2}[(a,b) \in E'] = \langle \mathbf{1}, P \rangle$.

We now want to find a connection between the inner products. If $P_a \neq 0$, then it defined as the conditional probability, and

$$\mathop{\mathbf{Pr}}_{b\sim N(a)}[b \in B', (a,b) \in E'] = \mathop{\mathbf{Pr}}_{b\sim N(a)}[b \in B'] \mathop{\mathbf{Pr}}_{b\sim N(a)}[(a,b) \in E' \mid b \in B'] = \mathop{\mathbf{Pr}}_{b\sim N(a)}[b \in B'] P_a.$$

If $P_a = 0$ then also $\mathbf{Pr}_{b\sim N(a)}[b \in B', (a,b) \in E'] = 0$, and the above equality still holds. We notice that $(M'\mathbf{1}_{B'})_a = \mathbf{Pr}_{b\in N(a)}[(a,b) \in E', b \in B']$ and $(M\mathbf{1}_{B'})_a = \mathbf{Pr}_{b\in N(a)}[b \in B']$, which means that for every $a \in A$, $(M'\mathbf{1}_{B'})_a = (M\mathbf{1}_{B'})_a P_a$ and

$$\langle M\mathbf{1}_{B'}, P \rangle = \langle \mathbf{1}, M'\mathbf{1}_{B'} \rangle.$$

Therefore we can express the difference between the two probabilities as

$$\begin{aligned}
\left| \mathop{\mathbf{Pr}}_{(a,b)\sim D_1}[(a,b) \in E'] - \mathop{\mathbf{Pr}}_{(a,b)\sim D_2}[(a,b) \in E'] \right| &= \left| \frac{1}{\mu} \langle \mathbf{1}, M'\mathbf{1}_{B'} \rangle - \langle \mathbf{1}, P \rangle \right| \\
&= \left| \frac{1}{\mu} \langle M\mathbf{1}_{B'}, P \rangle - \langle \mathbf{1}, P \rangle \right| \\
&= \frac{1}{\mu} \left| \langle M\mathbf{1}_{B'} - \mu\mathbf{1}, P \rangle \right| \\
&\leq \frac{1}{\mu} \left\| M\mathbf{1}_{B'} - \mu\mathbf{1} \right\| \left\| P \right\| \qquad \text{(By Cauchy Swartz)}
\end{aligned} \tag{13}$$

Since $P$ is a vector in $[0,1]$ and the inner product we use is expectation, $\|P\| \leq 1$. In order to finish the proof we need to bound the size of the vector

$$M\mathbf{1}_{B'} - \mu\mathbf{1} = M\mathbf{1}_{B'} - \mu M\mathbf{1} = M(\mathbf{1}_{B'} - \mu\mathbf{1}).$$

We notice that $\mathbf{1}_{B'}$ is a $\{0,1\}$ vector of measure $\mu$, so $\langle \mathbf{1}_{B'}, \mathbf{1} \rangle = \langle \mathbf{1}_{B'}, \mathbf{1}_{B'} \rangle = \mu$, and $(\mathbf{1}_{B'} - \mu\mathbf{1}) \perp \mathbf{1}_B$. By the definition of $\lambda(G)$, this means that

$$\left\| M(\mathbf{1}_{B'} - \mu\mathbf{1}) \right\| \leq \lambda(G) \left\| \mathbf{1}_{B'} - \mu\mathbf{1} \right\| \leq \lambda\sqrt{\mu}.$$

