6.893 Approximability of Optimization Problems	November 8, 1999
Lecture 16	
Lecturer: Madhu Sudan	*Scribe: Yevgeniy Dodis*

In this lecture will will complete the low-degree test (by talking about the bivariate polynomial test) and start talking about **PCP**'s again. The analysis of the bivariate test will be essentially complete (except for one algebraic result that we will take for granted).

1 Bivariate Polynomial Test

Recall, we reduced our problem to the *bivariate polynomial test.* The question is the following. We are given a function $f : \mathbf{F}^2 \to \mathbf{F}$. We know that for most of the horizontal and vertical lines, the restriction of f agrees with a univariate polynomial of degree at most d. We would like to say that this implies the existence of a *single bivariate* polynomial $T(x, y)$ of degree at most d in each variable that agrees with f "well". By "well" we always mean at least with constant probability.

We look at this problem from a slightly different angle. Let $H \subseteq \mathbf{F}$ be some subfield of $\mathbf{F}$ (H corresponds to a subset of rows/columns of f which are polynomials of degree at most d, and we know that we have a vast majority of those). Assume that for each row $y \in H$ we are given a degree at most d "row polynomial" $r_y(x)$, and for each column $x \in H$ we are given a degree at most d "column polynomial" $c_x(y)$. Furthermore, assume that

$$\Pr_{x,y\in H}[r_y(x) = c_x(y)] \geq 1 - \delta \tag{1}$$

for some (small) constant δ. We would like to argue that there exists a polynomial $T(x, y)$ of degree at most d in x and y (individually) such that

$$\Pr_{x,y\in H}[T(x, y) = r_y(x) = c_x(y)] \geq 1 - O(\delta) \tag{2}$$

Thus, we have completely eliminated f from the picture. The question is, how do we define T, or how do we even start?

Before we start, we make a convention that given a polynomial $Q(x, y)$ we denote by $q_x(y)$ the univariate in y polynomial $Q(x, \cdot)$, and by $q_y(x)$ — the univariate in x polynomial $Q(\cdot, y)$. We also let $|H| = n$ and assume that $n \gg d$ (in particular, $n = \omega(d)$).

1.1 Ideal Case

First, suppose that $\delta = 0$, i.e. for all $x, y \in H$, $r_y(x) = c_x(y)$. Then we can define $T(x, y)$ that perfectly agrees with both $c_x(y)$ and $r_y(x)$. Indeed, define $T(x, y) = r_y(x) = c_x(y)$. To show it is a polynomial of degree d in x and y, take any $x_0, \ldots, x_d \in H$. Using Lagrange interpolation for r_y, we get

$$T(x, y) = r_y(x) = \sum_{i=0}^{d} \alpha_i(x) \cdot r_y(x_i) = \sum_{i=0}^{d} \alpha_i(x) c_{x_i}(y) \tag{3}$$

where $\alpha_i(x)$ are polynomials of degree d in x. Note that $\alpha_i(\cdot)$ and x_i are independent from x and y. Thus, T is indeed a bivariate polynomial $\sum_{i=0}^{d} \alpha_i(x)c_{x_i}(y)$ of degree d in x and y.

1.2 Reducing to Finding a Large Square where $r_y(x) = c_x(y)$

Unfortunately, $\delta > 0$, but the analysis above tells us that if we could somehow find a "large square" inside $H \times H$ where $r_y(x) = c_x(y)$, we can define a consistent T inside that square. Then we hope that, provided the square is large enough (each side is at least a constant fraction of H), Equation (2) would hold. We show that this is indeed the case. In other words, one cannot have a bivariate polynomial T that agrees with $r_y(x)$ and $c_x(y)$ on a large square, but does not agree well overall.

So assume that $H_x, H_y \subseteq H$ are such that for all $(x, y) \in H_x \times H_y$ we have $r_y(x) = c_x(y)$. Furthermore, if we denote $|H_x| = |H_y| = k$, we have $k = \Omega(n)$ (actually, in our case k will be very close to n, but any constant fraction suffices). As we did in the previous section, we can define a degree d bivariate polynomial T such that $T(x, y) = r_y(x) = c_x(y)$ for all $(x, y) \in H_x \times H_y$. Fix any $y \in H_y$. Since we have $t_y(x) = r_y(x)$ for all $x \in H_x$, both r_y and t_y are polynomials of degree d and since $|H_x| = k > d$, the polynomials must agree everywhere on H. Hence, $T(x, y) = r_y(x)$ for all $(x, y) \in H \times H_y$.

