Mixed-Integer Programming for Stochastic Optimization

Simge Küçükyavuz

Northwestern University

Acknowledgments

Collaborators

- Binyuan Chen
- Dinakar Gade
- Nam Ho-Nguyen
- Ruiwei Jiang
- Fatma Kılınç-Karzan
- Dabeen Lee
- Xiao Liu
- Jim Luedtke
- Merve Meraklı
- Nilay Noyan
- Suvrajeet Sen
- Yongjia Song
- Hao-Hsiang Wu
- Minjiao Zhang

Grants

- National Science Foundation #1907463, #1732364, #1100383, #0917952
- Office of Naval Research #N00014-19-1-2321

Agenda

In the next two days, we will discuss

- Two-stage stochastic mixed-integer programs (MIPs):
  - Large-scale MIPs
  - How to decompose?
  - Desirable algorithmic properties: Finite convergence, scalability
- Other stochastic (continuous) optimization problems
  - Risk measures/distributional ambiguity modeled as MIPs
  - Exploit combinatorial structure for improved formulations
- Theory, algorithm design, computations, and (some) applications.

Outline

1. Two-Stage Stochastic Integer Programming
   - Two-Stage Stochastic Linear Programming
   - Classification Scheme for Two-Stage Stochastic Mixed-Integer Programming
   - Two-Stage Stochastic Pure Integer Programming
   - Two-Stage Stochastic Mixed-Integer Programming

2. Chance-Constrained Programming
   - Static Joint Chance-Constrained Programming
   - Two-stage (Dynamic) Chance-Constrained Programming
   - Distributionally Robust Chance-Constrained Programming

Outline

Motivation and Scope

Motivation:

- Large capital investment decisions must hedge against uncertain future
- First stage: Strategic decisions (Warehouse/data center/power generator locations)
- Second stage: Operational decisions (Shipments/routing/distribution)
- Applications: Energy, telecommunications, healthcare, supply chain, finance ...

Motivation and Scope

Motivation:

- Large capital investment decisions must hedge against uncertain future
- First stage: Strategic decisions (Warehouse/data center/power generator locations)
- Second stage: Operational decisions (Shipments/routing/distribution)
- Applications: Energy, telecommunications, healthcare, supply chain, finance ...

Scope:

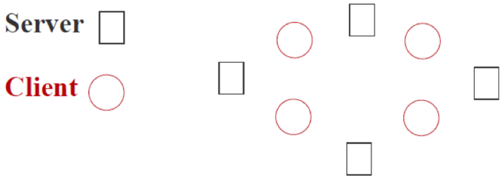
- Focus on Benders type methods
- Will not cover other methods such as Lagrangian relaxation, column generation, etc.

An Example: Stochastic Server Location and Sizing (SSLS)

Applications:

- Preparation and execution of disaster plans
- Location and sizing of data centers in cloud computing
- Supply chain planning with disruptions
- Battery charging infrastructure for electric vehicles

Planning Locations to Hedge Against Demand Uncertainty



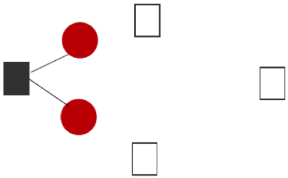
There are two sets of decisions:

- First stage: Determine data center locations (binary) and number of servers to locate (general integer)
- Second stage (once random demand is realized): Allocate servers to customers
- Constraints: capacity, demand satisfaction, etc.

Deterministic Server Location Problem

● Observed demand nodes, ■ Optimal server location

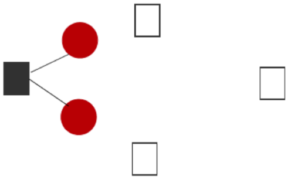
Scenario 1:



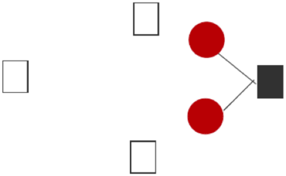
Deterministic Server Location Problem

● Observed demand nodes, ■ Optimal server location

Scenario 1:



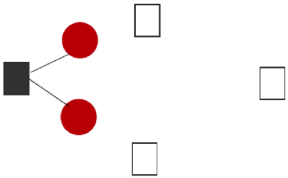
Scenario 2:



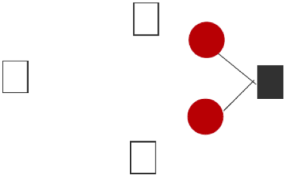
Deterministic Server Location Problem

● Observed demand nodes, ■ Optimal server location

Scenario 1:



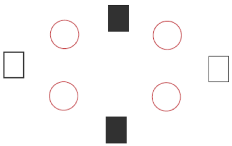
Scenario 2:



Suppose each scenario is equally likely? What is the optimal server location plan?

