Georgia Tech Fall'04
CS 8002 : PCPs and Hardness of Approximation

Lecture 4: Håstad's 3-Bit PCP and MAX 3 LIN

Lecturer: *Subhash Khot* Scribe: *Subrahmanyam K.*

In this class, we look at hardness of approximating the problem MAX-3LIN. We give an outline of Håstad's 3-bit PCP [1] from which the hardness result for MAX-3 LIN follows immediately. We look at two modules used in the Håstad's PCP construction, namely Linearity Testing and Junta Testing.

1 MAX-3LIN

Definition 1 (MAX-3LIN) *Given a system of linear equations over $GF(2)$ with variables $x_1, x_2, x_3, ..., x_n$ and equations of the form,*

$$x_i + x_j + x_k = c$$

where $1 \leq i, j, k \leq n$ and $c \in \{0, 1\}$, the goal is to find an assignment that satisfies maximum fraction of equations.

Note that, if there is a solution which satisfies all the equations, we can find it in polynomial time by Gaussian elimination. The interesting case is when there is no such solution. It is easy to see that a random $\{0, 1\}$ assignment to each x_i would satisfy (in expectation) half the equations. The following hardness result states that this is the best that we can achieve (in polynomial time).

Theorem 1 (Hardness of MAX-3LIN, [1])
*It is **NP**-hard to tell whether the optimal solution (call it OPT) for a system of linear equations, with three variables in each equation, satisfies*

$$\begin{aligned} OPT \ &\geq 1 - \epsilon, \ \text{or} \\ OPT \ &\leq \frac{1}{2} + \eta \end{aligned}$$

where ϵ, η are arbitrarily small constants.

The above theorem implies that MAX-3LIN is factor $2 - \epsilon$ hard to approximate. This is a tight hardness result, given that we have a way of assigning variables to satisfy half the equations. Let us now look at Håstad's 3-bit PCP construction from which the above theorem follows.

2 Håstad's 3-bit PCP construction

Theorem 2 (Håstad's 3-bit PCP)
$\forall \epsilon, \eta > 0$ ***NP** has PCP verifiers such that*

- *Verifier uses $O(\log n)$ random bits, queries three bits from the proof Π, say x, y, z.*
- *Acceptance condition is linear (i.e. $x + y + z = 1$ or $x + y + z = 0$).*
- *Completeness :*
$$x \in L \Rightarrow \exists \Pi,\ s.t. \mathbf{Pr}[V(\Pi) = accept] \geq 1 - \epsilon$$
- *Soundness :*
$$x \notin L \Rightarrow \forall \Pi, \mathbf{Pr}[V(\Pi) = accept] \leq \frac{1}{2} + \eta$$

The hardness result for MAX-3LIN follows by a standard reduction from PCPs to constraint satisfaction problems. We let the bits in the proof to be unknown boolean variables and the tests of the verifier (which happen to be linear in this case) as constraints. Now, the task of finding a proof that maximizes acceptance probability is same as the task of finding an assignment that satisfies the maximum fraction of constraints. Since the former is NP-hard (even to apprimate), so is the latter.

2.1 Overview of Håstad's PCP

Here is a brief overview of the construction of this PCP. We start with an instance of Label Cover $\mathcal{L}(G(V, W, E), [M], [N], \{\pi_{vw}\})$ (as seen in the last lecture). We ask as a PCP proof, the labeling to instance $\mathcal{L}$. The verification procedure works as follows. We pick one edge $(v, w) \in E$ at random and read the labels $l(v), l(w)$ from the proof. We accept if and only if $\pi(l(w)) = l(v)$. This PCP is known as the Outer PCP.

Clearly, we use $O(\log n)$ random bits and two queries. One query is from the alphabet $[N]$ and one from $[M]$, and hence the total number of bits queried is $\log M + \log N$. This is of course far from the desired goal of three query bits. However, what we gain from the Outer PCP is a large gap in the completeness and soundness probabilities. We know that it is NP-hard to tell whether a label cover instance $\mathcal{L}$ has $OPT(\mathcal{L}) = 1$ or $OPT(\mathcal{L}) \leq \delta$. This translates to saying that the Outer PCP has perfect completeness and soundness atmost δ with $|M|, |N| = 1/\delta^{O(1)}$.

