Mixed Integer Programming (MIP) for Causal Inference and Beyond

Juan Pablo Vielma
Massachusetts Institute of Technology

Columbia Business School
Traveling Salesman Problem (TSP): Visit Cities Fast
Traveling Salesman Problem (TSP): Visit Cities Fast

“A computer would have to check all these possible routes to find the shortest one.”

Paradoxes, Contradictions, and the Limits of Science

Many research results define boundaries of what cannot be known, predicted, or described. Classifying these limitations shows us the structure of science and reason.

Noson S. Yanofsky
MIP = Avoid Enumeration

- Number of tours for 49 cities = $\frac{48!}{2} \approx 10^{60}$
- Fastest supercomputer $\approx 10^{17}$ flops
- Assuming one floating point operation per tour:
  $> 10^{35}$ years $\approx 10^{25}$ times the age of the universe!
- How long does it take on an iphone?
  - Less than a second!
  - 4 iterations of cutting plane method!
  - Dantzig, Fulkerson and Johnson 1954 did it by hand!
  - For more info see tutorial in ConcordeTSP app
  - Cutting planes are the key for effectively solving (even NP-hard) MIP problems in practice.
Using IP to visit Germany

45 cities (1832)  
120 cities (1977)  

http://www.math.uwaterloo.ca/tsp/d15sol/dhistory.html
50+ Years of MIP = Significant Solver Speedups

- Algorithmic Improvements (Machine Independent):
  - Commercial, but free for academic use

- (Reasonably) effective free / open source solvers:
  - GLPK, CBC and SCIP (free only for non-commercial)

- Easy to use, fast and versatile modeling languages
  - Julia based JuMP modelling language

- Linear MIP solvers very mature and effective:
  - Convex nonlinear MIP getting there (quadratic nearly there)
Matching

Treated Units: \( \mathcal{T} = \{t_1, \ldots, t_T\} \)

Control Units: \( \mathcal{C} = \{c_1, \ldots, c_C\} \)

Observed Covariates: \( \mathcal{P} = \{p_1, \ldots, p_P\} \)

Covariate Values: \( \mathbf{x}^t = (x^t_p)_{p \in \mathcal{P}}, \quad t \in \mathcal{T} \)

\( \mathbf{x}^c = (x^c_p)_{p \in \mathcal{P}}, \quad c \in \mathcal{C} \)
Maximum Cardinality Exact Matching

\[
\begin{align*}
\text{max} & \quad \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} m_{t,c} \\
\text{s.t.} & \quad \sum_{t \in \mathcal{T}} m_{t,c} \leq 1, \quad \forall c \in \mathcal{C} \\
& \quad \sum_{c \in \mathcal{C}} m_{t,c} \leq 1, \quad \forall t \in \mathcal{T} \\
& \quad m_{t,c} = 0 \quad \forall t, c \quad x^t \neq x^c \\
0 & \leq m_{t,c} \leq 1 \quad m_{t,c} \in \{0,1\} \quad \forall t \in \mathcal{T}, \quad c \in \mathcal{C}.
\end{align*}
\]

- Solve time for truncated Lalonde CPS 429 = 0.001 s
Maximum Cardinality Marginal Means

\[
\max \sum_{t \in T} \sum_{c \in C} m_{t,c}
\]

\[s.t.
\]

\[
\sum_{t \in T} m_{t,c} \leq 1, \quad \forall c \in C
\]

\[
\sum_{c \in C} m_{t,c} \leq 1, \quad \forall t \in T
\]

\[
\sum_{t \in T} \sum_{c \in C} x_p^t m_{t,c} = \sum_{c \in C} \sum_{t \in T} x_p^c m_{t,c} \quad \forall p \in P
\]

\[
m_{t,c} \in \{0, 1\} \quad \forall t \in T, \quad c \in C.
\]