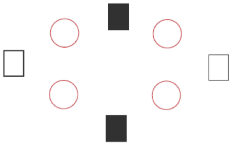
Stochastic Server Location Problem

Hedged Optimal Solution



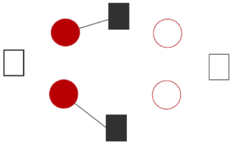
Stochastic Server Location Problem

Hedged Optimal Solution



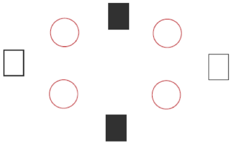
Dynamic Response to Demands/Threats

Scenario 1:



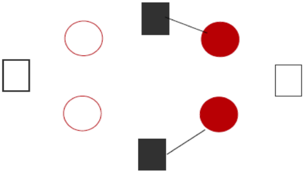
Stochastic Server Location Problem

Hedged Optimal Solution



Dynamic Response to Demands/Threats

Scenario 2:



Outline

Standard (Risk-Neutral) Stochastic Programming Formulation

- $\tilde{\omega}$: a random vector with support Ω
- Order of events:

Standard (Risk-Neutral) Stochastic Programming Formulation

- $\tilde{\omega}$: a random vector with support Ω
- Order of events: $x \rightarrow$

Standard (Risk-Neutral) Stochastic Programming Formulation

- $\tilde{\omega}$: a random vector with support Ω
- Order of events: $x \to \omega$

Standard (Risk-Neutral) Stochastic Programming Formulation

- $\tilde{\omega}$: a random vector with support Ω
- Order of events: $x \rightarrow \omega \rightarrow$

Standard (Risk-Neutral) Stochastic Programming Formulation

- $\tilde{\omega}$: a random vector with support Ω
- Order of events: $x \rightarrow \omega \rightarrow y(\omega)$

Standard (Risk-Neutral) Stochastic Programming Formulation

- $\tilde{\omega}$: a random vector with support Ω
- Order of events: $x \to \omega \to y(\omega)$
- $x \in X := \{x \in \mathbb{R}_+^{n-n_1} \times \mathbb{Z}_+^{n_1} : Ax \geq b\}$: first-stage decision vector
- $y(\omega) \in \mathbb{R}_+^{n_2}$: second-stage decision vector for each ω
- $\mathcal{X}, \mathcal{Y}$: integer, continuous and sign restrictions on x, y, resp.

Standard (Risk-Neutral) Stochastic Programming Formulation

- $\tilde{\omega}$: a random vector with support Ω
- Order of events: $x \to \omega \to y(\omega)$
- $x \in X := \{x \in \mathbb{R}_+^{n-n_1} \times \mathbb{Z}_+^{n_1} : Ax \geq b\}$: first-stage decision vector
- $y(\omega) \in \mathbb{R}_+^{n_2}$: second-stage decision vector for each ω
- $\mathcal{X}, \mathcal{Y}$: integer, continuous and sign restrictions on x, y, resp.
  A two-stage stochastic program:

$$
\begin{aligned}
\min \quad & c^\top x + \mathbb{E}_{\tilde{\omega}}(h(x, \tilde{\omega})) \\
s.t. \quad & Ax \geq b, \\
& x \in \mathcal{X},
\end{aligned}
$$