We substitute the norm of the vector in equation (13) and we are done. ∎

**Lemma B.2** (Restatement of Lemma 2.5)**.** *Let $D_3, D_4$ as defined in Definition 2.4. Let $G = (A \cup B, E)$ be a bi-regular bipartite graph, such that every two distinct $b_1, b_2 \in B$ have exactly the same number of common neighbors (i.e for all distinct $b_1, b_2 \in B$, $|N(b_1) \cap N(b_2)|$ is the same), and this number is non-zero. Then for every subset $B' \subset B$ of measure $\mu > 0$ and every $E' \subset E$*

$$\left| \mathop{\mathbf{Pr}}_{a,b_1,b_2\sim D_3}[(a,b_1)(a,b_2) \in E'] - \mathop{\mathbf{Pr}}_{a,b_1,b_2\sim D_4}[(a,b_1)(a,b_2) \in E'] \right| \leq \frac{2\lambda(G)}{\mu} + \frac{1}{\mu^2 d_A} + \frac{1}{\mu^2 |B|}$$

*Where is $D_4$ returned $\perp$, we treat is as it is not in $E'$ and $d_A$ is the degree on $A$ side.*

**Proof:** This proof is similar in spirit to the proof of Lemma 2.3, with more complication since the event contains two edges instead of a single one.

Let $M \in \mathbb{R}^{A \times B}$ the adjacency matrix of the graph $G$, normalized such that $M\mathbf{1} = \mathbf{1}$. We denote by $M'$ the matrix that represents the edges in $E'$, i.e for each $a \in A, b \in B$, $M'_{a,b} = M_{a,b} \cdot (\mathbf{1}_{E'})_{a,b}$.

Starting from $D_3$, we first write the conditional probability

$$\begin{aligned}\Pr_{\substack{b_1,b_2 \\ a\sim N(b_1)\cap N(b_2)}}[b_1, b_2 \in B', (a, b_1), (a, b_2) \in E'] &= \Pr_{b_1,b_2}[b_1, b_2 \in B'] \Pr_{a,b_1,b_2\sim D_3}[(a, b_1), (a, b_2) \in E'] \quad (14)\\ &= \mu^2 \Pr_{a,b_1,b_2\sim D_3}[(a, b_1), (a, b_2) \in E'].\end{aligned}$$

We want to express the left side as an inner product, we notice that for each $a \in A$:

$$(M'\mathbf{1}_{B'})_a = \mathop{\mathbf{E}}_{b\sim N(a)}[\mathbb{I}(b \in B', (a, b) \in E')].$$

Therefore the inner product satisfies

$$\begin{aligned}\langle M'\mathbf{1}_{B'}, M'\mathbf{1}_{B'}\rangle &= \mathop{\mathbf{E}}_{a\sim A}\left[\mathop{\mathbf{E}}_{b_1,b_2\sim N(a)}[\mathbb{I}(b_1, b_2 \in B', (a, b_1)(a, b_2) \in E')]\right] \quad (15)\\ &= \Pr_{a\sim A,b_1,b_2\sim N(a)}[b_1, b_2 \in B', (a, b_1)(a, b_2) \in E']\end{aligned}$$

Since each two $b_1, b_2 \in B$ has the same number of neighbors,

$$\Pr_{\substack{a\sim A \\ b_1\neq b_2\sim N(a)}}[b_1, b_2 \in B', (a, b_1)(a, b_2) \in E'] = \Pr_{\substack{b_1\neq b_2\sim B \\ a\sim N(b_1)\cap N(b_2)}}[b_1, b_2 \in B', (a, b_1)(a, b_2) \in E'].$$

We want to switch the expression in (15) by the one is (14), we know that they are equal when $b_1 \neq b_2$. But the probability of $b_1 = b_2$ is different between the two cases, it is $\frac{1}{d_A}$ if we pick neighbors of $a$ and $\frac{1}{|B|}$ if we pick two random vertices in $B$. If we add the probability of $b_1 = b_2$ as an error, we get that

$$\left|\mu^2 \Pr_{a,b_1,b_2\sim D_3}[(a, b_1)(a, b_2) \in E'] - \langle M'\mathbf{1}_{B'}, M'\mathbf{1}_{B'}\rangle\right| \leq \frac{1}{d_A} + \frac{1}{|B|} \quad (16)$$

Now we want to express the probability of $a, b_1, b_2 \sim D_4$ as an inner product. In order to do that, we define the vector $P$, for every $a \in A$

1. If $N(a) \cap B' = \emptyset$, then $P_a = 0$.
2. Else, $P_a = \mathbf{Pr}_{b_1,b_2\sim N(a)}[(a, b_1)(a, b_2) \in E' \mid b_1, b_2 \in B']$.

