CSE 533: The PCP Theorem and Hardness of Approximation **(Autumn 2005)**

Lecture 2: PCP Theorem and GAP-SAT; intro to expanders

Oct. 3, 2005

Lecturer: Ryan O'Donnell and Venkatesan Guruswami *Scribe: Atri Rudra*

1 The GAP-SAT problem and the PCP theorem

In the last lecture we said we will prove NP-hardness of approximation algorithms. To do this, we will follow the same approach that is used in NP-completeness — we convert an optimization problem into a decision problem. For a concrete example recall the $\mathrm{MAX\text{-}3ESAT}$ problem: each clause of the input boolean formula is an OR of exactly three literals and the goal is to find an assignment that maximizes the number of satisfied clauses. We now define the decision version of this problem.

Definition 1.1. $\mathrm{GAP\text{-}E3SAT}_{c,s}$ $(0 < s \leq c \leq 1)$*: Given an* $\mathrm{E3SAT}$ *formula on* m *clauses,*

- *output YES if* $OPT \geq cm$*;*
- *output NO if* $OPT < sm$*;*
- *output anything if* $sm \leq OPT < cm$*.*

Remark 1.2. *Recall that* $\mathrm{GAP\text{-}E3SAT}$ *being NP-hard means that there is a deterministic polynomial time reduction,* R*, from your favorite NP-complete language (say* $3\text{-}\mathrm{COLOR}$*) to* $\mathrm{E3SAT}$*, such that*

- *Completeness: given* $G \in 3\text{-}\mathrm{COLOR}$*,* $R(G)$ *gives an* $\mathrm{E3SAT}$ *formula with* $OPT \geq cm$*;*
- *Soundness: given* $G \notin 3\text{-}\mathrm{COLOR}$*,* $R(G)$ *gives an* $\mathrm{E3SAT}$ *formula with* $OPT < sm$*.*

Remark 1.3. $\mathrm{GAP\text{-}E3SAT}_{c,s}$ *being NP-hard implies that there is no polynomial time* $\left(\frac{s}{c}\right)$*-factor approximation algorithm for* $\mathrm{MAX\text{-}3ESAT}$ *unless* $\mathrm{P} = \mathrm{NP}$*. To see this implication, assume we have such an algorithm; we then show how to solve* $3\text{-}\mathrm{COLOR}$ *in polynomial time. To do this, given a graph* G *apply the reduction* R *reducing it to* $\mathrm{E3SAT}$ *and then run the supposed* $\left(\frac{s}{c}\right)$*-factor approximation algorithm. If* $G \in 3\text{-}\mathrm{COLOR}$ *then this will produce an assignment that satisfies at least* $\left(\frac{s}{c}\right) cm = sm$ *clauses in* $R(G)$*. If* $G \notin 3\text{-}\mathrm{COLOR}$*, the algorithm will be unable to produce an assignment that satisfies as many as* sm *clauses of* $R(G)$*. Thus, we can distinguish the two cases and get a polynomial time algorithm for* $3\text{-}\mathrm{COLOR}$*.*

In this lecture, we will show the following result.

Theorem 1.4. *The following two statement are equivalent:*

1. *The PCP theorem;*

2. *There exists an universal constant $s < 1$ such that* $\mathrm{GAP\text{-}E3SAT}_{1,s}$ *is* NP*-hard.*

Proof. We first show that the second statement implies the first. To this end assume that $\mathrm{GAP\text{-}E3SAT}_{1,s}$ is NP-hard. We will construct a PCP system for 3-COLOR. Given an instance G of 3-COLOR on n vertices, the verifier V runs the reduction from 3-COLOR to E3SAT — let ψ be the constructed formula on m clauses. The proof that the prover P will provide will be an assignment to the variables in ψ (note that ψ has polynomial (in n) many variables). Finally, V uses $\log m = O(\log n)$ random bits to choose a random clause (call this clause ϕ), queries the proof on the variables in the clause, and checks if the given assignment satisfies ϕ. Note that the number of positions probed here, C, is 3. We now show that the above PCP system has the required properties.

- Completeness: If $G \in$ 3-COLOR then ψ has $OPT = m$. In this case P can write down the optimal assignment, which implies that all the clauses are satisfied, and hence V accepts with probability 1.

- Soundness: If $G \notin$ 3-COLOR then ψ has $OPT < sm$. Thus for any assignment P provides, V picks a a satisfied clause with probability less than s; that is, V accepts with probability less than s. The soundness can be brought down to $1/2$ by repeating the check $O(1)$ many times independently in parallel.