- Solve time for truncated Lalonde CPS 429 = 444 s
- Why? One reason = relaxation has fractions.
Maximum Cardinality Fine Balance, Take 1

\[
\max \sum_{t \in T} \sum_{c \in C} m_{t,c}
\]

\[
s.t.
\]

\[
\sum_{t \in T} m_{t,c}^{p} \leq 1, \quad \forall c \in C, \ p \in \mathcal{P}
\]

\[
\sum_{c \in C} m_{t,c}^{p} \leq 1, \quad \forall t \in T, \ p \in \mathcal{P}
\]

\[
m_{t,c}^{p} = 0 \quad \forall t, c \quad x_{p}^{t} \neq x_{p}^{c}, \ p \in \mathcal{P}
\]

\[
\sum_{t \in T} m_{t,c}^{p} = \sum_{t \in T} m_{t,c}^{q} \quad \forall c \in C, \ p, q \in \mathcal{P}
\]

\[
\sum_{c \in C} m_{t,c}^{p} = \sum_{c \in C} m_{t,c}^{q} \quad \forall t \in T, \ p, q \in \mathcal{P}
\]

\[
m_{t,c}^{p} \in \{0, 1\} \quad \forall t \in T, \ c \in C, \ p \in \mathcal{P}.
\]

• Solve time for truncated Lalonde CPS 429 = 0.81 s

• Why? One reason = relaxation has “fewer” fractions.
Basic Branch-and-Bound

\[ \begin{align*}
\text{max } z := & \quad x_2 \\
& \quad 3x_1 + 2x_2 \leq 6 \\
& \quad -2x_1 + x_2 \leq 0 \\
& \quad x_1, x_2 \geq 0 \\
& \quad x_1, x_2 \in \mathbb{Z}
\end{align*} \]

Linear Programming (LP) Relaxation

(1) FRAC
\[ z = 1.71 \]

(2) INT
\[ z = 0 \]

(3) FRAC
\[ z = 1.5 \]

(4) INT
\[ z = 1 \]

(5) INF

\[ x_1 \leq 0 \]
\[ x_1 \geq 1 \]
\[ x_2 \geq 2 \]
\[ x_1 \leq 1 \]
Stronger Formulations = Faster Solves ?

\[
\begin{align*}
\text{(1) FRAC} & \quad z = 1.71 \\
\text{(2) INT} & \quad z = 0 \\
\text{(3) INT} & \quad z = 1 \\
\end{align*}
\]

\[
\begin{align*}
x_1 & \leq 0 \\
x_1 & \geq 1
\end{align*}
\]

\[
\begin{align*}
\max z := & x_2 \\
3x_1 + 2x_2 & \leq 6 \\
-2x_1 + x_2 & \leq 0 \\
x_1, x_2 & \geq 0 \quad x_1, x_2 \in \mathbb{Z}
\end{align*}
\]

\[
\begin{align*}
\max z := & x_2 \\
x_1 + x_2 & \leq 2 \\
-2x_1 + x_2 & \leq 0 \\
x_1, x_2 & \geq 0 \quad x_1, x_2 \in \mathbb{Z}
\end{align*}
\]
Cutting Plane Example: Chátal-Gomory Cuts

\[ P := \left\{ x \in \mathbb{R}^2 : \begin{array}{l}
  x_1 + x_2 \leq 3, \\
  5x_1 - 3x_2 \leq 3
\end{array} \right\} \]

\[ H := \{ x \in \mathbb{R}^2 : \begin{array}{l}
  4x_1 + 3x_2 \leq 10.5 \\
  \in \mathbb{Z}
\end{array} \} \cap \mathbb{Z}^2 \text{ if } x \in \mathbb{Z}^2 \]

\[ 4x_1 + 3x_2 \leq \lfloor 10.5 \rfloor \]

Valid for \( H \cap \mathbb{Z}^2 \)
Valid for \( P \cap \mathbb{Z}^2 \)
Branch-and-Bound and Cuts (Branch-and-Cut)

\[ \max z := x_2 \]
\[ 3x_1 + 2x_2 \leq 6 \quad x_2 \leq \lfloor 1.71 \rfloor = 1 \]
\[ -2x_1 + x_2 \leq 0 \]
\[ x_1, x_2 \geq 0 \]
\[ x_1, x_2 \in \mathbb{Z} \]

(1) \text{FRAC} \quad z = 1.71

(2) \text{FRAC} \quad z = 1

\[ x_1 \leq 0 \quad x_1 \geq 1 \]

