Optimization Problems and Approximation

We are unable to solve NP-complete problems efficiently, i.e., there is no known way to solve them in polynomial time.

Most of them are decision versions of optimization problems...

with a set of feasible solutions for each instance

with an associated quality measure

Why not looking for an approximate solution?

Is there a difference in complexity?

Optimization Problems and Approximation Example Knapsack revisited

All set $T \subseteq S$: $\sum_{i \in T} w(i) \leq W$ are feasible solutions. $\sum v(i)$ is the quality of the solution *T* wrt. to the instance *i*.

KNAPSACK = < I, sol, m, max > $I = \{< S, w, W, v > | S = \{1, ..., n\}, w, v : S \to N, W \in N, V \in N\}$

 $\operatorname{sol}(i) = \left\{ T \subseteq S : \sum_{i \in T} w(i) \le W \right\}$

 $\mathbf{m}(i,s) = \sum_{i \in T} v(i)$

Optimization Problems and Approximation Definition of Optimization Problems

OPTPROB =< *I*, sol, m, type >

I the instance set

sol(i) the set of feasible solutions for instance i

(sol(*i*) nonempty for $i \in I$) m(*i*, *s*) the measure of solution *s* wrt.instance *i*

(positive integer for $i \in I$ and $s \in sol(i)$)

 $opt(i) = type_{s \in sol(i)} m(i, s)$

Optimization Problems and Approximation The Class NPO

NPO is the class of optimization problems whose decision versions are in *NP*.

OPTPROB = < I, sol, m, type $> \in NPO$ iff

 $\begin{array}{l} \exists \mathsf{polynomial} \ p : \forall i \in I, s \in \mathsf{sol}(i) : \mid s \mid \leq p(\mid i \mid) \\ \\ \mathsf{deciding} \ s \in \mathsf{sol}(i) \ \mathsf{is} \ \mathsf{in} \ P \end{array}$

computing m(s, i) is in *FP*

Optimization Problems and Approximation Example Problem: MaxkSat

MaxkSat =< *I*, sol, m, max >

I = CNF - Formulas with at most k literals per clause sol(φ) = set of assignments to the vars. of φ m(φ , A) = the number of clauses which are satisfied by A

MaxSat has all CNF - Expressions as instances.

There is also a weighted version : Each clause has a weight -- the measure is the sum of the weights of the satisfied clauses.

Example Problem: MaxkSat NP-hardness

MaxkSat =< I, sol, m, max >

I = CNF - Formulas with at most k literals per clause sol(φ) = set of assignments to the vars. of φ m(φ , A) = the number of clauses which are satisfied by A

Max3Sat(D) is certainly NP – complete (thus Max3Sat is NP – hard): 3SAT is a special case

But also Max2Sat(D) is NP-complete....

Example Problem: MaxkSat NP-hardness

But also *Max2Sat(D)* is *NP* – *complete...*. a local replacement reduction from 3*SAT* :

 $(x \lor y \lor z) \text{ is replaced by}$ (x) (y) (z) (w) $(\neg x \lor \neg y) (\neg y \lor \neg z) (\neg x \lor \neg z)$ (x \lor \neg w) (y \lor \neg w) (z \lor \neg w) A(x) = A(y) = A(z) =true A(w) = true7 clauses satisfied

Example Problem: MaxkSat NP-hardness

But also *Max2Sat(D)* is *NP* – *complete...*. a local replacement reduction from *3SAT* :

 $(x \lor y \lor z) \text{ is replaced by}$ (x)(y)(z)(w) $(\neg x \lor \neg y)(\neg y \lor \neg z)(\neg x \lor \neg z)$ $(x \lor \neg w)(y \lor \neg w)(z \lor \neg w)$

 $A(x) \neq A(y) = A(z) =$ true A(w) = true7 clauses satisfied

Example Problem: MaxkSat NP-hardness

But also Max2Sat(D) is NP – complete.... a local replacement reduction from 3SAT :

 $(x \lor y \lor z) \text{ is replaced by}$ (x) (y) (z) (w) $(\neg x \lor \neg y) (\neg y \lor \neg z) (\neg x \lor \neg z)$ $(x \lor \neg w) (y \lor \neg w) (z \lor \neg w)$

 $A(x) = A(y) \neq A(z) =$ true A(w) =false 7 clauses satisfied

Example Problem: MaxkSat NP-hardness

But also Max2Sat(D) is NP - complete....... a local replacement reduction from 3SAT :

