6.859J/15.083J Integer Programming and Combinatorial Optimization

Professors: Dimitris Bertsimas,
Andreas Schulz,

1 Structure of Class

Slide 1

- Formulations, complexity and relaxations, Lec. 1-9
- Robust Discrete Optimization, Lec. 10-11
- Algebra and geometry of IO, Lec. 12-15
- Algorithms for IO, Lec. 16-23
- Mixed Integer Optimization, Lec. 24-25

2 Requirements

Slide 2

- Homeworks: 30%
- Midterm Exam: 30%
- Final Exam: 40%
- Contributions to class: An important tie breaker

Use of CPLEX for solving IO problems

3 Todays Lecture

Slide 3

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

4 Integer Optimization

4.1 Mixed IO

Slide 4

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

4.2 Pure IO

SLIDE 5

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

Important special case: Binary Optimization

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

4.3 LO

SLIDE 6

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

5 Modeling with Binary Variables

5.1 Binary Choice

SLIDE 7

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

Example 1: IO formulation of the knapsack problem

n : projects, total budget b

a_j : cost of project j

c_j : value of project j

SLIDE 8

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

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

5.2 Modeling relations

SLIDE 9

- At most one event occurs

$$\sum_j x_j \le 1$$

- 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

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

5.3 The assignment problem

SLIDE 10

n people
m jobs
c_{ij} : cost of assigning person j to job i.

$$x_{ij} = \begin{cases} 1 & \text{person } j \text{is assigned to job } i \\ 0 & \end{cases}$$

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

5.4 Multiple optimal solutions

SLIDE 11

- Generate all optimal solutions to a BOP.

$$\begin{aligned} \max \quad & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\ & \boldsymbol{x} \in \{0, 1\}^n \end{aligned}$$

- Generate third best?
- Extensions to MIO?

5.5 Modeling nonconvex objective functions

SLIDE 12

- How to model min $c(\boldsymbol{x})$, where $c(\boldsymbol{x})$ is piecewise linear but not convex?

6 What is a good formulation?

6.1 Facility Location

SLIDE 13

- Data

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

- 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 14

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

SLIDE 15

Consider an alternative formulation.

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

Are both valid?
Which one is preferable?

6.2 Observations

SLIDE 16

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

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

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

SLIDE 17

- Let

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

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

6.3 Implications

SLIDE 18

- 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 19

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

6.4 Ideal Formulations

SLIDE 20

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

SLIDE 21

- 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.
- <u>Message</u>: Quality of formulation is judged by closeness to $CH(H)$.
$$CH(H) \subseteq P_1 \subseteq P_2$$

7 Minimum Spanning Tree (MST)

SLIDE 22

- 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 23

- 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 24

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

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

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

7.1 Formulation

SLIDE 25

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

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

Is this a good formulation?

SLIDE 26

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

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

7.2 What is $CH(H)$?

SLIDE 27

Let

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

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

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

Why is this a valid IO formulation?

SLIDE 28

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

8 The Traveling Salesman Problem

SLIDE 29

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.

8.1 Formulation I

SLIDE 30

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

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

8.2 Formulation II

SLIDE 31

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

SLIDE 32

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

SLIDE 33

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

9 Minimum Matching

SLIDE 34

- 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 35

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

10 Observations

SLIDE 36

- 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.

11 Summary

Slide 37

1. Modeling with binary variables allows a lot of modelling power.

2. An IO formulation is better than another one if the polyhedra of their linear relaxations are closer to the convex hull of the IO.

3. A good formulation may have an exponential number of constraints.

4. Conjecture: Formulations characterize the complexity of problems. If a problem is solvable in polynomial time, then the convex hull of solutions is known.

15.083J / 6.859J Integer Programming and Combinatorial Optimization
Fall 2009