The vector $P$ is defined such that

$$\Pr_{a,b_1,b_2\sim D_4}[(a, b_1)(a, b_2) \in E'] = \mathop{\mathbf{E}}_{a}[P_a] = \langle \mathbf{1}, P\rangle.$$

We want to find a connection between this expression and the expression representing the probability $\mathbf{Pr}_{a,b_1,b_2\sim D_3}[(a, b_1)(a, b_2) \in E']$.

We use (16) and the triangle inequality to bound the difference between the two target probabilities

$$\left|\underset{a,b_1,b_2\sim D_3}{\mathbf{Pr}}[(a,b_1)(a,b_2)\in E'] - \underset{a,b_1,b_2\sim D_4}{\mathbf{Pr}}[(a,b_1)(a,b_2)\in E']\right| \leq \left|\frac{1}{\mu^2}\langle M'\mathbf{1}_{B'}, M'\mathbf{1}_{B'}\rangle - \langle \mathbf{1}, P\rangle\right| + \frac{1}{\mu^2 d_A} + \frac{1}{\mu^2 |B|} \tag{17}$$

We now need to bound the expression in (17), in order to do that, we will first show that

$$\langle M'\mathbf{1}_{B'}, M'\mathbf{1}_{B'}\rangle = \underset{a\sim A,b_1,b_2\sim N(a)}{\mathbf{Pr}}[(a_1,b)(a_2,b)\in E', b_1,b_2\in B'] = \underset{a}{\mathbf{E}}[P_a(M\mathbf{1}_{B'})_a^2]. \tag{18}$$

We notice that for $a$ such that $P_a > 0$, it equals the conditional probability and

$$\underset{b_1,b_2\sim N(a)}{\mathbf{Pr}}[(a_1,b)(a_2,b)\in E', b_1,b_2\in B'] = \underset{b_1,b_2\sim N(a)}{\mathbf{Pr}}[b_1,b_2\in B']P_a.$$

If $a$ is such that $P_a = 0$, then $\mathbf{Pr}_{b_1,b_2\sim N(a)}[(a_1,b)(a_2,b)\in E', b_1,b_2\in B'] = 0$ and the above equality still holds. We further notice that

$$(M\mathbf{1}_{B'})_a = \underset{b\sim N(a)}{\mathbf{E}}[\mathbb{I}(b\in B')].$$

If we substitute $\mathbf{Pr}_{b_1,b_2\sim N(a)}[b_1,b_2\in B']$ in $(M\mathbf{1}_{B'})_a^2$, we get (18).
In order to finish the proof, we upper bound

$$\left|\frac{1}{\mu^2}\langle M'\mathbf{1}_{B'}, M'\mathbf{1}_{B'}\rangle - \langle \mathbf{1}, P\rangle\right| = \left|\underset{a}{\mathbf{E}}\left[\frac{1}{\mu^2}P_a(M\mathbf{1}_{B'})_a^2 - P_a\right]\right| = \frac{1}{\mu^2}\left|\underset{a}{\mathbf{E}}[P_a((M\mathbf{1}_{B'})_a^2 - \mu^2)]\right|.$$

