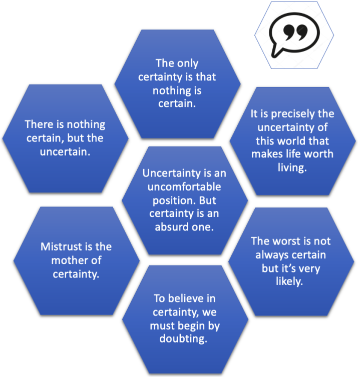
Methodological Advances in Two-stage Stochastic Programming



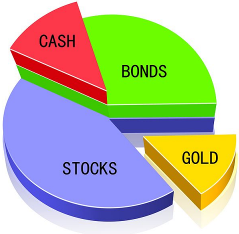
Merve Bodur

University of Edinburgh

School of Mathematics

Optimization Under Uncertainty

"The only certainty is that nothing is certain."

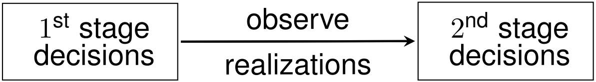


- Making decisions under uncertainty
- Integer/discrete decisions

$\Longrightarrow$ Stochastic programming

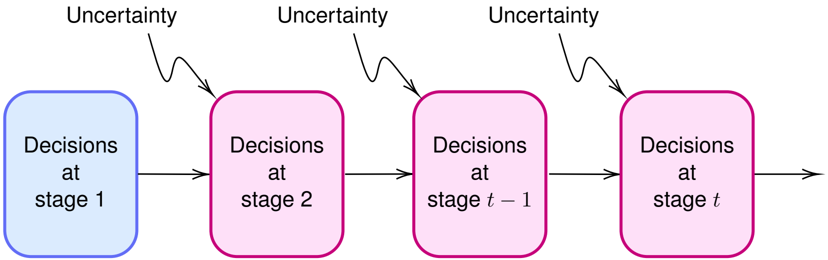
Two-Stage Stochastic Programs with Recourse

- There are uncertainties in some parameters (they are modeled as random variables)
- There are two decision stages



- Objective: min (1^{st} stage cost) + (Expected 2^{nd} stage cost)
- They have very wide range of applications

Sequential Decision-making Under Uncertainty



- Uncertainty is **gradually** observed
- Decisions are dynamically adapted

Can be approximated via **two-stage** models

General Two-Stage Stochastic Program

$$\min_{x} c^\top x + \mathbb{E}_\omega[Q(x,\omega)]$$
$$\text{s.t. } x \in \mathcal{X}$$

Example recourse/value function:

$$Q(x,\omega) = \begin{array}{ll} \min_{y} & q(\omega)^\top y \\ \text{s.t.} & T(\omega)x + W(\omega)y \geq h(\omega) \\ & y \geq 0 \end{array}$$

Challenges:

- Difficult to evaluate the expected value
  $\Rightarrow$ Use sample average approximation (SAA)
- **SAA problem** $\rightarrow$ Deterministic, but still difficult to solve

Assumption: Finitely many scenarios ($\mathcal{K}$ with $K = |\mathcal{K}|$)

Assumption: Relatively complete recourse

Extensive Form

A (very) large-scale (e.g., mixed-integer) program.

$$\begin{aligned} \min\ & c^\top x + \sum_{k \in \mathcal{K}} p_k q_k^\top y_k \\ \text{s.t. } & T_k x + W_k y_k \geq h_k \quad && \forall k \in \mathcal{K} \\ & x \in \mathcal{X} \\ & y_k \geq 0 && \forall k \in \mathcal{K} \end{aligned}$$

$\Rightarrow$ First classical approach: Try our favourite solver!
(Be prepared for some disappointment)

Extensive form does not scale well with K

Call center staffing and scheduling instances: [B. & Luedtke, 2016]

($I = 5,\ J = 5,\ T = 34,\ S = 333$, Time Limit = 1 hour)

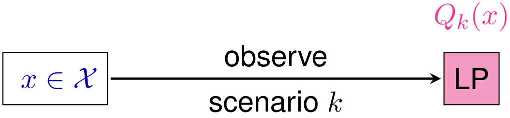
1835 integer variables + 30K continuous variables

K	Gap(%)	Nodes
100	2.8	1840
500	10.6	28
1000	16.6	0
1500	28.9	0
2000	32.8	0

- Usually solved via decomposition
- **This talk: Primal approach**
  (There are also dual approaches)

Continuous Recourse

(Recall: Finitely many scenarios, $k \in \mathcal{K}$)



- **Benders decomposition**
- Dual decomposition

Exploit the structure!