We now show that the first statement of the theorem implies the second. To this end, assume the PCP theorem. We will now give a deterministic polynomial time reduction from 3-COLOR to $\mathrm{GAP\text{-}E3SAT}_{1,s}$. We will think of the bits of the proof as variables $x_1, x_2, \cdots, x_{poly(n)}$ for anE3SAT formula. Given G, R will first run the verifier's polynomial time pre-computation steps. Then R enumerates all the $2^{O(\log n)} = \mathrm{poly}(n) = N$ random choices of V — each choice gives some C proof locations $(x_{i_1}, x_{i_2}, \cdots, x_{i_C})$ and a predicate ϕ on the C bits. R further canonically converts $\phi(x_{i_1}, x_{i_2}, \cdots, x_{i_C})$ to an equivalent E3CNF formula (in this step R may need to add some auxiliary variables, $y_1, y_2, \cdots, y_{C'}$). Without loss of generality we may assume that each equivalent E3CNF has exactly K clauses where $K = C \cdot 2^C$. Finally, R outputs the conjunction of all these $m = N \cdot K$ clauses. We now argue that this reduction works.

- Completeness: If $G \in$ 3-COLOR then we know there is a proof that satisfies all of the verifier's checks. Thus all of the E3CNF formulas the reduction outputs can be simultaneously satisfied; i.e., $OPT = m$ as needed.

- Soundness: If $G \notin$ 3-COLOR, then for every proof (assignment to the x_i's) and assignment to the auxiliary variables, at least half of the verifier's N checks must fail. Whenever a check fails, the corresponding E3CNF has at most $K - 1 = K(1 - 1/K)$ many satisfied clauses. Thus overall, the number of simultaneously satisfiable clauses is at most

$$\frac{N}{2}K(1-1/K) + \frac{N}{2}K = NK\left(1 - \frac{1}{2K}\right) = m\left(1 - \frac{1}{2K}\right).$$

Thus, $OPT \leq sm$, where $s = (1 - \frac{1}{2K})$, and this is an absolute constant less than 1 as needed.

□

2 The proof of the PCP theorem and expanders

Armed with Theorem 1.4, we will prove the PCP theorem by showing that $\mathrm{GAP\text{-}E3SAT}_{1,s}$ is NP-hard for some $s < 1$. Dinur's paper in fact proves this version of the PCP theorem. Her proof uses objects called expanders — in this and the next lecture, we will spend some time developing facts about expanders.

To give a rough idea of where expanders fit in the scheme of things, here is a brief overview of Dinur's proof. Note that it is easy to see from the proof of Theorem 1.4 that the PCP theorem is also implied by showing that $\mathrm{GAP3\text{-}COLOR}_{1,s}$ is NP-hard, where in this gap version, the quantity we are interested in is the number of edges in a 3-coloring that are colored properly. The way Dinur's proof work is to start with the fact that $3\text{-}\mathrm{COLOR}$ is NP-hard, from which one immediately deduces that $\mathrm{GAP3\text{-}COLOR}_{1,1-\frac{1}{m}}$ is NP-hard, where m is the number of edges. (This is because in any illegal 3-coloring, at least one edge must be violated.) The proof will try to amplify the "soundness gap" from $\frac{1}{m}$ up to some universal constant. At each stage the proof will be working with a *constraint graph* G (initially, the constraints in the input to $3\text{-}\mathrm{COLOR}$ is that the endpoints of each edge have differnt colors from $\{1, 2, 3\}$). In the most important step of the proof, a new graph G^t is constructed from G, where the constraints in G^t correspond to walks in G of length t. If the constraint graphs are nicely structured (i.e., are constant-degree *expanders*) then these walks in G will mix nicely.

3 Expanders

Roughly speaking, expanders are graphs that have no "bottlenecks". In other words, they are graphs with high connectivity. More formally, we will be interested in the following quantity:

Definition 3.1. *The* edge expansion *of a graph $G = (V, E)$, denoted by $\phi(G)$, is defined as*

$$\phi(G) = \min_{S \subseteq V, |S| \leq \frac{|V|}{2}} \frac{|E(S, \overline{S})|}{|S|},$$

where $\overline{S} = V \setminus S$ and $E(S, \overline{S}) = \{(u, v) \in E \mid u \in S \text{ and } v \in \overline{S}\}$.