 $\begin{array}{c} (x \lor y \lor z) \text{ is replaced by} \\ \hline (x) (y) (z) (w) \\ (\neg x \lor \neg y) (\neg y \lor \neg z) (\neg x \lor \neg z) \\ (x \lor \neg w) (y \lor \neg w) (z \lor \neg w) \end{array}$

A(x) = A(y) = A(z) = false A(w) = false 6 clauses satisfied

Example Problem: MaxkSat NP-hardness

But also Max2Sat(D) is NP - complete....... a local replacement reduction from 3SAT:

 $\begin{aligned} (x \lor y \lor z) \text{ is replaced by} \\ (x) (y) (z) (w) \\ (\neg x \lor \neg y) (\neg y \lor \neg z) (\neg x \lor \neg z) \end{aligned}$

Each 3-literal clause is replaced by a 10 clauses.

Iff the original clause was satisfied, then 7 in the replacement can be satisfied.

•

Set K = 7m where m is the number of clauses in the original.

Optimization Problems and Approximation Performance Ratio Approximation algorithms deliver solutions of guaranteed quality – they are not heuristics. But how to measure the quality of a solution? Let O =< I, sol, m, type > be an optimization problem. given $i \in I$ and a $s \in sol(i)$ we define opt(i) = m(i,s)

$$R(i, s) = \max\left\{\frac{1}{m(i, s)}, \frac{1}{opt(i)}\right\}$$
as the performance ratio.

 $s \in sol(i)$ is a an r – approximate solution if $R(i, s) \leq r$.

2





Optimization Problems and Approximation Approximation Problem

Let O = < I, sol, m, type > be an optimization problem and *r* a function $N \rightarrow [1, \infty)$.

Then the approximation problem < 0, r > is to find for all instances $i \in I$ an r(|i|) - *approximate* solution $s \in sol(i)$.

The question is which approximation problems < O, r > are located in *FP*.

And *how to prove* that they are not (under some assumption such as $P \neq NP$)..

Approximation Algorithm Example Problem: MaxkSat

 $approxMaxSat(\varphi)$

1. for i = 1 to n

2. $\operatorname{val} := \mathsf{E}(\mathsf{m}(\varphi, A \cup \{x_i = \mathsf{true}\})) > \mathsf{E}(\mathsf{m}(\varphi, A \cup \{x_i = \mathsf{false}\}));$ 3. $A := \mathsf{A} \cup \{x_i = \mathsf{val}\}; \ \varphi := \varphi[x_i = \mathsf{val}];$

4. return A;

 $\mathsf{E}(\varphi, \{\}) = \sum_{C \in \varphi} 1 - 2^{-|C|} \ge \sum_{C \in \varphi} 1 - 2^{-1} = \frac{1}{2} |\varphi|$

Thus, this algorithm is a 2-approximate algorithm or better.

Approximation Algorithm Example Problem: VertexCover

approxVertexCover(V,E)

1.C:=∅;

- 2. while $E \neq \emptyset$ do
- 3. pick $a < u, v > \in E$
- 4. $\mathbf{C} := \mathbf{C} \cup \{u, v\};$

5. remove {*u*,*v*} from *V*,*E*;

6. return C;

C is indeed a valid cover.

Every cover must cover all the edges picked in line 3.

Thus every cover must contain at least |C|/2 vertexes.

 $R(G,C) = \frac{\mathrm{m}(G,C)}{\mathrm{opt}(G)} \le 2$

Approximation Classes APX

We have two approximation problems, which can be solved within a constant performance ratio within polynomial time.

So it's time to define a corresponding class: APX.

Let O be an NPO problem.

 $O \in APX$ iff there exists an r - approximation algorithm for Owhich run in polynomial time for

some constant $r \ge 1$.

Approximation Classes Example Problem: <u>TSP (I)</u>

We will show that $TSP \in APX \iff P = NP$.

We use another *NP* – *complete* problem to reduce from : *HAMILTONIANCYCLE*

HAMILTONIANCYCLE : Given a graph $G = \langle V, E \rangle$, is there a cycle, which visits any node exactly once?