Standard (Risk-Neutral) Stochastic Programming Formulation

- $\tilde{\omega}$: a random vector with support Ω
- Order of events: $x \to \omega \to y(\omega)$
- $x \in X := \{x \in \mathbb{R}_+^{n-n_1} \times \mathbb{Z}_+^{n_1} : Ax \geq b\}$: first-stage decision vector
- $y(\omega) \in \mathbb{R}_+^{n_2}$: second-stage decision vector for each ω
- $\mathcal{X}, \mathcal{Y}$: integer, continuous and sign restrictions on x, y, resp.
  A two-stage stochastic program:

$$\begin{aligned} \min \quad & c^\top x + \mathbb{E}_{\tilde{\omega}}(h(x, \tilde{\omega})) \\ s.t. \quad & Ax \geq b, \\ & x \in \mathcal{X}, \end{aligned}$$

where

$$\begin{aligned} h(x, \omega) = \min \quad & y_0 \\ & y_0 - g(\omega)^\top y = 0 \\ & W(\omega) y \geq r(\omega) - T(\omega) x \\ & y \in \mathcal{Y}. \end{aligned}$$

- All second stage data can be random $(T(\omega), W(\omega), r(\omega), g(\omega))$

Finite sample space assumption

- We consider the setting where Ω is a finite sample space:

$$\Omega = \{\omega^1, \ldots, \omega^N\}$$

Finite sample space assumption

- We consider the setting where Ω is a finite sample space:

$$\Omega = \{\omega^1, \ldots, \omega^N\}$$

- Even if Ω is not finite, we can approximate it via an empirical distribution (see the theory of Sample Average Approximation (SAA), e.g., [Shapiro et al., 2009].

Finite sample space assumption

- We consider the setting where Ω is a finite sample space:

$$\Omega = \{\omega^1, \ldots, \omega^N\}$$

- Even if Ω is not finite, we can approximate it via an empirical distribution (see the theory of Sample Average Approximation (SAA), e.g., [Shapiro et al., 2009].

- Often, N is very large.

Finite sample space assumption

- We consider the setting where Ω is a finite sample space:

$$\Omega = \{\omega^1, \dots, \omega^N\}$$

- Even if Ω is not finite, we can approximate it via an empirical distribution (see the theory of Sample Average Approximation (SAA), e.g., [Shapiro et al., 2009].

- Often, N is very large.

- Let $p_i \in [0,1]$: probability of scenario $\omega^i \in \Omega$, where $\sum_{i\in[N]} p_i = 1$.

Deterministic Equivalent Formulation

$$
\begin{array}{llllll}
\min & c^\top x & +p_1 g^\top(\omega^1)y(\omega^1) & +p_2 g^\top(\omega^2)y(\omega^2) + \cdots & +p_N g^\top(\omega^N)y(\omega^N) & \\
\text{s.t} & Ax & & & & \geq b \\
& T(\omega^1)x & +W(\omega^1)y(\omega^1) & & & \geq r(\omega^1) \\
& T(\omega^2)x & & +W(\omega^2)y(\omega^2) & & \geq r(\omega^2) \\
& \vdots & & \ddots & & \vdots \\
& T(\omega^N)x & & & +W(\omega^N)y(\omega^N) & \geq r(\omega^N) \\
& x \in \mathcal{X}, & y(\omega^i) \in \mathcal{Y}, i \in [N]. & & &
\end{array}
$$

It's HUGE!!!

Review of Benders Decomposition Algorithm

Algorithms for two-stage stochastic program with continuous second-stage variables:

Benders' decomposition [Benders, 1962], L-shaped method [van Slyke and Wets, 1969]

Master Problem MP^k at iteration $k = 0, 1, \ldots,$

$$\begin{aligned} \mathrm{MP}^k: \quad \min \quad & c^\top x + \sum_{\omega^i \in \Omega} p_i \eta_{\omega^i} \\ \mathrm{s.t} \quad & A^k(x, \eta) \geq b^k, \\ & x \in \mathcal{X} \end{aligned}$$

where η_j approximates the second-stage value function of scenario j.

