CO327 Deterministic OR Models (2021-Spring)

p-median problem and introduction to mixed integer program

Instructor: Andersen Ang
Guest lecturer: Mariia Sobchuk msobchuk@uwaterloo.ca

Combinatorics and Optimization, U.Waterloo, Canada
ms**x**ang@uwaterloo.ca, where $\mathbf{x} = \lfloor \pi \rfloor$
Homepage: angms.science

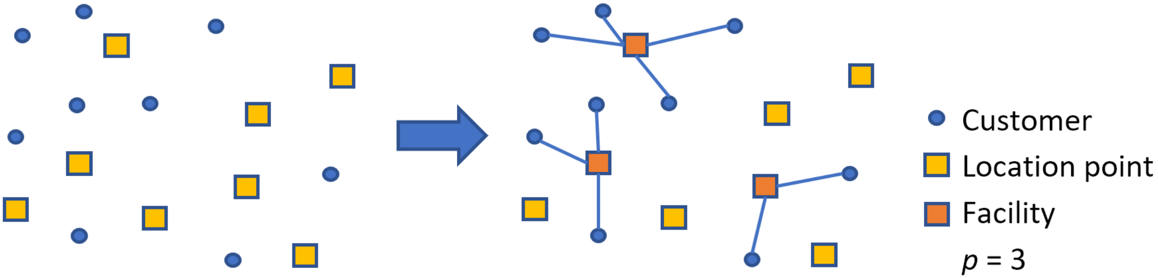
First draft: April 14, 2021 Last update: June 10, 2021

What is the p-median problem

- A facility location problem
- A mixed integer program
  - some variables are real number
  - some variables are integer
- A NP-hard problem: very difficult to solve efficiently for large problem size.
- What is the problem: determine where to place the facilities to minimize the total transportation cost of serving all customers.
- Related to previous lectures :
  - the lecture on "Tower defence".
  - the lecture on "Set covering, location decision and urban planning".

Problem setting

- There are n customers who are geographically apart.
- We want to determine the locations of p number of facilities to serve these customers.
- Each facility is assumed to have infinite service capacity. That is, we can simply assign each customer to exactly one facility. Our objective is to minimize the weighted sum of transportation cost to serve all customers.



Problem setting

You are given

- $\mathcal{I} = \{1, 2, \ldots, m\}$, the set of candidate locations for facilities
- $\mathcal{J} = \{1, 2, \ldots, n\}$, the set of customer locations
- p, the number of facilities to introduce
  Choose p out of m: $\frac{m!}{p!(m-p)!}$, NP-hard
- $d_1, d_2, \ldots, d_j$, the size of demand at customer $j \in \mathcal{J}$
- c_{ij}, the transportation cost from facility location $i \in \mathcal{I}$ to the customer location $j \in \mathcal{J}$

Model this problem as a Mixed Integer Program (MIP).

How to model: formulate the decision variables

- Identify the goal: "minimizing the total transportation cost from facility location i to the customer location"
- There are TWO decisions:
  - Which customer to go to WHICH facility
  - HOW MUCH the demand of the customer is served by the facility.
- Propose TWO decision variables
  - x_{ij}: the fraction of demand from customer $j \in \mathcal{J}$ being served by facility location $i \in \mathcal{I}$.
    - A continuous variable.
    - Fraction: $0 \leq x_{ij} \leq 1$
    - Note: total number of demand of each customer is d_j
  - y_i: location variable, $y_i = 1$ if a facility is introduced at candidate location $i \in \mathcal{I}$ and $y_i = 0$ otherwise.
    - $y_i = \{0, 1\}$
    - Binary decision.

