6.875 PCP and Hardness of Approximation MIT, Fall 2010

Lecture 5: Low Degree Testing

Lecturer: *Dana Moshkovitz* Scribe: *Gregory Minton and Dana Moshkovitz*

Having seen a probabilistic verifier for linearity of Boolean functions, we now move on to the goal of general low degree testing. Let us restate what we wish to show:

Theorem 1 (Low Degree Testing)**.** *There are constants $\delta, \gamma, \gamma' > 0$, such that given the table of a function $f : \mathbb{F}^m \to \mathbb{F}$ and a degree $d \leq \delta |\mathbb{F}|$, there is a probabilistic verifier for the statement "$\deg f \leq d$". The verifier is given access to f and to an auxiliary proof and satisfies the following:*

- Completeness: *If $\deg f \leq d$, then there is a proof that the verifier always accepts.*
- Soundness: *If f is γ-far from degree d, then for any proof, the verifier rejects with probability at least γ'.*

The verifier uses $O(\log(|\mathbb{F}|^m))$ random bits. It makes only $|\mathbb{F}|^{O(1)}$ queries to f.

1 The Line vs Point Test

If we were in the case $m = 1$, then there is a natural solution to the low degree testing problem: we simply ask for the prover to give us a polynomial P of degree d which supposedly equals f, and then we pick a random point x and test that $f(x) = P(x)$. For any P, the probability that this test fails is exactly $\Delta(f, P)$; thus we get the desired completeness and soundness conditions.

The following test generalizes this idea to higher dimensions by working with lines in $\mathbb{F}^m$.

Line vs. Point Test (Rubinfeld-Sudan)**.** *The auxiliary proof π consists of a list of polynomials of degree $\leq d$, one for each line ℓ in $\mathbb{F}^m$. Each polynomial is given by $d+1$ coefficients, and the polynomial $\pi(\ell)$ is supposed to represent f restricted to the line ℓ.*

For the test, pick a random line ℓ and a random point $x \in \ell$. Check that $\pi(\ell)(x) = f(x)$, where the left side is an evaluation of the polynomial $\pi(\ell)$ and the right side is a table value.

There is a technical issue in the statement of the Line vs. Point Test, namely in how we represent a function on ℓ. The natural way to do this is to parameterize ℓ as $\{x + ty\}$ for some $x, y \in \mathbb{F}^m$ ($y \neq 0$); then a function on $\mathbb{F}^m$ restricted to ℓ may be thought of as a univariate function of t. This has the caveat that the parameterization is not unique. However, this is not a concern for us; we simply choose a parameterization for each line arbitrarily.

Some basic properties of the Line vs. Point Test are obvious. We need to randomly pick a line and a point on that line; we can do that by just randomly picking two distinct points, so the randomness is $2 \log |\mathbb{F}^m|$. (In fact, this is an overestimate by about $\log |\mathbb{F}|$, but that observation is irrelevant for the asymptotics.) The test clearly has completeness 1, as the restriction of a low degree multivariate polynomial to a line is a low degree univariate polynomial. The number of queries is 2, if we think of π as being a table whose alphabet is of size $|\mathbb{F}|^{d+1}$ (so that we can look up all $d+1$ coefficients at once). If we instead think of the alphabet as being $\mathbb{F}$, then the test makes $d+2$ queries (1 for $f(x)$, and $d+1$ to look up $\pi(\ell)$).

The soundness of the Line vs. Point Test is the interesting part of the analysis. The original Rubinfeld-Sudan analysis proved that the Line vs. Point Test satisfies Theorem 1:

Theorem (Rubinfeld-Sudan)**.** *For some $\epsilon, \epsilon' > 0$, for all f, if there exists π such that*

$$\Pr[\textit{Line-Point accepts } f, \pi] \geq 1 - \epsilon,$$

then f is $1 - \epsilon'$-close to degree d.

If the probability of acceptance is close to 1, then we get a bound on the minimum distance of f to a low degree polynomial. Such results are sometimes referred to as "high-end" results. The theorem says nothing when the acceptance probability is not close to 1. However, it was later shown that any non-negligible acceptance probability corresponds to proximity of f to low degree:

Theorem (Arora-Sudan, '97)**.** *For all $f : \mathbb{F}^m \to \mathbb{F}$, $\max_\pi \Pr[\textit{Line-Point rejects } f, \pi]$ is the same as the distance of f from degree d, up to an additive error in $m^{O(1)} d^{O(1)} / |\mathbb{F}|^{\Omega(1)}$.*

2 The Plane vs Point Test

Instead of analyzing the Line vs. Point test, we shall analyze the following closely related and easier to analyze test, the Plane vs. Point Test. This is basically the same as the Line vs. Point Test; we have just replaced "line" with "plane".