- Fix the first-stage variables ⇒ the rest decomposes by scenario

$$
\begin{array}{lccccccccccc}
\min & c^\top x & + & p_1 q_1^\top y_1 & + & p_2 q_2^\top y_2 & + & \cdots & + & p_K q_K^\top y_K & & \\
\text{s.t.} & T_1 x & + & W_1 y_1 & & & & & & & \geq & h_1 \\
 & T_2 x & & & + & W_2 y_2 & & & & & \geq & h_2 \\
 & \vdots & & & & & & \ddots & & & & \vdots \\
 & T_K x & & & & & & & + & W_K y_K & \geq & h_K \\
 & x \in \mathcal{X} & & y_1 \geq 0 & & y_2 \geq 0 & & & & y_K \geq 0 & &
\end{array}
$$

Key Idea

Benders Decomposition: Characterize the optimal value of scenario LPs as a function of first-stage variables x, i.e., $Q_k(x)$

$$\min\{c^\top x + \sum_{k \in \mathcal{K}} p_k Q_k(x) : x \in \mathcal{X}\}$$

Key Observation: $Q_k(\cdot)$ is a piecewise-linear convex function

Benders Decomposition (L-shaped Method)

$$\text{(MP)}: \min_{\eta,x} c^\top x + \sum_{k\in\mathcal{K}} p_k \eta_k$$

$$\text{s.t. } x \in \mathcal{X}$$

Benders Cuts for $\eta_k \geq Q_k(x)$ $\eta_k \geq \hat{\pi}^\top (h_k - T_k x)$ $\forall k \in \mathcal{K}$

$$\eta \in \mathbb{R}^K$$

$$\text{(SP)}^k : Q_k(\hat{x}) = \min_{y_k} q_k^\top y_k$$

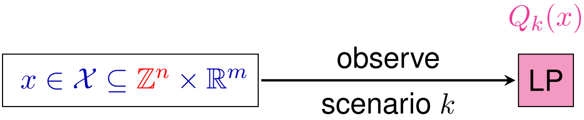
$$\text{s.t. } W_k y_k \geq h_k - T_k \hat{x} \quad (\hat{\pi})$$

$$y_k \in \mathbb{R}_+^J$$

- Subproblem decomposes by scenario $\rightarrow$ LPs

Single-cut version: $\sum_{k\in\mathcal{K}} p_k \eta_k \geq \sum_{k\in\mathcal{K}} p_k Q_k(x)$

Some Strengthening Ideas



Benders decomposition:

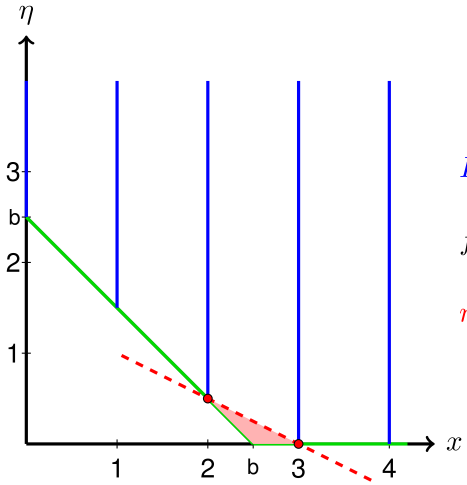
- Fast solution of LP subproblems :)
- Potentially weak bounds :(

Idea

Strengthen Benders with first-stage integrality-based cuts

- Add MIR cuts to MP [B. & Luedtke, 2016]
- Also add cuts to SPs [B., Dash, Günlük & Luedtke, 2016]

