Lecture 19: PCP Theorem and Hardness of Approximation II

Arijit Bishnu

28.04.2010

Outline

Outline

Constraint Satisfaction Problems

Max3SAT

Given a 3CNF φ, the problem is to find an assignment that satisfies the largest number of clauses. We denote $\mathcal{V}(\varphi)$ as the maximum fraction of the clauses satisfied over all assignments of values to the boolean variables.

Max$q$CSP

Let $q \in \mathbb{N}$. Given a system of boolean constraints $\mathcal{C}$ defined over boolean variables such that every constraint involves at most q variables. The problem is to find an assignment that satisfies as many constraints as possible. We denote $\mathcal{V}(\mathcal{C})$ as the maximum fraction of the total number of constraints satisfied over all assignments.

Inapproximability implication of PCP theorem

Theorem

The PCP theorem implies that $\exists$ a constant q such that there is no α-factor approximation algorithm for Max$q$CSP with $\alpha < 2$, unless NP $=$ P.

Proof

- Get hold of an NP-complete problem $L \in \mathrm{PCP}(\log n, q)$ where q is a constant and let V be the $(O(\log n), O(1))$-PCP verifier for L.
- Our goal is to make a reduction from L to Max$q$CSP.

The Proof Continued

- Given an instance z of L, we construct a Max$q$CSP instance I with $|I| = m = z^{O(1)}$ constraints such that

$$z \in L \Rightarrow I \text{ is satisfiable} \Rightarrow \mathcal{V}(I) = 1$$
$$z \notin L \Rightarrow \mathcal{V}(I) \leq \frac{1}{2}$$

  The RHS of the above equation means that every assignment of I contradicts at least half of the constraints.
- As V is an $(O(\log n), O(1))$-PCP verifier for L, the total number of all possible random sequences is equal to $2^{\log|z|} = |z|^{O(1)}$.
- For each random choice of length $O(\log|z|)$, V chooses q positions $i_1^R, \ldots, i_q^R$ from the witness π and a Boolean function $f_R : \{0,1\}^q \to \{0,1\}$ and accepts iff $f_R(\pi[i_1^R], \ldots, \pi[i_q^R]) = 1$.
- We want to simulate the actions of V as a Boolean formula/CSP.

The Proof Continued

- Introduce Boolean variables $x_1, \ldots, x_{|\pi|}$.
- For each R, we add the constraint $f_R(x_{i_1^R}, \ldots, x_{i_q^R}) = 1$.
- If $z \in L$, then there is a witness π that is accepted with probability 1.
  - Consider the assignment $x_i = \pi[i]$ where $\pi[i]$ is the i-th bit of π.
  - Such an assignment satisfies all constraints of I.
- If I has an assignment $x_i = a_i$ for which $\mathcal{V}(I) > \frac{1}{2}$, then the witness π defined as $\pi[i] = a_i$ is accepted with probability $> \frac{1}{2}$ by V that implies $z \in L$. Look at the contrapositive of the last statement to get the result.

Inapproximability implication of PCP theorem

Theorem

The PCP theorem implies that $\exists\epsilon > 0$ such that there is no $1 - \epsilon$-factor approximation algorithm for Max3SAT, unless NP $=$ P.

Proof

- Given an instance I of Max$q$CSP (where q is the constant of the PCP theorem) with m constraints using n variables $x_1, \ldots, x_n$, we construct an instance φ_I of Max3SAT with m' clauses such that .

$$\begin{aligned} I \text{ is satisfiable} \quad &\Rightarrow \quad \varphi_I \text{ is satisfiable.} \\ \mathcal{V}(I) \leq \frac{1}{2} \quad &\Rightarrow \quad \mathcal{V}(\varphi_I) \leq (1-\epsilon)m' \end{aligned}$$

The Proof Continued

- For every constraint $f(x_{i_1}, \ldots, x_{i_q}) = 1$ in I, construct an equivalent qCNF of size at most 2^q.
- On converting each qCNF to 3CNF, each CNF expands by a factor of q with introduction of auxiliary variables in addition to the original variables.
- With m constraints, the number of clauses in $\varphi_I = m' \leq q2^q m$.
- If I is satisfiable, then for the variables of I (because of which I had $\mathcal{V}(I) = 1$), set the same assignments to the original variables of φ_I and the auxiliary variables appropriately. This assignment satisfies all the clauses of φ_I to make φ_I satisfiable.

The Proof Continued

- If $\mathcal{V}(I) \leq \frac{1}{2}$, consider an arbitrary assigment to the variables of φ_I.
- Look at the original variables of φ_I. Carry over this assignment to I. This assignment should also contradict at least $\frac{1}{2}$ of the m constraints of I.
- This implies that at least $\frac{m}{2}$ clauses of φ_I are also contradicted.
- So, $\mathcal{V}(\varphi_I) \leq m' - \frac{m}{2}$. We want $m' - \frac{m}{2} \leq (1-\epsilon)m'$.
- Choose $\epsilon \leq \frac{m}{2m'}$. We already know $m' \leq q2^q m$. Therefore, $\epsilon \leq \frac{1}{2q2^q}$.

Moral of the Story

Theorem

There exists $\rho > 1$, such that for every $L \in \mathrm{NP}$, there is a polynomial time function f mapping strings to 3CNF formula such that

$$z \in L \;\Rightarrow\; \mathcal{V}(f(z)) = 1$$
$$z \notin L \;\Rightarrow\; \mathcal{V}(f(z)) < \frac{1}{\rho}$$

Outline

Inapproximability of Maximum Independent Set

Maximum Independent Set and Minimum Vertex Cover

- We know MVC $=n$ - MIS in a graph $G=(V,E)$ with $|V|=n$.
- Can we use the 2-factor algorithm of MVC for finding an approximation algorithm for MIS?
- Consider n to be even, and let the size of MVC be $\frac{n}{2}-1$.
- The approximation algorithm would return a VC of size at most $n-2$ making the IS size to be at least 2 whereas the MIS is of size $\frac{n}{2}+1$ making the approximation ratio $\frac{\frac{n}{2}+1}{2}$ arbitrarily worse.

Inapproximability of Maximum Independent Set

Lemma

There exists a polynomial time computable transformation f from the 3CNF formulas to graphs such that for every 3CNF formula φ, $f(\varphi)$ is an n-vertex graph whose $|\text{MIS}| = \mathcal{V}(\varphi) \cdot \frac{n}{3}$.

Proof

Look at the usual reduction.

Inapproximability of Maximum Independent Set

Lemma

There exists a constant $\rho > 1$ such that INDSET, the problem of independent set, cannot have a ρ-factor approximation algorithm, unless $\mathsf{P} = \mathsf{NP}$.

Proof

- Let $L \in \mathsf{NP}$. We know that the decision problem of L can be reduced to approximating MAX3SAT.
- The reduction produces an instance φ of MAX3SAT where $\mathcal{V}(\varphi) = 1$ (φ is satisfied) or $\mathcal{V}(\varphi) < \frac{1}{\rho}$ where $\rho > 1$ is some constant.
- Apply the reduction of the earlier lemma to φ to conclude that a ρ-factor approximation algorithm gives a ρ-factor approximation algorithm for MAX3SAT on φ.
- So, ρ-factor approximation algorithm for INDSET is NP-hard.

Inapproximability of Maximum Independent Set

Theorem

For every $\rho > 1$, computing a ρ-factor approximation algorithm for INDSET is not possible, unless P = NP.

Proof

- We have to amplify the approximation gap using graph product. See book.