2.1.1 Proof Composition

Now we build a new PCP on top of the Outer PCP. This PCP, known as the Inner PCP, is constructed as follows. Instead of asking for the labels of the vertices, we now ask for the *encodings* of these labels via some predefined encoding scheme. Now, as before, we pick an edge (v, w) at random. Let the encodings $ENC(v)$ and $ENC(w)$ be the supposed encodings of labels of v and w in the proof. Instead of checking the consistency of labels as before, we read *a few* bits from both $ENC(v)$ and $ENC(w)$. There are two things we need to verify : One is the *codeword test* where we need to verify that $ENC(v)$ and $ENC(w)$ are indeed correct encodings of some $i \in [N]$ and $j \in [M]$. And the second is the *consistency test* where we need to verify that $\pi(j) = i$.

In the rest of this lecture, we take a closer look at the codeword testing. Specifically, we look at the *Linearity Testing* and *Junta Testing*.

3 Linearity Testing

Given the truth table (of length 2^n) of a boolean function $A : \{\pm 1\}^n \to \{\pm 1\}$, *Linearity Testing Problem* asks for a probabilistic procedure that checks whether A represents a *linear function* by reading only a few bits from its truth table.

DEFINITION 2 *For a subset $\alpha \subseteq [n]$, the character function χ_α is defined as*

$$\chi_\alpha(x_1, x_2, \ldots, x_n) = \prod_{i \in \alpha} x_i$$

Note that if we use $\{0, 1\}$ notation instead of $\{-1, 1\}$ notation (and XOR instead of multiplication), the character functions are precisely the linear functions. However, $\{-1, 1\}$ notation will be useful in what follows.

The problem at hand is the following. Given the truth table $A : \{\pm 1\}^n \to \{\pm 1\}$, we have to test whether A is linear. We expect that the testing procedure satisfies :

- Read only $O(1)$ queries.
- **Completeness** : If A is linear (i.e. $A = \chi_\alpha$ for some $\alpha \subseteq [n]$), then $\mathbf{Pr}[\text{accept}] = 1$.
- **Soundness** : If A is *far* from being linear, then the the test rejects with reasonable probability.

Our goal is to prove Theorem 6 that gives a 3-query test for testing linearity. Before proceeding to the theorem, let us introduce the basics of discrete Fourier analysis.

3.1 Fourier Analysis

Consider the vector space of all functions $A : \{\pm 1\}^n \mapsto \mathbb{R}$. Here the addition of two functions is defined as pointwise addition and the dimension of this vector space is 2^n. One can define a natural inner product on this space by letting the inner product of two functions A and B to be

$$< A, B > = 2^{-n} \sum_x A(x)B(x) = \mathbf{E}_x[A(x)B(x)]$$

If A is boolean (i.e. $\{\pm 1\}$ valued) then $< A, A >= 1$. It is easy to see that the characters $\{\chi_\alpha \mid \alpha \subseteq [n]\}$ form an orthonormal basis for the vector space.

LEMMA 3

$$< \chi_\alpha, \chi_\beta >= 0 \qquad \textit{if } \alpha \neq \beta$$

PROOF:

$$\begin{aligned} < \chi_\alpha, \chi_\beta > &= \mathbf{E}_x[\chi_\alpha(x)\chi_\beta(x)] \\ &= \mathbf{E}_x[\prod_{i\in\alpha} x_i \prod_{j\in\beta} x_j] \\ &= \mathbf{E}_{x_1,x_2,\ldots,x_n}[\prod_{i\in\alpha\Delta\beta} x_i] \\ &= \prod_{i\in\alpha\Delta\beta} E_{x_i}[x_i] \end{aligned}$$

and the expectation is 0 since there is at least one $i_0 \in \alpha\Delta\beta$. $\square$

It follows that any function $A : \{\pm 1\}^n \mapsto \mathbb{R}$ can be written as a linear combination of characters, i.e.

$$A(x) \;=\; \sum_{\alpha\subseteq[n]} \hat{A}_\alpha \chi_\alpha(x) \qquad \forall\, x \in \{\pm 1\}^n$$

where $\hat{A}_\alpha$ are real numbers called Fourier coefficients; they are given by $\hat{A}_\alpha \;= < A, \chi_\alpha >$. By Parseval's identity, we have $\sum_\alpha \hat{A}_\alpha^2 =< A, A >$ ($= 1$ if A is boolean valued).

The Fourier coefficient $\hat{A}_\alpha$ is a measure of how close the function A is to χ_α.