Plane vs. Point Test (Raz-Safra)**.** *The auxiliary proof π consists of a list of bivariate polynomials of degree $\leq d$, one for each plane s in $\mathbb{F}^m$. Each polynomial is given by a list of $(d+1)(d+2)/2$ coefficients, and the polynomial $\pi(s)$ is supposed to represent f restricted to the plane s.*

For the test, pick a random plane s and a random point $x \in s$. Check that $\pi(s)(x) = f(x)$, where the left side is an evaluation of the polynomial $\pi(s)$ and the right side is a table value.

Note that the same technical issue of representation of the polynomials arises as in the Line vs. Point test; we resolve it by choosing an arbitrary parameterization for every plane.

As above, all of the properties except soundness are fairly obvious. The randomness is asymptotically $3 \log |\mathbb{F}^m|$, as we can just pick the point x and then two (linearly independent) vectors supporting the plane s. The number of queries is 2, if we think of the alphabet for π as being of size $|\mathbb{F}|^{(d+1)(d+2)/2}$; otherwise, if we use $\mathbb{F}$ as the alphabet, the number of queries is $(d+1)(d+2)/2 + 1$. The test certainly has completeness 1, as the restriction of a multivariate polynomial of low degree to a plane is a bivariate polynomial of low degree.

The following theorem shows the soundness of the Plane vs. Point Test.

Theorem (Raz-Safra, '97)**.** *For all f, $\max_\pi \Pr[\textit{Plane-Point rejects } f, \pi]$ is the same as the distance of f from degree d, up to an additive error in $m^{O(1)} (d/|\mathbb{F}|)^{\Omega(1)}$.*

Comparing with above, we see that the Line vs. Point Test has better randomness and alphabet size (or number of queries, depending on one's point of view). On the other hand, comparing the Raz-Safra Theorem with the Arora-Sudan Theorem, the soundness bound on the Plane vs. Point Test has better degree to field size tradeoff. These differences do not matter for our application.

3 Analyzing the Plane vs Point Test

The advantage of the Plane vs. Point Test is that it admits a nicer analysis than the Line vs. Point Test. Our plan is to give a form of the Raz-Safra analysis for $m = 3$. (The Plane vs. Point Test is only interesting for $m \geq 3$, as if $m = 2$ then there is exactly one plane, and if $m = 1$, then there are no planes) One can then extend the result via induction to general m. We will skip this induction here, although we shall feel free to use the general m result later. The interested reader may consult [1], where the analysis of the Plane vs. Point Test can be found as a special case of a more general test.

The remainder of this section is devoted to proving the Raz-Safra Theorem for $m = 3$.

First note that for every $\gamma > 0$, if f is γ-close to a polynomial p of degree at most d, then necessarily $\max_\pi \Pr[\text{Plane-Point rejects } f, \pi] \geq \gamma$. This is by taking $\pi(s)$ to be the restriction of p to s. Thus, to prove the Raz-Safra theorem, it is enough to show the following: Assume there exists π such that

$$\Pr[\text{Plane-Point accepts } f, \pi] \geq \gamma,$$

for some γ sufficiently large with respect to $m^{O(1)}(d/|\mathbb{F}|)^{\Omega(1)}$. Then, there exists a polynomial $p : \mathbb{F}^m \to \mathbb{F}$ of degree at most d, such that the agreement between f and p, $1 - \Delta(f, p)$, is $\gtrsim \gamma$, where the $\gtrsim$ hides an additive $m^{O(1)}(d/|\mathbb{F}|)^{\Omega(1)}$.

Our overall strategy for this proof is to define a graph encoding the agreement between different planes, and then analyze that graph's combinatorial properties. These properties will yield the polynomial we want.

The vertices in our graph are the planes in $\mathbb{F}^3$, and a pair (s_1, s_2) of planes has an edge between them if $\pi(s_1)$ and $\pi(s_2)$ agree on their intersection. Note that there always exists an edge between parallel planes, as distinct parallel planes are disjoint (so trivially agree on their intersection, as the intersection is empty). As the proof will be rather lengthy, we first summarize what we shall prove:

(a) The graph has a lot of edges, where "a lot" is related to the probability γ of acceptance of the Plane vs. Point Test. The edges correspond to agreement of π with f.

