DM554/DM545

Linear and Integer Programming

Lecture 10
IP Modeling
Formulations, Relaxations

Marco Chiarandini

Department of Mathematics & Computer Science
University of Southern Denmark

Outline

Outline

Outline

Outline

Outline

Matching

Definition (Matching Theory Terminology)

- Matching: set of pairwise non adjacent edges
- Covered (vertex): a vertex is covered by a matching M if it is incident to an edge in M
- Perfect (matching): if M covers each vertex in G
- Maximal (matching): if M cannot be extended any further
- Maximum (matching): if M covers as many vertices as possible
- Matchable (graph): if the graph G has a perfect matching

$$\begin{aligned} \max \sum_{v \in V} w_e x_e & \\ \sum_{e \in E : v \in e} x_e \leq 1 & \quad \forall v \in V \\ x_e \in \{0,1\} & \quad \forall e \in E \end{aligned}$$

Special case: bipartite matching $\equiv$ assignment problems

Vertex Cover

Select a subset $S \subseteq V$ such that each edge has at least one end vertex in S.

$$\begin{aligned} \min \quad & \sum_{v \in V} x_v \\ & x_v + x_u \geq 1 \qquad \forall u, v \in V, uv \in E \\ & x_v \in \{0,1\} \; \forall v \in V \end{aligned}$$

Approximation algorithm: set S derived from the LP solution in this way:

$$S_{LP} = \{v \in V : x^*_v \geq 1/2\}$$

(it is a cover since $x^*_v + x^*_u \geq 1$ implies $x^*_v \geq 1/2$ or $x^*_u \geq 1/2$)

Proposition

The LP rounding approximation algorithm gives a 2-approximation:
$|S_{LP}| \leq 2|S_{OPT}|$ *(at most as bad as twice the optimal solution)*

Proof: Let $\bar{x}$ be opt to IP. Then $\sum x^*_v \leq \sum \bar{x}_v$.
$|S_{LP}| = \sum_{v \in S_{LP}} 1 \leq \sum_{v \in V} 2x^*_v$ since $x^*_v \geq 1/2$ for each $v \in S_{LP}$
$|S_{LP}| \leq 2 \sum_{v \in V} x^*_v \leq 2 \sum_{v \in V} \bar{x}_v = 2|S_{OPT}|$

Maximum independent Set

Find the largest subset $S \subseteq V$ such that the induced graph has no edges

$$\begin{aligned} \max \quad & \sum_{v \in V} x_v \\ & x_v + x_u \leq 1 \qquad \forall u, v \in V, uv \in E \\ & x_v = \{0, 1\} \ \forall v \in V \end{aligned}$$

Optimal sol of LP relaxation sets $x_v = 1/2$ for all variables and has value $|V|/2$.

What is the value of an optimal IP solution of a complete graph?

LP relaxation gives an $O(n)$-approximation (almost useless)

Traveling Salesman Problem

- Find the cheapest movement for a drilling, welding, drawing, soldering arm as, for example, in a printed circuit board manufacturing process or car manufacturing process

- n locations, c_{ij} cost of travel

Variables:

$$x_{ij} = \begin{cases} 1 \\ 0 \end{cases}$$

Objective:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

Constraints:

-

$$\sum_{j:j\neq i} x_{ij} = 1 \qquad \forall i = 1,\ldots,n$$

$$\sum_{i:i\neq j} x_{ij} = 1 \qquad \forall j = 1,\ldots,n$$

- cut set constraints

$$\sum_{i\in S}\sum_{j\notin S} x_{ij} \geq 1 \qquad \forall S \subset N, S \neq \emptyset$$

- subtour elimination constraints

$$\sum_{i\in S}\sum_{j\in S} x_{ij} \leq |S| - 1 \qquad \forall S \subset N, 2 \leq |S| \leq n-1$$

Outline

Modeling Tricks

Objective function and/or constraints do not appear to be linear?

- Absolute values
- Minimize the largest function value
- Maximize the smallest function value
- Constraints include variable division
- Constraints are either/or
- A variable must take one of several candidate values

Modeling Tricks I

Minimize the largest of a number of function values:

$$\min \quad \max\{f(x_1), \ldots, f(x_n)\}$$

- Introduce an auxiliary variable z:

$$\begin{aligned} \min \quad & z \\ \text{s. t. } & f(x_1) \leq z \\ & f(x_2) \leq z \end{aligned}$$

Modeling Tricks II

Constraints include variable division:

- Constraint of the form

$$\frac{a_1x + a_2y + a_3z}{d_1x + d_2y + d_3z} \leq b$$

  - Rearrange:

    $$a_1x + a_2y + a_3z \leq b(d_1x + d_2y + d_3z)$$

    which gives:

    $$(a_1 - bd_1)x + (a_2 - bd_2)y + (a_3 - bd_3)z \leq 0$$

III "Either/Or Constraints"

In conventional mathematical models, the solution must satisfy all constraints.
Suppose that your constraints are "either/or":

$$a_1x_1 + a_2x_2 \leq b_1 \quad \text{or}$$
$$d_1x_1 + d_2x_2 \leq b_2$$

Introduce new variable $y \in \{0, 1\}$ and a large number M:

$$a_1x_1 + a_2x_2 \leq b_1 + My \qquad \text{if } y = 0 \text{ then this is active}$$
$$d_1x_1 + d_2x_2 \leq b_2 + M(1 - y) \qquad \text{if } y = 1 \text{ then this is active}$$

III "Either/Or Constraints"

Binary integer programming allows to model alternative choices:

- Eg: 2 feasible regions, ie, disjunctive constraints, not possible in LP. introduce y auxiliary binary variable and M a big number:

$$\begin{aligned} Ax &\leq b + My && \text{if } y = 0 \text{ then this is active} \\ A'x &\leq b' + M(1-y) && \text{if } y = 1 \text{ then this is active} \end{aligned}$$

