Complexity Theory

Jan Křetínský

Technical University of Munich

Summer 2019

June 3, 2019

Lecture 19

Hardness of Approximation

Recap: optimization

- many decision problems we have seen have optimization versions
- both minimization and maximization
- algorithms return best solution with respect to optimization parameter ρ

Recap: optimization

- many decision problems we have seen have optimization versions
- both minimization and maximization
- algorithms return best solution with respect to optimization parameter ρ

Examples

problem	min/max	parameter
3SAT	max	fraction of satisfiable clauses
Indset	max	size of independent set
VC	min	size of cover

Recap: approximation results

- vertex cover has a 2-approximation
  - possibly **NP**-hard to approximate to within $2 - \epsilon$ for all $\epsilon > 0$
  - currently known: **NP**-hard to approximate to within $10\sqrt{5} - 21$;
  - *I. Dinur, S. Safra*, The importance of being biased, STOC 2002.

Recap: approximation results

- vertex cover has a 2-approximation
  - possibly **NP**-hard to approximate to within $2 - \epsilon$ for all $\epsilon > 0$
  - currently known: **NP**-hard to approximate to within $10\sqrt{5} - 21$;
  - *I. Dinur, S. Safra*, The importance of being biased, STOC 2002.
- set cover has a $\ln n$ approximation
  - this is optimal; it is **NP**-hard to approximate to within $(1 - \epsilon)\ln n$
  - *U. Feige*, A threshold of $\ln n$ for approximating set cover, STOC 1996.

Recap: approximation results

- vertex cover has a 2-approximation
  - possibly **NP**-hard to approximate to within $2 - \epsilon$ for all $\epsilon > 0$
  - currently known: **NP**-hard to approximate to within $10\sqrt{5} - 21$;
  - *I. Dinur, S. Safra*, The importance of being biased, STOC 2002.
- set cover has a $\ln n$ approximation
  - this is optimal; it is **NP**-hard to approximate to within $(1 - \epsilon) \ln n$
  - *U. Feige*, A threshold of $\ln n$ for approximating set cover, STOC 1996.
- TSP also hard to approximate to within any $1 + \epsilon$

Polynomial time approximation schemes

A problem has a polynomial time approximation scheme if for all $\epsilon > 0$ it can be efficiently approximated to within a factor of $1 - \epsilon$ for maximization and $1 + \epsilon$ for minimization.

Polynomial time approximation schemes

A problem has a polynomial time approximation scheme if for all $\epsilon > 0$ it can be efficiently approximated to within a factor of $1 - \epsilon$ for maximization and $1 + \epsilon$ for minimization.

Examples

- knapsack
- bin packing
- subset sum
- a number of other scheduling problems

Polynomial time approximation schemes

A problem has a polynomial time approximation scheme if for all $\epsilon > 0$ it can be efficiently approximated to within a factor of $1 - \epsilon$ for maximization and $1 + \epsilon$ for minimization.

Examples

- knapsack
- bin packing
- subset sum
- a number of other scheduling problems

Which **NP**-complete problems do have PTAS? Which don't? How to prove results on previous slide?

Recap: gap – TSP$[|V|, h|V|]$

An algorithm to solve the gap problem needs to:

- if G has a shortest tour of length $< |V|$ then G is accepted by the gap algorithm
- if the shortest tour of G is $> h|V|$ then G is rejected
- otherwise: don't care

Recap: gap – TSP$[|V|, h|V|]$

An algorithm to solve the gap problem needs to:

- if G has a shortest tour of length $< |V|$ then G is accepted by the gap algorithm
- if the shortest tour of G is $> h|V|$ then G is rejected
- otherwise: don't care

Theorem: For any $h \geq 1$ gap – TSP$[|V|, h|V|]$ is **NP**-hard by reduction from Hamiltonian cycle

Recap: gap $-$ TSP$[|V|, h|V|]$

An algorithm to solve the gap problem needs to:

- if G has a shortest tour of length $< |V|$ then G is accepted by the gap algorithm
- if the shortest tour of G is $> h|V|$ then G is rejected
- otherwise: don't care

Theorem: For any $h \geq 1$ gap $-$ TSP$[|V|, h|V|]$ is **NP**-hard by reduction from Hamiltonian cycle

$\Rightarrow$ It is **NP**-hard to approximate TSP to within any factor $h \geq 1$.

Recap: gap – TSP$[|V|, h|V|]$

An algorithm to solve the gap problem needs to:

- if G has a shortest tour of length $< |V|$ then G is accepted by the gap algorithm
- if the shortest tour of G is $> h|V|$ then G is rejected
- otherwise: don't care

Theorem: For any $h \geq 1$ gap – TSP$[|V|, h|V|]$ is **NP**-hard by reduction from Hamiltonian cycle

$\Rightarrow$ It is **NP**-hard to approximate TSP to within any factor $h \geq 1$.