How to model: formulate the cost

- x_{ij}: the fraction of demand from customer $j \in \mathcal{J}$ being served by facility location $i \in \mathcal{I}$
- c_{ij}: the unit transportation cost from facility location $i \in \mathcal{I}$ to the customer location $j \in \mathcal{J}$
- $c_{ij}x_{ij}$: the transportation cost from facility location $i \in \mathcal{I}$ to the customer location $j \in \mathcal{J}$ for serving the fraction of demand
- d_j, the size of demand at customer $j \in \mathcal{J}$
- $d_j c_{ij} x_{ij}$: the transportation cost from facility location $i \in \mathcal{I}$ to the customer location $j \in \mathcal{J}$ for serving all the demand
- Total cost between all customer and all facility location on full demand

$$\sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} d_j c_{ij} x_{ij}$$

How to model: formulate the constraints

- Demand served cannot be negative

$$x_{ij} \geq 0 \text{ for all } i \in \mathcal{I}, j \in \mathcal{J}$$

- All facilities together has to satisfy all the demand of customer j

$$\sum_{i \in \mathcal{I}} x_{ij} = 1$$

- There are totally p facilities

$$\sum_{i \in \mathcal{I}} y_i = p.$$

- x cannot over y (demand served $\leq$ supply limit)

$$x_{ij} \leq y_i$$

The p-median problem

$$
\begin{aligned}
\min \quad & \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} d_j c_{ij} x_{ij} \\
\text{s.t.} \quad & x_{ij} \geq 0 \text{ for all } i \in \mathcal{I}, j \in \mathcal{J} \\
& y_i \in \{0, 1\} \text{ for all } i \\
& \sum_{i \in \mathcal{I}} x_{ij} = 1 \text{ for all } \cancel{i \in \mathcal{I}}, j \in \mathcal{J} \\
& \sum_{i \in \mathcal{I}} y_i = p \\
& x_{ij} \leq y_i \text{ for all } i \in \mathcal{I}, j \in \mathcal{J}
\end{aligned}
$$

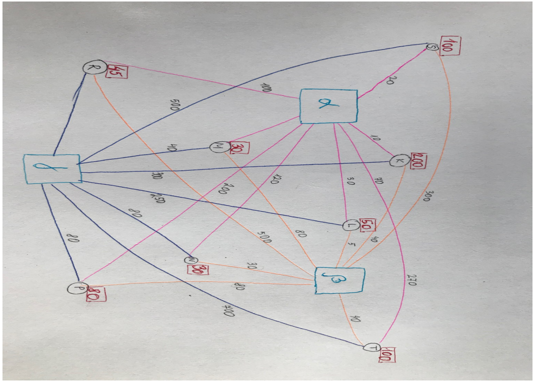
A linear mixed integer program !

Q: what does the last constraint do?

The p-median problem: why not this formulation?

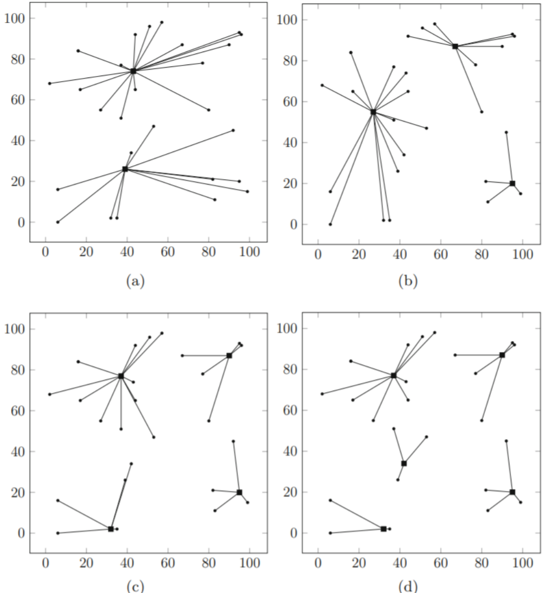
$$
\begin{aligned}
\min \quad & \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} d_j c_{ij} x_{ij} \\
\text{s.t.} \quad & x_{ij} \geq 0 \text{ for all } i \in \mathcal{I}, j \in \mathcal{J} \\
& y_i \in \{0, 1\} \text{ for all } i \\
& \sum_{i \in \mathcal{I}} x_{ij} = 1 \text{ for all } i \in \mathcal{I}, j \in \mathcal{J} \\
& \sum_{i \in \mathcal{I}} y_i = p \\
& \sum_{j} x_{ij} \leq |\mathcal{J}| y_i \text{ for all } i
\end{aligned}
$$

