# 15.093: Optimization Methods

## Lecture 12: Discrete Optimization

# 1 Todays Lecture

SLIDE 1

- Modeling with integer variables
- What is a good formulation?
- Theme: The Power of Formulations

# 2 Integer Optimization

## 2.1 Mixed IO

SLIDE 2

$$
\begin{array}{lll}
\text{(MIO)} & \max & \boldsymbol{c}'\boldsymbol{x} + \boldsymbol{h}'\boldsymbol{y} \\
& \text{s.t.} & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \leq \boldsymbol{b} \\
& & \boldsymbol{x} \in Z_+^n (\boldsymbol{x} \geq 0, \boldsymbol{x} \text{ integer}) \\
& & \boldsymbol{y} \in R_+^m (\boldsymbol{y} \geq 0)
\end{array}
$$

## 2.2 Pure IO

SLIDE 3

$$
\begin{array}{lll}
\text{(IO)} & \max & \boldsymbol{c}'\boldsymbol{x} \\
& \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\
& & \boldsymbol{x} \in Z_+^n
\end{array}
$$

Important special case: Binary Optimization

$$
\begin{array}{lll}
\text{(BO)} & \max & \boldsymbol{c}'\boldsymbol{x} \\
& \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\
& & \boldsymbol{x} \in \{0,1\}^n
\end{array}
$$

## 2.3 LO

SLIDE 4

$$
\begin{array}{lll}
\text{(LO)} & \max & \boldsymbol{c}'\boldsymbol{x} \\
& \text{s.t.} & \boldsymbol{B}\boldsymbol{y} \leq \boldsymbol{b} \\
& & \boldsymbol{y} \in R_+^n
\end{array}
$$

# 3 Modeling with Binary Variables

## 3.1 Binary Choice

SLIDE 5

$$
x \in \begin{cases} 1, & \text{if event occurs} \\ 0, & \text{otherwise} \end{cases}
$$

Example 1: IO formulation of the knapsack problem

SLIDE 6

$n$ : projects, total budget $b$
$a_j$ : cost of project $j$
$c_j$ : value of project $j$

$$
x_j = \begin{cases} 1, & \text{if project } j \text{ is selected.} \\ 0, & \text{otherwise.} \end{cases}
$$

$$
\begin{aligned}
\max \quad & \sum_{j=1}^{n} c_j x_j \\
\text{s.t.} \quad & \sum a_j x_j \leq b \\
& x_j \in \{0,1\}
\end{aligned}
$$

## 3.2 Modeling relations

SLIDE 7

- At most one event occurs

$$\sum_j x_j \leq 1$$

- Neither or both events occur

$$x_2 - x_1 = 0$$

- If one event occurs then, another occurs

$$0 \leq x_2 \leq x_1$$

- If $x = 0$, then $y = 0$; if $x = 1$, then $y$ is uncontrained

$$0 \leq y \leq Ux, \qquad x \in \{0,1\}$$

## 3.3 The assignment problem

SLIDE 8

$$
\begin{aligned}
n \quad & \text{people} \\
m \quad & \text{jobs} \\
c_{ij}: \quad & \text{cost of assigning person } j \text{ to job } i. \\
x_{ij} \quad & = \begin{cases} 1 & \text{person } j \text{is assigned to job } i \\ 0 & \end{cases} \\
\min \quad & \sum c_{ij} x_{ij} \\
\text{s.t.} \quad & \sum_{j=1}^{n} x_{ij} = 1 \quad \text{each job is assigned} \\
& \sum_{i=1}^{m} x_{ij} \leq 1 \quad \text{each person can do at most one job.} \\
& x_{ij} \in \{0,1\}
\end{aligned}
$$

# 4 What is a good formulation?

## 4.1 Facility Location

SLIDE 9

- Data

$$
\begin{aligned}
& N = \{1 \dots n\} && \text{potential facility locations} \\
& I = \{1 \dots m\} && \text{set of clients} \\
& c_j: && \text{cost of facility placed at } j \\
& h_{ij}: && \text{cost of satisfying client } i \text{ from facility } j.
\end{aligned}
$$

- Decision variables

$$x_j = \begin{cases} 1, & \text{a facility is placed at location } j \\ 0, & \text{otherwise} \end{cases}$$
$$y_{ij} = \text{fraction of demand of client } i \text{ satisfied by facility } j.$$

Slide 10

$$\begin{aligned} IZ_1 = \min \quad & \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} \sum_{j=1}^{n} h_{ij} y_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^{n} y_{ij} = 1 \\ & y_{ij} \leq x_j \\ & x_j \in \{0,1\}, 0 \leq y_{ij} \leq 1. \end{aligned}$$

Slide 11

Consider an alternative formulation.

$$\begin{aligned} IZ_2 = \min \quad & \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} \sum_{j=1}^{n} h_{ij} y_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^{n} y_{ij} = 1 \\ & \sum_{i=1}^{m} y_{ij} \leq m \cdot x_j \\ & x_j \in \{0,1\}, 0 \leq y_{ij} \leq 1. \end{aligned}$$

Are both valid?
Which one is preferable?

