1 Low Degree Testing

In this lecture, we introduce the notion of low degree testing. We have already studied multivariate polynomial codes, where codewords correspond to multivariate polynomials of bounded degree. Here, in the spirit of self-testing, we seek to devise algorithms to determine if a given function is "close" to a low degree polynomial, that is, if a particular string is close to a codeword from a multivariate polynomial code. Our algorithms will have black-box access to the function being investigated, and we seek algorithms that minimize the number of queries made to the function, as well as the number of random bits required.

Our notion of "closeness" will be the usual one, corresponding to the Hamming metric on strings. Formally,

Definition 1 *Let f and g be two functions mapping a discrete domain D to a range R. We say f is* ϵ-close *to g if*

$$\Pr_{x \in D}[f(x) \neq g(x)] \leq \epsilon.$$

A low degree test, for some degree d, given oracle access to a function $f : \mathbf{F}^m \to \mathbf{F}$, tests if f is ϵ-close to any polynomial g over $\mathbf{F}$ on m variables of total degree at most d. As we have before, we also allow the test to make use of an auxiliary oracle $\mathbf{O}$ that help it in its testing.

We now specify the desired behavior of a low degree test T as follows: If for some oracle $\mathbf{O}$, we have that:

$$\Pr[T^{f,\mathbf{O}} \text{ rejects}] < \delta_0$$

Then, letting $\delta = \Pr[T^{f,\mathbf{O}} \text{ rejects}]$, we have that there exists some polynomial g of degree at most d such that f is 2δ-close to g.

The cases of interest to us will be when the degree d satisfies $|\mathbf{F}|^\alpha \ll d \ll |\mathbf{F}|$ for some constant $\alpha > 0$.

1.1 Test Outline

All the low degree tests we will consider will follow the following outline:

1. Make one query to $\mathbf{O}$, which returns some element g from $\mathbf{F}^{d^{O(1)}}$ (which we will typically interpret as a low degree polynomial over few variables).

2. Make one query to f, which returns an element x from $\mathbf{F}$.

3. Check for consistency between x and g.

We will now review the history of low degree tests. We describe several low degree tests that fit into the above framework, which differ essentially in how the auxiliary oracle $\mathbf{O}$ works.

1.2 Axis-Parallel Line Test

This test has its origins in the paper of Babai, Fortnow, and Lund showing the $MIP = NEXPTIME$ [5]. In this test, a random axis-parallel line is chosen, and the auxiliary oracle is asked for the univariate polynomial g corresponding to f restricted to this line. Note that when a polynomial of total degree at most d is restricted to any line, this yields a univariate polynomial of degree at most d. A random point

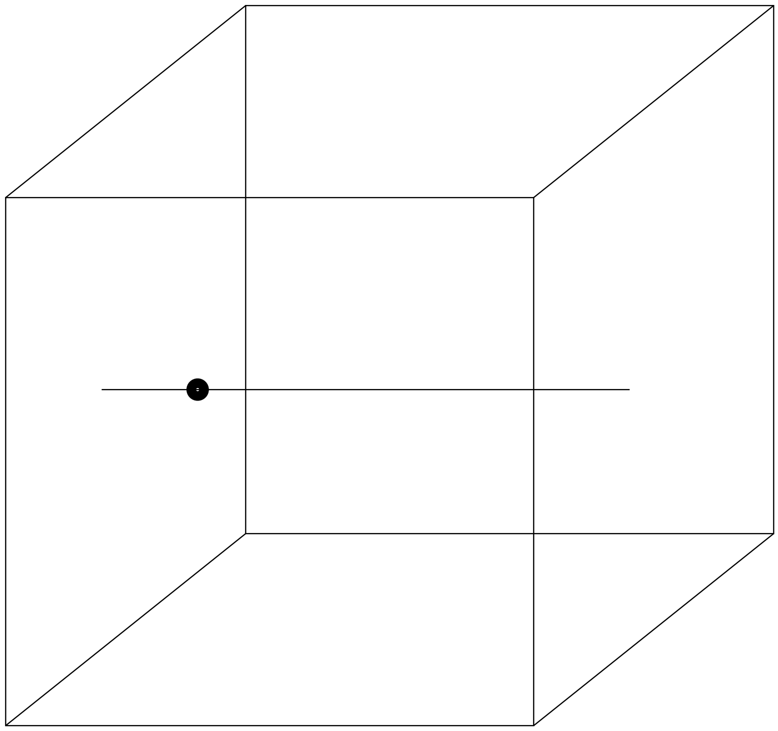

Figure 1: Axis Parallel Line Test

x on this line is selected, and a consistency check is performed between the value of f at this point, and the value of g on this point. More formally, the test is the following:

Axis-Parallel Line Test (See Figure 1):

1. Pick $i \in_R \{1, 2, \ldots, m\}$.
2. Pick $x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m \in_R \mathbf{F}$.
3. Query $\mathbf{O}(i, x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m)$ to receive degree at most d univariate polynomial g.
4. Pick $x_i \in_R \mathbf{F}$.
5. Query $y = f(x_1, \ldots, x_m)$.
6. If $y = g(x_i)$ then accept, otherwise reject.

