DD2445 COMPLEXITY THEORY

RECAP FROM LAST LECTURE

$\boxed{L \in PCP_{c,s}[r(n), q(n)]}$ if for some $K_1, K_2 > 0$ there is a verifier V that given $x \in \{0,1\}^n$ and $\pi \in \{0,1\}^*$

- runs in time poly$(|x|)$
- flips $\leq K_1 \cdot r(n)$ random coins
- makes $\leq K_2 \cdot q(n)$ nonadaptive oracle queries to π
- outputs decision $V^{\pi}(x) \in \{0,1\}$

such that

COMPLETENESS If $x \in L$, then $\exists \pi$ (and w.l.o.g. $|\pi| \leq K_2\, q(n)\, 2^{K_1 r(n)}$) such that

$$\Pr[V^{\pi}(x) = 1] \geq c$$

SOUNDNESS If $x \notin L$, then $\forall \pi'$

$$\Pr[V^{\pi'}(x) = 1] \leq s$$

PCP THEOREM, VERSION A

$$NP = PCP_{1,1/2}(\log n, 1)$$

For a CNF formula φ, define

$$\boxed{\mathrm{val}_N(\varphi) = \frac{\text{max \# satisfiable clauses}}{\text{total \# clauses}}} \in [0,1]$$

PCP THEOREM, VERSION B

II

There exists a $\rho^* < 1$ such that for every $L \in NP$ there is a polynomial-time function f_L mapping strings to 3-CNF formulas such that

$$x \in L \Rightarrow val_N(f_L(x)) = 1$$
$$x \notin L \Rightarrow val_N(f_L(x)) < \rho^*$$

This way of viewing the PCP theorem leads to hardness of approximation results

COROLLARY

There exists a constant $\rho^* < 1$ such that if there is a polynomial-time ρ^*-approximation algorithm for MAX-3SAT, then $P = NP$.

<u>Proof</u> Fix some NP-complete language L. Suppose A is a ρ^*-approximation algorithm for MAX-3SAT with ρ^* as in version B of the PCP theorem, and fix f_L as in version B.

That is, given φ for which T clauses can be satisfied, A finds assignment satisfying at least $\rho^* T$ clauses.

Then the following algorithm decides L in poly time

Compute $\varphi = f_L(x)$
Let m = # clauses in φ
If $A(\varphi)$ satisfies $\geq \rho^* m$ clauses, return "$x \in L$"
else return "$x \notin L$"

Analysis:

If $x \in L$, then $\varphi = f_L(x)$ is satisfiable. All m clauses can be satisfied. A will find an assignment satisfying at least $\rho^* m$ clauses since it is a ρ^*-approx.

If $x \notin L$, then no algorithm can do better than the optimal solution, which is $< \rho^* m$ clauses by the properties of f_L.

Observation

Algorithm A actually does not have to compute satisfying assignment. Just providing the numerical estimate of max # satisfiable clauses is enough.

TODAY we want to show that versions A and B of PCP theorem are equivalent.

Introduce notion of constraint satisfaction problems

DEFINITION 11.21 $\quad q \in \mathbb{N}^+, \quad u \in \{0,1\}^n$

φ_i q-ary constraint: - $f_i : \{0,1\}^q \to \{0,1\}$
- positions $j_{i,1}, j_{i,2}, \ldots, j_{i,q}$

$$\varphi_i(u) = f_i(u_{j_{i,1}}, u_{j_{i,2}}, \ldots, u_{j_{i,q}})$$

An instance of a q-ary CONSTRAINT SATISFACTION PROBLEM (qCSP) is a collection $\{\varphi_1, \ldots, \varphi_m\}$ of such constraints

Ex 3SAT is a qCSP problem where $q=3$ and all f_i's are disjunctions of at most 3 variables or negated variables.