We construct a distance matrix M as follows (for $r \ge 1$):

 $M(u,v) = \begin{cases} 1: < u, v > \in E \\ \left\lceil r \mid V \mid \right\rceil: \text{ otherwise} \end{cases}$

Approximation Classes Example Problem: TSP (II)

We construct a distance matrix *M* as follows $(r \ge 1)$: $M(u,v) = \begin{cases} 1: < u, v > \in E \\ \lceil r | V \rceil \end{cases}$ otherwise

If *G* is a positive instance, then opt(M) = |V|. Otherwise $opt(M) \ge \lceil r/V \rceil + |V| - 1$.

Now assume that there is an r – *approximate* algorithm for *TSP*.

Approximation Classes Example Problem: TSP (III)

If *G* is a positive instance, then opt(M) = |V|. Otherwise $opt(M) \ge \lceil r/V \rceil + |V| - 1$.

Now assume that there is an r – *approximate* algorithm apporx for *TSP* and let s = approx(M).

If $G \in HAMILTONIANCYCLE$, we find

 $r \ge R(M,s) = \frac{\mathrm{m}(M,s)}{\mathrm{opt}(M)} = \frac{\mathrm{m}(M,s)}{|V|} \text{ and so } |V| r \ge \mathrm{m}(M,s).$ But otherwise we have

 $\mathbf{m}(M,s) \ge \mathrm{opt}(\mathbf{M}) \ge \left\lceil r \mid V \mid \rceil + \mid V \mid -1 > \left\lceil r \mid V \mid \rceil \right\rceil$

Approximation Classes Example Problem: TSP (IV)

So we could prove that $TSP \notin APX$ (assuming $P \neq NP$) by giving a reduction from an NP-hard problem, which established a gap between positive and negative instances.

The gap was large enough to distinguish whether we reduced from a positive or a negative instance.

Wanted: A generic reduction from NP - hard problems, to approximation problems which produces gaps.

Approximation Classes Relationships

 $APX \subseteq NPO$

 $TSP \in APX \iff P = NP$

 $APX \subset NPO \Leftrightarrow P \neq NP$

Max3Sat and VertexCover are in APX.



Approximation Schemes The classes PTAS and FPTAS

 $O \in FPTAS$ if there is an approximation scheme A such that A(i, r) runs in DTIME(poly(|i|, 1/(r-1)))for all $i \in I$ and r > 1.

 $O \in PTAS$ if there is an approximation scheme A such that A(i, r) runs in DTIME(poly(/i/))for all $i \in I$ and any fixed r > 1.

Approximation Schemes Example Problem: KNAPSACK *KNAPSACK* =< *I*, sol, m, max > $I = \{ < S, w, W, v > | S = \{ 1, ..., n \}, w, v : S \to N, W \in N \}$ $\operatorname{sol}(i) = \left\{ T \subseteq S : \sum w(i) \le W \right\}$ $m(i,s) = \sum v(i)$ Let W(i, v) be the minimum weight attainable

by selecting among the first i items such that

that their total value is *exactly v*.

Example Problem: KNAPSACK A Pseudo-Polynomial Algorithm

Let W(i, v) be the minimum weight attainable by selecting among the first i items such that their total value is exactly v. W(0,0) = 0

 $W(0,v) = \infty \ v \neq 0$

 $W(i+1,v) = \min\{W(i,v), [W(i,v-v(i+1)) + w(i+1)]\}$

By building the table of the W(i, v) for $0 \le i \le n$ and $0 \le v \le V = \sum v(i)$ we can solve *KNAPSACK*.

This algorithm runs in DTIME(poly(n, V)) (pseudo - poly.)

Example Problem: KNAPSACK An FPTAS (I)

This algorithm runs in DTIME(poly(n, V)) (pseudo - poly.) Assume $\varepsilon > 0$ fixed.

Let $l = \lceil \log \max_{i \in S} v(i) \rceil$.

Choose k with $\frac{n}{n^k} < \varepsilon$.

Set $L = l - k \log n$.

Define I' with $v'(i) = v(i)/2^L 2^L$ We keep the most significant $k \log n$ bits.

The rest, i.e., $L = l - k \log n$, gets zeroized.

Example Problem: KNAPSACK An FPTAS (II)

This algorithm runs in *DTIME*(poly(*n*,*V*)) (pseudo - poly.)

Assume $\varepsilon > 0$ fixed. Let $l = \lceil \log \max_{i \in S} v(i) \rceil$.

Choose k with $\frac{n}{n^k} < \varepsilon$.