(3) \text{INT} \quad z = 0

(4) \text{INT} \quad z = 1
Partial Solves and GAP

\[
\begin{align*}
\text{max } z &= x_2 \\
0 &\leq x_1 \\ x_1 &\geq 1
\end{align*}
\]

\[
\begin{align*}
(1) \text{ FRAC} \\
z &= 1.71 \\
x_1 &\leq 0 \quad x_1 &\geq 1
\end{align*}
\]

\[
(2) \text{ INT} \\
z &= 0
\]

\[
(3) \text{ FRAC} \\
z &= 1.5
\]

\[
GAP = 100 \times \frac{|\text{bestnode} - \text{bestinteger}|}{10^{-10} + |\text{bestinteger}|}
\]

\[
BBGAP = 100 \times \frac{|1.5 - 0|}{10^{-10} + |0|} = 1.5 \times 10^{12}
\]

\[
BB^{+}GAP = 100 \times \frac{|1.5 - 1|}{10^{-10} + |1|} = 50%
\]

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Cuts/</th>
<th>Objective</th>
<th>Best Integer</th>
<th>Best Bound</th>
<th>ItCnt</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node</td>
<td>Left</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1415.0000</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1446.4286</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>0+</td>
<td>3130.0000</td>
<td>1446.4286</td>
<td>317</td>
<td>53.79%</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>3130.0000</td>
<td>1446.4286</td>
<td>317</td>
<td>53.79%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ZeroHalf: 33</td>
<td>317</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>28</td>
<td>integral</td>
<td>0</td>
<td>1449.7619</td>
<td>1168</td>
<td>52.62%</td>
</tr>
<tr>
<td>*</td>
<td>34</td>
<td>integral</td>
<td>0</td>
<td>1449.7619</td>
<td>1320</td>
<td>52.39%</td>
</tr>
</tbody>
</table>

Elapsed real time = 0.18 sec. (tree size = 0.01 MB, solutions = 1)
Maximum Cardinality Marginal Means

\[
\max \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} m_{t,c}
\]

\[
\text{s.t.}
\]

\[
\sum_{t \in \mathcal{T}} m_{t,c} \leq 1, \quad \forall c \in \mathcal{C}
\]

\[
\sum_{c \in \mathcal{C}} m_{t,c} \leq 1, \quad \forall t \in \mathcal{T}
\]

\[
\sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} x_p^t m_{t,c} = \sum_{c \in \mathcal{C}} \sum_{t \in \mathcal{T}} x_p^c m_{t,c} \quad \forall p \in \mathcal{P}
\]

\[
m_{t,c} \in \{0, 1\} \quad \forall t \in \mathcal{T}, \quad c \in \mathcal{C}.
\]

- Solve time for truncated Lalonde CPS 429 = 444 s
- Optimal = 17, LP Relaxation = 19.35
Maximum Cardinality Marginal Means

Root relaxation: objective $1.935024\times 10^1$, 1947 iterations, 0.16 seconds

| Nodes Expl Unexpl | Current Node Obj Depth IntInf | Objective Bounds Incumbent BestBd Gap | Work It/Node Time |
|-------------------|-----------------------------|-----------------------------|-------------------|------------------|
| 0     0           | 19.35024                   | 0 10                        | -0.00000         | 19.35024        | -     -     0s |
| 0     0           | 19.30273                   | 0 20                        | -0.00000         | 19.30273        | -     -     1s |
|       .           |                             |                             |                   |                 |       .     . |
| 0     2           | 19.29748                   | 0 25                        | -0.00000         | 19.29748        | -     -     4s |
| 4     5           | 19.20830                   | 3 23                        | -0.00000         | 19.29273        | - 1407 5s |
|       .           |                             |                             |                   |                 |       .     . |
| 1140  763        | 18.56894                   | 18 23                       | -0.00000         | 19.16557        | - 340   51s|
| H 1149  726      |                             |                             | 4.00000000       | 19.16557        | 379%   344 51s|
|       .           |                             |                             |                   |                 |       .     . |
| *23876 7827      | 17.00000000               | 71                          | 18.00366         | 18.00366        | 5.90%  150 443s|

Cutting planes:  Gomory: 1  Flow cover: 1
Explored 25301 nodes (3601039 simplex iterations) in 443.89 seconds

Optimal solution found (tolerance 1.00e-04)
Best objective $1.700000000000\times 10^1$, best bound $1.700000000000\times 10^1$, gap 0.0%

“Black Magic”
Maximum Cardinality Fine Balance, Take 1

\[
\max \sum_{t \in T} \sum_{c \in C} m_{t,c}
\]

s.t.