DEF 11.12 (continued)
An assignment $u \in \{0,1\}^n$ satisfies φ_i if $\varphi_i(u) = 1$. The fraction of constraints satisfied by u is

$$\frac{\sum_{i=1}^{m} \varphi_i(u)}{m}$$

Let us denote

$$\boxed{val_N(\varphi) = \max_u \frac{\sum_{i=1}^{m} \varphi_i(u)}{m}}$$

(where we will omit the suffix N from now on)
φ is satisfiable if $val_N(\varphi) = 1$.

φ has ARITY q and SIZE m
Can assume $n \leq qm$ [variables not mentioned are redundant]

Any qCSP instance φ can be described using $O(2^q mq \log n)$ bits

We will always have q constant independent of n and m

The greedy approximation algorithm for 3SAT we discussed last lecture can be generalized to an algorithm satisfying $\frac{val(\varphi)}{2^q} m$ constraints for a qCSP instance φ.

DEFINITION 11.13 For $q \in \mathbb{N}^+$, $\rho \le 1$, let $\boxed{\rho\text{-GAP}_q\text{CSP}}$ be the problem of deciding for a given qCSP instance φ whether

(1) $\mathrm{val}(\varphi) = 1$ (yes instance), or
(2) $\mathrm{val}(\varphi) < \rho$ (no instance)

given the promise that one of these two cases apply (this is known as a PROMISE PROBLEM)

We say that ρ-GAP$_q$CSP is NP-hard if $\forall L \in NP$ there is a polynomial-time function f_L mapping strings to qCSP instances such that

COMPLETENESS: $x \in L \Rightarrow \mathrm{val}(f(x)) = 1$
SOUNDNESS: $x \notin L \Rightarrow \mathrm{val}(f(x)) < \rho$

PCP THEOREM, VERSION C (Thm 11.14)
There exist constants $q \in \mathbb{N}^+$, $\rho \in (0,1)$ such that ρ-GAP$_q$CSP is NP-hard

We now want to show that both versions A and B of the PCP theorem are equivalent to version C.

Version A $\Rightarrow$ Version C

Assume $NP \subseteq PCP_{1,1/2}(\log n, 1)$.
Fix some NP-complete language L.
There is a PCP-verifier V for L such that

- if $x \in L$, then $\exists \pi$ s.t. $V^{\pi}(x)$ accepts with probability 1
- if $x \notin L$, then $\forall \pi$ $Pr[V^{\pi}(x) \text{ accepts}] \le 1/2$

Finding a "best proof" π that makes $V^{\pi}(x)$ maximally likely to accept can be viewed as a CSP

PCP verifier makes $O(1)$ queries, say q.
Given input x and random string r of length $O(\log n)$ (say $\le c \cdot \log n$), let $V_{x,r}(\pi)$ be function that outputs 1 iff verifier accepts x after having queried π as determined by x and r

$V_{x,r}(\pi)$ depends on (at most) q locations in π — q-ary constraint.

Hence $\varphi_x = \{V_{x,r}\}_{r \in \{0,1\}^{c \cdot \log n}}$
is a polynomial-size qCSP instance for every x.

Since V runs in polynomial time, we can compute φ_x from x in polynomial time

If $x \in L$, then $\exists \pi$ s.t. $\Pr[V^{\pi}(x)=1]=1$, meaning that $\mathrm{val}(\varphi_x)=1$

If $x \notin L$, then $\forall \pi$ $\Pr[V^{\pi}(x)=1] \le 1/2$, so $\mathrm{val}(\varphi_x) \le 1/2$.

This proves the PCP theorem, version C. $\blacksquare$

Version C $\Rightarrow$ Version A

Suppose that ρ-GAP qCSP is NP-hard for some constants $q \in \mathbb{N}^+$, $\rho < 1$.

Translate into PCP verifier with q queries, completeness 1, soundness error ρ, and logarithmic randomness for any $L \in NP$

Given x, verifier runs reduction f to obtain qCSP instance $\varphi_x = \{\varphi_i\}_{i=1}^{m}$

Proof π considered as assignment to variables in φ_x. Notice $m = \mathrm{poly}(|x|)$

Pick random $i \in [m]$ using $O(\log(|x|))$ bits

Read positions $j_{i,1}, j_{i,2}, \ldots, j_{i,q}$ in π.