Set $L = l - k \log n$. Define I' with

 $v'(i) = v(i)/2^{L} 2^{L}$

 $\sum_{i \in T} v(i) \le \sum_{i \in T} v'(i) + |T| 2^{L}$

 $opt(I) \le opt(I') + n2^L$

 $\frac{\operatorname{opt}(I)}{\operatorname{opt}(I')} \le 1 + \frac{n2^L}{\max_{i \in S} v'(i)}$ $\leq 1 + \frac{n2^L}{2^l} < 1 + \varepsilon$

Solving I' optimally yields an $1 + \varepsilon$ approximate solution for I

Example Problem: KNAPSACK An FPTAS (III)

This algorithm runs in DTIME(poly(n, V)) (pseudo - poly.) Assume $\varepsilon > 0$ fixed.

Let $l = \lceil \log \max_{i \in S} v(i) \rceil$

Choose k with $\frac{n}{n^k} < \varepsilon$.

Set $L = l - k \log n$. Define I' with

 $v'(i) = v(i) / 2^L 2^L$

We can solve I' in DTIME(poly(n,V')) $= DTIME(poly(n, n2^{k \log n}))$

 $= DTIME(poly(n, 1/\varepsilon))$

Solving I' optimally yields an $1 + \varepsilon$ approximate solution for I within $DTIME(poly(/I/,1/\varepsilon))$. $KNAPSACK \in FPTAS$.





Polynomially Bound Problems Permit no FPTAS (II)

Set $r(i) = 1 + \frac{1}{p(|i|)}$, where p is the poly. - bound.

An r(i) – approximate solution s for i is optimal since, $\frac{p(|i|)+1}{n(|i|)} = r(i) \ge \frac{\text{opt}(i)}{m(i|s)} \text{ gives}$

$$\mathbf{m}(i,s) \ge \operatorname{opt}(i) \frac{p(|i|)}{p(|i|)+1} = \operatorname{opt}(i) - \frac{\operatorname{opt}(i)}{p(|i|)+1} > \operatorname{opt}(i) - 1$$

If *O* would be in *FPTAS* then we can solve *O* optimally in DTIME(poly(|i|,1/(r(|i|)-1)) = DTIME(poly(|i|)).

0



Approximation Classes Problems in PTAS-FPTAS

PLANAR INDEPENDENTSET is in NPO – PB and is NP – hard. PLANAR INDEPENDENTSET \in FPTAS \Rightarrow P = NP.

Unproven : *PLANAR INDEPENDENTSET* ∈ *PTAS*.



Hardness in Approximation

Wanted: A generic reduction from NP-hard problems, to approximation problems which produces gaps.

Remember the reduction to TSP ...

Relies on the so-called PCP-Theorem – an alternative formulation of NP.

It allows to reduce *NP* – *complete* languages to approximation problems.



Hardness in Approximation PCP-Theorem (I)

A language *L* is in PCP(r(n), q(n))if there is a polynomial time PCP(r(n), q(n)) - *Verifier V* such that

 $\forall x \in L \ \exists \Pi : \Pr_{\overline{R}} [V(x, \Pi, \overline{R}) = \operatorname{accept}] = 1$

 $\forall x \notin L \forall \Pi : \Pr_{\overline{n}}[V(x, \Pi, \overline{R}) = \operatorname{aceppt}] \leq 1/2$

with $\overline{R} = O(r(|x|))$, and V reading O(q(n)) bits non - adaptively from Π .

Easy : $NP \supseteq PCP(\log n, 1)$

 $\mathsf{Hard}: NP \subseteq PCP(\log n, 1)$

Hardness in Approximation PCP-Theorem (I)

A language *L* is in PCP(r(n), q(n))

if there is a polynomial time PCP(r(n),q(n)) - Verifier V such that

 $\forall x \in L \ \exists \Pi : \Pr_{\overline{R}} \left[V(x, \Pi, \overline{R}) = \operatorname{accept} \right] = 1$

 $\forall x \notin L \ \forall \Pi : \Pr_{\overline{R}}[V(x, \Pi, \overline{R}) = \operatorname{aceppt}] \le 1/2$

with $|\overline{R}| = O(r(|x|))$, and V reading O(q(n)) bits non-adaptively from Π .

PCP - Theorem : $NP = PCP(\log n, 1)$

Hardness in Approximation **PCP**-Theorem (II)

PCP - Theorem : $NP = PCP(\log n, 1)$

How to use?

Reduce the verification process to an approximation problem such that the gap of the acceptance probability of the PCP-Verifier translates into a gap in the measure of the optimal solution(s).