Let us now examine the rectangle $R = (H \backslash H_x) \times H_y$. Note, $R \subseteq H \times H_y$, where we have $T(x, y) = r_y(x)$. Let S_x be the set of points $x_0 \in H \backslash H_x$ such that $|\{y \in H_y \mid r_y(x) = c_x(y)\}| \leq d$ (i.e. $r_y(x)$ and $c_x(y)$ agree in at most d places in H_y) and let $|S_x| = \alpha$. Then $r_y(x) \neq c_x(y)$ for at least $(k - d)\alpha$ points (x, y). But from Equation (1), this number cannot be more than δn^2. Hence $(k - d)\alpha \leq \delta n^2$, and thus, $\alpha = O(\delta n)$ (since $(k - d) = \Omega(n)$). But now, for any $x \notin S_x$, we have that $c_x(y)$ and $r_y(x) = T(x, y) = t_x(y)$ agree in at least $d + 1$ places (for $x \in H_x$ this is clear by assumption, and for other x follows from the definition of S_x). Being polynomials of degree d in y, $c_x(y)$ and $t_x(y)$ must be identical over all of H, i.e. $c_x(y) = t_x(y) = T(x, y)$ for all $(x, y) \in (H \backslash S_x) \times H$.

Similarly, there exists a set $S_y \subseteq H \backslash H_y$ of cardinality $\beta = O(\delta n)$ such that $T(x, y) = c_x(y)$ for all $(x, y) \in H \times (H \backslash S_y)$. Hence, $T(x, y) = r_y(x) = c_x(y)$ for all $(x, y) \in (H \backslash S_x) \times (H \backslash S_y)$, so indeed

$$\Pr_{x,y \in H}[T(x, y) = r_y(x) = c_x(y)] \geq (1 - \alpha/n)(1 - \beta/n) = (1 - O(\delta))^2 = 1 - O(\delta) \tag{4}$$

1.3 Reducing to Finding Two Bivariate Polynomials

We saw that it suffices to find a large enough square where $r_y(x) = c_x(y)$. Assume instead that we found two bivariate polynomials A, B of degree at most $D < n/3$ such that $B \not\equiv 0$ and for all $(x, y) \in H \times H$ we have

$$B(x, y) r_y(x) = B(x, y) c_x(y) = A(x, y) \tag{5}$$

We now invoke the following powerful theorem in algebra, whose proof we omit.

Theorem 1 *Assume $A(x, y)$ and $B(x, y)$ are degree D polynomials such that*

- *For more than $2D$ values of $x_0 \in H$, we have $b_{x_0}(y)$ is not identically zero and divides $a_{x_0}(y)$.*
- *For more than $2D$ values of $y_0 \in H$, we have $b_{y_0}(x)$ is not identically zero and divides $a_{y_0}(x)$.*

Then $B(x, y)$ divides $A(x, y)$, i.e. for some $T(x, y)$ we have $A(x, y) = B(x, y)T(x, y)$.

We claim that our A and B satisfy the statement of the theorem. Indeed, let Z_x be the set of $x_0 \in H$ such that $b_{x_0}(y)$ is identically zero. Then $|Z_x| \leq D$, since $B(x, y)$ has degree D is x. Thus, for at least $n - D > 2D$ values of $x_0 \in H$ we have that $b_{x_0}(y)$ is not identically zero and divides $a_{x_0}(y)$ (yielding $c_{x_0}(y)$). Defining similarly Z_y, for at least $n - D > 2D$ values of $y_0 \in H$ (all y_0 not in Z_y) we have that $b_{y_0}(x)$ is not identically zero and divides $a_{y_0}(x)$ (yielding $r_{y_0}(x)$). Using the above theorem, we get that there exists a bivariate polynomial $T(x, y)$ such that for all $(x, y) \in H \times H$ we have

$$B(x, y) r_y(x) = B(x, y) c_x(y) = B(x, y) T(x, y) \tag{6}$$

In particular, $b_{x_0}(y) c_{x_0}(y) = b_{x_0}(y) t_{x_0}(y)$. Assume that $x_0 \notin Z_x$, i.e. $b_{x_0}(y) \not\equiv 0$. Then $b_{x_0}(y)$ can be zero in at most D points $y \in H$, so in at least $n - D > D$ points y we have $b_{x_0}(y) \neq 0$, and thus $c_{x_0}(y) = t_{x_0}(y)$. Being polynomials of degree D, we get that $c_{x_0}(y) = t_{x_0}(y)$ for all $y \in H$ and all $x_0 \notin Z_x$. Similarly, $r_{y_0}(x) = t_{y_0}(x)$ for all $x \in H$ and all $y_0 \notin Z_y$. But then for all $(x, y) \in (H \backslash Z_x) \times (H \backslash Z_y)$ we have

$$c_x(y) = t_x(y) = T(x, y) = t_y(x) = r_y(x) \tag{7}$$

Since both $|Z_x|, |Z_y| \leq D < |H|/3$, we get that $r_y(x) = c_x(y)$ holds in a large enough square (of size at least $(2n/3) \times (2n/3)$), as needed.

1.4 Finding the Polynomials

It remains to find A, B of degree $D < n/3$ satisfying Equation (5). The good news is that this simply amounts to solving a *homogeneous linear system* in the coefficients of A and B. Unfortunately, we can not argue immediately that this system has a non-zero solution (which is true as we will show). Indeed, the number of variables (coefficients of A, B) is $2D^2 < 2(n/3)^2 < n^2$. And the number of equations is at least n^2. However, just a slightly more difficult argument works.

