PCPs and HDX - Lecture 4

November 29, 2016

In the previous lecture we described the Rubinfeld-Sudan low degree test. We described it as a system of constraints on functions $f : \mathbb{F}^m \to \mathbb{F}$. Today we will prove that this system of constraints is expanding with constant expansion. This is just another way of saying that if the low degree test succeeds with probability $1 - \varepsilon$ then the function must be $1 - O(\varepsilon)$ close to a low degree function.

1 Low Degree Test

Let $f : \mathbb{F}^m \to \mathbb{F}$ be a function. We will consider the following low degree test, introduced by Rubinfeld and Sudan. Let us describe it as a randomized procedure

1. Choose a random line by choosing $x \in \mathbb{F}^m$ and $h \in \mathbb{F}^m$ (if $h = 0$ this line is trivial, but we keep going).
2. Compute via interpolation the unique function $p_{x,h} : \mathbb{F} \to \mathbb{F}$ of degree at most d for which $p(t) = f(x + th)$ for all $t = 1, \dots, d + 1$. Accept iff $p(0) = f(x)$.

Fact 1.1. *There are constants $\alpha_0, \dots, \alpha_{d+1} \in \mathbb{F} - \{0\}$ such that for each x, h, $p_{x,h}(0) = f(x)$ iff $\sum_i \alpha_i f(x + ih) = 0$.*

The test naturally defines a constraint system $(\mathbb{F}^m, \mathbf{LDT}_{\mathbb{F},d,m}, \mathbb{F})$ that has a constraint for each choice of x, h. The constraint looks at $f(x_0), \dots f(x_{d+1})$ and accepts iff $\sum_i \alpha_i f(x_i) = 0$. This correspondence is general: every randomized testing procedure can be equivalently viewed as a system of constraints.

We have mentioned that the set of functions for which the test succeeds with probability 1 (equivalently, that satisfy all of the constraints) is the set of functions of degree at most d, as long as $d < q/2$,

$$\mathrm{SAT}(\mathbf{LDT}_{\mathrm{\mathbb{F},d,m}}) = \{\mathrm{f : \mathbb{F}^m \to \mathbb{F} \mid f\ has\ degree\ at\ most\ d}\}.$$

We have mentioned that this is not always true when $d > q/2$ although clearly the $\supseteq$ direction always holds. The more challenging part is to prove soundness

Theorem 1.2. *For every finite field $\mathbb{F}$ of size q and integer $1 \le d < q/2$ and every $m \ge 2$, we have*

$$\gamma(\mathbf{LDT}_{\mathbb{F},d,m}) \ge 1/2(d + 2)^2 =: \gamma.$$

In other words, if $f : \mathbb{F}^m \to \mathbb{F}$ is δ-far from any degree d low degree function, then the test will reject with probability at least $\delta/2(d + 2)^2$.

2 Proof of theorem

Fix $f : \mathbb{F}^m \to \mathbb{F}$ and let $\varepsilon = \text{rej}(f)$. If $\varepsilon > 1/2(d+2)^2$ then $\delta \le 1 \le \varepsilon/\gamma$. So assume $\varepsilon < 1/2(d+2)^2$. We will show that in this case,

$$\text{dist}(f, \text{SAT}(\textbf{LDT})) \le 2\varepsilon < \varepsilon/\gamma.$$

Majority decoding. For each $y \in \mathbb{F}^m$, we let $g(y)$ be the value that would satisfy the maximal number of constraints that look at y. I.e., we look at all possible $h \in \mathbb{F}^m$ tuples, and define a function $g : \mathbb{F}^m \to \mathbb{F}$ to be

$$g(x) = \text{popular}_h[p_{x,h}(0)]$$

breaking the ties arbitrarily.

First observe that $\text{dist}(f, g) \le 2\varepsilon$. Indeed, for each x where $f(x) \ne g(x)$ half of the constraints with involving x and a random h are unsatisfied. So this can happen at most 2ε fraction of x's. We next proceed in a way similar to the proof that the linearity testing constraints are expanding. The two key claims will be

Claim 2.1. *For all $y \in \mathbb{F}^m$, $\Pr_h[p_{y,h}(0) = g(y)] \ge 1 - 2(d+1)\delta$.*

Claim 2.2. *$g \in \text{SAT}(\textbf{LDT})$ i.e. for every x, h, g satisfies the interpolation constraint associated with x, h. Namely, g is a low degree polynomial.*

Since g is close to f the theorem follows. We now proceed with proving these two claims.