The basic mixed integer rounding inequality



$$H = \{(\eta, x) \in \mathbb{R}_+ \times \mathbb{Z}_+ \mid \eta + x \geq b\}$$

$$f = b - \lfloor b \rfloor > 0$$

$\eta \geq f(\lceil b \rceil - x)$ is valid for H

Mixed integer rounding (MIR)

Exact form of Benders cuts

$$H := \{(\eta, x) \in \mathbb{R}_+ \times \mathbb{Z}_+^I : \eta \geq d_0^1 - \sum_{i \in \mathcal{I}} d_i^1 x_i,\ \eta \geq d_0^2 - \sum_{i \in \mathcal{I}} d_i^2 x_i\}$$

MIR Cut [B. & Luedtke, 2016]

For any constant $\beta > 0$ with $\bar{f}_0 > 0$,

$$\eta \geq d_0^1 + \frac{\bar{f}_0 \lceil \beta(d_0^2 - d_0^1) \rceil}{\beta} - \sum_{i \in \mathcal{I}} \frac{\min\{\bar{f}_0 \lceil \beta(d_i^2 - d_i^1) \rceil, \bar{f}_i + \bar{f}_0 \lfloor \beta(d_i^2 - d_i^1) \rfloor\} + \beta d_i^1}{\beta} x_i$$

is valid for H where

$\bar{f}_0 := \beta(d_0^2 - d_0^1) - \lfloor \beta(d_0^2 - d_0^1) \rfloor$

$\bar{f}_i := \beta(d_i^2 - d_i^1) - \lfloor \beta(d_i^2 - d_i^1) \rfloor,\ \forall i \in \mathcal{I}.$

Applying MIR

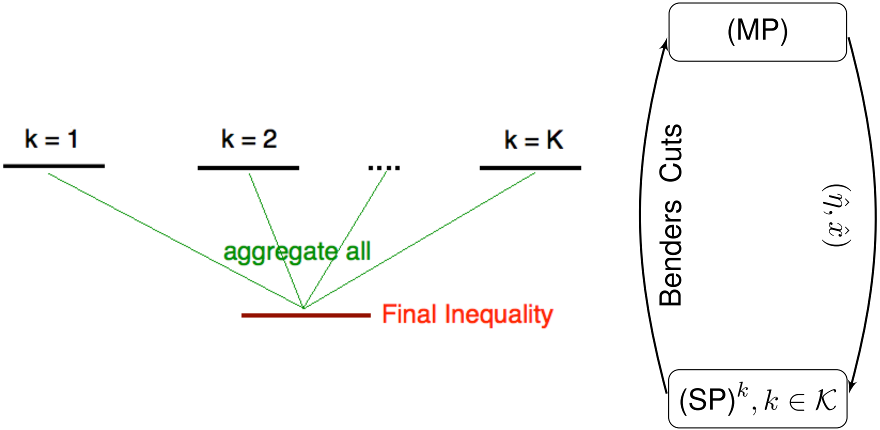
How we obtain MIR inequalities:

- Keep a pool of previously found Benders cuts
- Pair the current Benders cut with each previously found Benders cut and apply MIR

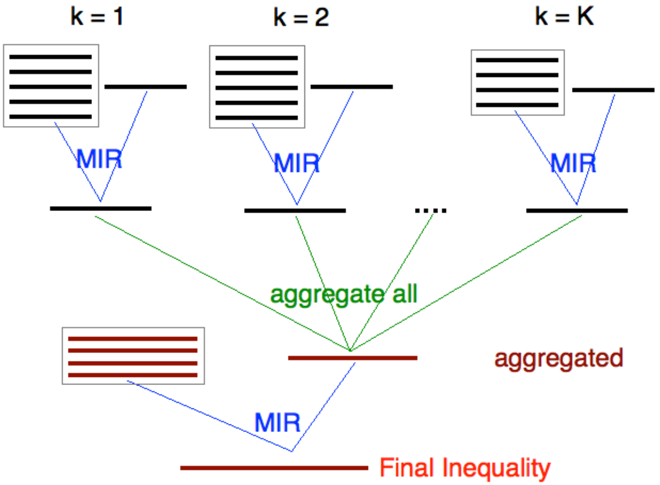
Can apply MIR in two different places:

1. Scenario level
2. Aggregated level

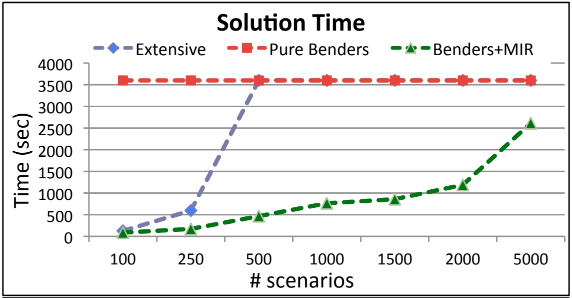
Benders single cut generation



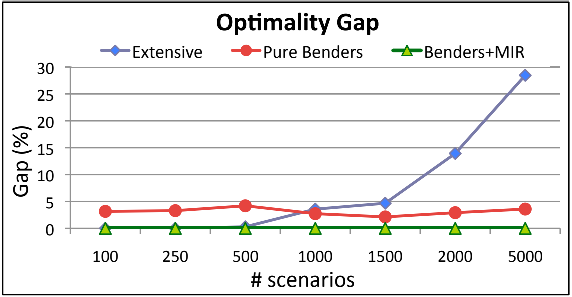
Cut generation using MIR



Call Center Staffing and Scheduling Instances



variables in Extensive Form:
534 + 42 · (# scenarios)



Average # nodes
in Benders+MIR = 750

Two Options for Using Integrality-based Cuts

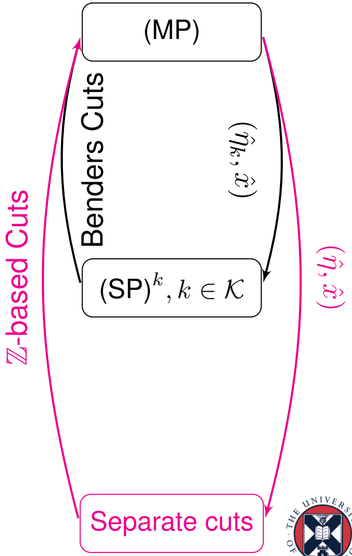
Strengthen Benders decomposition algorithm by:

- Project-and-cut: Add cuts to the master problem
- Cut-and-project: Add cuts to the subproblems

Project-and-cut

$$
\text{(MP)} : \min_{\eta, x} c^\top x + \sum_{k \in \mathcal{K}} p_k \eta_k
$$
$$
\text{s.t. } x \in \mathcal{X}
$$
Benders cuts
$$
\boxed{C\eta + Dx \geq g}
$$
$$
\eta \in \mathbb{R}^K
$$

$$
\text{(SP)}^k : Q_k(\hat{x}) := \min_{y_k} q_k^\top y_k
$$
$$
\text{s.t. } W_k y_k \geq h_k - T_k \hat{x}
$$
$$
y_k \in \mathbb{R}_+^J
$$



