LECTURE 9 26/11/2020

SET COVER – RECAP

GREEDY ALG. FOR SC

(1) $I := \emptyset$, $E := \emptyset$

(2) while $E \neq \{1,2,\dots,n\}$ do

for $j \in \{1,\dots,m\}$ s.t. $S_j \not\subseteq E$, define $p_j = \dfrac{c_j}{|S_j \setminus E|}$

$\ell = \operatorname{argmin}_j p_j$

$\rightarrow I := I \cup \{\ell\}$, $E := E \cup S_\ell$, $\boxed{y_e := p_j}$ $\forall e \in S_j \setminus E$

(3) OUTPUT I

LP Relaxation for SET COVER

(P) $\min \sum_{j=1}^{m} c_j x_j$

$\sum_{j: e \in S_j} x_j \geq 1 \quad \forall e$

$x_j \geq 0$

(D) $\max \sum_{e=1}^{n} y_e$

$\sum_{e \in S_j} y_e \leq c_j \quad \forall j$

$y_e \geq 0$

Define $y' = \dfrac{1}{H_g} y$ where $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} \sim \ln n$

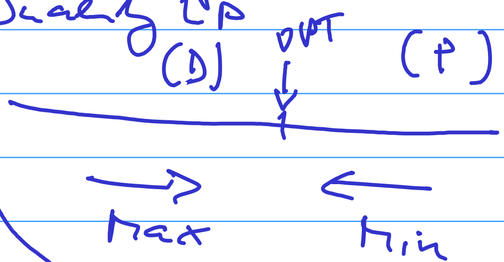
y' is a feasible solution for (D).

THM: GREEDY-SC is H_g-approximation algorithm.

Proof:

$$\sum_{j \in I} c_j = \sum_{e=1}^{n} y_e = H_g \sum_{e=1}^{n} y'_e \leq H_g \cdot OPT\_LP \leq H_g \cdot OPT$$

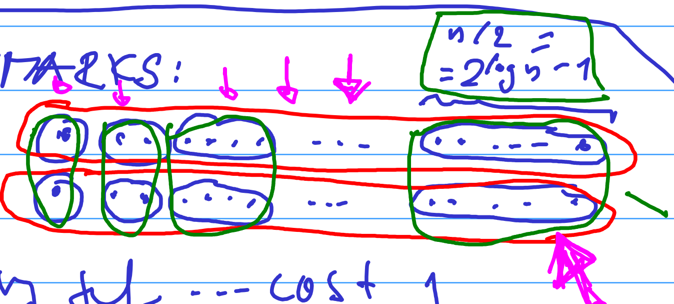
feasibility of y' & Duality LP



□

REMARKS:

Eg



every set ... cost 1

GREEDY ... $(\log n - 1)$ sets

OPT ... 2

ANALYSIS asymp. tight

THM (2014): For any $c<1$, there is no $(c\cdot \ln n)$-approx. algorithm for SET COVER, unless $P=NP$.

---

recall: $f = \max\limits_{i=\{1,\dots,n\}} |\{j \mid i\in S_j\}|$

$\boxed{\begin{array}{l}\text{PRIMAL LP-SC ALG.}\\ (1)\ \text{solve LP (P)} \dots x^*\\ (2)\ I=\{j \mid x_j^* \ge \frac{1}{f}\}\\ (3)\ \text{output } I\end{array}}$

THM: PRIMAL LP-SC is an f-approximation alg.

Proof: feasibility

as for each e: $\sum\limits_{j: e\in S_j} x_j \ge 1$

$\Rightarrow$ in feasible x^*, at least one set S_j has $x_j^* \ge \frac{1}{f}$

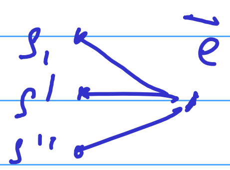
cost:

$$\sum_{j\in I} c_j \le f\cdot \sum_{j\in I} c_j\cdot x_j^* \le f\cdot LP\text{-}OPT \le f\cdot OPT$$

$\boxed{\begin{array}{l}\text{DUAL LP-SC ALG.}\\ (1)\ \text{solve LP (D)} \dots y^*\\ (2)\ I=\{j \mid \sum\limits_{e\in S_j} y_e^* = c_j\}\\ (3)\ \text{output } I\end{array}}$

THM: DUAL LP-SC is an f-approximation algorithm.