- $A^k(x, \eta) \geq b^k$ includes:
  - $Ax \geq b$
  - Optimality cuts generated from the subproblems in iterations $j = 1, \ldots, k-1$
  - Feasibility cuts generated from the subproblems in iterations $j = 1, \ldots, k-1$

Subproblems

Subproblem $\mathrm{SP}^k(x,\omega)$, $\omega \in \Omega$ at iteration $k = 0, 1, \ldots,$

Given (x, η), the solution of the master problem at iteration k, solve for each ω:

$$\begin{aligned} \mathrm{SP}^k(x,\omega): \quad h^k(x,\omega) := \min \quad & g(\omega)^\top y(\omega) \\ \text{s.t} \quad & W(\omega)y(\omega) \geq r(\omega) - T(\omega)x, \\ & y(\omega) \in \mathbb{R}_+^{n_2}, \end{aligned}$$

Let ψ_ω^k be the dual vector of the subproblem $\mathrm{SP}^k(x,\omega)$.

Subproblems

Subproblem $\mathrm{SP}^k(x, \omega)$, $\omega \in \Omega$ at iteration $k = 0, 1, \ldots,$

Given (x, η), the solution of the master problem at iteration k, solve for each ω:

$$\begin{aligned} \mathrm{SP}^k(x,\omega): \quad h^k(x,\omega) := \min \quad & g(\omega)^\top y(\omega) \\ \text{s.t} \quad & W(\omega)y(\omega) \geq r(\omega) - T(\omega)x, \\ & y(\omega) \in \mathbb{R}_+^{n_2}, \end{aligned}$$

Let ψ_ω^k be the dual vector of the subproblem $\mathrm{SP}^k(x, \omega)$.

- If $\mathrm{SP}^k(x, \omega)$ is feasible, but $\eta_\omega < h^k(x, \omega)$, then add the optimality cut

$$\eta_\omega \geq {\psi_\omega^k}^\top (r(\omega) - T(\omega)x)$$

- If $\mathrm{SP}^k(x, \omega)$ is infeasible, then its dual is unbounded, so using the corresponding dual ray ψ_ω^k, add the feasibility cut

$$0 \geq {\psi_\omega^k}^\top (r(\omega) - T(\omega)x)$$

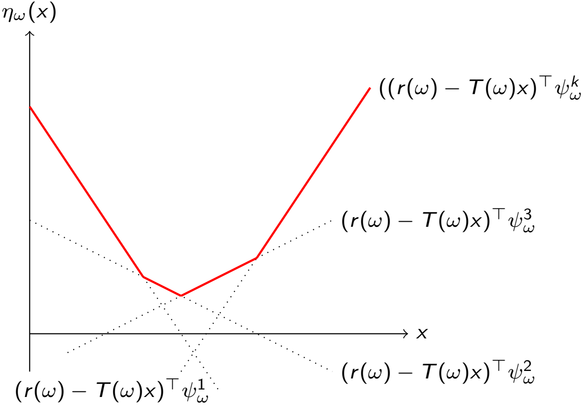


Figure 1: Piecewise-linear function, $\eta_\omega(x)$, for continuous recourse

Outline

1. Two-Stage Stochastic Integer Programming
   - Two-Stage Stochastic Linear Programming
   - Classification Scheme for Two-Stage Stochastic Mixed-Integer Programming
   - Two-Stage Stochastic Pure Integer Programming
   - Two-Stage Stochastic Mixed-Integer Programming

2. Chance-Constrained Programming
   - Static Joint Chance-Constrained Programming
   - Two-stage (Dynamic) Chance-Constrained Programming
   - Distributionally Robust Chance-Constrained Programming

Classification Scheme For Stochastic MIPs

$$
\begin{aligned}
B &= \text{Stages with Binary decision variables} \\
C &= \text{Stages with Continuous decision variables} \\
D &= \text{Stages with Discrete (general integer) decision variables.}
\end{aligned}
$$

For example, two-stage stochastic MIP with continuous recourse has: $B = D = \{1\}, C = \{1, 2\}$.