## 4.2 Observations

Slide 12

- $IZ_1 = IZ_2$, since the integer points both formulations define are the same.
-

$$P_1 = \left\{ (\boldsymbol{x}, \boldsymbol{y}) : \sum_{j=1}^{n} y_{ij} = 1, y_{ij} \leq x_j, \begin{array}{l} 0 \leq x_j \leq 1 \\ 0 \leq y_{ij} \leq 1 \end{array} \right\}$$

$$P_2 = \left\{ (\boldsymbol{x}, \boldsymbol{y}) : \sum_{j=1}^{n} y_{ij} = 1, \sum_{i=1}^{m} y_{ij} \leq m \cdot x_j, \begin{array}{l} 0 \leq x_j \leq 1 \\ 0 \leq y_{ij} \leq 1 \end{array} \right\}$$

Slide 13

- Let

$$Z_1 = \min_{(\boldsymbol{x},\boldsymbol{y}) \in P_1} \boldsymbol{cx} + \boldsymbol{hy}, \qquad Z_2 = \min_{(\boldsymbol{x},\boldsymbol{y}) \in P_2} \boldsymbol{cx} + \boldsymbol{hy}$$

- $Z_2 \leq Z_1 \leq IZ_1 = IZ_2$

## 4.3 Implications

SLIDE 14

- Finding $IZ_1 (= IZ_2)$ is difficult.
- Solving to find $Z_1, Z_2$ is a LOP. Since $Z_1$ is closer to $IZ_1$ several methods (branch and bound) would work better (actually much better).
- Suppose that if we solve $\min \boldsymbol{cx} + \boldsymbol{hy}, (\boldsymbol{x}, \boldsymbol{y}) \in P_1$ we find an integral solution. Have we solved the facility location problem?

SLIDE 15

- Formulation 1 is better than Formulation 2. (Despite the fact that 1 has a larger number of constraints than 2.)
- What is then the criterion?

## 4.4 Ideal Formulations

SLIDE 16

- Let $P$ be a linear relaxation for a problem
- Let
$$H = \{(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{x} \in \{0,1\}^n\} \cap P$$
- Consider Convex Hull (H)
$$= \{\boldsymbol{x} : \boldsymbol{x} = \sum_i \lambda_i x^i, \sum_i \lambda_i = 1, \lambda_i \geq 0, x^i \in H\}$$

SLIDE 17

- The extreme points of $CH(H)$ have $\{0,1\}$ coordinates.
- So, if we know $CH(H)$ explicitly, then by solving $\min \boldsymbol{cx} + \boldsymbol{hy}, (\boldsymbol{x}, \boldsymbol{y}) \in CH(H)$ we solve the problem.
- Message: Quality of formulation is judged by closeness to $CH(H)$.
$$CH(H) \subseteq P_1 \subseteq P_2$$

# 5 Minimum Spanning Tree (MST)

SLIDE 18

- How do telephone companies bill you?
- It used to be that rate/minute: Boston $\rightarrow$ LA proportional to distance in MST
- Other applications: Telecommunications, Transportation (good lower bound for TSP)

SLIDE 19

- Given a graph $G = (V, E)$ undirected and Costs $c_e$, $e \in E$.
- Find a tree of minimum cost spanning all the nodes.
- Decision variables $x_e = \begin{cases} 1, & \text{if edge } e \text{ is included in the tree} \\ 0, & \text{otherwise} \end{cases}$

SLIDE 20

- The tree should be connected. How can you model this requirement?
- Let $S$ be a set of vertices. Then $S$ and $V \setminus S$ should be connected
- Let $\delta(S) = \{e = (i, j) \in E : \begin{matrix} i \in S \\ j \in V \setminus S \end{matrix} \}$
- Then,

$$\sum_{e \in \delta(S)} x_e \geq 1$$

- What is the number of edges in a tree?
- Then, $\sum_{e \in E} x_e = n - 1$

## 5.1 Formulation

SLIDE 21

$$IZ_{MST} = \min \sum_{e \in E} c_e x_e$$

$$H \left\{ \begin{array}{ll} \sum_{e \in \delta(S)} x_e \geq 1 & \forall\, S \subseteq V, S \neq \emptyset, V \\ \sum_{e \in E} x_e = n - 1 & \\ x_e \in \{0, 1\}. & \end{array} \right.$$

Is this a good formulation?