Sample 2-median problem



$$
\begin{aligned}
\text{min} \quad & 100 \cdot 20 x_{\alpha S} + 200 \cdot 10 x_{\alpha K} + \ldots \\
\text{s.t.} \quad & x_{\alpha S} \geq 0, x_{\alpha K} \geq 0, \ldots. \\
& y_{\alpha}, y_{\boldsymbol{\eta}}, y_{\gamma} \in \{0, 1\} \\
& \sum_{i \in \mathcal{I}} x_{ij} = 1 \text{ for all } i \in \mathcal{I}, j \in \mathcal{J} \\
& y_{\alpha} + y_{\boldsymbol{\eta}} + y_{\gamma} = 2 \\
& x_{\alpha S} \leq y_{\alpha}, x_{\alpha K} \leq y_{\alpha}, \ldots
\end{aligned}
$$

Examples of other optimal solutions



Special case: the 1-median problem

$$
\begin{aligned}
\min \quad & \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} d_j c_{ij} x_{ij} \\
\text{s.t.} \quad & x_{ij} \geq 0 \text{ for all } i \in \mathcal{I}, j \in \mathcal{J} \\
& y_i \in \{0, 1\} \text{ for all } i \\
& \sum_{i \in \mathcal{I}} x_{ij} = 1 \text{ for all } \sout{i \in \mathcal{I}}, j \in \mathcal{J} \\
& \sum_{i \in \mathcal{I}} y_i = 1 \\
& x_{ij} \leq y_i \text{ for all } i \in \mathcal{I}, j \in \mathcal{J}
\end{aligned}
$$

- $p = 1$: only one facility.
- Application of this model: solve location problem for things that at most 1 can be build
  - Decide where to place the capital
  - Decide where to live (in this case the "facility" is your home)

Some applications of p-median problem

- Decide where to live
- Facility location
- Clustering
- Telecommunication industry: antenna location setup
- Politics and government administration
- Vehicle routing
- Computer network

Problem twists (more in assignment 3)

- We assumed all facilities have infinite capacity.
  What if this is not true? $\rightarrow$ capacited p-median problem
- What if each location can build more than one facility?
- What if a demand can only be serviced by one facility?
  For example, you hate to go to a specific hospital.
- What if we want to minimize the "maximum distance between the customer and facility instead"?

Capacitated p-median problem

- u_i: the upper bound on the capacity that facility i can fulfill
- binary y_j indicating if a facility is opened at location $j \in \mathcal{J}$
- x_{ij} fraction of demand of customer $j \in \mathcal{J}$ served by facility $i \in \mathcal{I}$
- c_{ij}, d_j

How to modify out p-median problem?

- Each facility cannot serve above its capacity

$$\sum_{j \in J} d_j x_{ij} \leq u_i y_i$$

The formulation of capacitated p-median problem

$$
\begin{aligned}
\min \quad & \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} d_j c_{ij} x_{ij} \\
\text{s.t.} \quad & x_{ij} \geq 0 \text{ for all } i \in \mathcal{I}, j \in \mathcal{J} \\
& y_i \in \{0, 1\} \text{ for all } i \\
& \sum_{i \in \mathcal{I}} x_{ij} = 1 \text{ for all } \sout{i \in \mathcal{I},}\ j \in \mathcal{J} \\
& \sum_{i \in \mathcal{I}} y_i = p \text{ for all } i \\
& \sum_{j \in \mathcal{J}} d_j x_{ij} \leq u_j y_j \text{ for all } i \in \mathcal{I}, j \in \mathcal{J}
\end{aligned}
$$

What variant of the problem does this program model?