Literature Overview

	First-stage	Second-stage
Laporte and Louveaux (1993) Sen and Sherali (2006)	Binary	Mixed-integer
Carøe and Tind (1997) Sherali and Zhu (2007)	Mixed-binary	Mixed-binary
Carøe and Tind (1998)	Mixed-integer	Integer
Schultz et al. (1998)	Continuous	Integer
Ahmed et al. (2004)	Mixed-binary	Integer
Sherali and Fraticelli (2002) Sen and Higle (2005) Ntaimo and Sen (2005, 2008) Ntaimo (2009)	Binary	Mixed-binary
Gade, K., Sen (2012)	Binary	Integer
Kong et al. (2006) Trapp et al. (2013) Zhang and K. (2014)	Integer	Integer
Qi and Sen (2017, 2021+)	Mixed-Integer	Mixed-Integer

Outline

A Two-Stage Stochastic Integer Program

Consider binary first stage and general integer second stage variables (i.e., B=$\{1,2\}$, D=$\{2\}$, C=$\emptyset$)

$$\begin{aligned} \min \quad & c^\top x + \mathbb{E}\left[h(x, \tilde{\omega})\right] \\ \text{s.t.} \quad & Ax \geq b \\ & x \in \mathbb{B}^n, \end{aligned}$$

where for a particular realization (scenario) ω of $\tilde{\omega}$, $h(x, \omega)$ is defined as

$$\begin{aligned} h(x, \omega) = \min \quad & y_0 \\ \text{s.t.} \quad & y_0 - g(\omega)^\top y = 0 \\ & W(\omega) y \geq r(\omega) - T(\omega) x \\ & y_0 \in \mathbb{Z}, y \in \mathbb{Z}_+^{n_2} \end{aligned}$$

A Two-Stage Stochastic Integer Program

Consider binary first stage and general integer second stage variables (i.e., B={1,2}, D={2}, C=$\emptyset$)

$$\begin{aligned} \min \quad & c^\top x + \mathbb{E}\left[h(x, \tilde{\omega})\right] \\ \text{s.t.} \quad & Ax \geq b \\ & x \in \mathbb{B}^n, \end{aligned}$$

where for a particular realization (scenario) ω of $\tilde{\omega}$, $h(x, \omega)$ is defined as

$$\begin{aligned} h(x, \omega) = \min \quad & y_0 \\ \text{s.t.} \quad & y_0 - g(\omega)^\top y = 0 \\ & W(\omega)y \geq r(\omega) - T(\omega)x \\ & y_0 \in \mathbb{Z}, y \in \mathbb{Z}_+^{n_2} \end{aligned}$$

- $\tilde{\omega}$ is defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and has finite support
- $Y(x, \omega) := \{y_0 \in \mathbb{Z}, y \in \mathbb{Z}_+^{n_2} : y_0 - g(\omega)^\top y = 0, W(\omega)y \geq r(\omega) - T(\omega)x\}$.

A Two-Stage Stochastic Integer Program

Consider binary first stage and general integer second stage variables (i.e., B={1,2}, D={2}, C=∅)

$$\begin{aligned} \min \quad & c^\top x + \mathbb{E}\left[h(x,\tilde{\omega})\right] \\ \text{s.t.} \quad & Ax \geq b \\ & x \in \mathbb{B}^n, \end{aligned}$$

where for a particular realization (scenario) ω of $\tilde{\omega}$, $h(x,\omega)$ is defined as

$$\begin{aligned} h(x,\omega) = \min \quad & y_0 \\ \text{s.t.} \quad & y_0 - g(\omega)^\top y = 0 \\ & W(\omega)y \geq r(\omega) - T(\omega)x \\ & y_0 \in \mathbb{Z}, y \in \mathbb{Z}_+^{n_2} \end{aligned}$$

- $\tilde{\omega}$ is defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and has finite support
- $Y(x,\omega) := \{y_0 \in \mathbb{Z}, y \in \mathbb{Z}_+^{n_2} : y_0 - g(\omega)^\top y = 0, W(\omega)y \geq r(\omega) - T(\omega)x\}$.
- Relatively complete recourse

A Two-Stage Stochastic Integer Program

Consider binary first stage and general integer second stage variables (i.e., B={1,2}, D={2}, C=∅)

$$\begin{aligned} \min \quad & c^\top x + \mathbb{E}\left[h(x, \tilde{\omega})\right] \\ \text{s.t.} \quad & Ax \geq b \\ & x \in \mathbb{B}^n, \end{aligned}$$

where for a particular realization (scenario) ω of $\tilde{\omega}$, $h(x, \omega)$ is defined as

$$\begin{aligned} h(x, \omega) = \min \quad & y_0 \\ \text{s.t.} \quad & y_0 - g(\omega)^\top y = 0 \\ & W(\omega) y \geq r(\omega) - T(\omega) x \\ & y_0 \in \mathbb{Z}, y \in \mathbb{Z}_+^{n_2} \end{aligned}$$

