Lecture 13: Facility location Problems (Cont'd).

Last time: local search for k-median, constant approx.

Today: LP rounding algorithms for facility location / k-median.

recall: same setup (almost) for both probs.

metric (V,d). clients $C \subseteq V$

fac loc: open $F \subseteq V$, cost $\sum_{i \in F} f_i + \sum_{j \in C} d(j,F)$

kmed: open $|F| = k$, $F \subseteq V$ cost $= \sum_{j \in C} d(j,F)$.

Local search for fac. loc can also give $O(1)$ approx.

- Today see LP-based solutions that do better, also more versatile. and give better apx. (comparable to LP relaxation as opposed to OPT integer sol^n).

Focus on Fac.Loc for now.

How to solve?

1. Cast as set cover. Each set given by a center and some subset of clients to serve. ~~[illegible]~~ See Exercise in HW1.

   gets $O(\log n)$ approx!

Better? Yes.

Today: write an LP. round it (relax-and-round again).

$$\text{LP:}\quad \boxed{\begin{aligned} \min\ & \sum_i f_i y_i + \sum d_{ij} x_{ij} \\ \text{st}\ & \sum_i x_{ij} = 1 \quad \forall j \in C \\ & x_{ij} \le y_i \quad \forall ij \in V \\ & x_{ij}, y_i \ge 0. \end{aligned}}$$

Check: ILP version is exact formulation of the fac loc problem!

Thm: Can round and get solution $F \subseteq V$ st. $\text{cost}(F) = \sum_{i \in F} f_i + \sum_{j \in C} d(j, F)$

$\le \underbrace{\text{constant}}_{\text{best known: } 1.488 \text{ or so.}} \cdot$ LP value.

Thm: Integrality gap is at least 1.463

our small gap here. But the focus will be on the alg. ideas, not the actual numerical values.

In previous applications we've not considered the dual program, but that gives a lot of information as well.

<u>Dual</u>: $\max \sum_{j \in C} \alpha_j$ ← dual variable per client

dual variable per client-facility pair

$$\text{s.t.} \quad \alpha_j - \beta_{ij} \le d_{ij} \qquad \forall ij \quad i \in V\ j \in C$$

$$\sum_j \beta_{ij} \le f_i \qquad \forall i \in V.$$

$$\beta_{ij} \ge 0, \quad \alpha_i \text{ unconstrained.}$$

↑ but really set $\alpha_i = \min_{j \in C} \{ d_{ij} + \beta_{ij} \}$ (since max problem) so nonnegative.

We will interpret this LP in greater detail soon, but for right now, ~~not~~ recall basic facts about LP duality.

---

e.g.

$$(P)\quad \begin{array}{l} \min c^T x \\ Ax \ge b \\ x \ge 0 \end{array} \quad \Rightarrow \quad (D)\quad \begin{array}{l} \max b^T y \\ A^T y \le c \\ y \ge 0 \end{array}$$

<u>Thm</u> (Weak duality) If x, y are feasible LP solutions to (P), (D) then $c^T x \ge b^T y$.

<u>Pf</u> $c^T x \underset{\substack{\uparrow \\ \text{dual } \cancel{\text{primal}} \text{ feas} \\ \text{and } x \ge 0}}{\ge} (A^T y)^T x = y^T A x \underset{\substack{\uparrow \\ \text{primal feasibility} \\ \text{and } y \ge 0.}}{\ge} y^T b.$ ☺

<u>Thm</u> (Strong duality). If ∃ feasible primal & dual solutions, then ∃ x^*, y^* optimal feasible, ~~st feasible~~ ~~and~~ s.t. $c^T x^* = b^T y^*$. (equality).

Not giving proof for now, see, e.g. Schrijver or Matoušek - Gärtner or ....

<u>Corollary</u>: (Complementary slackness). for x^*, y^* optimal primal/dual solns.

$$x^T (A^T y - c) = 0 \quad \text{and} \quad y^T (Ax - b) = 0.$$

<u>Pf</u>: both inequalities here must be tight if $c^T x = b^T y$. ☺

In other words, if a dual variable is non-zero then the corresponding primal ~~a~~ constraint is tight.

non zero dual ⇒ constraint is "important"

(and the same for primal var/dual constraint)

each var in dual corresps to constraint in primal

As example: sps (x,y), (α,β) optimal solutions

$x_{ij} > 0$ then $\alpha_j - \beta_{ij} = d_{ij}$ ⇒ $\alpha_j = \beta_{ij} + d_{ij}$

BTW: other implication of ~~strong~~ weak duality.