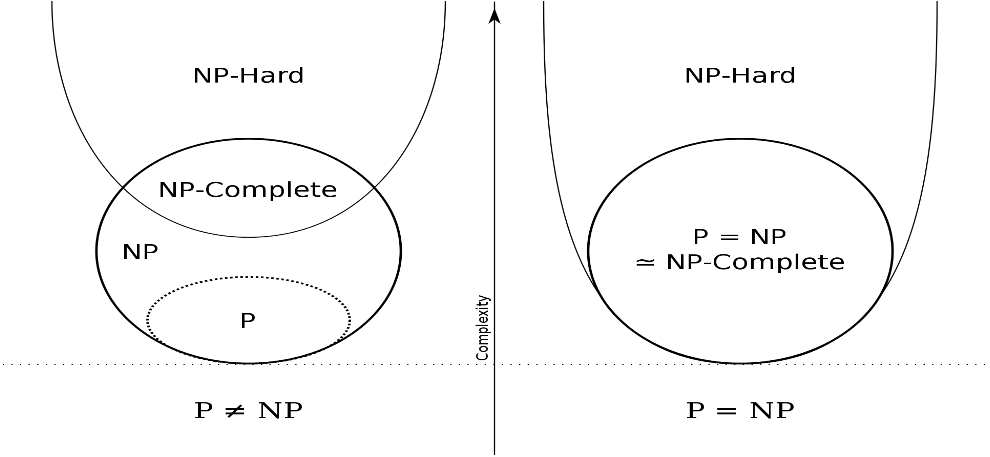
$$
\begin{aligned}
\min \quad & \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} d_j c_{ij} x_{ij} \\
\text{s.t.} \quad & x_{ij} \in \{0,1\} \text{ for all } i, j \\
& y_i \in \{0,1\} \text{ for all } i \\
& \sum_{i \in \mathcal{I}} x_{ij} = 1 \text{ for all } i, j \\
& \sum_{i \in \mathcal{I}} y_i = p \text{ for all } i \\
& \sum_{j \in \mathcal{J}} d_j x_{ij} \leq u_j y_j \text{ for all } i \in \mathcal{I}, j \in \mathcal{J}
\end{aligned}
$$

Binary capacitated p-median problem

- Demand of each customer must be served by a unique facility
- u_i, x_{ij}, c_{ij}, d_j

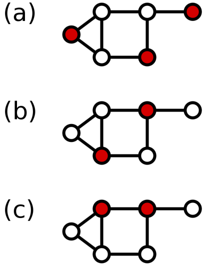
NP-hardness of the problem

- A problem is in **NP** if a solution is polynomial-time verifiable.



Problem is **NP-hard** if an algorithm for solving it can be used to solve any problem in NP with at most polynomial overhead.

NP-complete dominating set problem



Problem is **NP-complete** it is in NP and any other problem in NP can be reduced to it in polynomial time.

The Dominating set problem: Given a graph $G = (V, E)$ and a positive integer $1 \leq p \leq |V(G)|$, does there exist a subset V_p^* of p vertices, such that each vertex of G is either in V_p^* or is adjacent to a vertex in V_p^*?

[Garey and Johnson, 1979]: Let G be the graph that is planar of maximum degree 3 and $1 < p < n$. Finding if there is a dominating-set of cardinally p is NP-complete.

Reduction of p-median problem to Dominating set problem

- p-median problem is NP-hard even when the graph is planar of maximum degree 3, with all edges of weight 1 and all demands 1.
- Can show that on such graph a problem of determining if there exists a dominating set of size p is polynomial time reducible to the problem of finding a p-median in G.
  - There exists a dominating set of cardinality p in G if and only if the otimal solution to the p-median problem gives objective value $n - p$.