- $\tilde{\omega}$ is defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and has finite support
- $Y(x, \omega) := \{y_0 \in \mathbb{Z}, y \in \mathbb{Z}_+^{n_2} : y_0 - g(\omega)^\top y = 0, W(\omega) y \geq r(\omega) - T(\omega) x\}.$
- Relatively complete recourse
- SIP has a finite optimum

Problem Structure

Deterministic Equivalent of SIP

$$
\begin{array}{llllll}
\min & c^\top x & +p_1 g(\omega^1)^\top y(\omega^1) + p_2 g(\omega^2)^\top y(\omega^2) & + \cdots + & p_N g(\omega^N)^\top y(\omega^N) & \\
& Ax & & & & \geq b \\
& T(\omega^1)x & +W(\omega^1)y(\omega^1) & & & \geq r(\omega^1) \\
& T(\omega^2)x & \quad + W(\omega^2)y(\omega^2) & & & \geq r(\omega^2) \\
& \vdots & & \ddots & & \vdots \\
& T(\omega^N)x & & & +W(\omega^N)y(\omega^N) & \geq r(\omega^N)
\end{array}
$$

$$x \in \mathbb{B}^n, y(\omega) \in \mathbb{Z}^{n_2}, \omega \in \Omega.$$

- Large-scale integer program
- For a fixed $x \in X$, SIP decomposes by scenario

Value Function Reformulation and Challenges

- Recall $X \cap \mathcal{X} = \{x \in \mathbb{B}^n : Ax \geq b\}$.
- Standard approach in L-shaped decomposition is the value function reformulation of SIP:

$$\min_{x \in X \cap \mathcal{X}} \{c^\top x + \eta : \eta \geq \mathcal{Q}(x)\}, \qquad \mathcal{Q}(x) := \mathbb{E}(h(x, \tilde{\omega}))$$

Value Function Reformulation and Challenges

- Recall $X \cap \mathcal{X} = \{x \in \mathbb{B}^n : Ax \geq b\}$.

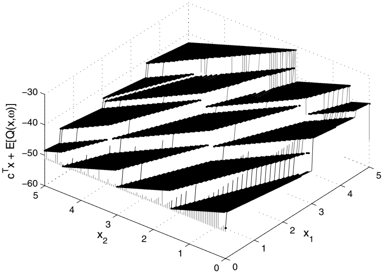
- Standard approach in L-shaped decomposition is the value function reformulation of SIP:

$$\min_{x \in X \cap \mathcal{X}} \{c^\top x + \eta : \eta \geq \mathcal{Q}(x)\}, \qquad \mathcal{Q}(x) := \mathbb{E}(h(x, \tilde{\omega}))$$

- If second stage is a linear program $\rightarrow h(\cdot, \omega)$, $\omega \in \Omega$: value function of an LP. It is piecewise linear and convex. Benders' decomposition and L-Shaped decomposition exploit this property.

Challenge for SIP

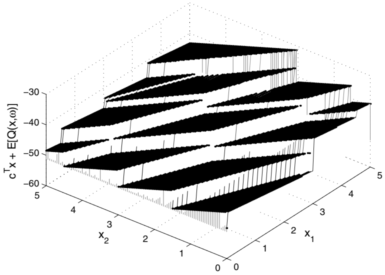
If second stage is an integer program, then $h(\cdot, \omega)$: value function of an integer program [Blair and Jeroslow, 1982]. It is non-linear & non-convex.



From [Ahmed et al., 2004]

Challenge for SIP

If second stage is an integer program, then $h(\cdot, \omega)$: value function of an integer program [Blair and Jeroslow, 1982]. It is non-linear & non-convex.



From [Ahmed et al., 2004]

How to create "good" lower bounding approximations practically?