\[
\sum_{t \in T} m^p_{t,c} \leq 1, \quad \forall c \in C, p \in \mathcal{P}
\]

\[
\sum_{c \in C} m^p_{t,c} \leq 1, \quad \forall t \in T, p \in \mathcal{P}
\]

\[
m^p_{t,c} = 0 \quad \forall t, c \quad x^t_p \neq x^c_p, p \in \mathcal{P}
\]

\[
\sum_{t \in T} m^p_{t,c} = \sum_{t \in T} m^q_{t,c} \quad \forall c \in C, p, q \in \mathcal{P}
\]

\[
\sum_{c \in C} m^p_{t,c} = \sum_{c \in C} m^q_{t,c} \quad \forall t \in T, p, q \in \mathcal{P}
\]

\[
m^p_{t,c} \in \{0, 1\} \quad \forall t \in T, c \in C, p \in \mathcal{P}.
\]

- Solve time for truncated Lalonde CPS 429 = 0.81 s
- Optimal = 10, LP Relaxation = 11
Maximum Cardinality Fine Balance, Take 1

Root relaxation: objective 1.100000e+01, 1490 iterations, 0.03 seconds

<table>
<thead>
<tr>
<th>Nodes</th>
<th>Expl Unexpl</th>
<th>Current Node</th>
<th>Objective Bounds</th>
<th>Work</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>11.00000</td>
<td>0 158 -0.00000</td>
<td>11.00000 - - 0s</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>11.00000</td>
<td>0 164 -0.00000</td>
<td>11.00000 - - 0s</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>11.00000</td>
<td>0 164 -0.00000</td>
<td>11.00000 - - 0s</td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>11.00000</td>
<td>0 111 10.00000</td>
<td>11.00000 10.0% - 0s</td>
</tr>
</tbody>
</table>

Cutting planes: Zero half: 2

Explored 8 nodes (9933 simplex iterations) in 0.78 seconds
Optimal solution found (tolerance 1.00e-04)
Best objective 1.000000000000e+01, best bound 1.000000000000e+01, gap 0.0%

Feasible Solution \rightarrow Optimal Solution \rightarrow Bound
Maximal Cardinality Fine Balance, Take 1

$$\max \sum_{t \in T} \sum_{c \in C} m_{t,c}$$

s.t.

$$\sum_{t \in T} m_{t,c}^p \leq 1, \quad \forall c \in C, \; p \in P$$

$$\sum_{c \in C} m_{t,c}^p \leq 1, \quad \forall t \in T, \; p \in P$$

$$m_{t,c}^p = 0 \quad \forall t, c \quad x_t^p \neq x_c^p, \; p \in P$$

$$\sum_{t \in T} m_{t,c}^p = \sum_{t \in T} m_{t,c}^q \quad \forall c \in C, \; p, q \in P$$

$$\sum_{c \in C} m_{t,c}^p = \sum_{c \in C} m_{t,c}^q \quad \forall t \in T, \; p, q \in P$$

$$m_{t,c}^p \in \{0, 1\} \quad \forall t \in T, \; c \in C, \; p \in P.$$

- Solve time for truncated Lalonde CPS 429 = 0.81 s
- Scalability? Size = $|T| \times |C| \times |P|$
Maximum Cardinality Fine Balance, Take 2

\[
\begin{align*}
\max & \quad \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} m_{t,c} \\
\text{s.t.} & \quad \sum_{t \in \mathcal{T}} m_{t,c} \leq 1, \quad \forall c \in \mathcal{C} \\
& \quad \sum_{c \in \mathcal{C}} m_{t,c} \leq 1, \quad \forall t \in \mathcal{T} \\
& \quad \sum_{t \in \mathcal{T}_{p,k}} \sum_{c \notin \mathcal{C}_{p,k}} m_{t,c} = \sum_{t \notin \mathcal{T}_{p,k}} \sum_{c \in \mathcal{C}_{p,k}} m_{t,c} \quad \forall p \in \mathcal{P}, k \in \mathcal{K}(p) \\
& \quad m_{t,c} \in \{0, 1\} \quad \forall t \in \mathcal{T}, \ c \in \mathcal{C}.
\end{align*}
\]

