AM 121: Intro to Optimization Models and Methods

Fall 2016

Lecture 12: IP Modeling and Exercises

David C. Parkes

SEAS

Lesson Plan

- Group Exercise1. Airline scheduling
- Group Exercise 2. Firehouse location
- Solutions
- Formulation Strength

Exercise 1. Airline scheduling

- Need to operate a number of flights
  - e.g., LAX to ORD, 9am
- Routes are feasible sequences of flights for a single plane considering factors such as turnaround time.
- Find an assignment of plane equipment to a set of possible routes (e.g. LAX-ORD, ORD-JFK, JFK-LAX).
- Each route has an associated cost (e.g., based on distance.)
- **Goal**: select a set of routes so that each flight is operated **exactly** once.

Example set of possible routes

Route

Flight	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
SFO-LAX	1			1			1			1		
SFO-DEN	✓	1			1			1			1	
SFO-SEA			1			1			1			1
LAX-ORD				2			2		3	1		3
LAX-SFO	2					3				5	5	
ORD-DEN				3	3				4			
ORD-SEA							3	3		3	3	4
DEN-SFO		2		4	4				5			
DEN-ORD					2			2			2	
SEA-SFO			2				4	4				5
SEA-LAX						2			2	4	4	2
COST ($1000's)	**2**	**3**	**4**	**6**	**7**	**5**	**7**	**8**	**9**	**9**	**8**	**9**

e.g., route 6 does SFO-SEA then SEA-LAX then LAX-SFO.

Exercise 1: Two variations

- **Equipment scheduling**:
  - Equipment routes $j \in J$ and flights $i \in I$. Let c_j denote the cost of using route j.
  - **Goal:** Minimize total cost, subject to each flight covered by exactly one plane.
- **Crew scheduling**:
  - Once planes scheduled, can now schedule crews to “crew routes.” Set of crew routes may be different from equipment routes because turnaround time between flights different. Also: it is ok to have more than one crew on a flight.
  - Crew routes $k \in K$, flights $i \in I$. Cost c_k.
  - **Goal:** Minimize total cost, subject to each flight covered by one or more crew.

Exercise 2: Firehouse location

- J potential sites (m total), I districts (n total).
- Each site can serve every district, but distance $d_{ij} \geq 0$ from site j to district i. Serve every district.
- Each district $i \in \mathrm{I}$ has population p_i.
- Cost $f_j(s_j) = K_j + c_j s_j$ to use firehouse at j to service s_j people in total. K_j fixed cost, c_j variable cost.
- Budget B.

- **Goal**: determine firehouse sites and assignment of districts to these sites to minimize max distance to a firehouse. Don’t exceed budget.

Solution: Airline Scheduling

- **Equipment**. Routes j, flights i. x_j=1 if select route j. Let a_{ij}=1 if flight i in route j, a_{ij}=0 o.w.
  min $\sum_{j\in J} c_j x_j$
  s.t. $\sum_{j\in J} a_{ij} x_j = 1$ for all i
  $x_j \in \{0,1\}$
- **Crew**. Routes k. x_k =1 if select route k, x_k=0 o.w. a_{ik}=1 if flight i in route k.
  min $\sum_{k\in K} c_k x_k$
  s.t. $\sum_{k\in K} a_{ik} x_k \geq 1$ for all i
  $x_k \in \{0,1\}$

Solution: Firehouse location

- x_{ij} = 1 if district i assigned to site j, x_{ij} = 0 o.w.
- y_j = 1 if select site j, y_j = 0 o.w.
- D = max distance
- s_j = total population served by site j
- min D
  s.t. $\sum_j x_{ij} = 1$, for all i
  $\sum_i x_{ij} \leq n y_j$, for all j
  $D \geq \sum_j d_{ij} x_{ij}$, for all i
  $\sum_j (K_j y_j + c_j s_j) \leq B$
  $s_j = \sum_i p_j x_{ij}$, for all j
  $x_{ij} \in \{0,1\}$, $y_j \in \{0,1\}$, $s_j \geq 0$, $D \geq 0$

Correct formulations

- **Definition.** A formulation is **correct** if the set of feasible values for decision variables corresponds to the set of feasible decisions in the problem to be modeled.

- Correctness => optimal solution will solve intended problem for any choice of objective.
- A "true model" of the problem.

