15.081J/6.251J Introduction to Mathematical Programming

Lecture 24: Discrete Optimization
1 Outline

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

2 Integer Programming

2.1 Mixed IP

\[
\begin{align*}
\text{(MIP)} & \quad \max & & c'x + h'y \\
\text{s.t.} & & & Ax + By \leq b \\
& & & x \in Z^n_+ (x \geq 0, x \text{ integer}) \\
& & & y \in R^n_+ (y \geq 0)
\end{align*}
\]

2.2 Pure IP

\[
\begin{align*}
\text{(IP)} & \quad \max & & c'x \\
\text{s.t.} & & & Ax \leq b \\
& & & x \in Z^n_+
\end{align*}
\]

Important special case: Binary IP

\[
\begin{align*}
\text{(BIP)} & \quad \max & & c'x \\
\text{s.t.} & & & Ax \leq b \\
& & & x \in \{0, 1\}^n
\end{align*}
\]

2.3 LP

\[
\begin{align*}
\text{(LP)} & \quad \max & & c'x \\
\text{s.t.} & & & By \leq b \\
& & & y \in R^n_+
\end{align*}
\]

3 Modeling with Binary Variables

3.1 Binary Choice

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

Example 1: IP formulation of the knapsack problem

- \(n\): projects, total budget \(b\)
- \(a_j\): cost of project \(j\)
- \(c_j\): value of project \(j\)

\[
x_j = \begin{cases} 
1, & \text{if project } j \text{ is selected.} \\
0, & \text{otherwise.}
\end{cases}
\]
\[ \text{max} \quad \sum_{j=1}^{n} c_j x_j \\
\text{subject to} \quad \sum_{j=1}^{n} a_j x_j \leq b \\
x_j \in \{0,1\} \]

### 3.2 Modeling relations

- 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 unconstrained
  \[ 0 \leq y \leq U x, \quad x \in \{0,1\} \]

### 3.3 The assignment problem

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

### 3.4 Multiple optimal solutions

- Generate all optimal solutions to a BOP.
  \[ \text{max} \quad c' x \]
  \[ \text{subject to} \quad A x \leq b \]
  \[ x \in \{0,1\}^n \]

- Generate third best?
- Extensions to MIO?
3.5 Nonconvex functions

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

4 What is a good formulation?

4.1 Facility Location

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

\[
IZ_1 = \min \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} \sum_{j=1}^{n} h_{ij} y_{ij}
\]

\[
s.t. \quad \sum_{j=1}^{n} y_{ij} = 1
\]

\[
y_{ij} \leq x_j
\]

\[
x_j \in \{0, 1\}, 0 \leq y_{ij} \leq 1
\]

Consider an alternative formulation.

\[
IZ_2 = \min \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{m} \sum_{j=1}^{n} h_{ij} y_{ij}
\]

\[
s.t. \quad \sum_{j=1}^{n} y_{ij} = 1
\]

\[
\sum_{i=1}^{m} y_{ij} \leq m \cdot x_j
\]

\[
x_j \in \{0, 1\}, 0 \leq y_{ij} \leq 1
\]

Are both valid?
Which one is preferable?

4.2 Observations

- \( IZ_1 = IZ_2 \), since the integer points both formulations define are the same.

- \[ P_1 = \{(x, y) : \sum_{j=1}^{n} y_{ij} = 1, y_{ij} \leq x_j, \quad 0 \leq x_j \leq 1, \quad 0 \leq y_{ij} \leq 1 \} \]
\[ P_2 = \{(x, y) : \sum_{j=1}^{n} y_{ij} = 1, \sum_{i=1}^{m} y_{ij} \leq m \cdot x_j, \]
\[ 0 \leq x_j \leq 1 \}
\[ 0 \leq y_{ij} \leq 1 \} \]

- Let \( Z_1 = \min cx + hy, Z_2 = \min cx + hy \)  
  \((x, y) \in P_1 \)
  \((x, y) \in P_2 \)
- \( Z_2 \leq Z_1 \leq IZ_1 = IZ_2 \)

### 4.3 Implications

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

### 4.4 Ideal Formulations

- Let \( P \) be an LP relaxation for a problem
- Let \( H = \{(x, y) : x \in \{0, 1\}^n \} \cap P \)
- Consider Convex Hull (H)
  \[ = \{x : x = \sum_{i} \lambda_i x^i, \sum_{i} \lambda_i = 1, \lambda_i \geq 0, x^i \in H \} \]
- The extreme points of \( CH(H) \) have \( \{0, 1\} \) coordinates.
- So, if we know \( CH(H) \) explicitly, then by solving \( \min cx + hy, (x, y) \in CH(H) \) 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)

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

- 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 $x_e = \begin{cases} 1, & \text{if edge } e \text{ is included in the tree} \\ 0, & \text{otherwise} \end{cases}$
- 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 : j \in V \setminus S \}
- 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

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

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

Is this a good formulation?

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

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

Let

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

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

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

Why is this a valid IP formulation?

- 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

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

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

$$\min \sum_{e \in E} c_e x_e$$

s.t. $$\sum_{e \in \delta(S)} x_e \geq 2, \quad S \subseteq E$$

$$\sum_{e \in \delta(i)} x_e = 2, \quad i \in V$$

$$x_e \in \{0, 1\}$$

6.2 Formulation II

$$\min \sum_{e \in E(S)} c_e x_e$$

s.t. $$\sum_{e \in \delta(i)} x_e \leq |S| - 1, \quad S \subseteq E$$

$$\sum_{e \in \delta(i)} x_e = 2, \quad i \in V$$

$$x_e \in \{0, 1\}$$
\[ \begin{align*}
PT_{\text{cut}} &= \{ x \in \mathbb{R}^{|E|} : \sum_{e \in \delta(S)} x_e \geq 2, \sum_{e \in \delta(i)} x_e = 2 \\
PT_{\text{sub}} &= \{ x \in \mathbb{R}^{|E|} : \sum_{e \in \delta(i)} x_e = 2 \\
&\quad \sum_{e \in \delta(S)} x_e \leq |S| - 1 \\
&\quad 0 \leq x_e \leq 1 \}
\end{align*} \]  

Slide 31

- Theorem: \( PT_{\text{cut}} = PT_{\text{sub}} \not\supseteq CH(H) \)

- Nobody knows \( CH(H) \) for the TSP

7 Minimum Matching

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

- Formulation:

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

- Is the LP relaxation \( CH(H) \)?

Let

\[ PMAT = \{ x \in \mathbb{R}^{|E|} : \sum_{e \in \delta(S)} x_e = 2k + 1, S \neq \emptyset \}
\]

Theorem: \( PMAT = CH(H) \)

8 Observations

- 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.

- Conjecture: The convex hull of problems that are polynomially solvable are explicitly known.
9 Summary

1. An IP formulation is better than another one if the polyhedra of their LP relaxations are closer to the convex hull of the IP.

2. A good formulation can 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.