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

Lecture 4: Hastad's 3-bit PCP

Lecturer: *Subhash Khot* Scribe: *Aranyak Mehta*

We will first state the consistency test we use in the PCP construction, and provide its proof later.

Given tables $A : \{\pm 1\}^n \to \{\pm 1\}$ and $B : \{\pm 1\}^m \to \{\pm 1\}$, and a map $\pi : [m] \to [n]$, we say A, B, π satisfy Condition$(*)$ if:
(i) A is a character $\chi_{\{i_0\}}$, for some $i_0 \in [n]$
(ii) B is a character $\chi_{\{j_0\}}$, for some $j_0 \in [m]$
(iii) Consistency: $\pi(j_0) = i_0$

Theorem 1
For every $\epsilon, \eta > 0$, there is a consistency test with the following properties:
1. It reads 3 bits (one from table A, two from Table B).
2. The acceptance condition is a linear predicate of the three bits read.
3. Completeness: If A and B satisfy Condition $(*)$ above, then Pr[accept]$\geq 1 - \epsilon$
4. Soundness: If Pr[accept]$\geq \frac{1}{2} + \eta$ then

$$\sum_{\substack{\alpha \subseteq [n], \beta \subseteq [m], \\ |\alpha|, |\beta| \leq O(\frac{1}{\epsilon}\log(\frac{1}{\eta})), \\ \alpha \subseteq \pi(\beta)}} \hat{A}_\alpha^2 \hat{B}_\beta^2 \geq \eta^2$$

where $\pi(\beta) \subseteq [n]$ is defined as $\pi(\beta) = \{\pi(j) \mid j \in \beta\}$.

We now describe the 3-bit PCP, and prove its properties using Theorem 1. The proof of Theorem 1 will be provided later.

The 3-bit PCP Construction:
We start with an instance of the Label-Cover problem $\mathcal{L}(G(V, W, E), [m], [n], \{\pi_{vw}\})$, where $m = 7^k, n = 2^k$.
By the PCP Theorem and Raz's Parallel Repetition Theorem, we know that it is *NP*-hard to tell if the instance is a *YES* instance (i.e. $OPT(\mathcal{L}) = 1$) or a *NO* instance ($OPT(\mathcal{L}) \leq \delta = 2^{-\Omega(k)}$).

The proof of our 3-bit PCP consists of, for all $v \in V, w \in W$, a special encoding of $l(v)$ and $l(w)$, the supposed labels in a supposed labeling to the Label Cover instance (here $l(v) \in [n], l(w) \in [m]$). The encodings are the character tables $\chi_{\{l(v)\}}$ and $\chi_{\{l(w)\}}$ respectively. This encoding scheme is called the long code: for every $i \in [n]$, i is encoded by a 2^n bit string which is the truth table of the character $\chi_{\{i\}}$. The long code was introduced by Bellare, Goldreich and Sudan [1].

The verifier works as follows:

- Pick a random edge $(u, v) \in E$.
- Let A be the table "supposed to be" $\chi_{\{l(v)\}}$ and let B be the table "supposed to be" $\chi_{\{l(w)\}}$.
- Run Consistency-Test(A, B, π_{vw}) (of Theorem 1).

Observe from Theorem 1 that the verifier reads three bits of the proof, and performs a linear test. We next prove the completeness and soundness of the proof system.

LEMMA 2
Completeness: *If* $\mathrm{OPT}(\mathcal{L}) = 1$ *then there exists a proof* Π *s.t. Pr[*Π *is accepted]*$\geq 1 - \epsilon$.

PROOF: Since $OPT(\mathcal{L}) = 1$ there is a correct labeling $l(v), l(w)$ for $v \in V, w \in W$, s.t. for every edge (v, w), $\pi_{vw}(l(w)) = l(v)$. The proof Π consists of the encodings of these labels, i.e., for $v \in V, w \in W$, Π has the truth tables of $\chi_{\{l(v)\}}, \chi_{\{l(w)\}}$. Now, for each edge (v, w), in proof Π we have $A = \chi_{\{l(v)\}}$, $B = \chi_{\{l(w)\}}$ and $\pi_{vw}(l(w)) = l(v)$. Hence, by Theorem 1, Consistency-Test(A, B, π_{vw}) accepts with probability at least $1 - \epsilon$. $\square$

LEMMA 3
Soundness: *If* $OPT(\mathcal{L}) \leq \delta$ *and* δ *is sufficiently small (as a function of* ϵ, η*), then for every proof* Π*, Pr[*Π *is accepted]*$\leq \frac{1}{2} + 2\eta$.