We now upper bound the expectation as follows,

$$\begin{aligned}
\underset{a}{\mathbf{E}}[P_a((M\mathbf{1}_{B'})_a^2 - \mu^2)] &= \underset{a}{\mathbf{E}}[P_a((M\mathbf{1}_{B'})_a - \mu)((M\mathbf{1}_{B'})_a + \mu)] \\
&\leq \max_a\{|P_a|\}\,\underset{a}{\mathbf{E}}[|((M\mathbf{1}_{B'})_a - \mu)((M\mathbf{1}_{B'})_a + \mu)|] \\
&\leq \|M\mathbf{1}_{B'} - \mu\mathbf{1}\|\,\|M\mathbf{1}_{B'} + \mu\mathbf{1}\| \qquad (19)\\
&\leq \lambda\sqrt{\mu}\sqrt{4\mu}, \qquad (20)
\end{aligned}$$

where (19) is due to Cauchy-Schwarz inequality and using $|P_a| \leq 1$. In (20), we bound $\|M\mathbf{1}_{B'} - \mu\mathbf{1}\|$ like in the previous proof,

$$\|M\mathbf{1}_{B'} - \mu\mathbf{1}\| = \|M\mathbf{1}_{B'} - \mu M\mathbf{1}\| = \|M(\mathbf{1}_{B'} - \mu\mathbf{1})\| \leq \lambda\,\|\mathbf{1}_{B'}\| \leq \lambda\sqrt{\mu}.$$

Finally, we bound $\|M\mathbf{1}_{B'} + \mu\mathbf{1}\|$:

$$\begin{aligned}
\|M\mathbf{1}_{B'} + \mu\mathbf{1}\|^2 &= \langle M\mathbf{1}_{B'} + \mu\mathbf{1}, M\mathbf{1}_{B'} + \mu\mathbf{1}\rangle \\
&= \langle M\mathbf{1}_{B'}, M\mathbf{1}_{B'}\rangle + 2\langle M\mathbf{1}_{B'}, \mu\mathbf{1}\rangle + \langle \mu\mathbf{1}, \mu\mathbf{1}\rangle \\
&\leq \|\mathbf{1}_{B'}\|^2 + 2\mu + \mu^2\,\|\mathbf{1}\|^2 \\
&\leq \mu + 2\mu + \mu^2 \leq 4\mu.
\end{aligned}$$

■

# C Rubinfeld-Sudan Characterization

In this section, we present a proof of Theorem 3.12. The proof uses the following fact from [vdWANB49]:

**Fact C.1.** *Let $f : \mathbb{F}^m \to \mathbb{F}$ be a function, and let $N_{y,h} = \{y + ih \mid i \in \{0, \ldots, d+1\}\}$. $f$ is degree $d$ iff it satisfies the following identity for all $y$ and $h$:*

$$\sum_{i=0}^{d+1} \alpha_i f(y + ih) = 0,$$

*where $\alpha_i = \binom{d+1}{i}(-1)^{i+1}$.*

Throughout this section we let $\alpha_i = \binom{d+1}{i}(-1)^{i+1}$ as in the above fact.

**Theorem C.2** (Restatement of Theorem 3.12)**.** *Let $f : \mathbb{F}^m \to \mathbb{F}$ be a function, and let $N_{y,h} = \{y + ih \mid i \in \{0, \ldots, d+1\}\}$, if $f$ satisfies*

$$\Pr_{y,h \in \mathbb{F}^m}[\exists \deg d \text{ polynomial } p \textbf{ s.t. } p_{|_{N_{y,h}}} = f_{|_{N_{y,h}}}] \geq 1 - \delta, \tag{21}$$

*for $\delta \leq \frac{1}{2(d+2)^2}$, then there exists a degree $d$ polynomial $g$ such that $f \overset{2\delta}{\approx} g$.*

**Proof:** Define a function $g : \mathbb{F}^m \to \mathbb{F}$ to be $g(y) = \mathtt{maj}_{h \in \mathbb{F}^m}\{\sum_{i=1}^{d+1} \alpha_i f(y + ih)\}$ breaking the ties arbitrarily. Next we argue that $g$ is very close to $f$ and $g$ itself is a degree $d$ function.