IV "Either/Or Constraints"

Generally:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1m}x_m &\leq d_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ldots + a_{2m}x_m &\leq d_2 \\
&\vdots \\
a_{m1}x_1 + a_{N2}x_2 + a_{N3}x_3 + \ldots + a_{Nm}x_m &\leq d_N
\end{aligned}
$$

Exactly K of the N constraints must be satisfied.
Introduce binary variables $y_1, y_2, \ldots, y_N$ and a large number M

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1m}x_m &\leq d_1 + My_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ldots + a_{2m}x_m &\leq d_2 + My_2 \\
&\vdots \\
a_{m1}x_1 + a_{N2}x_2 + a_{N3}x_3 + \ldots + a_{Nm}x_m &\leq d_N + My_N
\end{aligned}
$$

$$y_1 + y_2 + \ldots y_N = N - K$$

K of the y-variables are 0, so K constraints must be satisfied

IV "Either/Or Constraints"

At least $h \leq k$ of $\sum_{j=1}^{n} a_{ij}x_j \leq b_i$, $i = 1, \ldots, k$ must be satisfied

introduce y_i, $i = 1, ..., k$ auxiliary binary variables

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i + My_i$$

$$\sum_{i} y_i \leq k - h$$

V "Possible Constraints Values"

A constraint must take on one of N given values:

$$
\begin{aligned}
a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_mx_m &= d_1 \text{ or} \\
a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_mx_m &= d_2 \text{ or} \\
&\vdots \\
a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_mx_m &= d_N
\end{aligned}
$$

Introduce binary variables $y_1, y_2, \ldots, y_N$:

$$a_1x_1 + a_2x_2 + a_3x_3 + \ldots + a_mx_m = d_1y_1 + d_2y_2 + \ldots d_Ny_N$$

$$y_1 + y_2 + \ldots y_N = 1$$

Outline

Outline

Uncapacited Facility Location (UFL)

Given:

- depots $N = \{1, \dots, n\}$
- clients $M = \{1, \dots, m\}$
- f_j fixed cost to use depot j
- transport cost for all orders c_{ij}

Task: Which depots to open and which depots serve which client

Variables: $y_j = \begin{cases} 1 & \text{if depot open} \\ 0 & \text{otherwise} \end{cases}$, x_{ij} fraction of demand of i satisfied by j

Objective:

$$\min \sum_{i \in M} \sum_{j \in N} c_{ij} x_{ij} + \sum_{j \in N} f_j y_j$$

Constraints:

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad \forall i = 1, \dots, m$$

$$\sum_{i \in M} x_{ij} \le m y_j \qquad \forall j \in N$$

Outline

Good and Ideal Formulations

Definition (Formulation)

A polyhedron $P \subseteq \mathbb{R}^{n+p}$ is a formulation for a set $X \subseteq \mathbb{Z}^n \times \mathbb{R}^p$ if and only if $X = P \cap (\mathbb{Z}^n \times \mathbb{R}^p)$

That is, if it does not leave out any of the solutions of the feasible region X.

There are infinite formulations

Definition (Convex Hull)

Given a set $X \subseteq \mathbb{Z}^n$ the convex hull of X is defined as:

$$\text{conv}(X) = \Big\{ \mathbf{x} : \mathbf{x} = \sum_{i=1}^{t} \lambda_i \mathbf{x}^i, \sum_{i=1}^{t} \lambda_i = 1, \lambda_i \geq 0, \text{for } i = 1, \dots, t, \\ \text{for all finite subsets } \{\mathbf{x}^1, \dots, \mathbf{x}^t\} \text{ of } X \Big\}$$

Proposition

$conv(X)$ is a polyhedron (ie, representable as $A\mathbf{x} \leq \mathbf{b}$)

Proposition

Extreme points of $conv(X)$ all lie in X

Hence:

$$\max\{\mathbf{c}^T\mathbf{x} : \mathbf{x} \in X\} \equiv \max\{\mathbf{c}^T\mathbf{x} : \mathbf{x} \in \text{conv}(X)\}$$

However it might require exponential number of inequalities to describe $\text{conv}(X)$

What makes a formulation better than another?

$$X \subseteq \text{conv}(X) \subseteq P_1 \subset P_2$$

$$P_1 \text{ is better than } P_2$$

Definition

Given a set $X \subseteq \mathbb{R}^n$ and two formulations P_1 and P_2 for X, P_1 is a better formulation than P_2 if $P_1 \subset P_2$

Example

$P_1 =$ UFL with $\sum_{i \in M} x_{ij} \le m y_j \quad \forall j \in N$
$P_2 =$ UFL with $x_{ij} \le y_j \quad \forall i \in M, j \in N$

$$P_2 \subset P_1$$

- $P_2 \subseteq P_1$ because summing $x_{ij} \le y_j$ over $i \in M$ we obtain $\sum_{i \in M} x_{ij} \le m y_j$
- $P_2 \subset P_1$ because there exists a point in P_1 but not in P_2: $m = 6 = 3 \cdot 2 = k \cdot n$

$$x_{10} = 1,\ x_{20} = 1,\ x_{30} = 1, \qquad \sum_i x_{i0} \le 6 y_0 \quad y_0 = 1/2$$
$$x_{41} = 1,\ x_{51} = 1,\ x_{61} = 1 \qquad \sum_i x_{i1} \le 6 y_1 \quad y_1 = 1/2$$

Resume

1. Modeling
    - Assignment Problem
    - Set Problems
    - Graph Problems
    - Modeling Tricks

2. Formulations
    - Uncapacited Facility Location
    - Alternative Formulations