PROOF: We prove the contrapositive, that if on an instance $\mathcal{L}$ of label cover, there is a proof Π such that the verifier accepts Π with probability more than $\frac{1}{2} + 2\eta$, then $\mathcal{L}$ has a labeling that satisfies $\Omega(\epsilon^2 \log^{-2}(\frac{1}{\eta})\eta^3)$ fraction of its edges. This gives a contradiction provided δ was chosen sufficiently small.

Suppose Pr[Π is accepted]$\geq \frac{1}{2} + 2\eta$, then for at least η fraction of the edges, the probability that Consistency-Test accepts the tables for that edge is at least $\frac{1}{2} + \eta$ (call these edges "good"). By Theorem 1, this means that for every good edge (v, w):

$$\sum_{\substack{\alpha \subseteq [n], \beta \subseteq [m], \\ |\alpha|, |\beta| \leq O(\frac{1}{\epsilon} \log(\frac{1}{\eta})), \\ \alpha \subseteq \pi(\beta)}} \hat{A}_\alpha^2 \hat{B}_\beta^2 \geq \eta^2 \tag{1}$$

where A and B are the tables for $l(v)$ and $l(w)$ respectively. Using this we gives a randomized labeling for $\mathcal{L}$ that satisfies a significant fraction of edges in expectation. It follows that there must exist a labeling that satisfies at least that much fraction of edges.

For $v \in V$, $l(v)$ is defined as follows:

Look at Table A.

Pick $\alpha \subseteq [n]$ with probability $\hat{A}_\alpha^2$ (note that $\sum_{\alpha \subseteq [n]} \hat{A}_\alpha^2 = 1$ and therefore the squares of Fourier coefficients define a probability distribution).
If $|\alpha| \geq O(\frac{1}{\epsilon}\log(\frac{1}{\eta}))$ then give up.
Else, define $l(v)$ to be a random element of the set α.

For $w \in W$, $l(w)$ is defined as follows:

Look at Table B.
Pick $\beta \subseteq [m]$ w.p. $\hat{B}_\beta^2$.
If $|\beta| \geq O(\frac{1}{\epsilon}\log(\frac{1}{\eta}))$ then give up.
Else, define $l(w)$ to be a random element of the set β.

Note that we are assuming that $\alpha, \beta \neq \phi$. We will come back to this in the proof of the consistency test.
Now, from (1) we see that for a good edge (v, w), with probability at least η^2 we pick for v a set α and for w a set β s.t. $|\alpha|, |\beta| < O(\frac{1}{\epsilon}\log(\frac{1}{\eta}))$ and s.t. $\alpha \subseteq \pi(\beta)$.
In particular, $\exists\ i_0 \in \alpha, j_0 \in \beta$ s.t. $\pi(j_0) = i_0$. With probability $\Omega(\epsilon^2 \log^{-2}(\frac{1}{\eta}))$ we pick i_0 to be $l(v)$ and j_0 to be $l(w)$, and then the edge (v, w) is satisfied. Recall that the fraction of good edges is at least η, and for each good edge we select a good pair of sets α, β with probability at least η^2. Hence it follows that the fraction of edges satisfied by the labeling described above is at least $\Omega(\epsilon^2 \log^{-2}(\frac{1}{\eta})\eta^3)$. $\Box$

It remains to prove Theorem 1.

Proof: (of Theorem 1)
For $x \in \{\pm 1\}^n$, define $x \circ \pi \in \{\pm 1\}^m$ to be the string s.t. $(x \circ \pi)_j = x_{\pi(j)}$, for $j = 1...m$.

Test:
Pick at random $x \in \{\pm 1\}^n$, $y \in \{\pm 1\}^m$
Pick $\mu \in \{\pm 1\}^m$, s.t. $\forall j = 1...m$,

$$\mu_j = \begin{cases} +1 & w.p.\ 1 - \epsilon \\ -1 & w.p.\ \epsilon \end{cases}$$

Let $z = y(x \circ \pi)\mu$
Accept iff $B(z) = A(x)B(y)$

Completeness: Assume $A = \chi_{\{i_0\}}, B = \chi_{\{j_0\}}, \pi(j_0) = i_0$. Then
$A(x) = x_{i_0}$,
$B(y) = y_{j_0}$, and
$B(z) = z_{j_0} = y_{j_0}(x \circ \pi)_{j_0}\mu_{j_0}$.
But $(x \circ \pi)_{j_0} = x_{\pi(j_0)} = x_{i_0}$, so the test reduces to accepting iff $\mu_{j_0} = +1$, which holds $w.p.$ $1 - \epsilon$.