To see that $g$ is $(1-2\delta)$ close to $f$, consider the set of all $y$ for which $\mathbf{Pr}_h[f(y) = \sum_{i=1}^{d+1} \alpha_i f(y+ih)] > 1/2$. For all these $y$, $f(y) = g(y)$ as $g$ was the majority vote. It is easy to see that fraction of $y$ for which the probability is at most $1/2$ is at most $2\delta$ as otherwise it will contradict the hypothesis (21). The rest of the proof will be proving the following two claims.

**Claim C.3.** *For all $y \in \mathbb{F}^m$, $\mathbf{Pr}_h[g(y) = \sum_{i=1}^{d+1} \alpha_i f(y + ih)] \geq 1 - 2(d+1)\delta$.*

**Claim C.4.** *For all $y$ and $h$ in $\mathbb{F}^m$, we have $\sum_{i=0}^{d+1} \alpha_i g(y + ih) = 0$.*

Claim C.4 and Fact C.1 imply that $g$ is in fact a degree $d$ function and hence the theorem follows. We now proceed with proving these two claims.

**Proof of Claim C.3**: We will show that for all $y \in \mathbb{F}^m$,

$$\Pr_{h_1,h_2}\left[\sum_{i=1}^{d+1} \alpha_i f(y + ih_1) = \sum_{j=1}^{d+1} \alpha_j f(y + jh_2)\right] \geq 1 - 2(d+1)\delta. \tag{22}$$

Note that this is enough to prove the claim. To see this, let $p_a = \mathbf{Pr}_h[\sum_{i=1}^{d+1} \alpha_i f(y + ih) = a]$ for $a \in \mathbb{F}$. Then (22) becomes $\sum_{a \in \mathbb{F}} p_a^2 \geq 1 - 2(d+1)\delta$. Since $g(y)$ was the majority vote, we have $\mathbf{Pr}_h[g(y) = \sum_{i=1}^{d+1} \alpha_i f(y + ih)] = \max_{a \in \mathbb{F}} p_a \geq \sum_{a \in \mathbb{F}} p_a^2 \geq 1 - 2(d+1)\delta$.

To prove (22), consider the following $(d+2) \times (d+2)$ matrix $Z$ with $(i,j)^{th}$ entry $Z_{i,j} = \alpha_i \alpha_j f(y + ih_1 + jh_2)$, for $i, j \in \{0, \ldots, d+1\}$.

$$Z = \begin{bmatrix} f(y) & \ldots & \alpha_0\alpha_j f(y + jh_2) & \ldots \\ \vdots & \ddots & \vdots & \ddots \\ \alpha_i\alpha_0 f(y + ih_1) & \ldots & \alpha_i\alpha_j f(y + ih_1 + jh_2) & \ldots \\ \vdots & \ddots & \vdots & \ddots \end{bmatrix}$$

If $h_1 \in \mathbb{F}^m$ u.a.r then for any $i \in \{1, 2, \ldots, d+1\}$, $ih_1$ is distributed uniformly in $\mathbb{F}^m$. Same is true for $h_2$ and $jh_2$. Consider the following events:

- For every $i \in \{1, 2, \ldots, d+1\}$, $R_i$ be the event that the sum of the $i$'th row is $zero$, i.e $\sum_{j=0}^{d+1} Z_{i,j} = 0$.
- For every $j \in \{1, 2, \ldots, d+1\}$, $C_j$ be the event that sum of the $j$'th column is $zero$, i.e $\sum_{i=0}^{d+1} Z_{i,j} = 0$.