L-Shaped Algorithms for 2-Stage SMIP - Literature

- Integer L-shaped method [Laporte and Louveaux, 1993]: Binary first stage, mixed-integer second stage - First stage B&B and linear optimality cuts. Solve second stage MIPs to optimality. Improved in [Angulo et al., 2016]

L-Shaped Algorithms for 2-Stage SMIP - Literature

- Integer L-shaped method [Laporte and Louveaux, 1993]: Binary first stage, mixed-integer second stage - First stage B&B and linear optimality cuts. Solve second stage MIPs to optimality. Improved in [Angulo et al., 2016]

- Disjunctive Cuts for mixed-binary second stage: e.g., [Carøe and Tind, 1997], [Sherali and Fraticelli, 2002], [Sen and Higle, 2005], [Sen and Sherali, 2006], [Ntaimo and Sen, 2007], [Ntaimo, 2009].

L-Shaped Algorithms for 2-Stage SMIP - Literature

- Integer L-shaped method [Laporte and Louveaux, 1993]: Binary first stage, mixed-integer second stage - First stage B&B and linear optimality cuts. Solve second stage MIPs to optimality. Improved in [Angulo et al., 2016]

- Disjunctive Cuts for mixed-binary second stage: e.g., [Carøe and Tind, 1997], [Sherali and Fraticelli, 2002], [Sen and Higle, 2005], [Sen and Sherali, 2006], [Ntaimo and Sen, 2007], [Ntaimo, 2009].

- Computations: e.g., [Laporte et al., 2002], [Ntaimo and Sen, 2005, 2008], [Yuan and Sen, 2009], [Ntaimo and Tanner, 2008].

L-Shaped Algorithms for 2-Stage SMIP - Literature

- Integer L-shaped method [Laporte and Louveaux, 1993]: Binary first stage, mixed-integer second stage - First stage B&B and linear optimality cuts. Solve second stage MIPs to optimality. Improved in [Angulo et al., 2016]
- Disjunctive Cuts for mixed-binary second stage: e.g., [Carøe and Tind, 1997], [Sherali and Fraticelli, 2002], [Sen and Higle, 2005], [Sen and Sherali, 2006], [Ntaimo and Sen, 2007], [Ntaimo, 2009].
- Computations: e.g., [Laporte et al., 2002], [Ntaimo and Sen, 2005, 2008], [Yuan and Sen, 2009], [Ntaimo and Tanner, 2008].
- Global Optimization and other approaches for pure integer second stage: e.g., [Ahmed et al., 2004], [Kong et al., 2006], [Schultz et al., 1998], [Schultz and Hemmecke, 2003],[Klein, 2020]
- Gomory cuts for SMIP: [Carøe and Tind, 1998]

Gomory Fractional Cuts (GFC) for Deterministic Pure IPs

- Given first-stage vector $\bar{x}$, solve the LP relaxation of the second-stage IP with simplex.
- Let $\mathcal{B}, \mathcal{N}$ - Basic and nonbasic column index sets of LP.

Gomory Fractional Cuts (GFC) for Deterministic Pure IPs

- Given first-stage vector $\bar{x}$, solve the LP relaxation of the second-stage IP with simplex.
- Let $\mathcal{B}, \mathcal{N}$ - Basic and nonbasic column index sets of LP.
- Re-write **source** row, with $\nu_i \notin \mathbb{Z}$, as

$$y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j = \nu_i(\bar{x}),$$

Gomory Fractional Cuts (GFC) for Deterministic Pure IPs

- Given first-stage vector $\bar{x}$, solve the LP relaxation of the second-stage IP with simplex.
- Let $\mathcal{B}, \mathcal{N}$ - Basic and nonbasic column index sets of LP.
- Re-write **source** row, with $\nu_i \notin \mathbb{Z}$, as

$$y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j = \nu_i(\bar{x}),$$

-

$$\underbrace{y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \lceil \bar{w}_{ij} \rceil y_j}_{\in \mathbb{Z}} \geq y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j = \underbrace{\nu_i(\bar{x})}_{\notin \mathbb{Z}}.$$

Gomory Fractional Cuts (GFC) for Deterministic Pure IPs

- Given first-stage vector $\bar{x}$, solve the LP relaxation of the second-stage IP with simplex.
- Let $\mathcal{B}, \mathcal{N}$ - Basic and nonbasic column index sets of LP.
- Re-write **source** row, with $\nu_i \notin \mathbb{Z}$, as

$$y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j = \nu_i(\bar{x}),$$

- $$\underbrace{y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \lceil \bar{w}_{ij} \rceil y_j}_{\in \mathbb{Z}} \geq y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j = \underbrace{\nu_i(\bar{x})}_{\notin \mathbb{Z}}.$$

- Let $\xi(\beta) := \lceil \beta \rceil - \beta$.