Accept iff $\varphi_i(\pi) = f_i(\pi_{j_{i,1}}, \pi_{j_{i,2}}, \ldots, \pi_{j_{i,q}}) = 1$

If $x \in L$, then $\exists$ satisfying assignment π, so $\Pr[\text{accept}] = 1$

If $x \notin L$, then at most fraction ρ of constraints

satisfied, so $\Pr[\text{accept}] \leq \rho$.
Verifier can repeat this test K times for K such that $\rho^K \leq 1/2$

$$K \sim 1/\log(1/\rho) = O_\rho(1) \text{ enough}$$

Query complexity $K \cdot q = O(1)$. $\blacksquare$

Views of the PCP theorem

Locally checkable proof	Hardness of approximation
PCP verifier V run on x	CSP instance φ_x
PCP proof π	Assignment to $u = Vars(\varphi_x)$
Length $\lvert\pi\rvert$	$n = \lvert Vars(\varphi_x)\rvert$ # variables
# queries q	Arity q of constraints
# random bits r	$\log$ (# constraints m)
Soundness error s	Maximum $val(\varphi_x)$ for no instance x
$NP \subseteq PCP_{1,1/2}(\log n, 1)$	ρ-GAP qCSP is NP-hard

Version B $\Rightarrow$ Version C

This is immediate – 3-CNF formulas are a particular form of 3CSP instances

Version C $\Rightarrow$ Version B

Suppose that $q \in \mathbb{N}^+$ and $\rho \in (0,1)$ are such that ρ-GAP qCSP is NP-hard.
Let $\varepsilon = 1 - \rho > 0$.
Let φ be a qCSP instance over n variables with m constraints

Each constraint

$$\varphi_i(u) = f_i(u_{j_{i,1}}, u_{j_{i,2}}, \ldots, u_{j_{i,q}})$$

can be expressed as a CNF formula at most 2^q clauses of size q

Let φ_i' denote this q-CNF formula

Let $\varphi' = \bigwedge_{i=1}^{m} \varphi_i'$ denote the q-CNF formula corresponding to the collection of clauses φ_i' for all constraints $\varphi_i \in \varphi$.

Then φ' has at most $m \cdot 2^q$ clauses.

φ yes instance $\Rightarrow$ φ' satisfiable

φ no instance $\Rightarrow$ Any assignment violates ε-fraction of constraints φ_i
$\Rightarrow$ Violates at least $\boxed{\dfrac{\varepsilon}{2^q}}$-fraction of clauses in φ'

Any q-clause can be turned into $\le q$ 3-clauses using (unique) extension variables $\overline{X}$

$$a_1 \vee a_2 \vee \dots \vee a_{q-1} \vee a_q \qquad (1)$$

$$\begin{gathered}
a_1 \vee a_2 \vee y_1 \\
\overline{y_1} \vee a_3 \vee y_2 \\
\overline{y_2} \vee a_4 \vee y_3 \\
\vdots \\
\overline{y_{q-4}} \vee a_{q-2} \vee y_{q-3} \\
\overline{y_{q-3}} \vee a_{q-1} \vee a_q
\end{gathered} \qquad (2)$$

Let φ'' be φ' turned into 3-CNF formula in this way

Any assignment violating (1) has to violate at least one clause in (2)

φ satisfiable $\Rightarrow$ φ' satisfiable $\Rightarrow$ φ'' satisfiable

At least fraction ε of constraints in φ violated $\Rightarrow$

$\Rightarrow$ –"– $\varepsilon/2^q$ –"– φ' –"– $\Rightarrow$

$\Rightarrow$ –"– $\dfrac{\varepsilon}{q \cdot 2^q}$ –"– φ''

And φ'' is a CNF formula with $\le q m 2^q$ clauses over $\le n + q m 2^q$ variables