- Solve time for truncated Lalonde CPS 429 = 0.61 s
- Scalability? Size = $|\mathcal{T}| \times |\mathcal{C}| + |\mathcal{P}|$
Take 1 Revisited

\[
\begin{align*}
\max & \quad \sum_{t \in \mathcal{T}} \sum_{c \in \mathcal{C}} m_{t,c} \\
\text{s.t.} & \quad \sum_{t \in \mathcal{T}} m_{t,c}^p \leq 1, \quad \forall c \in \mathcal{C}, \ p \in \mathcal{P} \\
& \quad \sum_{c \in \mathcal{C}} m_{t,c}^p \leq 1, \quad \forall t \in \mathcal{T}, \ p \in \mathcal{P} \\
& \quad m_{t,c}^p = 0 \quad \forall t, c \quad x_t^p \neq x_c^p, \ p \in \mathcal{P} \\
& \quad \sum_{t \in \mathcal{T}} m_{t,c}^p = \sum_{t \in \mathcal{T}} m_{t,c}^q \quad \forall c \in \mathcal{C}, \ p, q \in \mathcal{P} \\
& \quad \sum_{c \in \mathcal{C}} m_{t,c}^p = \sum_{c \in \mathcal{C}} m_{t,c}^q \quad \forall t \in \mathcal{T}, \ p, q \in \mathcal{P} \\
& \quad m_{t,c}^p \in \{0, 1\} \quad \forall t \in \mathcal{T}, \ c \in \mathcal{C}, \ p \in \mathcal{P}.
\end{align*}
\]

• Solve time for truncated Lalonde CPS 429 = 0.81 s

• Scalability? Size = $|\mathcal{T}| \times |\mathcal{C}| \times |\mathcal{P}|$
Take 1 Revisited

\[
\begin{align*}
\text{max} \quad & \sum_{t \in \mathcal{T}} x_t \\
\text{s.t.} \quad & \sum_{t \in \mathcal{T}_{p,k}} m_{t,c}^p = y_c, \quad \forall p \in \mathcal{P}, \quad k \in \mathcal{K}(p), \quad c \in \mathcal{C}_{p,k} \\
& \sum_{c \in \mathcal{C}_{p,k}} m_{t,c}^p = x_t, \quad \forall \mathcal{S} \in \mathcal{F}, \quad k \in \mathcal{K}(p), \quad t \in \mathcal{T}_{p,k} \\
& m_{t,c}^p \in \{0, 1\} \quad \forall p \in \mathcal{P}, \quad k \in \mathcal{K}(p), \quad t \in \mathcal{T}_{p,k}, \quad c \in \mathcal{C}_{p,k} \\
& x_t \in \{0, 1\} \quad \forall t \in \mathcal{T} \\
& y_c \in \{0, 1\} \quad \forall c \in \mathcal{C}.
\end{align*}
\]
Maximum Cardinality Fine Balance, Take 3

\[
\begin{align*}
\max \quad & \sum_{t \in \mathcal{T}} x_t \\
\text{s.t.} \quad & \sum_{t \in \mathcal{T}} x_t = \sum_{c \in \mathcal{C}} y_c, \\
& \sum_{t \in \mathcal{T}_{p,k}} x_t = \sum_{c \in \mathcal{C}_{p,k}} y_c, \quad \forall p \in \mathcal{P}, \quad k \in \mathcal{K}(p) \\
& x_t \in \{0, 1\}, \quad \forall t \in \mathcal{T} \\
& y_c \in \{0, 1\}, \quad \forall c \in \mathcal{C}.
\end{align*}
\]

\[
\mathcal{K}(p) = \{ x_p^c \}_{c \in \mathcal{P}} \cup \{ x_t^t \}_{t \in \mathcal{T}}
\]

\[
\mathcal{C}_{p,k} = \{ c \in \mathcal{C} : x_p^c = k \}
\]

\[
\mathcal{T}_{p,k} = \{ t \in \mathcal{T} : x_t^t = k \}
\]

- Solve time for truncated Lalonde CPS 429 = 0.006 s
- Scalability? Size = \(|\mathcal{P}| \times (|\mathcal{T}| + |\mathcal{C}|)\)
2. Modeling Piecewise Linear Functions

An appropriate way of modeling a piecewise linear function $f: \mathbb{D} \to \mathbb{R}^n$ is to model its epigraph given by $\text{epi}(f) = \{(x, z) \in \mathbb{D} \times \mathbb{R}: f(x) \leq z\}$. For example, the epigraph of the function in Figure 2(a) is depicted in Figure 2(b).