(b) By removing a small number of the edges, we are left with a disjoint union of cliques. From this we deduce that most edges are in large cliques.

(c) Each large clique corresponds to a single low degree polynomial that agrees with all of the planes in the clique.

(d) One of these polynomials must be close to f.

Whenever convenient, we shall refer to the graph as $G = (V, E)$; so two edges $s_1, s_2 \in V$ agree on their intersection iff $(s_1, s_2) \in E$. In an abuse of notation, we shall write that $(s, s) \in E$ for any plane s, even though we do not think of the graph G as having self-loops. (This just gives us a convenient notation for when two planes agree on their intersection.)

3.1 Step (a): The Consistency Graph is Dense

The first bound we will establish relies on Jensen's Inequality, a basic but useful result in integration theory.

Theorem (Jensen)**.** *If f is a convex (real) function and X is a random (real-valued) variable, $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$.*

Corollary 3.1. *If X is a random (real-valued) variable, then $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$.*

Proof. Take $f(x) = x^2$ in Jensen's Inequality. Note that this result has the simple description of "the variance is positive", as one formula for the variance is $\operatorname{Var} X = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. □

Claim 3.2. $\Pr_{s_1,s_2,x\in s_1\cap s_2}[\pi(s_1)(x) = f(x) = \pi(s_2)(x)] \geq \gamma^2$.

Proof. Note that the left side above requires some explanation: we think of it as the probability of the event $\pi(s_1)(x) = f(x) = \pi(s_2)(x)$ when we sample x at random, and then randomly (and independently) choose planes s_1 and s_2 which contain x.

For a plane s and a point $x \in s$, let $I_{s,x}$ denote the indicator variable for the event $\pi(s)(x) = f(x)$. Then the left side is exactly

$$\Pr_{s_1,s_2,x\in s_1\cap s_2}[\pi(s_1)(x) = f(x) = \pi(s_2)(x)] = \mathbb{E}_x \mathbb{E}_{s_1,s_2\ni x}[I_{s_1,x} I_{s_2,x}].$$

Because s_1 and s_2 are independent and identically distributed, we can rewrite this as

$$\mathbb{E}_x\Big[(\mathbb{E}_{s\ni x} I_{s,x})^2\Big].$$

Applying Corollary 3.1, this is $\geq (\mathbb{E}_x[\mathbb{E}_{s\ni x} I_{s,x}])^2 = (\mathbb{E}_{s,x\in s}[I_{s,x}])^2 = \gamma^2$, as desired. □

To extend this claim, note that there are only three possibilities for how two planes in $\mathbb{F}^3$ intersect: they may be disjoint, they may be equal, or they may intersect in a line. (This observation is a homework exercise.) Thus if we have $x \in s_1 \cap s_2$, then either $s_1 \cap s_2$ is a line or $s_1 = s_2$. With this, we can show the following.

Claim 3.3. $\Pr_{s_1,s_2,x\in s_1\cap s_2}[(s_1,s_2) \in E, \pi(s_1)(x) = \pi(s_2)(x) = f(x)] \geq \gamma^2 - \dfrac{d}{|\mathbb{F}|}$.

Proof. Suppose we have planes $s_1, s_2 \in V$ such that $(s_1, s_2) \notin E$. Then $s_1 \neq s_2$ and $s_1 \cap s_2 \neq \phi$, so $s_1 \cap s_2$ is a line on which $\pi(s_1)$ and $\pi(s_2)$ do not agree. The restrictions of $\pi(s_1)$ and $\pi(s_2)$ to the line $s_1 \cap s_2$ are univariate polynomials of degree $\leq d$. Thus if these polynomials are not equal identically, then they are equal in at most d points, and so $\Pr_{x\in s_1\cap s_2}[\pi(s_1)(x) = \pi(s_2)(x)] \leq d/|\mathbb{F}|$.