Cut-and-project

$$\text{(MP)} : \min_{\eta, x} c^\top x + \sum_{k \in \mathcal{K}} p_k \eta_k$$

$$\text{s.t. } x \in \mathcal{X}$$

Benders cuts

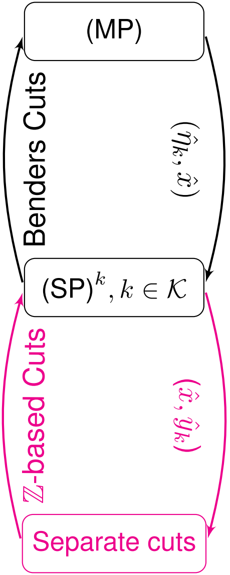
$$\eta \in \mathbb{R}^K$$

$$\text{(SP)}^k : Q_k(\hat{x}) := \min_{y_k} q_k^\top y_k$$

$$\text{s.t. } W_k y_k \geq h_k - T_k \hat{x}$$

$$C_k y_k \geq g_k - D_k \hat{x}$$

$$y_k \in \mathbb{R}^J_+$$



Add integrality-based cuts to (SP)^k, even though it is an LP

Capacitated Facility Location Instances

- $K = 500$, Time limit = 4 hours

	Avg Time (# unsolved)			
CAP #	EXT	BEN	MP	SP
101-104	1171	- (4)	- (4)	149
111-114	10787(3)	- (4)	- (4)	957
121-124	10935(3)	- (4)	- (4)	4738(1)
131-134	9512(3)	- (4)	- (4)	1527
Mean Time	6020	-	-	1008
Avg Opt Gap	1.64%	14.87%	15.53%	0.02%

$\therefore$ Cut-and-project has far more impact

- App on last-mile delivery with crowd-shipping and mobile depots: Also for the risk-averse (CVaR) case [Mousavi, B., & Roorda, 2022]

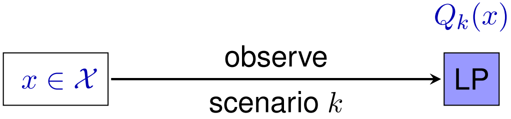
Network Interdiction Instances

- $K = 456$, Time limit = 4 hours

Budget	EXT	BEN	SP	MP	MP+SP
30	- (5)	639	442	183	415
40	- (5)	7915(3)	2253	784	830
50	- (5)	8626(3)	2328	512	867
60	- (5)	10599(4)	2425(1)	906	1121
70	- (5)	- (5)	4435(1)	1402	1389
80	- (5)	- (5)	10096(4)	1938	1579
90	- (5)	- (5)	13283(4)	4794	4050
Mean Time	-	7536	3188	980	1169
Avg Opt Gap	25.7%	2.6%	0.4%	-	-

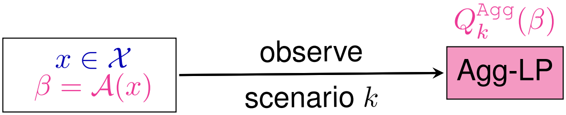
$\therefore$ Project-and-cut is very effective

Aggregation Cuts



Add Benders cuts to represent $\eta_k \geq Q_k(x)$

Aggregate second-stage constraints $\to$ change of variables: x to β



First add Benders cuts to represent $\eta_k \geq Q_k^{\text{Agg}}(\beta)$

Telecommunications Application

Stochastic RWA and Lightpath Rerouting [Daryalal & B., 2022]

$$x_{\ell,w}^{(s,d)} \xrightarrow[\text{future}]{\text{observe}} y_{\ell,w}^{(s,d)}$$

Which wavelinks are used to serve existing requests

Which wavelinks are used to serve future requests

Substitute

$$\sum_{(s,d)\in\mathcal{SD}_k^2} y_{\ell,w}^{(s,d)} \leq 1 - \sum_{(s,d)\in\mathcal{SD}^1} x_{\ell,w}^{(s,d)} \quad \forall w \in \mathcal{W}, \ell \in \mathcal{L}$$

via aggregation over wavelengths, with

$$\sum_{w\in\mathcal{W}} \sum_{(s,d)\in\mathcal{SD}_k^2} y_{\ell,w}^{(s,d)} \leq (|\mathcal{W}| - \beta_\ell) \quad \forall \ell \in \mathcal{L}$$

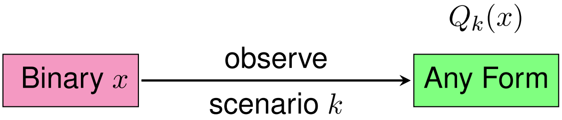
Optical Network Instances

Time limit = 10 minustes

K	EXTENSIVE		BENDERS-x			BENDERS-$x\beta$			
	time (s)	gap (%)	time (s)	gap (%)	$\#x-$cuts	time (s)	gap (%)	$\#\beta-$cuts	$\#x-$cuts
10	200	0	187	3	1124	17	0	8	15
20	TL	NA	465	2	1370	42	0	19	61
30	TL	NA	536	3	1538	157	0	31	204
40	TL	NA	TL	3	1785	164	0	34	203
50	TL	NA	TL	3	1980	206	0	58	209
100	TL	NA	TL	3	2164	305	0	99	400