The reduction is called gap-producing.

Agenda

- $\text{gap} - 3\text{SAT}[\rho, 1]$
- 7/8 approximation for max3SAT
- PCP theorem: hardness of approximation view
- gap-preserving reductions
- hardness of approximating Indset and VC

gap-3SAT$[\rho, 1]$

- gap $-$ 3SAT$[\rho, 1]$ is the gap version of max3SAT which computes the largest fraction of satisfiable clauses
- a 3CNF with m clauses is accepted if it is satisfiable
- it is rejected if $< \rho \cdot m$ clauses are satisfiable
- until 1992 it was an open problem whether max3SAT could be approximated to within any factor $> 7/8$
- why 7/8?

A 7/8 approximation of max3SAT

Theorem

For all 3CNF with exactly three independent literals per clause, there exists an assignment that satisfies $\geq 7/8$ *of the clauses.*

A 7/8 approximation of max3SAT

Theorem

For all 3CNF with exactly three independent literals per clause, there exists an assignment that satisfies $\geq 7/8$ of the clauses.

Proof

- for a random assignment let Y_i be the random variable that is true if clause C_i is true under the assignment
- then $N = \Sigma_{i=1}^{m} Y_i$ is the number of satisfied clauses
- $E[Y_i] = 7/8$ for all i

$\Rightarrow$ $E[N] = 7/8 \cdot m$

- by the law of average (probabilistic method basic principle) there must exist an assignment that makes 7/8 of the clauses true

A 7/8 approximation of max3SAT

Theorem

For all 3CNF with exactly three independent literals per clause, there exists an assignment that satisfies $\geq 7/8$ of the clauses.

Proof

- for a random assignment let Y_i be the random variable that is true if clause C_i is true under the assignment
- then $N = \Sigma_{i=1}^{m} Y_i$ is the number of satisfied clauses
- $E[Y_i] = 7/8$ for all i

$\Rightarrow$ $E[N] = 7/8 \cdot m$

- by the law of average (probabilistic method basic principle) there must exist an assignment that makes 7/8 of the clauses true

Can we do any better than 7/8?

No!

Theorem

For every $\epsilon > 0$ $\mathrm{gap} - 3\mathrm{SAT}[7/8 + \epsilon, 1]$ *is* **NP**-*hard.*

- this is a PCP theorem by *J. Håstad*, Some optimal inapproximability results, STOC 1997.
- as a consequence, if there exists a $7/8 + \epsilon$ approximation of max3SAT then $\mathbf{P} = \mathbf{NP}$
- we will later prove a much weaker PCP theorem

Agenda

- gap – 3SAT$[\rho, 1]$ ✓
- 7/8 approximation for max3SAT ✓
- PCP theorem: hardness of approximation view
- gap-preserving reductions
- hardness of approximating Indset and VC

THE PCP theorem

Håstads result is one in a series of inapproximability results based on the PCP theorem.

Theorem (PCP: hardness of approximation)

There exists a $\rho < 1$ *such that* $\mathrm{gap}-3\mathrm{SAT}[\rho, 1]$ *is* **NP***-hard.*

- Safra: *One of the deepest and most complicated proofs in computer science with a matching impact.*
- original proof in two papers:
  - *Arora, Safra*, Probabilistic checking of proofs, FOCS 92
  - *Arora, Lund, Motwani, Sudan, Szegedy*, Proof verification and the hardness of approximations, FOCS 92.
- virtually all inapproximability results depend on the PCP theorem and the notion of gap preserving reductions by Papadimitriou and Yannakakis

Probabilistically checkable proofs

- the PCP theorem is equivalent to the statement **NP** $=$ **PCP**$[\log n, 1]$
- PCP stands for probabilistically checkable proofs and is related to interactive proofs and **MIP** $=$ **NEXP**
- equivalence of two views shown in next lecture
- **NP** $=$ **PCP**$[poly(n), 1]$ shown after that

Agenda

- gap – 3SAT$[\rho, 1]$ ✓
- 7/8 approximation for max3SAT ✓
- PCP theorem: hardness of approximation view ✓
- gap-preserving reductions
- hardness of approximating Indset and VC

Gap-producing and preserving reductions

PCP theorem states that for every $L \in \mathbf{NP}$ there exists a gap-producing reduction f to $\text{gap} - 3\text{SAT}[\rho, 1]$:

- $x \in L \implies f(x)$ is satisfiable
- $x \notin L \implies$ less than ρ of the $f(x)$'s clauses can be satisfied at the same time

Gap-producing and preserving reductions

PCP theorem states that for every $L \in \mathbf{NP}$ there exists a gap-producing reduction f to $\mathrm{gap}-\mathrm{3SAT}[\rho, 1]$:

- $x \in L \implies f(x)$ is satisfiable
- $x \notin L \implies$ less than ρ of the $f(x)$'s clauses can be satisfied at the same time

Observation

- in order to show inapproximability of other problems, we want to preserve gaps by reductions

$\text{gap} - \text{3SAT}[\rho, 1] \leq_{gap} \text{gap} - \text{IS}[\rho, 1]$

Consider the proof of 3SAT $\leq_p$ Indset (nodes are satisfying assignments for each clause, edges between incompatible ones).