This test underwent a great deal of analysis in the works of [5, 4, 6] (with some help from a mysterious and influential character named Shen), culminating in a proof that this test works for a value of δ_0 roughly equal to $1/md$. Arora and Safra [2] gave a new analysis, showing that this test actually works for $\delta_0 = \Omega(1/m)$, removing the dependence on the degree entirely. The dependence on the number of variables, however, is unavoidable in the axis-parallel test: Consider a function that consists of $|\mathbf{F}|$ different polynomials p_x on $(m-1)$ variables, where $f(x_1, \ldots, x_m) = p_{x_1}(x_2, \ldots, x_m)$. Now, such an f can be very far from any degree d polynomial over m variables, and yet the test will only fail if i is chosen to be 1. [8] improve the parameters in the result of Arora and Safra somewhat, but of course cannot remove the dependence on m.

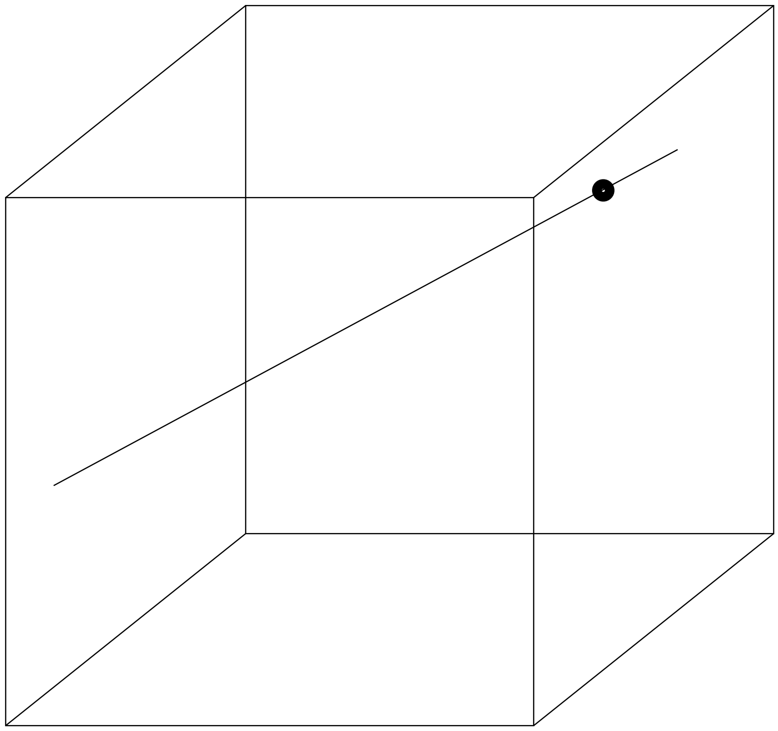

Figure 2: Point and Line Test

1.3 Point and Line Test

To overcome this dependence on m, we move to a more general test, where a random line is chosen rather than an axis-parallel line. This test was implicit in the work of Gemmel, Lipton, Rubinfeld, Sudan, and Wigderson [7], and further analyzed by [10, 1, 3].

We first introduce the notation $\ell_{x,h}$ to mean the parametric description of the line $\ell_{x,h}(t) = x + th$. Here, x and h are from $\mathbf{F}^m$, and $t \in \mathbf{F}$.

Point and Line Test (See Figure 2):

1. Pick $x, h \in_R \mathbf{F}^m$.
2. Query $\mathbf{O}(x, h)$ to receive degree at most d univariate polynomial g.
3. Pick $t_0 \in_R \mathbf{F}$.
4. Query $y = f(\ell_{x,h}(t_0)) = f(x + t_0 h)$.
5. If $y = g(t_0)$ then accept, otherwise reject.

This test was analyzed by [7, 10] yielding a result with threshold $\delta_0 = \Omega(1/d)$, independent of the dimension m. [1] combined the analyses of [10, 2] to obtain the first result where $\delta_0 = \Omega(1)$. Using more powerful techniques, Arora and Sudan [3] were able to show that this test in fact achieves $\delta_0 = 1 - o(1)$.

1.4 Point and Plane Test

Raz and Safra [9] proposed an even more general test, where a random line is replaced by a random plane. They were able to give a very elegant analysis for this test, which yields $\delta_0 = 1 - o(1)$.