Now, suppressing the subscript $s_1, s_2, x \in s_1 \cap s_2$ for all probabilities,

$$\Pr[(s_1,s_2) \in E, \pi(s_1)(x) = \pi(s_2)(x) = f(x)]$$

is at least

$$\Pr[\pi(s_1)(x) = \pi(s_2)(x) = f(x)] - \Pr[\pi(s_1)(x) = \pi(s_2)(x) = f(x), (s_1,s_2) \notin E].$$

By the combination of the bounds from the last paragraph and from the previous claim, we see that this is $\geq \gamma^2 - d/|\mathbb{F}|$, as desired. □

The error bound of $m^{O(1)}(d/|\mathbb{F}|)^{\Omega(1)}$ allows us to absorb the $d/|\mathbb{F}|$ from the claim above. In addition, since we are working with $m = 3$, a uniformly random pair of planes intersect each other with high probability $> 1 - 1/|\mathbb{F}|$. Thus, the error bound also allows us to absorb the error $< 1/|\mathbb{F}|$ coming from the difference between picking a uniformly random pair of intersecting planes (as in Claim 3.3) and picking a uniformly random pair of planes. Let us explain what we mean by this. Note that in the expression in the claim above, we end up sampling from pairs (s_1, s_2) with a not-quite-uniform distribution. Ignoring the case $s_1 = s_2$, which is rare enough to be absorbed into the error bound, we will never sample pairs (s_1, s_2) that are parallel. Because all other pairs of planes intersect at a line, i.e. they intersect at the same number of points, we end up uniformly sampling from the set of pairs of intersecting planes. As parallel planes are rare (in particular, they make up $< 1/|\mathbb{F}|$ of the total pairs of planes), we can absorb this non-uniformity into the error bound. Thus when we write $\gtrsim$, we need not worry about the exact probability distribution being used to sample pairs of edges.

With this technical discussion out of the way, one high-level interpretation of the above claim is that $\Pr[(s_1, s_2) \in E] \gtrsim \gamma^2$, i.e. G contains at least (roughly) γ^2 of the possible edges. We can be more precise, though. Define the weight of a pair (s_1, s_2) of planes to be

$$w(s_1, s_2) = \Pr_{x \in s_1 \cap s_2}[\pi(s_1)(x) = f(x) = \pi(s_2)(x)].$$

(If $s_1 \cap s_2 = \phi$, then we arbitrarily choose to define $w(s_1, s_2) = 1$, say. This is not important, as $s_1 \cap s_2 = \phi$ is rare.) Then the claim tells us that the total weight of all of the edges in the graph G is $\gtrsim \gamma^2$ of its maximum possible value (which equals the total number of pairs of vertices).

4 Step (b): The Consistency Graph is Close to a Disjoint Union of Cliques

We have shown that G contains a lot of edges, completing step (a) in our roadmap. We next begin work on step (b), showing that the graph is almost a disjoint union of cliques. The key here is to show that G satisfies an "almost transitive" property.

Definition 2. *A graph $G = (V, E)$ is β-almost transitive if, for all pairs of distinct vertices $(a, b) \notin E$, $\Pr_{c \in V}[(c, a), (c, b) \in E] \leq \beta$.*

To help clarify this definition, note that 0-almost transitivity is just regular transitivity, which corresponds to the graph being a disjoint union of cliques.

Lemma 4.1. *Our graph G is $\frac{d+1}{|\mathbb{F}|}$-almost transitive.*

Proof. Suppose we have a pair $(a, b) \notin E$. Then, by definition, the planes a and b intersect in a line, say ℓ, and the polynomials $\pi(a)$ and $\pi(b)$ do not agree on ℓ. Pick a random plane $c \in V$; we want to bound the probability that c is adjacent to both a and b. There are two cases to consider.

First, c be parallel to the line ℓ (but not containing it). But for this to happen, any normal to c must be perpendicular to ℓ, and this happens with probability $< 1/|\mathbb{F}|$. (A random normal of c is uniform on the nonzero vectors of $\mathbb{F}^m$. If we sample uniformly from $\mathbb{F}^m$, then the dot product with the direction of ℓ is uniform on $\mathbb{F}$, so vanishes with probability $1/|\mathbb{F}|$. If we exclude the zero vector, then the probability of vanishing decreases by a tiny amount.)

In the second case, c is not parallel to ℓ, so it intersects ℓ at exactly one point. It is clear that the intersection point is uniform on ℓ. As $\pi(a)$ and $\pi(b)$ do not restrict to the same polynomial on ℓ, they can agree on at most d points. Thus if we randomly sample a point $x \in \ell$, then the probability that $\pi(a)(x) = \pi(b)(x)$ is at most $d/|\mathbb{F}|$. This of course upper bounds the probability that $\pi(a)(x) = \pi(c)(x) = \pi(b)(x)$, which in turn upper bounds the probability that c is adjacent to both a and b in our graph G.