Hardness in Approximation Example Problem: Max3Sat (I)

Observe that the once the O(q(n)) bits have been read from the proof Π , the decision of V is only depending on them.

Thus we can define a set of Boolean Expressions $\varphi[x, \overline{R}](\overline{p})$ where x is the input, \overline{R} is the random string of length $O(\log n)$, \overline{p} are the bits read in Π , $\varphi[x, \overline{R}](\overline{p}) = 1 \Leftrightarrow V(x, \Pi, \overline{R}) = \text{accept.}$

Hardness in Approximation Example Problem: Max3Sat (II)

Each $\varphi[x, \overline{R}](\overline{p})$ can be expressed by *d* clauses, where *d* is constant (since $|\overline{p}|$ is constant).

Let φ be the conjunction of the expressions $\varphi[x, \overline{R}](\overline{p})$ for all \overline{R} ($|\overline{R}| = c \log n$).

 $x \in L \Rightarrow \exists \Pi : \Pr_{\overline{R}} [V(x, \Pi, \overline{R}) = \operatorname{accept}] = 1$

- \Rightarrow all $\varphi[x, \overline{R}]$ can be satisified satisfied simultansously
 - $\Rightarrow \phi$ satisfiable.

Let A be

Hardness in Approximation Example Problem: Max3Sat (III)

Each $\varphi[x, \overline{R}](\overline{p})$ can be expressed by *d* clauses, where *d* is constant (since $|\overline{p}|$ is constant).

Let φ be the conjunction of the expressions $\varphi[x, \overline{R}](\overline{p})$ for all \overline{R} ($|\overline{R}| = c \log n$).

$$\begin{split} x \notin L \Rightarrow \forall \Pi : \Pr_{\overline{R}} \Big| V(x, \Pi, \overline{R}) &= \operatorname{accept} \Big| \le 1/2 \\ \Rightarrow \text{ each assignment must leave } 1/2 \\ \text{ of the expressions } \varphi[x, \overline{R}] \text{ unsatisified.} \\ \Rightarrow \frac{\operatorname{opt}(\varphi)}{|\varphi|} \le \frac{1}{2} + \frac{1}{2} \frac{d-1}{d} \coloneqq f < 1 \end{split}$$

Hardness in Approximation Example Problem: Max3Sat (IV)

$$x \in L \Rightarrow \frac{1}{|\varphi|} = 1$$
$$x \notin L \Rightarrow \frac{\operatorname{opt}(\varphi)}{|\varphi|} \le \frac{1}{2} + \frac{1}{2} \frac{d-1}{d} \coloneqq f < 1$$
an $1 < r < \frac{1}{2}$ approximate solution for φ .

 $\frac{\mathrm{m}(\varphi, A)}{\mathrm{opt}(\varphi)} \ge \frac{1}{r} > f \qquad \qquad f = m(\varphi, A) > f = f |\varphi|$ $x \notin L \Rightarrow \mathrm{m}(\varphi, A) \le \mathrm{opt}(\varphi) \le f |\varphi| \quad \text{(for all } A)$

r – approximating Max3Sat is NP – hard (constant r > 1).

Hardness in Approximation Remark: Decoding of PCP-Proofs

 $\forall x \in L \; \exists \Pi : \Pr_{\overline{R}} [V(x, \Pi, \overline{R}) = \operatorname{accept}] = 1$ $\forall x \notin L \; \forall \Pi : \Pr_{\overline{R}} [V(x, \Pi, \overline{R}) = \operatorname{accept}] \leq 1/2$

Given a proof Π with $\Pr_{\overline{R}}[V(x, \Pi, \overline{R}) = \operatorname{accept}] > 1/2$ a proof Π ' with $\Pr_{\overline{R}}[V(x, \Pi', \overline{R}) = \operatorname{accept}] = 1$ can be reconstructed efficiently (in FP).

Π is basically encoded for error - correction -thus it possible to find the corresponding "usually encoded" proof efficiently.







Approximation Classes More Example Problems				
PLANAR INDE	EPSET SETC	ÇOVER	[TSP
$FPTAS \subseteq PTAS \subseteq APX \subseteq \log - APX \subseteq \operatorname{poly} - APX \subseteq \exp - APX \subseteq NPO$				
<i>KNAPSACK</i>	MAX 3SAT	COLC	DRING	MAXONESSAT