DEFINITION 3 *Given* $A, B : \{\pm 1\}^n \to \{\pm 1\}$, *the* Hamming distance Δ *is defined as*

$$\begin{aligned} \Delta(A, B) &= \frac{\#x \textit{ such that } A(x) \neq B(x)}{2^n} \\ &= \mathbf{Pr}_x[A(x) \neq B(x)] \end{aligned}$$

LEMMA 4

$$\Delta(A, \chi_\alpha) = \tfrac{1}{2} - \tfrac{1}{2}\hat{A}_\alpha$$

PROOF:

$$\begin{aligned} \hat{A}_\alpha =& \langle A, \chi_\alpha \rangle \\ =& \mathbf{E}_x[A(x)\chi_\alpha(x)] \\ =& \mathbf{Pr}_x[A(x) = \chi_\alpha(x)] - \mathbf{Pr}_x[A(x) \neq \chi_\alpha(x)] \\ =& 1 - 2\mathbf{Pr}_x[A(x) \neq \chi_\alpha(x)] \\ =& 1 - 2\Delta(A, \chi_\alpha) \end{aligned}$$

In particular, as $\hat{A}_\alpha \to 1, \Delta(A, \chi_\alpha) \to 0$. $\square$

We note these two identities for future reference.

LEMMA 5

$$\chi_\alpha(x)\chi_\beta(x) = \chi_{\alpha\Delta\beta}(x)$$
$$\chi_\alpha(xy) = \chi_\alpha(x)\chi_\alpha(y)$$

where xy is a string obtained by co-ordinatewise product of two strings $x, y \in \{\pm 1\}^n$.

DEFINITION 4 *A is θ-*far from being linear *if $\forall \alpha \subseteq [n], \Delta(A, \chi_\alpha) \geq \theta$.*

We are now ready to prove the following theorem.

THEOREM 6
There is a 3 query test for linearity testing such that,

- ***Completeness** : If A is linear, then $\mathbf{Pr}[accept] = 1$.*
- ***Soundness** : If A is δ-far from being linear, then $\mathbf{Pr}[accept] \leq 1 - \delta$. In other words, if $\mathbf{Pr}[accept] \geq 1 - \delta$, then $\exists \alpha_0$ such that $\Delta(A, \alpha_0) \leq \delta$.*

The test was proposed and analyzed by Blum, Luby and Rubinfeld [2]. Given a truth table $A : \{\pm 1\}^n \to \{\pm 1\}$, we pick $x, y \in \{\pm 1\}^n$ at random and accept if $A(x)A(y) = A(xy)$.

Clearly, if A is linear, this test accepts with probability 1. We analyze the soundness.
PROOF: Note that the expression $(1 + A(x)A(y)A(xy))/2$ equals 1 if the test accepts and 0 otherwise. Hence the probability of accepting the test equals the expectation of this expression.

$$\begin{aligned}
\mathbf{Pr}[\text{accept}] =& \mathbf{E}_{x,y}\left[\frac{1 + A(x)A(y)A(xy)}{2}\right] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\mathbf{E}_{x,y}[A(x)A(y)A(xy)] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\mathbf{E}_{x,y}[(\sum_{\alpha} \hat{A}_\alpha \chi_\alpha(x)) \cdot (\sum_{\beta} \hat{A}_\beta \chi_\beta(y)) \cdot (\sum_{\gamma} \hat{A}_\gamma \chi_\gamma(xy))] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\mathbf{E}_{x,y}[(\sum_{\alpha,\beta,\gamma} \hat{A}_\alpha \hat{A}_\beta \hat{A}_\gamma \chi_\alpha(x)\chi_\beta(y)\chi_\gamma(xy)] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\sum_{\alpha,\beta,\gamma} \hat{A}_\alpha \hat{A}_\beta \hat{A}_\gamma \mathbf{E}_{x,y}[\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(xy)] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\sum_{\alpha,\beta,\gamma} \hat{A}_\alpha \hat{A}_\beta \hat{A}_\gamma \mathbf{E}_{x,y}[\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(x)\chi_\gamma(y)] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\sum_{\alpha,\beta,\gamma} \hat{A}_\alpha \hat{A}_\beta \hat{A}_\gamma \mathbf{E}_{x,y}[\chi_{\alpha\Delta\gamma}(x)\chi_{\beta\Delta\gamma}(y)] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\sum_{\alpha,\beta,\gamma} \hat{A}_\alpha \hat{A}_\beta \hat{A}_\gamma \mathbf{E}_{x}[\chi_{\alpha\Delta\gamma}(x)]\mathbf{E}_y[\chi_{\beta\Delta\gamma}(y)]
\end{aligned}$$