Adding the bounds of the last two paragraphs, we see that c is adjacent to both a and b with probability $\leq (d+1)/|\mathbb{F}|$, as desired. □

The almost transitivity property is exactly what we need to deduce that our graph is almost a disjoint union of cliques. For this we use the following graph-theoretic lemma.

Lemma 4.2. *Given a graph on n vertices which is β-almost transitive, the graph can be made transitive by removing $\leq O(n^2\sqrt{\beta})$ edges.*

Proof. We prove this result by presenting an algorithm which does not cut too many edges but terminates with a transitive graph. The algorithm is simple, but the fact that it has the desired properties is somewhat delicate. We just repeat the following procedure until the graph is transitive:

(1) If there is a vertex which is adjacent to $\leq \sqrt{\beta}$ of the vertices, then remove all edges incident upon that vertex (i.e. isolate that vertex).

(2) If there are no vertices satisfying (1), then pick a vertex v such that there exist elements in v's connected component which are not adjacent to v. Remove all edges between the set of vertices adjacent to v and the rest of the connected component.

It is clear that this process terminates in a transitive graph. (In particular, if the graph is not transitive then a vertex v can always be found in step (2). Also, every step removes at least one edge, so there can only be finitely many steps.) We just need to show that this removes $O(n^2\sqrt{\beta})$ edges.

Let n be the number of vertices in the graph. Each time we are in step (1), we remove at most $n\sqrt{\beta}$ edges. We of course do not apply step (1) to any vertex twice, so at the end of the algorithm the total number of edges removed in step (1) must be $\leq n^2\sqrt{\beta}$.

Suppose we are in the situation of step (2). Let $N(v)$ be the set of neighbors of v, and let $D(v)$ be the rest of the connected component of v. (That is, the connected component of v is the disjoint union of $\{v\}$, $N(v)$, and $D(v)$.) Choose $u \in D(v)$. Then u is not adjacent to v, so by the almost transitivity property, there are at most βn common neighbors of u and v. Rephrasing this, there are at most βn edges from u to $N(v)$. Summing over all u, we see that there are at most $|D(v)|\,\beta n$ edges removed in this application of step (2). There are $|D(v)| \cdot |N(v)|$ potential edges between the vertices in $D(v)$ and the vertices in $N(v)$, so the fraction we remove is at most

$$\frac{|D(v)|\,\beta n}{|D(v)| \cdot |N(v)|} = \frac{\beta n}{|N(v)|} \leq \frac{\beta n}{n\sqrt{\beta}} = \sqrt{\beta}.$$

To restate this, every time we apply step (2), we are looking at a set of $|D(v)| \cdot |N(v)|$ pairs of vertices, and remove edges from at most $\sqrt{\beta}$ of them. These pairs of vertices should be considered unordered as we are not double-counting any pair (t, u) and (u, t).

Now observe that if the unordered pair (t, u) is considered in a given application of step (2), then in this step we remove all edges between the connected components of t and u. Thus afterwards t and u will not be in the same connected component, so cannot be considered again in a future application of step (2). Hence every (unordered) pair of vertices is considered at most once in the course of all applications of step (2). Every time we consider a set of pairs, we remove edges from at most $\sqrt{\beta}$ of them. Thus the total number of edges removed by step (2) is $\leq \sqrt{\beta} \cdot \binom{n}{2} < \frac{1}{2}n^2\sqrt{\beta}$.

Putting this together, the algorithm removes at most $n^2\sqrt{\beta} + \frac{1}{2}n^2\sqrt{\beta} = \frac{3}{2}n^2\sqrt{\beta}$ edges, which is certainly $O(n^2\sqrt{\beta})$. □

Note that since we can make any graph a disjoint union of cliques by removing all of the edges, for the lemma to be meaningful the graph must be dense. As we showed in step (a), our graph is dense.

To complete step (b) in our proof roadmap, we just need to observe the corollary that most of the edges in our graph lie in large cliques. The proof is basically that small cliques contain very few edges. The reason for our precise choice of "large" in the corollary will become clear in the next step.