Any dual solution ≤ optimal dual LP solⁿ ≤ optimal primal LP solⁿ ≤ optimal ~~p~~ ILP solⁿ before ~~re~~ relaxing.

dual is max problem for us

weak duality

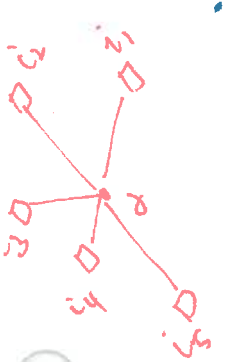
LP is relaxation of IP

⇒ if we can show that our solution ALG has cost $\leq \alpha \cdot$ dual,
$\leq \alpha \cdot OPT$.

$\boxed{\text{"dual" is like an accounting device.}}$

So suffices to show: ∃ solution to facility location (and dual solution) with cost ≤ c. Dual value.

- Solve the LP, and dual optimally. (x,y) (α,β) solutions.



- For each client $j \in C$,

  define its "~~neighbors~~ cluster" $N(j) = \{ i \in V \mid x_{ij} > 0 \}$.

  potential facilities that LP is sending j to fractionally

  Fact: if $i \in N(j)$ then $\alpha_j = \beta_{ij} + d_{ij}$.

- What if we open the cheapest of these facilities and send j there?

disjoint clusters

Just for a moment, ~~assume~~ consider some ~~disjoint~~ set C' of clients s.t. $N(j) \cap N(j') = \emptyset$ for $j, j' \in C'$

then opening these cheapest facilities in each cluster gives cost

$$\sum_{j \in C'} (\text{cost of cheapest facility in } N(j) + \text{dist of } j \text{ to this facility}).$$

Hmm.... how to account for this cost? ~~[illegible]~~

~~Instead sps~~ Sps $i(j)$ is facility cheapest in $N(j)$. then

$$f_{i(j)} \underset{\text{min} \leq \text{average}}{\leq} \sum_{i \in N(j)}' f_i x_{ij} \underset{x_{ij} \leq y_i \text{ constraint}}{\leq} \sum_{i \in N(j)}' f_i y_i = \text{LP cost inside that cluster.}$$

$\Rightarrow$ b/c clusters are disjoint,

$$\sum f_{i(j)} \leq \text{LP cost for opening facilities.}$$

What about connection costs?

if $i \in N(j)$ then $\alpha_j = \beta_{ij} + d_{ij} \geq d_{ij}$ ($\beta_{ij} \geq 0$.) $\Rightarrow d_{ij} \leq \alpha_j$

connection cost of j — dual contribution of j

$\Rightarrow$ for centers of these clusters, ~~a connection cost of~~ ~~[illegible]~~
total connection cost $\leq$ their dual contribution.

But how to do this when clusters are <u>not</u> <u>disjoint</u> ??

- Pick an independent / disjoint set of clusters of "small" radius.
- ~~[illegible]~~ Open cheapest facility within each selected cluster (has low cost, already seen)
- Show that rerouting clients of unselected clusters can be sent to these open facilities also with small cost.

Algorithm: Solve LP and dual.

- Sort clients st $\alpha_1 \le \alpha_2 \le \dots \le \alpha_n$.
- $S \leftarrow \emptyset$.
- for $j = 1$ to n

  if $N(j)$ is disjoint from all $N(j')$, $j' \in S$

  $S \leftarrow S \cup \{j\}$.

  open cheapest facility in $N(j)$. ← call this set F of open facilities

---

disjoint by construction

So open cheapest facilities in $\{N(j)\}_{j \in S}$.

- facility costs $= \sum_{j \in S} \min_{i \in N(j)} f_i \le \sum_{j \in S} \sum_{i \in N(j)} f_i x_{ij} \le \sum_{ij} f_{ij} x_{ij}$

  min ≤ avg

  $N(j)$ disjoint

  $\le$ LP primal

- connection costs for $j \in S$.

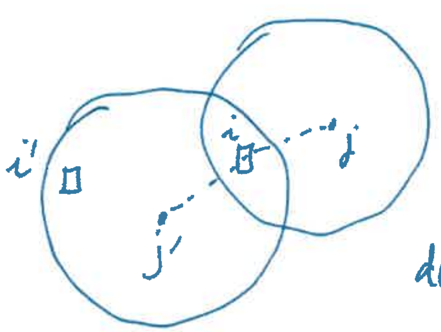
  $d_{ij} \le \alpha_j$ because distance from j to F

  $\le$ distance from j to open fac in $N(j)$

  $\le \alpha_j$. (as above).