Unless $\alpha = \gamma$, $\mathbf{E}_x[\chi_{\alpha\Delta\gamma}(x)] = 0$. Similarly, unless $\beta = \gamma$, $\mathbf{E}_y[\chi_{\beta\Delta\gamma}(y)] = 0$. Thus

$$\mathbf{Pr}[\text{accept}] = \tfrac{1}{2} + \tfrac{1}{2}\sum_{\alpha} \hat{A}_\alpha^3$$

If $\mathbf{Pr}[\text{accept}] \geq 1 - \delta$, we have

$$\begin{aligned} \tfrac{1}{2} + \tfrac{1}{2}\sum_{\alpha} \hat{A}_{\alpha}^{3} &\geq 1 - \delta \\ \Rightarrow \quad \sum_{\alpha} \hat{A}_{\alpha}^{3} &\geq 1 - 2\delta \\ \Rightarrow \quad \sum_{\alpha} \hat{A}_{\alpha}^{2} \cdot \hat{A}_{\alpha} &\geq 1 - 2\delta \end{aligned}$$

Since $\sum_{\alpha} \hat{A}_{\alpha}^{2} = 1$, $\exists \alpha_0$ such that, $\hat{A}_{\alpha_0} \geq 1 - 2\delta$. So by Lemma 4, we get $\Delta(A, \chi_{\alpha_0}) \leq \delta$
□

4 Junta Testing

Of all the linear functions, we are specifically interested in the following type of functions called *dictatorships*. These are defined as follows.

DEFINITION 5 *A character χ_α is called* dictator *if $|\alpha| = 1$.*

Thus dictatorships are boolean functions of the form $A(x_1, x_2, \ldots, x_n) = x_i$.

Our goal here is, given a function $A : \{\pm 1\}^n \to \{\pm 1\}$, to test if A is a dictatorship. Ideally speaking, we want the verifier to

- Read just $O(1)$ queries.
- Perfect completeness. (dictatorship is accepted with probability 1)
- If A is far from being a dictatorship, then the test rejects with a reasonable probability.

Dictatorships are boolean functions that depend on only one co-ordinate. Juntas are relaxed form of dictatorships that depend on a *few* co-ordinates. We define an even more relaxed form of juntas that we call (δ, h)-juntas.

DEFINITION 6 *A is called (δ, h)-junta if $\exists \alpha \subseteq [n]$ such that $|\alpha| \leq h$ and $|\hat{A}_\alpha| \geq \delta$.*

Note that if A is a dictatorship, then A is a $(1, 1)$-junta.

Håstad proves the following theorem.

THEOREM 7
For every $\epsilon, \eta > 0$, there is a linear test that :

- *Reads 3 queries.*
- *If A is a dictatorship, $\mathbf{Pr}[accept] \geq 1 - \epsilon$.*
- *If $\mathbf{Pr}[accept] \geq \frac{1}{2} + \eta$, then A is $(2\eta, O(\frac{1}{\epsilon}\log(\frac{1}{\eta})))$-junta.*

Note that the parameters in the above statements are independent of n. Also note that in the soundness case, we can claim A to be only a (δ, h)-junta instead of claiming it to be a dictatorship (as we would have liked).

Note further that if we use the test $A(x)A(y) = A(xy)$ as earlier, this would fail, because even for χ_α such that $|\alpha|$ is large, the verifier would accept with probability 1.

4.1 Håstad's test

- Pick x, y at random.
- Define $z = xy\mu$ where μ is "perturbation function" defined as $\forall i \in [n]$,

$$\mu_i = \begin{cases} 1 & \text{w.p. } 1-\epsilon \\ -1 & \text{w.p. } \epsilon \end{cases}$$

- Test if $A(x)A(y) = A(z)$.

The intuition behind this can be described as follows. A linear function χ_α passes the test if

$$\prod_{i\in\alpha} x_i \cdot \prod_{i\in\alpha} y_i = \prod_{i\in\alpha} x_i y_i \mu_i$$

i.e. if $\prod_{i\in\alpha} \mu_i = 1$. For characters of large size ($|\alpha| \gg 1/\epsilon$), the probability of this happening is $\approx \frac{1}{2}$. Thus the test "punishes" linear functions with large size.