Proof: feasibility: assume that there is $\bar{e}\in\{1,\dots,n\}$ that is not covered by I



For each S_j containing $\bar{e}$: $\sum\limits_{e\in S_j} y_e^* < c_j$

let $\varepsilon = \min\limits_{j:\bar{e}\in S_j} (c_j - \sum\limits_{e\in S_j} y_e^*)$ Note: $\varepsilon>0$

Define $y'_e = \begin{cases} y^*_e & \text{for } e \neq \bar{e} \\ y^*_{\bar{e}} + \varepsilon \end{cases}$

$\therefore$ y' is a feasible solution for (D).

a contradiction!

· As $y'_{\bar{e}} > y^*_{\bar{e}}$, $\sum_{e=1}^{n} y'_e > \sum_{e=1}^{n} y^*_e$ (OPT)

cost: $\underline{\sum_{j\in I} c_j} = \sum_{j\in I}\sum_{e\in S_j} y^*_e = \sum_{e=1}^{n} \underbrace{|\{j \mid j\in I : e\in S_j\}|}_{\leq f} \cdot y^*_e$

$\leq f \left(\sum_{e=1}^{n} y^*_e\right) \leq \underline{f \cdot OPT}$ ∎

PRIMAL-DUAL SC ALG. Invariant: y is a feasible solution for (D)

(1) $y_e = 0$ $\forall e\in\{1,\dots,n\}$; $I := \emptyset$; $E := \emptyset$

E - covered elements

(2) while there exists $\bar{e} \notin E$ do

$$\varepsilon = \min_{j:\bar{e}\in S_j} \left(c_j - \sum_{e\in S_j} y_e\right)$$

$y_{\bar{e}} := y_e + \varepsilon$

for each j s.t. $\left(\bar{e}\in S_j \text{ and } \sum_{e\in S_j} y_e = c_j\right)$: $I := I \cup \{j\}$, $E := E \cup S_j$

(3) OUTPUT I

THM: Primal-Dual SC alg. is an f-approximation alg.

Proof: · time — in each iteration, at least one new element covered $\Rightarrow$ poly

· feasibility of I: as in the previous analysis, from the description of the alg.

· cost $\underline{\sum_{j\in I} c_j} = \sum_{j\in I}\sum_{e\in S_j} y_e \leq f \cdot \sum_{e=1}^{n} y_e \leq f \cdot LP\_OPT \leq \underline{f \cdot OPT}$ ∎

INDEPENDENT SET PROBLEM

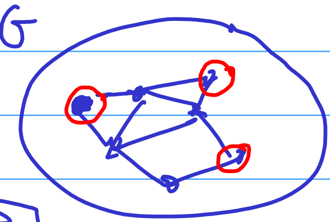
$G=(V,E)$: find $I \subset V$ s.t. I is ind. set

MAXIMUM IND. SET PROBLEM - find I of maximum size.

It holds: $\forall \varepsilon > 0$, there is no $n^{1-\varepsilon}$-approximation algorithm for the maximum ind. set problem, unless P=NP.

MAXIMAL IND. SET - an ind. set that cannot be extended

G



MODEL:

PRAM --- Parallel RAM

processors $P_1, \ldots, P_k$

each local memory

global memory

$\boxed{P_1}$ $\boxed{P_2}$ - - $\boxed{P_k}$

$\boxed{\text{GLOBAL MEM.}}$

d_v ... degree of v

synchronized computation:

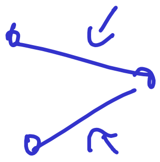
each step: global read

local computation

global write (no conflicts)

PARAL - MAXIMAL IS

1. $I := \emptyset$; $V' := V$; $G' := G$
2. while $(V' \neq \emptyset)$ do
3. in parallel for $\forall v \in V'$: if $d_v = 0$, then $I := I \cup \{v\}$
$V' := V' \setminus \{v\}$
4. in parallel for $\forall v \in V'$ mark v with prob $\frac{1}{2d_v}$
(independ. for different vertices & iterations)
5. in parallel for $\forall \{u,v\} \in E(G')$: if both u,v marked, then unmark the lower degree vertex of u,v



(break ties arbitrarily)

let S be the set of marked vertices and $N(S)$ be their neighbours

6. $I := I \cup S$; $V' := V' \setminus (S \cup N(S))$; $G' := G[V']$ induced by V'

7. OUTPUT I

Let $G_j = (V_j, E_j)$ --- the graph G' after iteration j.