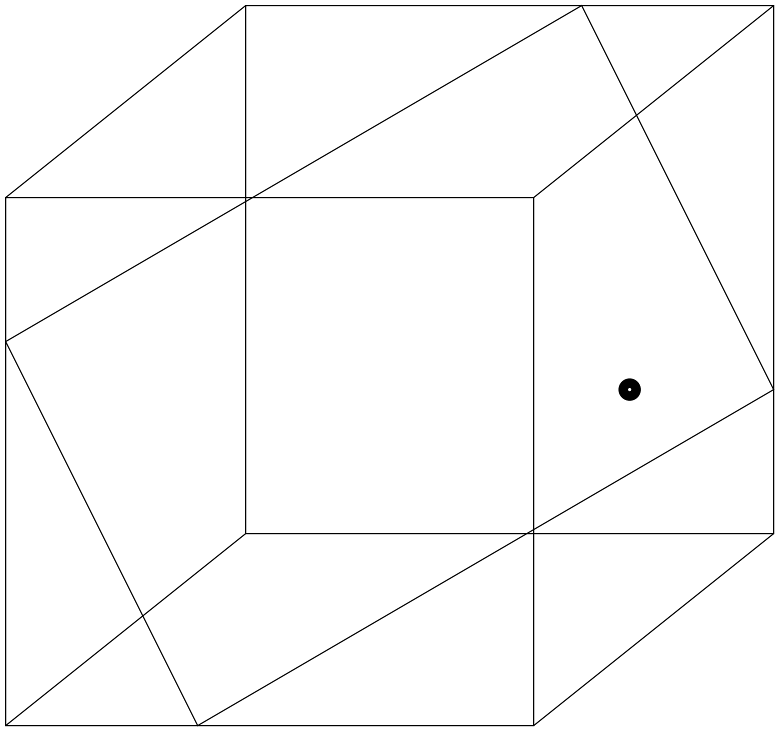

Figure 3: Point and Plane Test

We again introduce some notation: By P_{x,h_1,h_2} we will parametrically represent the plane $P_{x,h_1,h_2}(t_1,t_2) = x + h_1t_1 + h_2t_2$. Note again that any degree d polynomial over m variables, when restricted to such a plane P, will yield a degree d bivariate polynomial, which can have at most d^2 field elements. The test then formally proceeds as follows:

Point and Plane Test (See Figure 3):

1. Pick $x, h_1, h_2 \in_R \mathbf{F}^m$.
2. Query $\mathbf{O}(x, h_1, h_2)$ to receive degree at most d bivariate polynomial g.
3. Pick $t_a, t_b \in_R \mathbf{F}$.
4. Query $y = f(P_{x,h_1,h_2}(t_a, t_b)) = f(x + t_a h_1 + t_b h_2)$.
5. If $y = g(t_a, t_b)$ then accept, otherwise reject.

Over the next few lectures, we will present the analyses from [10, 2, 1] of the axis-parallel line test and the point and line test.

2 Analysis of Point and Line Test

We begin with the analysis from [10], showing that the point and line test achieves $\delta_0 = \Omega(1/d)$.

We will call the auxiliary oracle $f \mid_{lines}$, since it is supposed to provide us with f restricted to various lines. Recall δ was defined to be the probability of rejection with respect to some fixed f and $f \mid_{lines}$. Then we see that

$$\delta = \Pr_{x,h \in_R \mathbf{F}^m; t \in_R \mathbf{F}} \left[(f \mid_{lines}(x,h))(t) \neq f(x+th) \right].$$

It is immediate from this equation that the adversarially optimal choice of $f\,|_{lines}$ for any given f, in order to minimize δ, would be:

$$q_{x,h} = \operatorname*{argmax}_{g:\deg(g)\leq d}\left(\Pr_{t\in_R \mathbf{F}}[g(t) = f(x+th)]\right)$$

Note that since for any values of $x, h \in \mathbf{F}^m$ and $s \in \mathbf{F}$, the lines $\ell_{x,h}$ and $\ell_{x+sh,h}$ define the same lines, by the above definition, we can assume that $q_{x,h}(t) = q_{x+sh,h}(t-s)$ for all $t \in \mathbf{F}$. Hence, the condition that $q_{x,h}(t) = f(x+th)$ is equivalent to $q_{x+th,h}(0) = f(x+th)$. So we may re-write the expression for δ as:

$$\delta = \Pr_{x,h\in_R \mathbf{F}^m}[q_{x,h}(0) \neq f(x)]\,.$$

We now define the "self-correction" of f to be the function g, defined as follows:

$$g(x) = \operatorname*{majority}_{h\in \mathbf{F}^m}(q_{x,h}(0)).$$

Here, $\text{majority}_x(f(x))$ takes on the value y for which $y = f(x)$ for at least half of the possible x. If no such value exists, then the function is undefined.