Clearly, if $A = \chi_\alpha$ with $|\alpha| = 1, \alpha = \{i_0\}$ then the test accepts iff $\mu_{i_0} = 1$. Thus the test accepts dictatorships with probability $1 - \epsilon$. Now we analyze the soundness using Fourier methods as before.

$$\begin{aligned}
\mathbf{Pr}[\text{accept}] =& \mathbf{E}_{x,y,\mu}[\tfrac{1}{2}(1 + A(x)A(y)A(xy\mu))] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\mathbf{E}_{x,y,\mu}[\sum_{\alpha,\beta,\gamma} \hat{A}_\alpha \hat{A}_\beta \hat{A}_\gamma \chi_\alpha(x)\chi_\beta(y)\chi_\gamma(xy\mu)] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\sum_{\alpha,\beta,\gamma} \hat{A}_\alpha \hat{A}_\beta \hat{A}_\gamma \mathbf{E}_{x,y,\mu}[\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(x)\chi_\gamma(y)\chi_\gamma(\mu)] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\sum_{\alpha,\beta,\gamma} \hat{A}_\alpha \hat{A}_\beta \hat{A}_\gamma \mathbf{E}_{x,y,\mu}[\chi_{\alpha\Delta\gamma}(x)\chi_{\beta\Delta\gamma}(y)\chi_\gamma(\mu)] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\sum_{\alpha,\beta,\gamma} \hat{A}_\alpha \hat{A}_\beta \hat{A}_\gamma \mathbf{E}_x[\chi_{\alpha\Delta\gamma}(x)]\mathbf{E}_y[\chi_{\beta\Delta\gamma}(y)]\mathbf{E}_\mu[\chi_\gamma(\mu)] \\
=& \tfrac{1}{2} + \tfrac{1}{2}\sum_{\alpha} \hat{A}^3_\alpha \mathbf{E}_\mu[\chi_\alpha(\mu)]
\end{aligned}$$

LEMMA 8

$$\mathbf{E}_\mu[\chi_\alpha(\mu)] = (1-2\epsilon)^{|\alpha|}$$

PROOF:

$$\mathbf{E}_\mu[\chi_\alpha(\mu)] = \mathbf{E}_\mu[\prod_{i\in\alpha}\mu_i] = \prod_{i\in\alpha}\mathbf{E}_{\mu_i}[\mu_i] = (1-2\epsilon)^{|\alpha|}$$

□

Hence we get the probability of acceptance

$$\mathbf{Pr}[\text{accept}] = \tfrac{1}{2} + \tfrac{1}{2}\sum_\alpha \hat{A}_\alpha^3(1-2\epsilon)^{|\alpha|}$$

If $\mathbf{Pr}[\text{accept}] \geq \frac{1}{2} + \eta$, we have,

$$\begin{aligned} \sum_\alpha \hat{A}_\alpha^3(1-2\epsilon)^{|\alpha|} &\geq 2\eta \\ \Rightarrow \quad \sum_\alpha \hat{A}_\alpha^2 \cdot \hat{A}_\alpha(1-2\epsilon)^{|\alpha|} &\geq 2\eta \end{aligned}$$

Since $\sum_\alpha \hat{A}_\alpha^2 = 1$, $\exists \alpha_0$ such that $\hat{A}_{\alpha_0}(1-2\epsilon)^{\alpha_0} \geq \; 2\eta$

$$\Rightarrow \hat{A}_{\alpha_0} \geq 2\eta \text{ and } (1-2\epsilon)^{\alpha_0} \geq \; 2\eta$$

In particular

$$|\alpha_0| \leq O(\tfrac{1}{\epsilon}\log\tfrac{1}{\eta})$$

This finishes the proof of Theorem 7.

In the next class, we will look at a variant of the junta test that works on *two* given tables $A : \{\pm1\}^n \mapsto \{\pm1\}$ and $B : \{\pm1\}^m \mapsto \{\pm1\}$. The verifier is given in addition a map $\pi : [m] \mapsto [n]$. The verifier wants to check (in possibly some weak sense) that B is a dictatorship corresponding to some $j_0 \in [m]$ and A is a dictatorship corresponding to some $i_0 \in [n]$ and in addition, $\pi(j_0) = i_0$. We will use a modification of the junta testing, where we read 2 bits from A and 1 bit from B. This test forms the core of Håstad's 3-bit PCP.

References

[1] Johan Håstad. Some optimal inapproximability results, *Journal of the ACM*, 2001.

[2] Blum, Luby, Rubinfeld. Self-testing and self-correcting programs, with applications to numerical programs. Journal of Computer and System Sciences, 47:549–595, 1993.