Soundness: Note the following expression for the probability that the test accepts A, B, π:

$$\begin{aligned}
\Pr[\text{accept}] &= \underset{x,y,\mu}{E}\left[\frac{1 + A(x)B(y)B(y(x\circ\pi)\mu)}{2}\right] \\
&= \frac{1}{2} + \frac{1}{2}\underset{x,y,\mu}{E}\left[\sum_{\substack{\alpha\subseteq[n]\\ \beta,\gamma\subseteq[m]}} \hat{A}_\alpha \hat{B}_\beta \hat{B}_\gamma \chi_\alpha(x)\chi_\beta(y)\chi_\gamma(y(x\circ\pi)\mu)\right] \\
&= \frac{1}{2} + \frac{1}{2}\sum_{\alpha,\beta,\gamma} \hat{A}_\alpha \hat{B}_\beta \hat{B}_\gamma \underset{x,y,\mu}{E}[\chi_\alpha(x)\chi_\beta(y)\chi_\gamma(y)\chi_\gamma(x\circ\pi)\chi_\gamma(\mu)]
\end{aligned} \tag{2}$$

We need the following lemma to proceed:

Lemma 4
$\chi_\gamma(x\circ\pi) = \chi_{\pi_2(\gamma)}(x)$, where $\pi_2(\gamma) \subseteq [n]$ is defined as $\pi_2(\gamma) = \{i : i \in [n], |\pi^{-1}(i) \cap \gamma|$ is odd$\}$

Proof:

$$\chi_\gamma(x\circ\pi) = \prod_{l\in\gamma}(x\circ\pi)_l = \prod_{l\in\gamma} x_{\pi(l)} = \chi_{\pi_2(\gamma)}(x)$$

□

Hence

$$\Pr[\text{accept}] = \frac{1}{2} + \frac{1}{2}\sum_{\alpha,\beta,\gamma} \hat{A}_\alpha \hat{B}_\beta \hat{B}_\gamma \underset{x,y,\mu}{E}[\chi_{\beta\triangle\gamma}(y)\chi_{\alpha\triangle\pi_2(\gamma)}(x)\chi_\gamma(\mu)]$$

The sum is zero unless $\beta = \gamma, \alpha = \pi_2(\gamma)$. Also, we have $E_\mu[\chi_\gamma(\mu)] = (1-2\epsilon)^{|\gamma|}$. Hence,

$$\Pr[\text{accept}] = \frac{1}{2} + \frac{1}{2}\sum_{\substack{\alpha,\beta\\ \alpha=\pi_2(\beta)}} \hat{A}_\alpha \hat{B}_\beta^2 (1-2\epsilon)^{|\beta|}$$

If the probability of acceptance is at least $\frac{1}{2} + \eta$, we get:

$$\begin{aligned}
\frac{1}{2} + \eta &\leq \frac{1}{2} + \frac{1}{2}\sum_{\substack{\alpha,\beta\\ \alpha=\pi_2(\beta)}} \hat{A}_\alpha \hat{B}_\beta^2 (1-2\epsilon)^{|\beta|} \\
\Rightarrow 2\eta &\leq \sum_{\substack{\alpha,\beta\\ \alpha=\pi_2(\beta)}} \hat{A}_\alpha \hat{B}_\beta^2 (1-2\epsilon)^{|\beta|}
\end{aligned}$$

We now use the Cauchy-Swartz inequality to get the following:

$$4\eta^2 \le \Big(\sum_{\substack{\alpha,\beta \\ \alpha=\pi_2(\beta)}} \hat{A}_\alpha \hat{B}_\beta^2 (1-2\epsilon)^{|\beta|} \Big)^2 \le \sum_{\substack{\alpha,\beta \\ \alpha=\pi_2(\beta)}} \hat{A}_\alpha^2 \hat{B}_\beta^2 (1-2\epsilon)^{2|\beta|} \sum_\beta \hat{B}_\beta^2$$

But we have $\sum_\beta \hat{B}_\beta^2 = 1$, hence

$$\sum_{\substack{\alpha,\beta \\ \alpha=\pi_2(\beta)}} \hat{A}_\alpha^2 \hat{B}_\beta^2 (1-2\epsilon)^{2|\beta|} \ge 4\eta^2$$

Now, for $|\beta| > O(\frac{1}{\epsilon}\log(\frac{1}{\eta}))$, we get that the "damping factor" $(1-2\epsilon)^{2|\beta|} = (1-2\epsilon)^{O(\frac{1}{\epsilon}\log(\frac{1}{\eta}))} \le \eta^2$. Hence we may constrain the sum further by $|\beta| \le O(\frac{1}{\epsilon}\log(\frac{1}{\eta}))$, and still be at least $3\eta^2$. Finally, we may replace $\alpha = \pi_2(\beta)$ by $\alpha \subseteq \pi(\beta)$ and only increase the sum. This yields the soundness condition as required.