For simplicity, we assume that the function domain $\mathbb{D}$ is bounded and $f$ is only used in a constraint of the form $f(x) \leq 0$ or as an objective function that is being minimized. We then need a model of $\text{epi}(f)$ since $f(x) \leq 0$ can be modeled as $(x, z) \in \text{epi}(f), z \leq 0$ and the minimization of $f$ can be achieved by minimizing $z$ subject to $(x, z) \in \text{epi}(f)$. For continuous functions we can also work with its graph, but modeling the epigraph will allow us to extend most of the results to some discontinuous functions and will simplify the analysis of formulation properties.

### Simple Formulation for Univariate Functions

$$z = f(x)$$

$$\begin{pmatrix} x \\ z \end{pmatrix} = \sum_{j=1}^{5} \begin{pmatrix} d_j \\ f(d_j) \end{pmatrix} \lambda_j$$

$$1 = \sum_{j=1}^{5} \lambda_j, \quad \lambda_j \geq 0$$

$$y \in \{0, 1\}^4, \quad \sum_{i=1}^{4} y_i = 1$$

$$0 \leq \lambda_1 \leq y_1$$

$$0 \leq \lambda_2 \leq y_1 + y_2$$

$$0 \leq \lambda_3 \leq y_2 + y_3$$

$$0 \leq \lambda_4 \leq y_3 + y_4$$

$$0 \leq \lambda_5 \leq y_4$$
Modeling Piecewise Linear Functions

An appropriate way of modeling a piecewise linear function $f: \mathbb{D} \to \mathbb{R}^n$ is to model its epigraph given by $\text{epi}(f) = \{(x, z) \in \mathbb{D} \times \mathbb{R} : f(x) \leq z\}$. For example, the epigraph of the function in Figure 2(a) is depicted in Figure 2(b).

For simplicity, we assume that the function domain $\mathbb{D}$ is bounded and $f$ is only used in a constraint of the form $f(x) \leq 0$ or as an objective function that is being minimized. We then need a model of $\text{epi}(f)$ since $f(x) \leq 0$ can be modeled as $(x, z) \in \text{epi}(f), z \geq 0$ and the minimization of $f$ can be achieved by minimizing $z$ subject to $(x, z) \in \text{epi}(f)$.

For continuous functions we can also work with its graph, but modeling the epigraph will allow us to extend most of the results to some discontinuous functions and will simplify the analysis of formulation properties.

Advanced Formulation for Univariate Functions

\[
\begin{align*}
z &= f(x) \\
\begin{bmatrix} x \\ z \end{bmatrix} &= \sum_{j=1}^{5} \begin{bmatrix} d_j \\ f(d_j) \end{bmatrix} \lambda_j \\
1 &= \sum_{j=1}^{5} \lambda_j, \quad \lambda_j \geq 0
\end{align*}
\]

\[
y \in \{0, 1\}^2
\]

\[
\begin{align*}
0 &\leq \lambda_1 + \lambda_5 \leq 1 - y_1 \\
0 &\leq \lambda_3 \leq y_1 \\
0 &\leq \lambda_4 + \lambda_5 \leq 1 - y_2 \\
0 &\leq \lambda_1 + \lambda_2 \leq y_2
\end{align*}
\]

Size $= O(\log_2 \text{# of segments})$

Ideal: Integral Extreme Points
Extended Formulation for PWL Functions

\[ S = \text{gr} (f) = \bigcup_{i=1}^{k} \left\{ (x, z) \in \mathbb{R}^2 : \begin{array}{l} d_i \leq x \leq d_{i+1} \\ m_i x + c_i = z \end{array} \right\} \]

MC Formulation:

\[
\begin{align*}
    d_i y_i & \leq x^i \leq d_{i+1} y_i & \forall i \in [k] \\
    m_i x^i + c_i y_i & = z^i & \forall i \in [k] \\
    \sum_{i=1}^{k} x^i & = x \\
    \sum_{i=1}^{k} z^i & = z \\
    \sum_{i=1}^{k} y_i & = 1 \\
    y & \in \{0, 1\}^k
\end{align*}
\]
## Abstracting Univariate Functions

A continuous piecewise linear function and its epigraph as the union of polyhedra.