HARDNESS OF APPROXIMATION FOR VERTEX COVER AND INDEPENDENT SET

$|V| = n$

Given undirected graph $G = (V, E)$

A VERTEX COVER $S \subseteq V$ satisfies

$$\forall (u,v) \in E \quad S \cap \{u,v\} \neq \emptyset$$

An INDEPENDENT SET $I \subseteq V$ satisfies

$$\forall (u,v) \in E \quad \{u,v\} \not\subseteq I.$$

Let $\boxed{VC(G)} = \min \{ |S| : S \text{ vertex cover of } G \}$
Let $\boxed{IS(G)} = \max \{ |I| : I \text{ independent set of } G \}$

We have

$$\boxed{VC(G) = n - IS(G)} \qquad (*)$$

since any complement of a vertex cover is an independent set and vice versa.

Approximation-wise, problems can be very different.

Suppose $\quad VC(G) = IS(G) = n/2$

$1/2$-approximation algorithm for MIN VERTEX COVER will find set S of size $|S| \leq n-1$.

Complement $I = V \setminus S$ can be of size $|I| = 1$, although $IS(G) = n/2$!

Approximation factor $\frac{1}{n/2} \to 0 \ldots$

This is inherent!

THEOREM 11.15
There is some $\gamma \in (1/2, 1)$ such that computing a γ-approximation to MINVERTEXCOVER is NP-hard
For every $\rho \in (0,1)$ it holds that computing a ρ-approximation to MAXINDEPENDENTSET is NP-hard

Recall reduction from INDEPENDENTSET to 3SAT in Thm 2.15

Clause $C \rightsquigarrow$ clique of 7 satisfying (partial) assignments
Edges between inconsistent partial assignments in different clusters

LEMMA 11.16
The polynomial-time reduction from 3-CNF formula φ to graph $G(\varphi)$ in (proof of) Thm 2.15 is such that

$$\boxed{IS(G(\varphi)) = val(\varphi) \cdot \frac{|V(G(\varphi))|}{7}}$$

Proof Left as an exercise — any independent set corresponds to a (partial) truth value assignment satisfying that many clauses.

COROLLARY 11.17
If $P \neq NP$, then there exist constants $\rho_{IS}, \rho_{VC} < 1$ such that it is not possible to ρ_{IS}-approximate MAX INDEPENDENT SET or ρ_{VC}-approximate MIN VERTEX COVER in polynomial time.

<u>Proof</u>

Let L be any NP-complete language. Let f_L be a poly-time reduction from L to 3SAT as in PCP Theorem, version B such that if $x \in L$ then $\varphi = f_L(x)$ has $val(\varphi) = 1$ and if $x \notin L$ then $val(\varphi) < \rho^*$

Run the reduction in Lemma 11.16 to obtain graph $G(\varphi)$. Applying a ρ^*-approximation algorithm for MAX INDEPENDENT SET will then allow us to decide whether $x \in L$ (if a independent set of size $\geq \rho^* |V(G(\varphi))|/7$ is found) or $x \notin L$ (if the independent set found is smaller). Hence, we can pick $\rho_{IS} = \rho^*$.

$\boxed{n := |V(G(\varphi))|}$

For vertex cover, we have (by (*)) that

$$VC(G(\varphi)) = n - val(\varphi) \cdot \frac{n}{7}$$

Suppose that MIN VERTEX COVER has a ρ'-approximation algorithm for

$$\rho' = \frac{6}{7 - \rho^*} \in (0, 1)$$

If $x \in L$, then φ is satisfiable and $VC(G(\varphi)) = n - \frac{n}{7}$

A ρ'-approximation algorithm would return a vertex cover of size

$$\leq \frac{1}{\rho'}\left(n - \frac{n}{7}\right)$$

$$= \frac{7-\rho^*}{6} \frac{6n}{7} = n - \rho^* \frac{n}{7}$$

If $x \notin L$, then an optimal vertex cover has size

$$= n - val(\varphi)\frac{n}{7} > n - \rho^* \frac{n}{7}$$

Hence, we would be able to decide L, and the lemma is true for

$$\rho_{VC} = \frac{6}{7-\rho^*}$$

To complete the proof of Thm 11.5, need to amplify approximation gap for independent set. One standard trick, that works also in this case, is to use a kind of graph product as defined next.