Remark: In the proof so far, we assumed that $\hat{A}_\phi, \hat{B}_\phi = 0$. This is not a valid assumption - indeed the proof in which A and B are all $+1$s is always accepted. Hence one needs a further mechanism to ensure that $\hat{A}_\phi, \hat{B}_\phi = 0$. We use the following fact:

Lemma 5
If table A is such that for all x, $A(-x) = -A(x)$, then $\hat{A}_\phi = 0$. Furthermore, for all $i = 1,..,n$, the character $\chi_{\{i\}}$ satisfies $\chi(-x) = -\chi(x)$ for all x.

Proof: If $A(-x) = -A(x)$ for all x, then

$$\hat{A}_\phi = E_x[A(x)\chi_\phi(x)] = E_x[A(x)] = E_x[A(-x)] = E_x[-A(x)] = -\hat{A}_\phi$$

Also, for any i, $\chi_{\{i\}}(-x) = (-x)_i = -\chi_{\{i\}}(x)$. □

This suggests the following mechanism to force the tables A to have $\hat{A}_\phi = 0$: In the proof we don't ask for the whole table, but for each x, we ask for exactly one of the entries $A(x)$ or $A(-x)$. Suppose w.l.o.g. that we asked for the bit $A(x)$. When we need the bit $A(x)$, we read it from the proof and when we need the bit $A(-x)$, we read $A(x)$ instead and negate the result. In this manner, the prover is effectively forced to provide a table A with $\hat{A}_\phi = 0$. □

Remark : Imperfect Completeness The completeness condition is that if the optimum of the label cover instance is 1, then there exists a proof which is accepted by the verifier with probability at least $1-\epsilon$. We note that this is the most we can expect with a linear acceptance condition. In other words, PCPs with linear acceptance condition and completeness equal to 1 cannot capture NP-hard problems. This is because in this case one can always find a correct proof for a YES instance through Gaussian elimination.

Implications: The 3 bit PCP has the following corollaries. The first is immediate from the PCP, we show a gadget-based reduction for the second, the rest are obtained by similar reductions (see Håstad's paper).

THEOREM 6
MAX-3-LIN is $(\frac{1}{2}+\epsilon)$*-hard to approximate, i.e. given an instance of MAX-3-LIN, it is* NP*-hard to tell if* $OPT \geq 1-\epsilon$ *or* $OPT \leq \frac{1}{2}+\epsilon$.

THEOREM 7
MAX-3-SAT is $(\frac{7}{8}+\epsilon)$*-hard to approximate.*

PROOF: Consider the following reduction from MAX-3-LIN to MAX-3-SAT. Given an instance of MAX-3-LIN, obtain an instance of MAX-3-SAT by having the same set of variables, and having 4 clauses for each equation. If the equation is $x \oplus y \oplus z = 0$, then the corresponding 4 clauses are: $\bar{x} \vee \bar{y} \vee \bar{z}$, $\bar{x} \vee y \vee z$, $x \vee \bar{y} \vee z$, and $x \vee y \vee \bar{z}$. Consider any $\{0,1\}$ assignment to the variables of the MAX-3-LIN instance. The corresponding assignment to the variables of the MAX-3-SAT instance is from {False, True}, with 0 mapped to False and 1 mapped to True. Then one can check that if a setting of the variables satisfies an equation, then the corresponding setting satisfies all 4 clauses. But if a setting of the variables does not satisfy an equation, then the corresponding setting can satisfy only 3 out of the 4 clauses. This implies that if the optimum of the MAX-3-LIN instance is at least $1-\epsilon$, then the optimum of the MAX-3-SAT instance is also at least $1-\epsilon$, and if the optimum of the MAX-3-LIN instance is at most $\frac{1}{2}+\epsilon$, then the optimum of the MAX-3-SAT instance is at most $\frac{7}{8}+\epsilon$. $\Box$

THEOREM 8
MAX-CUT is $(\frac{16}{17}+\epsilon)$*-hard to approximate.*

THEOREM 9
MAX-2-SAT is $(\frac{21}{22}+\epsilon)$*-hard to approximate.*

THEOREM 10
Vertex Cover is $(\frac{7}{6}-\epsilon)$*-hard to approximate.*

References

[1] Bellare, Goldreich and Sudan. Free bits, PCPs and non-approximability – towards tight results. SIAM Journal on Computing, 27(3):804–915, 1998.