For simplicity, we assume that the function domain $D$ is bounded and $f$ is only used in a constraint of the form $f(x) \leq 0$ or as an objective function that is being minimized. We then need a model of $\text{epi}(f)$ since $f(x) \leq 0$ can be modeled as $(x, z) \in \text{epi}(f), z \leq 0$ and the minimization of $f$ can be achieved by minimizing $z$ subject to $(x, z) \in \text{epi}(f)$. For continuous functions we can also work with its graph, but modeling the epigraph will allow us to extend most of the results to some discontinuous functions and will simplify the analysis of formulation properties.

### General Formulations

- **$P_i$**
  - $P_i : = \{ \lambda \in \Delta^5 : \lambda_j = 0 \ \forall j \notin T_i \} d_j \lambda_j$
  - $T_i : = \{ i, i + 1 \} \ i \in \{ 1, \ldots, 4 \}$

- **$\Delta^5$**
  - $\Delta^5 = \{ f(d_1), f(d_2), f(d_3), f(d_4), f(d_5) \}$
  - $\lambda \in \bigcup_{i=1}^{4} P_i \subseteq \Delta^5$

- **Abstracting $z$**
  - $z = f(x)$
  - $\lambda_j = 0 \ \forall j \notin T_i$

- **Example**
  - $\sum_{j=1}^{5} (f(d_j)) \lambda_j = 1$ \(\lambda_j \geq 0\)
Abstraction Works for Multivariate Functions

\[ P_i := \left\{ \lambda \in \Delta^m : \lambda_j = 0 \quad \forall \nu_j \notin T_i \right\} \]

\[
\lambda \in \bigcup_{i=1}^{n} P_i \subseteq \Delta^m
\]
Complete Abstraction

- $\Delta^V := \left\{ \lambda \in \mathbb{R}^V_+ : \sum_{v \in V} \lambda_v = 1 \right\}$,
- $P_i = \left\{ \lambda \in \Delta^V : \lambda_v = 0 \quad \forall v \notin T_i \right\}$
- $\lambda \in \bigcup_{i=1}^{n} P_i$
- $T_i = \text{cliques of a graph}$
From Cliques to (Complement) Conflict Graph

SOS2
From Conflict Graph to Bi-clique Cover

Let $T_1, T_2, T_3, T_4$ be the conflict graph.

SOS2
From Bi-clique Cover to Formulation

SOS2

\[
0 \leq \lambda_1 + \lambda_5 \leq 1 - y_1 \\
0 \leq \lambda_3 \leq y_1 \\
0 \leq \lambda_4 + \lambda_5 \leq 1 - y_2 \\
0 \leq \lambda_1 + \lambda_2 \leq y_2
\]
Ideal Formulation from Bi-clique Cover

- Conflict Graph \( G = (V, E) \)

\[
E = \{ (u, v) : u, v \in V, u \neq v, \exists i \text{ s.t. } u, v \in T_i \}
\]

- Bi-clique cover \( \{(A^j, B^j)\}_{j=1}^t \), \( A^j, B^j \subseteq V \)

\[
\forall \{u, v\} \in E \exists j \text{ s.t. } u \in A^j \land v \in B^j
\]

- Formulation

\[
\sum_{v \in A^j} \lambda_v \leq 1 - y_j \quad \forall j \in [t]
\]

\[
\sum_{v \in B^j} \lambda_v \leq y_j \quad \forall j \in [t]
\]

\[
y \in \{0, 1\}^t
\]
Recursive Construction of Cover for SOS2, Step 1

Base case $n=2^1$:

Step 1 recursion:

Repeat for all bi-cliques from $2^{k-1}$ to cover all edges within first and last half of conflict graph.
Recursive Construction of Cover for SOS2, Step 2

Only edges missing are those between first and last half of conflict graph

Step 2: Add one more bi-clique

Cover has $\log_2 n$ bi-cliques.

For non-power of two just delete extra nodes.
Grid Triangulations: Step 1 = SOS2 for Inter-Box

Covers all arcs between boxes
Grid Triangulations: Step 2 = Ad-hoc Intra-Box

Covers all arcs within boxes

Sometimes 1 additional cover
Grid Triangulations: Step 2 = Ad-hoc Intra-Box

Sometimes 2 additional covers

Sometimes more, but always less than 9

Simple rules to get (near) optimal in Fall ‘16