Proof of Claim 2.1. : We will show that for all $y \in \mathbb{F}^m$,

$$\Pr_{h_1,h_2}[p_{y,h_1}(0) = p_{y,h_2}(0)] \ge 1 - 2(d+1)\delta. \tag{1}$$

Note that this is enough to prove the claim. To see this, let $\beta_{y,a} = \Pr_h[p_{y,h}(0) = a]$ for $a \in \mathbb{F}$. Then (1) implies

$$1 - 2(d+1)\delta \le \sum_{a\in\mathbb{F}} (\beta_{y,a})^2 \le \sum_{a\in\mathbb{F}} \beta_{y,a} \cdot \max_a \beta_{y,a} = 1 \cdot \Pr_h[p_{y,h}(0) = g(y)].$$

where the last equality is because $g(y)$ was defined to be the most popular value.

To prove (1), choose $h_1, h_2 \in \mathbb{F}^m$ and consider

$$x_{ij} = y + ih_1 + jh_2 \qquad i, j = 0, \ldots, d+1.$$

- For each $i > 0$ or $j > 0$ the corresponding line or column is a random $d+2$-tuple of points on a line.
- The constraint that the j-th column ($j > 0$) must satisfy is given by $\sum_i \alpha_i f(x_{ij}) = 0$. Similarly for each row $i > 0$ the constraint is by $\sum_j \alpha_j f(x_{ij}) = 0$.
- By definition of p_{y,h_2}, for $i = 0$, $\sum_{j=1}^{d+1} \alpha_j f(x_{0,j}) = -\alpha_0 p_{y,h_2}(0)$.
- By definition of p_{y,h_1}, for $j = 0$, $\sum_{i=1}^{d+1} \alpha_j f(x_{i,0}) = -\alpha_0 p_{y,h_1}(0)$.
- The probability of all constraints to be satisfied is (by union bound) at least $1 - 2(d+1)\varepsilon$.

Consider the $(d+2) \times (d+2)$ matrix Z with $(i,j)^{th}$ entry

$$Z_{i,j} = \alpha_i \alpha_j f(x_{ij}), \qquad i,j \in \{0,\dots,d+1\}.$$

Now the magic square argument kicks in: sum up all the entries in Z. If all row and column constraints $(i,j>0)$ are satisfied then the sum of each row $i = 1,\dots,d+1$ is zero, and the sum of each column $j = 1,\dots,d+1$ is zero. So we are left with

$$0 + \alpha_0 \cdot \sum_{j=0}^{d+1} \alpha_j f(x_{0,j}) = \sum_{i,j=0}^{d+1} \alpha_i \alpha_j f(x_{i,j}) = \alpha_0 \sum_{i=0}^{d+1} \alpha_i f(x_{i,0}) + 0$$

subtracting $\alpha_0 \alpha_0 f(x_{0,0})$ from both sides and dividing by $\alpha_0 \neq 0$ gives us $p_{y,h_1}(0)$ on the left and $p_{y,h_2}(0)$ on the right, and the desired equality. $\square$

Proof of Claim 2.2. Fix x and h arbitrarily. We will show that

$$\sum_{i=0}^{d+1} \alpha_i g(x_{i,0}) = 0.$$

Since this is an arbitrary constraint in **LDT** we will conclude that $g \in \mathrm{SAT}(\mathbf{LDT})$. Choose randomly h_1, h_2 and define

$$x_{ij} = x + ih + j(h_1 + ih_2).$$

Consider the $(d+2)\times(d+2)$ matrix Y that is $Y_{i,j} = \alpha_i \alpha_j f(x_{ij})$ when $j > 0$ and $Y_{i,0} = \alpha_i \alpha_0 g(x_{i,0})$ (i.e. replacing f by g on that column).

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

We will prove that $\Pr[C_0] > 0$ where the probability is taken over h_1, h_2. Since this event is independent of h_1, h_2 we deduce that C_0 occurs with probability 1.

For each row $i \in \{0,1,2,\dots,d+1\}$ we apply Claim 2.1 with $y = x_{i,0}$, noticing that the remaining elements in this row are distributed as in the claim. We get $\Pr_{r'}[\neg R_i] \leq 2(d+1)\delta$.

In addition, since the columns $j > 0$ are distributed as in the test and independently of x and r, we have for all columns except $j = 0$, $\Pr_{r_1,r_2,t_1,\dots,t_{d+1}}[\neg C_j] \leq \delta$. Using union bound, we get

$$\Pr_{r_1,r_2}\left[\bigwedge_{i=0}^{d+1} R_i \bigwedge_{j=1}^{d+1} C_j\right] \geq 1 - 2(d+1)(d+2)\delta - (d+1)\delta > 0.$$

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*. $\square$