Integer L-shaped Method

[Laporte & Louveaux, 1993]



- MP provides a candidate: $(\hat{x}, \hat{\eta})$
- $(\text{SP})^k$ evaluates $Q_k(\hat{x})$
- Integer L-shaped cut:

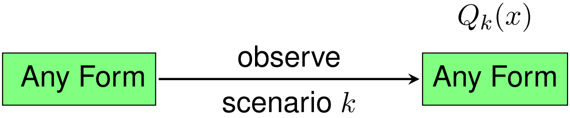
$$x = \hat{x} \Rightarrow \eta_k \geq Q_k(\hat{x})$$

$$x \neq \hat{x} \Rightarrow \eta_k \geq Q_k^{\text{LB}} \qquad \text{(i.e., redundant)}$$

$$\eta_k \geq Q_k(\hat{x}) + \left(Q_k(\hat{x}) - Q_k^{\text{LB}}\right)\left(\sum_{i:\hat{x}_i=1}(x_i - 1) - \sum_{i:\hat{x}_i=0} x_i\right)$$

- Cuts might be strengthened using problem-specific structure

Logic-based Benders Decomposition



- LBBD cuts from the inference dual [Hooker & Ottosson, 2003]

- Very successful applications in IP
- Few applications in SP

Distributed Operating Room Scheduling

[Guo et al., 2021]

1st stage Assign patients to (hospital, room, day)	→ observe surgery durations →	2nd stage Cancel patients if there is overtime

Objective: min (Operational cost) + (Expected cancellation cost)

$$Q^k_{hdr}(\hat{x}) = \min_z \sum_{p \in \mathcal{P}} c_p^{\text{cancel}} (\hat{x}_{hdpr} - z^k_{hdpr})$$

$$\text{s.t.} \quad z^k_{hdpr} \leq \hat{x}_{hdpr} \qquad p \in \mathcal{P}$$

$$\sum_{p \in \mathcal{P}} T^k_p z^k_{hdpr} \leq B_{hd}$$

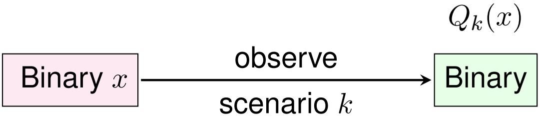
$$z^k_{hdpr} \in \{0,1\} \qquad p \in \mathcal{P}$$

LBBD cut:

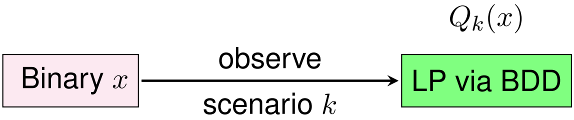
$$\eta^k_{hdr} \geq Q^k_{hdr}(\hat{x}) - \sum_{p \in \hat{\mathcal{P}}_{hdr}} c_p^{\text{cancel}} \left(1 - x_{hdpr}\right)$$

where $\hat{\mathcal{P}}_{hdr} = \{p \in \mathcal{P} | \hat{x}_{hdpr} = 1\}$

Convexification via Binary Decision Diagrams



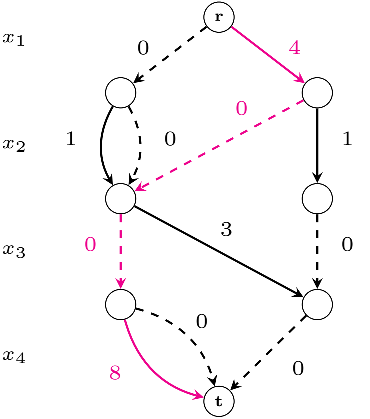
- Represent the second-stage problem via BDDs