Corollary 4.3. *By removing $O(|V|^2\sqrt{d/|\mathbb{F}|})$ edges from our graph G, we obtain a graph which is the disjoint union of cliques, all of which are either singletons or of relative size more than $(2d+1)/|\mathbb{F}|$.*

Proof. As the original graph is $(d+1)/|\mathbb{F}|$-almost transitive, by applying Lemma 4.2, we obtain a disjoint union of cliques by removing $O(|V|^2\sqrt{d/|\mathbb{F}|})$ edges. Next, we remove all edges present in cliques of relative size $\le (2d+1)/|\mathbb{F}|$. Let these small cliques have sizes $c_1, \dots, c_r$, and suppose without loss of generality that c_1 is the largest. Then the number of edges present in these cliques is

$$\sum_{i=1}^{r} \binom{c_i}{2} < \frac{1}{2}\sum_{i=1}^{r} c_i^2 \le \frac{1}{2}\sum_{i=1}^{r} c_1 c_i = \frac{1}{2}c_1\left(\sum_{i=1}^{r} c_i\right) \le \frac{(2d+1)\,|V|^2}{2\,|\mathbb{F}|}.$$

This is in $O(|V|^2\, d/|\mathbb{F}|)$. As $d/|\mathbb{F}| < 1$, it is also in $O(|V|^2\sqrt{d/|\mathbb{F}|})$, as desired. □

5 Step (c): Large Cliques Correspond to Low Degree Polynomials

The following interpolation lemma addresses step (c) in our roadmap, and at the same time answers why we made the particular choice of $(2d+1)/|\mathbb{F}|$ above.

Lemma 5.1. *Suppose $C \subseteq V$ is a clique containing more than $(2d+1)/|\mathbb{F}|$ of the vertices (planes). Then there is a polynomial P of degree $\le 2d$ such that P agrees with $\pi(s)$ for every plane $s \in C$.*

Proof. We can partition all planes into classes, where each class is identified with linear plane in $\mathbb{F}^3$ (i.e., a plane through $\vec{0}$). A plane in $\mathbb{F}^3$ is in the class if it is an affine shift of the linear plane. Each class contains $|\mathbb{F}|$ planes. Averaging over all classes, the expected proportion of planes from the class in C is more than $(2d+1)/|\mathbb{F}|$. Thus there certainly exists some linear plane s_1 (with normal y_1, say) such that $2d+1$ affine shifts $\{c_1^i y_1 + s_1\}_{i=0}^{2d}$ are in C.

There are $(|\mathbb{F}|^3 - 1)/(|\mathbb{F}| - 1) = |\mathbb{F}|^2 + |\mathbb{F}| + 1$ total linear planes, because we can specify each by choosing a normal vector, which is well-defined up to scaling. There could theoretically be as many as $|\mathbb{F}|$ affine shifts of s_1 in C. If every other linear plane $s \neq s_1$ had $\le 2d$ affine shifts in C, then we would have

$$|C| \le |\mathbb{F}| + (|\mathbb{F}|^2 + |\mathbb{F}|)2d = 2d\,|\mathbb{F}|^2 + (2d+1)\,|\mathbb{F}|\,.$$

Now there are $|\mathbb{F}|\,(|\mathbb{F}|^2 + |\mathbb{F}| + 1)$ total planes, so our assumption says that

$$|C| \ge \frac{2d+1}{|\mathbb{F}|} \cdot |\mathbb{F}|\,(|\mathbb{F}|^2 + |\mathbb{F}| + 1) = (2d+1)(|\mathbb{F}|^2 + |\mathbb{F}| + 1).$$

The last two inequalities are contradictory. Thus we deduce that there is at least other linear plane s_2 (with normal y_2, say) such that $2d+1$ affine shifts $\{c_2^i y_2 + s_2\}_{i=0}^{2d}$ are in C.

Now we have a collection of $2d+1$ polynomials $\{\pi(c_1^i y_1 + s_1)\}_{i=0}^{2d}$, each of degree $\leq d$ and each on a distinct (parallel) plane. By an interpolation exercise, we can find a polynomial P_1 on $\mathbb{F}^3$ of degree $\leq 2d$ which agrees with each $\pi(c_1^i y_1 + s_1)$ on the plane $c_1^i y_1 + s_1$. Similarly, we can find a polynomial P_2 on $\mathbb{F}^3$ of degree $\leq 2d$ which agrees with each $\pi(c_2^i y_2 + s_2)$ on the plane $c_2^i y_2 + s_2$.