Given undirected graph $G=(V,E)$
$|V|=n$ and $k \in \mathbb{N}^+$

Define G^k by

$V(G^k) = \{ S \mid S \subseteq V, |S|=k \}$

$E(G^k) = \{ (S_1, S_2) \mid S_1 \cup S_2$ is not an independent set in $G \}$

G^k has $\binom{n}{k}$ vertices

If $I^k = \{S_1, S_2, \dots, S_j\}$ is an independent set in G^k, then $\bigcup_{i=1}^{j} S_i$ is an independent set in G.

If I is an independent set of size t in G then $\{S \mid S \subseteq I, |S|=k\}$ is an independent set of size $\binom{t}{k}$ in G^k

Hence

$$\boxed{IS(G^k) = \binom{IS(G)}{k}}$$

Let us go back to reductions from L to 3SAT and from 3SAT to INDEPENDENT SET and compose with a k-wise graph product to obtain $G(\varphi)^k$. This is a poly-time reduction for any constant k.

If $x \in L$, then

$$IS(G(\varphi)^k) = \binom{n/7}{k} \qquad (i)$$

If $x \notin L$, then

$$IS(G(\varphi)^k) < \binom{\rho^* n/7}{k} \qquad (ii)$$

The quotient of (ii) and (i) is

$$\binom{\rho^* n/7}{k} \Big/ \binom{n/7}{k} =$$

$$\frac{(\rho^* n/7)(\rho^* n/7 - 1)\cdots(\rho^* n/7 - k + 1)}{(n/7)(n/7 - 1)\cdots(n/7 - k + 1)} <$$

$$\left(\frac{\rho^* n/7}{n/7}\right)^k = (\rho^*)^k$$

Thus, if we can approximate MAXINDEPENDENTSET to within factor $(\rho^*)^k$, then we can distinguish cases "$x \in L$" and "$x \notin L$".

Let $\rho' > 0$ be any constant.

Picking $k = O(1)$, $\left[k = \left\lceil \frac{\log(1/\rho')}{\log(1/\rho^*)} \right\rceil\right]$ so that $(\rho^*)^k > \rho'$, shows that a ρ'-approximation of MAXINDEPENDENTSET would show $P = NP$.

This concludes the proof of Thm 11.15.

WHAT DID WE DO TODAY?

- Introduced constraint satisfaction problems (CSP) and $\boxed{\rho\text{-GAP }q\text{ CSP}}$
- Saw that

  PCP THEOREM as locally checkable proof $\Longleftrightarrow$ PCP THEOREM as hardness of approximation

- Key insight: $x \in L$ if exists proof π which verifier V is likely to accept

  Finding proof that makes V accept

  $\Updownarrow$

  Solving constraint satisfaction problem

- Decision versions of VERTEXCOVER and INDEPENDENTSET are equivalent
- MINVERTEXCOVER has $1/2$-approximation but cannot be approximated arbitrarily well
- MAXINDEPENDENTSET has no constant-factor approximation algorithm !!

WHAT IS UP NEXT?

- Proof of weaker version of PCP Theorem:

$$\boxed{NP \subseteq PCP_{1,1/2}(\text{poly}(n), 1)}$$

- Will use linearity test extensively
- Given almost linear function f, will need to evaluate f at any x (even if $f(x)$ is one of distorted values)
- Will need that following problem is NP-complete:

Variables $u_1, u_2, \ldots, u_n$

Equation E_ℓ

$$\sum_{i=1}^{n} \sum_{j=i}^{n} a_{\ell,ij}\, u_i u_j = b_\ell \qquad \begin{array}{l} a_{\ell,ij} \in \{0,1\} \\ b_\ell \in \{0,1\} \end{array}$$

QUADEQ

Given equations $\{E_1, \ldots, E_m\}$, is there a $\{0,1\}$-assignment to u_i:s satisfying all equations?