$\Rightarrow$ amenable to Benders decomposition

Knapsack BDD Example

$$\max_x \; 4x_1 + x_2 + 3x_3 + 8x_4$$
$$\text{s.t. } 2x_1 + x_2 + 3x_3 + 3x_4 \le 5$$
$$x \in \{0,1\}^4$$



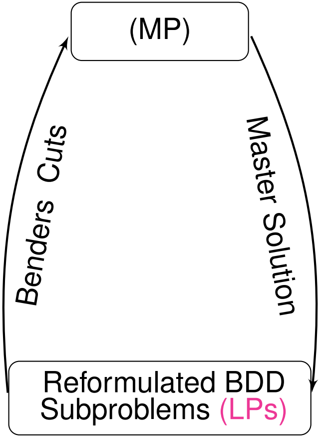
$$\max_f \sum_{a \in \mathcal{A}} w_a f_a$$
$$\text{s.t. } \sum_{a | s(a) = r} f_a = 1$$
$$\sum_{a | s(a) = i} f_a - \sum_{a | d(a) = i} f_a = 0 \quad \forall i \in \mathcal{N} \setminus \{r, t\}$$
$$\sum_{a | d(a) = t} f_a = -1$$
$$f_a \ge 0 \qquad \forall a \in \mathcal{A}$$

BDD-based Approach

[Lozano & Smith, 2018]

- Assume special structure: Each recourse constraint is impacted by at most one first-stage binary variable
- Transform recourse problem to a capacitated shortest path problem
- Derive classical Benders cuts

  (can be easily strengthened)



The Transformed BDD Subproblem

$$\min \sum_{p\in\mathcal{P}} c_p^{\text{cancel}}(\hat{x}_{hdpr} - z^k_{hdpr})$$

$$\text{s.t.} \sum_{p\in P} T^k_p z^k_{hdpr} \le B_{hd}$$

$$z^k_{hdpr} \le \hat{x}_{hdpr} \qquad p\in\mathcal{P}$$

$$z^k_{hdpr} \in \{0,1\} \qquad p\in\mathcal{P}$$

→ A knapsack problem

$$\min \sum_{a\in\mathcal{A}^k} g^k_a f_a$$

$$\text{s.t.} \sum_{a|s(a)=r} f_a = 1$$

$$\sum_{a|s(a)=i} f_a - \sum_{a|d(a)=i} f_a = 0 \quad i\in\mathcal{N}^k\setminus\{r,t\}$$

$$\sum_{a|d(a)=t} f_a = -1$$

$$f_a \le \hat{x}_{hdpr} \qquad a\in\mathcal{A}^{sp}_1$$

$$f_a \ge 0 \qquad a\in\mathcal{A}^k$$

→ A shortest path problem

Further Leveraging Binary Decision Diagrams

- Previously: [Lozano & Smith, 2018]

$$Q_k(x) = \min \ q_k^\top y$$
$$\text{s.t. } y \in \mathcal{Y}_k \subseteq \{0,1\}^{n_\text{Y}}, \quad x_i^\text{B} = 0 \Longrightarrow y \in \mathcal{Y}_i^{\texttt{logical},k} \quad \forall i = 1, \dots, n_\text{x}^\text{B}$$

- More recently: [MacNeil & B., 2024]

$$Q_k(x) = \min q_k^\top y$$
$$y \in \mathcal{Y}_k \subseteq \{0,1\}^{n_\text{Y}}, \quad \mathbb{I}(L_j^k(x)) = 1 \Longrightarrow y \in \mathcal{Y}_j^{\texttt{logical},k} \quad \forall j = 1, \dots, m$$

and

$$Q_k(x) = \min \ (q_k^1 + q_k^2)^\top y$$
$$y \in \mathcal{Y}_k \subseteq \{0,1\}^{n_\text{Y}}, \quad \mathbb{I}(L_j^k(x)) = 1 \Longrightarrow q_{k,\sigma(j)=0}^1 \quad \forall j = 1, \dots, m$$

New BDD-based Approaches

Model 1:

- Generalizes the existing BDD-based decomposition approach
- **Arc capacities** in the BDDs are parametrized by x