- What about $j \notin S$. must be b/c $N(j) \cap N(j') \neq \emptyset$ for $j' \in S$

  say i lies in



and cheapest fac i' in $N(j')$ opened

Δ inequality

$$d(j,i') \le d(j,i) + d(j',i) + d(j',i')$$

$$\le \alpha_j + \alpha_{j'} + \alpha_{j'} \le 3\alpha_j$$

by comp slackness.

sorted order

Overall: Facility opening cost $\leq$ LP value

Connection cost $\leq$ 3. Dual value $\leq$ 3.LP value

$\Rightarrow$ 4-apx

Moral of Story: properties of dual solution / relationship to primal allowed us to get quick apx-algo for Facility Location

- Deterministic rounding
- Clustering ensured that facility opening costs small.

  Sines ensured that connection costs for non-cluster-centers small.

— x —

What is the integrality gap? Showed upper bound of 4. Better? Yes.

Here's a $\boxed{\text{randomized rounding}}$ idea.

- Having picked a set S of cluster centers, open one facility randomly.

  ie $\forall j \in S$, Pick $i \in N(j)$ wp x_{ij} (recall $\sum_i x_{ij} = 1$ so prob. distrib).

  so pick one facility in each such set $N(j)$, $j \in S$.

- For all i ~~$\notin \bigcup_{j \in S} N(j)$~~, pick indep wp $y_i - \sum_{j \in S} x_{ij}$

- $\mathbb{E}[\text{facility cost}] = \sum_{j \in S} \sum_{i \in N(S)} \mathbb{E}[f_i] = \sum_{j \in S, i} f_i x_{ij} \leq$ LP value.

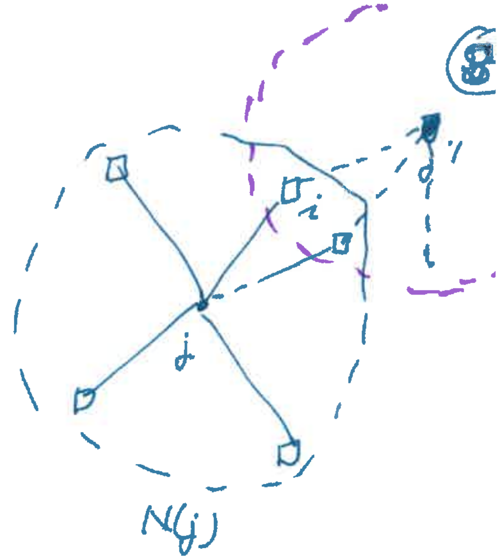
- Again distance of each $j \in S$ to closest fac $\leq \alpha_j$ = its dual contribution.

What about $j \notin S$.

Since $j \notin S$, $\exists\, j' \in S$ s.t. $N(j') \cap N(j) \neq \emptyset$

and $\alpha_{j'} \leq \alpha_j$

maybe 1 maybe more ~~facilities in~~ points in here



<u>Idea</u>: if $\exists$ a facility open in $N(j)$, can just go there, else go to open facility in $N(j')$.

<u>Show</u>: ~~$\mathbb{E}$[cost of this fac | such fac exists] $\leq \alpha_j$~~

• ~~cost~~ $\Pr[\text{such fac exists}] \geq (1 - 1/e)$ (show soon)

• $d(\text{such fac to } j) \leq \alpha_j$

cost $\leq 3\alpha_j$ with remaining prob.

$$\Rightarrow \mathbb{E}[\text{cost}] \leq \alpha_j (1 - 1/e) + 3\alpha_j (1/e) = (1 + 2/e)\alpha_j$$

$\Rightarrow$ overall connection cost (in expectation) $\leq (1 + 2/e) \sum_j' \alpha_j = (1 + 2/e)$ dual value

$$\Rightarrow \text{Integrality gap} \leq \underbrace{1}_{\text{facility cost.}} + \underbrace{(1 + 2/e)}_{\text{connection cost}} \leq 2.7\text{something}$$

——— x ———

Can save the "1 + .." loss as well to get $(1 + 2/e)$. See the [WS10] book.

And more ideas, but not today.

Let's see a different "primal-dual" way to use duality next time.