SLIDE 22

$$P_{cut} = \{\boldsymbol{x} \in R^{|E|} : \ 0 \leq \boldsymbol{x} \leq \boldsymbol{e},$$
$$\sum_{e \in E} x_e = n - 1$$
$$\sum_{e \in \delta(S)} x_e \geq 1 \ \forall\, S \subseteq V, S \neq \emptyset, V\}$$

Is $P_{cut}$ the $CH(H)$?

## 5.2 What is $CH(H)$?

SLIDE 23

Let

$$P_{sub} = \{\boldsymbol{x} \in R^{|E|} : \sum_{e\in E} x_e = n-1$$

$$\sum_{e\in E(S)} x_e \leq |S| - 1 \;\forall\; S \subseteq V,\; S \neq \emptyset, V\}$$

$E(S) = \left\{ e = (i,j) : \begin{array}{l} i \in S \\ j \in S \end{array} \right\}$

Why is this a valid IO formulation?

SLIDE 24

- Theorem: $P_{sub} = CH(H)$.
- $\Rightarrow P_{sub}$ is the best possible formulation.
- MESSAGE: Good formulations can have an exponential number of constraints.

# 6 The Traveling Salesman Problem

SLIDE 25

Given $G = (V,E)$ an undirected graph. $V = \{1,\ldots,n\}$, costs $c_e \;\forall\; e \in E$. Find a tour that minimizes total length.

## 6.1 Formulation I

SLIDE 26

$x_e = \begin{cases} 1, & \text{if edge } e \text{ is included in the tour.} \\ 0, & \text{otherwise.} \end{cases}$

$$\begin{array}{ll} \min & \sum_{e\in E} c_e x_e \\ \text{s.t.} & \sum_{e\in\delta(S)} x_e \geq 2, \quad S \subseteq E \\ & \sum_{e\in\delta(i)} x_e = 2, \quad i \in V \\ & x_e \in \{0,1\} \end{array}$$

## 6.2 Formulation II

SLIDE 27

$$\begin{array}{ll} \min & \sum c_e x_e \\ \text{s.t.} & \sum_{e\in E(S)} x_e \leq |S| - 1, \quad S \subseteq E \\ & \sum_{e\in\delta(i)} x_e = 2, \quad i \in V \\ & x_e \in \{0,1\} \end{array}$$

SLIDE 28

$$
\begin{aligned}
P_{cut}^{TSP} = & \{x \in R^{|E|}; \sum_{e \in \delta(S)} x_e \geq 2, \sum_{e \in \delta(i)} x_e = 2 \\
& 0 \leq x_e \leq 1\} \\
P_{sub}^{TSP} = & \{x \in R^{|E|}; \sum_{e \in \delta(i)} x_e = 2 \\
& \sum_{e \in \delta(S)} x_e \leq |S| - 1 \\
& 0 \leq x_e \leq 1\}
\end{aligned}
$$

SLIDE 29

- Theorem: $P_{cut}^{TSP} = P_{sub}^{TSP} \not\supseteq CH(H)$
- Nobody knows $CH(H)$ for the TSP

# 7 Minimum Matching

SLIDE 30

- Given $G = (V, E)$; $c_e$ costs on $e \in E$. Find a matching of minimum cost.
- Formulation:

$$
\begin{aligned}
\min \quad & \sum c_e x_e \\
\text{s.t.} \quad & \sum_{e \in \delta(i)} x_e = 1, \quad i \in V \\
& x_e \in \{0, 1\}
\end{aligned}
$$

- Is the linear relaxation $CH(H)$?

SLIDE 31

Let

$$
\begin{aligned}
P_{MAT} = & \{x \in R^{|E|} : \sum_{e \in \delta(i)} x_e = 1 \\
& \sum_{e \in \delta(S)} x_e \geq 1 \quad |S| = 2k + 1, S \neq \emptyset \\
& x_e \geq 0\}
\end{aligned}
$$

Theorem: $P_{MAT} = CH(H)$

# 8 Observations

SLIDE 32

- For MST, Matching there are efficient algorithms. $CH(H)$ is known.
- For TSP $\not\exists$ efficient algorithm. TSP is an $NP - hard$ problem. $CH(H)$ is not known.
- Conjuecture: The convex hull of problems that are polynomially solvable are explicitly known.

## 9 Summary

Slide 33

1. An IO formulation is better than another one if the polyhedra of their linear relaxations are closer to the convex hull of the IO.
2. A good formulation may have an exponential number of constraints.
3. Conjecture: Formulations characterize the complexity of problems. If a problem is solvable in polynomial time, then the convex hull of solutions is known.