We say G is an *expander* if $\phi(G)$ is "large" (at least some positive constant). Note however that it is not hard to find such a graph; for example, the complete graph has $\phi(G) \geq \Omega(n)$. The challenge is to find *sparse* expanders, especially d-regular expanders for some constant d. In fact such sparse expanders exist and can be constructed explicitly.

Theorem 3.2. *There exist constants $d > 1$ and $\phi_0 > 0$ and an explicit family of d-regular graphs $\{G_n\}_{n \geq 1}$ such that $\phi(G_n) \geq \phi_0$.*

3.1 Alternate definition of expanders

We now consider an alternate way of looking at expanders. For any d-regular graph G, let A^G denote the *adjacency matrix*, that is, A^G_{ij} is 1 if $(i,j) \in E(G)$ and 0 otherwise (one could also work with multi-graphs in which case for an edge (i,j), A^G_{ij} would be the multiplicity of that edge).

For the all-ones vector, $\vec{v} = \vec{1}$, $A^G\vec{v} = d \cdot \vec{v}$; that is, $\vec{v}$ is an *eigenvector* with *eigenvalue* d. If A is a real and symmetric $n \times n$ matrix (as A^G is) then A has n real-valued eigenvalues $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$. For A^G, $\lambda_1 = d$ is easily seen to be the largest eigenvalue.

Definition 3.3. *A d-regular graph G is an* (n,d,λ)-expander *if* $\lambda = \max\{|\lambda_i(G)| : i \neq 1\} = \max\{\lambda_2(G), |\lambda_n(G)|\}$ *and* $\lambda < d$.

This definition of an expander is closely related to the definition we saw before.

Theorem 3.4. *If G is a (n,d,λ)-expander then*

$$\frac{\phi(G)^2}{2d} \leq d - \lambda \leq 2\phi(G).$$

In other words, large expansion is equivalent to large *spectral gap* (that is, $d - \lambda$). We will see the proof of the upper bound (which is the direction we actually need) next lecture. Explicit constructions of expanders tend to work with this spectral definition:

Theorem 3.5. *There exist explicit constants $d \geq 3$ and $\lambda < d$ and an explicit (polynomial-time computable) family of (n,d,λ)-expanders.*

The second-largest eigenvalue λ of a real symmetric $n \times n$ is characterized as follows (via a "Rayleigh quotient"):

$$\lambda = \max_{x \in \mathbb{R}^n, x \cdot \vec{1} = 0, x \neq 0} \frac{|\langle Ax, x\rangle|}{\langle x, x\rangle}. \tag{1}$$

Let us show this. As A is a real symmetric matrix, there exists an orthonormal basis $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n$ where each $\vec{v}_i$ is an eigenvector of A. Letting $x = \vec{v}_2$ yields a ratio in (1) of $|\lambda_2|$; similarly, we can let $x = \vec{v}_n$ and get a ratio of $|\lambda_n|$. Thus, we certainly have $\leq$ in (1). For the other direction, write any x as $\sum_{i=1}^n a_i \vec{v}_i$. Using $\langle x, \vec{v}_1\rangle = 0$, we conclude $a_1 = 0$. Now $Ax = \sum_{=1}^n a_i \lambda_i \vec{v}_i$ and thus,

$$|\langle Ax, x\rangle| = \left|\sum_{i=2}^n a_i^2 \lambda_i\right| \leq \sum_{i=2}^n |\lambda_i| a_i^2 \leq \lambda \sum_{i=2}^n a_i^2 = \lambda \langle x, x\rangle$$

as required.

Finally, we conclude this lecture by proving a simple lemma that will be used in the proof of the PCP theorem.

Lemma 3.6. *If G is a d-regular graph on the vertex set V and H is a d'-regular graph on V then $G' = G \cup H = (V, E(G) \cup E(H))$[1] is a $d + d'$-regular graph such that*

$$\lambda(G') \leq \lambda(G) + \lambda(H)$$

[1] Here the union of edges results in a mutliset.

Proof. Choose x such that $\|x\| = 1$, $x \cdot \vec{1} = 0$ and $\lambda(G') = \langle A^{G'}x, x\rangle$. Now

$$\begin{aligned} \langle A^{G'}x, x\rangle &= \langle A^{G}x, x\rangle + \langle A^{H}x, x\rangle \\ &\leq \lambda(G) + \lambda(H). \end{aligned}$$

The equality follows from the definition of G' and the inequality follows from (1). □