Alternate Formulations

Formulation (P_1):

•If m=2 sites, and n=4 districts:

•$x_{11} + x_{21} + x_{31} + x_{41} <= 4\ y_1$

•$x_{12} + x_{22} + x_{32} + x_{42} <= 4\ y_2$

•m constraints

Formulation (P_2):

•$x_{11} <= y_1$, $x_{21} <= y_1$, $x_{31} <= y_1$, $x_{41} <= y_1$

•$x_{12} <= y_2$, $x_{22} <= y_2$, $x_{32} <= y_2$, $x_{42} <= y_2$

•mn constraints

• Which is easier to solve? Why?

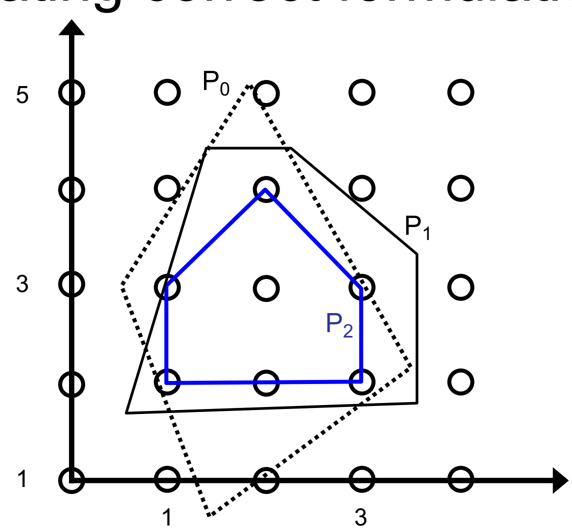
LP Relaxations

- A **polyhedron** is a set that can be described in the form $P = \{x \in R^n : Ax \leq b\}$
- Fact: the feasible set of any LP can be described as a polyhedron.

- **Definition**. The **LP relaxation** of an integer program replaces all integer variables with continuous variables.

- **Definition**. The **polyhedron (P) of an IP** is the feasible region for the LP relaxation.

Illustrating correct formulations



Suppose P_0 is a correct formulation. Why are P_1 and P_2 also correct? Which is "ideal"?

Formulation Strength

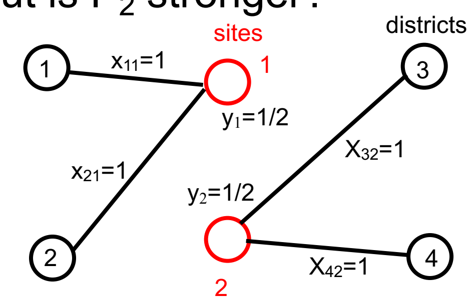
Definition. Given two correct IP formulations, formulation 2 is **stronger** than formulation 1 if $P_2 \subset P_1$.

Equivalently: formulation 2 is “tighter”.

A stronger formulation is preferred because it eliminates some fractional solutions. Tends to makes the IP easier to solve.

Example: Facility Location

- P_1: $\sum_i x_{ij} \leq ny_j$, for each j
- P_2: $x_{ij} \leq y_j$, for each i, j
- We have $P_2 \subseteq P_1$ because P_2 implies P_1. By summing over districts i, then $\sum_i x_{ij} \leq ny_j$.
- But is P_2 stronger?



Fractional solution in P_1 but not P_2.
Conclude $P_2 \subset P_1$.

P_1:
$x_{11}+x_{21}+x_{31}+x_{41} \leq 4y_1$
$x_{21}+x_{22}+x_{32}+x_{42} \leq 4y_2$
P_2:
$x_{11} \leq y_1$
$x_{21} \leq y_1$
…
$x_{32} \leq y_2$
$x_{42} \leq y_2$

On Good IP Formulations

Correctness is most important!

But: to make IPs faster to solve:

- Prefer formulations that are “strong” and exclude fractional solutions. This may involve using more constraints.
- Generally try to minimize the number of integer variables.
- If using “big Ms” make them as small as possible.

Summary: Modeling via IPs

- IPs are very expressive. Able to model lots of real-world problems.
- Modeling tricks often use indicator variables and “big M” constants.
- There can be multiple correct formulations, and they may differ in speed to solve.
- Prefer stronger formulations, i.e. those with tighter LP relaxations.