We can finish the analysis by establishing the following three claims:

Claim 2 *If g is well defined, then g is 2δ-close to f.*

Claim 3 *If $\delta < \delta_0$, then for all $x \in \mathbf{F}^m$, we have that*

$$\Pr_{h_1,h_2\in \mathbf{F}^m}[q_{x,h_1}(0) \neq q_{x,h_2}(0)] < \frac{1}{2}$$

Claim 4 *g is a polynomial of degree at most d.*

We need Claim 3 to show that g is well-defined.

Claim 2 has a straightforward proof:

Proof We want to show that $\Pr_{x\in \mathbf{F}^m}[g(x) \neq f(x)] \leq 2\delta$. Let us define the set B (for Bad) to be $\{x \in \mathbf{F}^m \mid \Pr_{h\in \mathbf{F}^m}[f(x) \neq q_{x,h}(0)] \geq \frac{1}{2}\}$. Observe that if $x \notin B$, then by the definition of g, we have that $f(x) = g(x)$. Thus, it only remains to show that $\Pr_{x\in \mathbf{F}^m}[x \in B] \leq 2\delta$.

This, however, follows immediately, since we have:

$$\Pr_{h\in \mathbf{F}^m}[q_{x,h}(0) \neq f(x) \mid x \in B] \geq 1/2.$$

■

We will not establish Claims 4 and 3 fully until the next lecture, but we will build some of the machinery necessary to do so now. We will begin with Claim 3.

To this end, let us recall the definition $P_{x,h_1,h_2}(t_1,t_2) = x + h_1t_1 + h_2t_2$. We will focus on this plane P when x is fixed, while h_1 and h_2 are chosen uniformly and independently at random.

We make two observations:

- For any $(t_1,t_2) \neq (0,0)$ in $\mathbf{F} \times \mathbf{F}$, the point $P_{x,h_1,h_2}(t_1,t_2)$ is a uniformly random point in $\mathbf{F}^m$.
- For any $(a,b) \neq (0,0)$ in $\mathbf{F} \times \mathbf{F}$, the lines ℓ_{x+ah_1,h_2} and ℓ_{x+bh_2,h_1} are independent and uniformly random lines in $\mathbf{F}^m$.

Now for any setting of x, h_1, h_2, define the bivariate function $Q(t_1,t_2) = f(x + t_1h_1 + t_2h_2)$, and the polynomial families $C_{t_1}(\cdot) = P_{x+t_1h_1,h_2}(\cdot)$ (for "columns") and $R_{t_2}(\cdot) = P_{x+t_2h_2,h_1}(\cdot)$ (for "rows").

In the next lecture, we will see how these functions will allow us to use the results known for the axis-parallel line test to finish the analysis for the point and line test.

References

[1] Arora, Lund, Motwani, Sudan, and Szegedy. Proof verification and intractability of approximation problems. In *Proc. 33nd IEEE Symp. on Foundations of Computer Science,* pp. 13-22, 1992.

[2] Arora and Safra. Probabilistic checking of proofs: A new characterization of NP. In *Proc. 33nd IEEE Symp. on Foundations of Computer Science,* pp. 2-13, 1992.

[3] Arora and Sudan. Improved low-degree testing and its applications. In *Proc. 29th ACM Symp. on Thoery of Computing*, 1997.

[4] Babai, Fortnow, Levin, and Szegedy. Checking computations in polylogarithmic time. In *Proc. 23rd ACM Symp. on Theory of Computing,* pp. 21-31, 1991.

[5] Babai, Fortnow, and Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1:3-40, 1991.

[6] Feige, Goldwasser, Lovasz, Safra, and Szegedy. Approximating clique is almost NP-complete. In *Proc. 32nd IEEE Symp. on Foundations of Computer Science,* pp. 2-12, 1991.

[7] Gemmel, Lipton, Rubinfeld, Sudan, and Wigderson. Self-Testing/correcting for polynomials and for approximate functions. In *Proc. 23rd ACM Symp. on Theory of Computing,* pp. 32-42, 1991.

[8] Polishchuk and Spielman. Nearly linear-size holographics proofs. In *Proc. 26th ACM Symp. on Theory of Computing,* pp. 194-203, 1994.

[9] Raz and Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proc. 29th ACM Symp. on Theory of Computing*, pp. 475-484, 1997.

[10] Rubinfeld and Sudan. Testing polynomial functions efficiently and over rational domains. In *Proc. 3rd ACM-SIAM Symp. on Discrete Algorithms*, pp. 23-32, 1992.