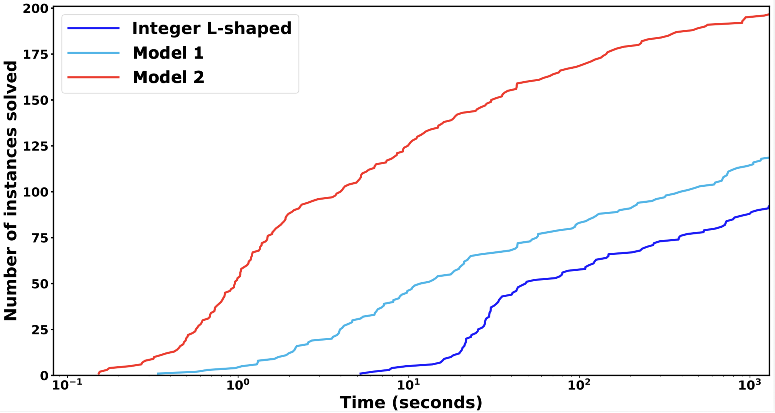
Model 2:

- Novel; might be more natural for certain applications
- **Arc costs** in the BDDs are parametrized by x

(They are extended to a risk-averse (CVaR) setting as well)

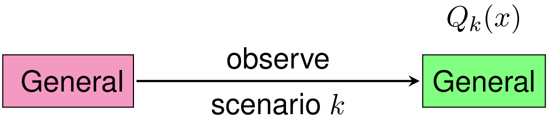
Dominating Set Instances

- Up to 50 vertices, varying edge densities, 850 scenarios
- Solution time limit = 1 hour (i.e., after BDD generation)



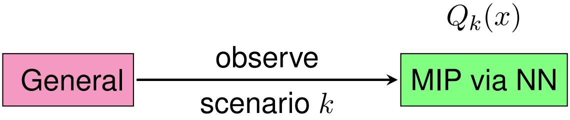
Neur2SP

[Dumouchelle, Patel, Khalil, & B., 2022]



The idea: Learn to get a monolithic formulation

- Learn $Q_k(x)$ or even better $\mathbb{E}_k[Q_k(x)]$ via supervised learning
- MIPify the obtained neural network (NN)
- Solve the combined surrogate model



Scenario Reduction

[Keutchayan, Ortmann, & Rei, 2021]

- **Distribution-driven**
  - Cluster scenario vectors: $\{\xi_1, \xi_2, \ldots, \xi_K\}$
  - For each cluster, pick a representative

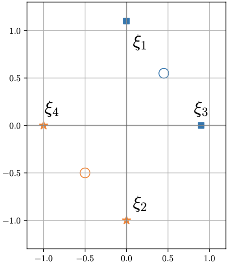
- **Problem-driven**
  - Solve single-scenario problems

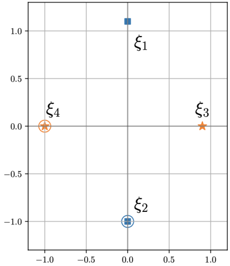
$$x_k^* \in \arg\min_{x \in \mathcal{X}} c^\top x + Q(x, \xi_k)$$

  - Evaluate obtained first-stage solutions under the other individual scenarios:

$$V_{k,k'} := c^\top x_k^* + Q(x_k^*, \xi_{k'})$$

  - Solve a clustering MIP model, minimizing a discrepancy measure based on the V values





Summary

- Benders-like decomposition techniques are effective
- Can be strengthened/leveraged further by incorporating:
  - Some IP technology, e.g.:
    - Integrality-based cuts
    - Logic-based cuts
  - New tools, e.g.:
    - Decision diagrams
    - Machine learning

- An introduction to two-stage stochastic mixed-integer programming [Küçükyavuz & Sen, 2017]
- Stochastic mixed-integer programming: A survey [Romeijnders, Zhang & Sen, 2025]
- A review on the performance of linear and mixed integer two-stage stochastic programming software [Torres, Li, Apap & Grossmann, 2022]