Note that $R_i, C_j$ are not defined for the first row and column ($i = 0$ and $j = 0$). Using the hypothesis (21) of the theorem and Fact C.1, we have

$$\begin{aligned} \mathop{\mathbf{Pr}}_{h_1,h_2}[R_i] &\geq 1 - \delta, \qquad \forall i \in \{1, 2, \ldots, d+1\} \\ \mathop{\mathbf{Pr}}_{h_1,h_2}[C_j] &\geq 1 - \delta, \qquad \forall j \in \{1, 2, \ldots, d+1\} \end{aligned}$$

The event in (22) is same as $\sum_{i=1}^{d+1} Z_{i,0} = \sum_{j=1}^{d+1} Z_{0,j}$ (note that the sums don't include the first element, $Z_{0,0}$). If all the above events $R_i, C_j$ happen then $\sum_{i=1}^{d+1} Z_{i,0} = \sum_{j=1}^{d+1} Z_{0,j} = -\sum_{i,j=1}^{d+1} Z_{i,j}$. By using union bound we get $\mathbf{Pr}[\wedge_{i=1}^{d+1} R_i \wedge_{j=1}^{d+1} C_j] \geq 1 - 2(d+1)\delta$ which implies (22).

**Proof of Claim C.4**: In this case, consider the following $(d+2) \times (d+2)$ matrix $Y$ whose $(i,j)^{th}$ entry is $Y_{i,j} = \alpha_i \alpha_j f(y + ih + j(h_1 + ih_2))$ except when $j = 0$. When $j = 0$, $Y_{i,0} = \alpha_i \alpha_0 g(y + ih)$.

$$Y = \begin{bmatrix} \alpha_0\alpha_0 g(y) & \ldots & \alpha_0\alpha_j f(y + jh_1) & \ldots \\ \vdots & \ddots & \vdots & \ddots \\ \alpha_i\alpha_0 g(y + ih) & \ldots & \alpha_i\alpha_j f(y + ih + j(h_1 + ih_2)) & \ldots \\ \vdots & \ddots & \vdots & \ddots \end{bmatrix}$$

Define the following set of events:

- For $i \in \{0, 1, \ldots, d+1\}$, $R_i$ be the event that the sum of all elements from row $i$ is $zero$, i.e $\sum_{i=0}^{d+1} Y_{i,j} = 0$.
- For $j \in \{0, 1, \ldots, d+1\}$, $C_j$ be the event that the sum of all elements from column $j$ is $zero$, i.e $\sum_{j=0}^{d+1} Y_{i,j} = 0$.

Let $h_1, h_2$ are independent and distributed u.a.r in $\mathbb{F}^m$. As the event $C_0$ is independent of $h_1$ and $h_2$, in order to prove the claim it is enough to show that $\mathbf{Pr}_{h_1,h_2}[C_0] > 0$.

For each row $i \in \{0, 1, 2, \ldots, d+1\}$ we apply Claim C.3 with $y' = y + ih$ and $h' = h_1 + ih_2$, and get $\mathbf{Pr}_{h_1,h_2}[\neg R_i] \leq 2(d+1)\delta$ (note that $\alpha_0 = -1$). If $h_1, h_2$ are independent and distributed u.a.r in $\mathbb{F}^m$ then so are $(y + jh_1)$ and $(h + h_2)$. Therefore, using the hypothesis (21) of the theorem and Fact C.1, we have for all columns except $j = 0$, $\mathbf{Pr}_{h_1,h_2}[\neg C_j] \leq \delta$. Using union bound, we get

$$\mathop{\mathbf{Pr}}_{h_1,h_2}\left[\bigwedge_{i=0}^{d+1} R_i \bigwedge_{j=1}^{d+1} C_j\right] \geq 1 - 2(d+1)(d+2)\delta + (d+1)\delta > 0.$$

The claim now follows using the observation that the event $C_0$ is implied by the event $\wedge_{i=0}^{d+1} R_i \wedge_{j=1}^{d+1} C_j$. To see this, the event $\wedge_{i=0}^{d+1} R_i$ implies that the sum of all entries in $Y$ is $zero$ whereas $\wedge_{j=1}^{d+1} C_j$ implies that the sum of all elements from the submatrix $(Y_{i,j})_{j=1}^{d+1}$ is $zero$. Hence, if both these events happen then the sum of all elements from column 0 must be $zero$. ■



ECCC ISSN 1433-8092
http://eccc.hpi-web.de