The reduction f used there is actually gap-preserving, we write

$$\text{gap} - \text{3SAT}[\rho, 1] \leq_{gap} \text{gap} - \text{IS}[\rho, 1]$$

- if 3CNF ψ with m clauses is satisfiable then graph $f(\psi)$ has an independent set of size m
- if less than ρ of ψ's clauses can be satisfied, the largest independent set has less than $\rho \cdot m$ nodes
- hence: if we can approximate Indest to within ρ, then we can approximate max3SAT to within ρ, then we can decide any $L \in$ **NP**

What about vertex cover?

The same reduction f from independent set can be used to show hardness of approximating vertex cover to within $(7-\rho)/6$ for the same ρ used in max3SAT and Indset.

What about vertex cover?

The same reduction f from independent set can be used to show hardness of approximating vertex cover to within $(7-\rho)/6$ for the same ρ used in max3SAT and Indset.

- ψ satisfiable

$\Rightarrow$ $f(\psi)$ has i.s. of size m

$\Rightarrow$ $f(\psi)$ has a v.c. of size $6m$

What about vertex cover?

The same reduction f from independent set can be used to show hardness of approximating vertex cover to within $(7-\rho)/6$ for the same ρ used in max3SAT and Indset.

- ψ satisfiable

$\Rightarrow$ $f(\psi)$ has i.s. of size m

$\Rightarrow$ $f(\psi)$ has a v.c. of size $6m$

- only $\rho \cdot m$ of ψ's clauses satisfiable

$\Rightarrow$ $f(\psi)$ has largest i.s. smaller than ρm

$\Rightarrow$ $f(\psi)$ has smallest v.c. of size larger than $(7-\rho)m$

Independent set vs. vertex cover

- For both independent set and vertex cover, we know that there exist a $\rho < 1$ such that neither can be approximated to within ρ (resp. $1/\rho$)

Independent set vs. vertex cover

- For both independent set and vertex cover, we know that there exist a $\rho < 1$ such that neither can be approximated to within ρ (resp. $1/\rho$)
- optimal solutions are intimately related: if vc is the smallest vertex cover and is the largest independent set then $vc = is - n$

Independent set vs. vertex cover

- For both independent set and vertex cover, we know that there exist a $\rho < 1$ such that neither can be approximated to within ρ (resp. $1/\rho$)
- optimal solutions are intimately related: if vc is the smallest vertex cover and is the largest independent set then $vc = is - n$
- but: approximation is different; using the ρ app. for independent set, yields a $\frac{n-\rho \cdot is}{n-is}$ approximation for set cover

Independent set vs. vertex cover

- For both independent set and vertex cover, we know that there exist a $\rho < 1$ such that neither can be approximated to within ρ (resp. $1/\rho$)
- optimal solutions are intimately related: if vc is the smallest vertex cover and is the largest independent set then $vc = is - n$
- but: approximation is different; using the ρ app. for independent set, yields a $\frac{n-\rho\cdot is}{n-is}$ approximation for set cover
- for independent set we can show **NP**-hardness of approximation to within any factor $\rho < 1$ by gap amplification

Gap amplification

- given instance $G = (V, E)$
- construct $G' = (V \times V, E')$ where

$$E' = \{(u, v), (u', v') \mid (u, u') \in E \vee (v, v') \in E\}$$

- if $I \subseteq V$ is an i.s. of G then $I \times I$ is an i.s. of G'; hence $opt(G') \geq opt(G)^2$
- if I' is an optimal i.s. in G' with vertices $(u_1, v_1), \ldots, (u_j, v_j)$ then the u_i and the v_i are each i.s. in G with at most $opt(G)$ vertices; hence $opt(G') \leq opt(G)^2$
- hence i.s. is also hard to approximate within ρ^2
- this can be done any constant k times to obtain the result

What have we learnt?

- 7/8 approximation for max3SAT
- PCP theorem: hardness of approximating max3SAT
- gap-preserving reductions to obtain more inapproximability results
- **NP**-hardness of approximating Indset to within any $\rho < 1$
- **NP**-hardness of approximating VC to within some $\rho > 1$ (yet unknown)
- but: many **NP**-complete problems can still be approximated to within any factor $1 + \epsilon$

Up next

- PCP: hardness of approximation vs. prob. checkable proofs
- proof of a weaker PCP theorem