The idea is that $2d+1$ shifts are enough to ensure that any plane in the clique agrees with the polynomial P_1 (and, symmetrically, with P_2 as well). To demonstrate this idea, we first show that $P_1 = P_2$. Towards this end, fix $i \in \{0, 1, \ldots, 2d\}$ and consider the plane $t_1^i = c_1^i y_1 + s_1$. We shall show that P_1 and P_2 have the same restriction to t_1^i. To do this, let $j \in \{0, 1, \ldots, 2d\}$ be arbitrary and consider also the plane $t_2^j = c_2^j y_2 + s_2$. As t_1^i and t_2^j are not parallel, they intersect at a line $\ell = t_1^i \cap t_2^j$. Now both t_1^i and t_2^j are in the clique C, so they are adjacent in the graph G, and thus the polynomials $\pi(t_1^i)$ and $\pi(t_2^j)$ agree on the intersection ℓ. Now $\pi(t_1^i)$ is the restriction of P_1 and $\pi(t_2^j)$ is the restriction of P_2, by definition. Thus we have identified $|\ell| = |\mathbb{F}|$ points $x \in t_1^i$ on which $P_1(x) = P_2(x)$.

Repeating this for all j, we get a total of $|\mathbb{F}|\,(2d+1)$ points. (Note that we do indeed get all distinct points this way, as the planes $c_2^j y_2 + s_2$ are disjoint for distinct values of j.) As P_1 and P_2 both have degree $\leq 2d$, the restrictions of P_1 and P_2 to the plane t_1^i both have degree $\leq 2d$. Hence by Schwartz-Zippel, the fact that they agree on $|\mathbb{F}|\,(2d+1)$ of the $|\mathbb{F}|^2$ points in the plane tells us that they agree on the entire plane.

We have shown that P_1 and P_2 have the same restriction to the plane t_1^i, which consists of $|\mathbb{F}|^2$ points. Now repeating this for varying i, we find that P_1 and P_2 have the same restriction to every plane $c_1^i y_1 + s_1$, which is a total of $|\mathbb{F}|^2\,(2d+1)$ points. Again applying Schwartz-Zippel (this time in three dimensions), we deduce that $P_1 = P_2$, as claimed. With this in mind, we now define P to be the polynomial $P_1 = P_2$.

At this point we are almost done. Indeed, let $s \in C$ be any plane in the clique C. Then s cannot be parallel to both s_1 and s_2. Assume without loss of generality that it is not parallel to s_1. Then s intersects each affine shift $c_1^i y_1 + s_1$ at a line, and the clique condition gives us consistency on those lines, so we can apply the same argument as above to deduce that the restriction of P to s agrees with $\pi(s)$. That completes the lemma. □

6 Step (d): Deducing f is Close to a Low Degree Polynomial

By the corollary above, we can remove a small number of edges and obtain a graph consisting of a disjoint union of large cliques. Let $C_1, \ldots, C_t$ be the resulting cliques. Applying the interpolation lemma, we obtain a polynomial P_i for each clique C_i such that $\pi(s)$ agrees with the restriction of P_i to s for every plane $s \in C_i$. For each i, let $A_i = \{x \in \mathbb{F}^3 : f(x) = P_i(x)\}$. Our goal in step (d) is to show that at least one of the sets A_i is large.

For the analysis we use the following lemma proved in the homework:

Lemma 6.1 (Sampling). *Let $A \subseteq \mathbb{F}^3$. Let $\varepsilon > 0$. Then, if we select uniformly at random a line l in $\mathbb{F}^3$,*

$$\Pr_l\left[\left|\frac{|l \cap A|}{|\mathbb{F}|} - \frac{|A|}{|\mathbb{F}|^3}\right| \geq \varepsilon\right] \leq \frac{1}{\varepsilon^2}\frac{1}{|\mathbb{F}|}\frac{|A|}{|\mathbb{F}|^3}.$$

Corollary 6.2. *Let $\varepsilon > 0$. If we select uniformly at random a line l in $\mathbb{F}^3$,*

$$\Pr_l \left[\exists 1 \leq i \leq t, \ \left| \frac{|l \cap A_i|}{|\mathbb{F}|} - \frac{|A_i|}{|\mathbb{F}|^3} \right| \geq \varepsilon \right] \leq \frac{1}{\varepsilon^2} \frac{1}{|\mathbb{F}|} \left(1 + \frac{d}{|\mathbb{F}|} \right).$$