Let us ignore A for now and consider the system $B(x,y)r_y(x) = B(x,y)c_x(y)$. Since $r_y(x) \neq c_x(y)$ in at most δn^2 points (from Equation (1)), the system has at most δn^2 non-trivial equations. On the other hand, it has D^2 unknowns (coefficients of B). Hence, if we let $\epsilon = \sqrt{\delta}$, it suffices to set $D = \epsilon n$ to insure that the number of equations is less than the number of unknowns, i.e. the system has a non-zero solution $B(x,y)$. We note that we are allowed to set $D = \epsilon n$, since this is indeed much smaller than $n/3$ (assuming δ is small enough).

Let $D' = \epsilon n + d \ll n/3$. To complete the analysis, we claim that we can find A of degree D' satisfying Equation (5). Indeed, define $r'_y(y) = b_y(x)r_y(x)$ and $c'_x(y) = b_x(y)c_x(y)$. Then we have that $r'_y(x)$ and $c'_x(y)$ are polynomials of degree D' such that for all $(x,y) \in H \times H$ we have

$$r'_y(x) = b_y(x)r_y(x) = B(x,y)r_y(x) = B(x,y)c_x(y) = b_x(y)c_x(y) = c'_x(y) \tag{8}$$

But now we are exactly in the situation "$\delta = 0$" considered in section Section 1.1 with $r_y(x)$ replaced with $r'_y(x)$, $c_x(y)$ replaced with $c'_x(y)$ and d replaced by D'. Thus, using Equation (3) we can indeed define A of degree D' satisfying Equation (5).

This completes the bivariate polynomial test, as well as our study of the low-degree testing.

2 Summary of Error-Correction/Error-Detection Techniques

The following table briefly summarizes our error-correction and error-detection tests for the Hadamard code (linearity testing) and multivariate polynomials code (low-degree testing). In the table below we always assume constant probability of error.

Code	Error-Correction (Self-Correction)	Error-Detection (Self-Testing)
Hadamard	2 queries ($\tilde{f}(x) = f(x+y) - f(y)$)	same
Multivariate Polynomials	$O(d)$ queries to the oracle for f OR 2 queries: one to f, one to the prover	same

3 Back to the PCPs

We now start applying the machinery we have developed to construct efficient Probabilistically Checkable Proofs (**PCPs**) for **NP**. For that we define the the following **NP**-complete problem, for which we will develop a somewhat efficient **PCP**.

INSTANCE: n variables $z_1, \ldots, z_n$ over $GF(2)$, m polynomials $p_1, \ldots, p_m$ of degree 3.

QUESTION: Does there exist an assignment $\vec{a} = a_1, \ldots, a_n$ such that $p_i(\vec{a}) = 0$ for all $i \in [m]$.

The **NP**-completeness follows immediately by a reduction from **3-SAT**. Indeed, given boolean variables $x_1, \ldots, x_n$ and 3-clauses $c_1, \ldots, c_m$, where $c_j = x_{j1}^{b_{j1}} \vee x_{j2}^{b_{j2}} \vee x_{j3}^{b_{j3}}$ (here $x^0 = \overline{x}$ and $x^1 = x$), we define polynomial $p_j(z_{j1}, z_{j2}, z_{j3}) = (z_{j1} - b_{j1})(z_{j2} - b_{j2})(z_{j3} - b_{j3})$. We see that $a_{j1}, a_{j2}, a_{j3} \in \{0,1\}$ satisfy c_j if and only iff $p_j(a_{j1}, a_{j2}, a_{j3}) = 0$.

We briefly outline the **PCP** we will construct. We ask the prover (who allegedly knows $\vec{a}$ satisfying all our polynomials) to prepare the table $T = \{t(\vec{a}) \mid t \text{ is a polynomial of degree at most } 3\}$. For convenience, we split this table into 3 sub-tables we will use.

- A-table: $\{\ell(\vec{a}) \mid \ell \text{ is a linear polynomial (of degree 1)}\}$.
- B-table: $\{q(\vec{a}) \mid q \text{ is a quadratic polynomial (of degree 2)}\}$.
- C-table: $\{c(\vec{a}) \mid c \text{ is a cubic polynomial (of degree 3)}\}$.

We note in particular that the C-table contains the entries for $C(p_j)$. However, there is no guarantee that they are correct. Moreover, we cannot afford to read m bits. Here is what the verifier will do instead.

1. Check that A, B, C are written "correctly" (i.e. close to consistent tables for some $\vec{a}$).
2. Pick random $r_1, \ldots, r_m$ in $GF(2)$ and check that $C(\sum_{j=1}^{m} r_j p_j(\cdot)) = 0$.

In the next lecture we will fill in the missing details and prove that this test works.