Proof. We apply a union bound to Lemma 6.1 to get:

$$\Pr_l \left[\exists 1 \leq i \leq t, \ \left| \frac{|l \cap A_i|}{|\mathbb{F}|} - \frac{|A_i|}{|\mathbb{F}|^3} \right| \geq \varepsilon \right] \leq \sum_{i=1}^{t} \frac{1}{\varepsilon^2} \frac{1}{|\mathbb{F}|} \frac{|A_i|}{|\mathbb{F}|^3}.$$

The corollary follows from bounding $\sum_{i=1}^{t} \frac{|A_i|}{|\mathbb{F}|^3} \leq 1 + d/|\mathbb{F}|$: By the inclusion-exclusion formula together with the Schwartz-Zippel Lemma,

$$|\mathbb{F}|^3 = \left| \bigcup_{i=1}^{t} A_i \right| \leq \sum_{i=1}^{t} |A_i| - \sum_{1 \leq i < j \leq t} |A_i \cap A_j| \leq \sum_{i=1}^{t} |A_i| - d \, |\mathbb{F}|^2 .$$

□

Next we prove the main statement of Step (d):

Claim 6.3. *For some $1 \leq i \leq t$, we have $|A_i|/|\mathbb{F}|^3 \gtrsim \gamma^2/4$.*

Proof. Combining Claim 3.3, Corollary 4.3 and Lemma 5.1, we arrive at the conclusion that with probability $\gtrsim \gamma^2$, if we pick uniformly at random a pair of planes s_1, s_2 and a point $x \in s_1 \cap s_2$, the following hold:

(i) The intersection $s_1 \cap s_2$ is a line in $\mathbb{F}^3$;

(ii) $\pi(s_1), \pi(s_2)$ agree with a polynomial P_i;

(iii) $P_i(x) = f(x)$.

Let $\delta > 0$ be a parameter we set shortly. Then, if we pick uniformly at random a pair of planes s_1, s_2, with probability at least δ the following hold:

(i) The intersection $s_1 \cap s_2$ is a line l in $\mathbb{F}^3$;

(ii) $\pi(s_1), \pi(s_2)$ agree with a polynomial P_i;

(iii) $\frac{|l \cap A_i|}{|\mathbb{F}|} \gtrsim \gamma^2 - \delta$.

Since l is distributed uniformly over all lines, we deduce that if we pick uniformly at random a line l, with probability at least δ there exists $1 \leq i \leq t$ such that $\frac{|l \cap A_i|}{|\mathbb{F}|} \gtrsim \gamma^2 - \delta$. By Corollary 6.2 invoked with $\varepsilon = \gamma^2/2$, with probability at least $\delta - \frac{4}{\gamma^4} \frac{1}{|\mathbb{F}|}(1 + \frac{d}{|\mathbb{F}|})$, both $\frac{|l \cap A_i|}{|\mathbb{F}|} \gtrsim \gamma^2 - \delta$ and $\frac{|l \cap A_i|}{|\mathbb{F}|} \leq \frac{|A_i|}{|\mathbb{F}|^3} + \gamma^2/2$; hence, $\frac{|A_i|}{|\mathbb{F}|^3} \gtrsim \gamma^2/2 - \delta$. Take $\delta = \gamma^2/4$. For sufficiently large γ with respect to $m^{O(1)}(d/|\mathbb{F}|)^{\Omega(1)}$, we have that $\delta - \frac{4}{\gamma^4} \frac{1}{|\mathbb{F}|}(1 + \frac{d}{|\mathbb{F}|}) > 0$, and, thus, $\frac{|A_i|}{|\mathbb{F}|^3} \gtrsim \gamma^2/4$ for some $1 \leq i \leq t$.

□

As A_i is a set on which f agrees with a polynomial of degree $\leq 2d$, we have shown that f is $\gtrsim \gamma^2/4$-close to degree $2d$. This is not quite what the Raz-Safra Theorem asserts – we wanted to show that f is $\gtrsim \gamma$-close to degree d. Thus to get the theorem we should reduce the degree from $2d$ to d, and replace $\gamma^2/4$ with γ. We leave it to homework to show how to obtain these stronger conclusions.

References

[1] D. Moshkovitz and R. Raz. Sub-constant error low degree test of almost-linear size. *SIAM Journal on Computing*, 38(1):140–180, 2008.