Gomory Fractional Cuts (GFC) for Deterministic Pure IPs

- Given first-stage vector $\bar{x}$, solve the LP relaxation of the second-stage IP with simplex.
- Let $\mathcal{B}, \mathcal{N}$ - Basic and nonbasic column index sets of LP.
- Re-write **source** row, with $\nu_i \notin \mathbb{Z}$, as

$$y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j = \nu_i(\bar{x}),$$

- $$\underbrace{y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \lceil \bar{w}_{ij} \rceil y_j}_{\in \mathbb{Z}} \geq y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j = \underbrace{\nu_i(\bar{x})}_{\notin \mathbb{Z}}.$$

- Let $\xi(\beta) := \lceil \beta \rceil - \beta$.
- Derive a GFC : $y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \lceil \bar{w}_{ij} \rceil y_j \geq \lceil \nu_i(\bar{x}) \rceil$.
  or equivalently,

$$\sum_{j \in \mathcal{N}} \xi(\bar{w}_{ij}) y_j \geq \xi(\nu_i(\bar{x})).$$

Gomory Fractional Cuts (GFC) for Deterministic Pure IPs

- Given first-stage vector $\bar{x}$, solve the LP relaxation of the second-stage IP with simplex.
- Let $\mathcal{B}, \mathcal{N}$ - Basic and nonbasic column index sets of LP.
- Re-write **source** row, with $\nu_i \notin \mathbb{Z}$, as

$$y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j = \nu_i(\bar{x}),$$

- $$\underbrace{y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \lceil \bar{w}_{ij} \rceil y_j}_{\in \mathbb{Z}} \geq y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j = \underbrace{\nu_i(\bar{x})}_{\notin \mathbb{Z}}.$$

- Let $\xi(\beta) := \lceil \beta \rceil - \beta$.
- Derive a GFC : $y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \lceil \bar{w}_{ij} \rceil y_j \geq \lceil \nu_i(\bar{x}) \rceil$.
  or equivalently,

$$\sum_{j \in \mathcal{N}} \xi(\bar{w}_{ij}) y_j \geq \xi(\nu_i(\bar{x})).$$

- A pure cutting plane algorithm using GFC is finitely convergent if one chooses the source row as the variable with the smallest index and use lexicographic dual simplex [Gomory, 1963]

Gomory Cuts for SMIP: [Carøe and Tind, 1998]

Continuous first stage, pure integer second stage.

- Solve the second stage problem using Gomory cuts to optimality for each x, ω

Gomory Cuts for SMIP: [Carøe and Tind, 1998]

Continuous first stage, pure integer second stage.

- Solve the second stage problem using Gomory cuts to optimality for each x, ω
- Construct the optimal subadditive dual function $\mathcal{C}_\omega$ (Chvàtal function - nonlinear and nonconvex)
  $\mathcal{C}_\omega(d) = V\lceil M_t \lceil M_{t-1} \cdots \lceil M_2 \lceil M_1 d \rceil \rceil \cdots \rceil$, where M_j, V are rational matrices

Gomory Cuts for SMIP: [Carøe and Tind, 1998]

Continuous first stage, pure integer second stage.

- Solve the second stage problem using Gomory cuts to optimality for each x, ω
- Construct the optimal subadditive dual function $\mathcal{C}_\omega$ (Chvàtal function - nonlinear and nonconvex)
  $\mathcal{C}_\omega(d) = V\lceil M_t \lceil M_{t-1} \cdots \lceil M_2 \lceil M_1 d \rceil \rceil \cdots \rceil$, where M_j, V are rational matrices
- First-stage optimality cuts:

$$\eta \geq \sum_{\omega \in \Omega} p_\omega \mathcal{C}_\omega((r(\omega) - T(\omega)x))$$

Gomory Cuts for SMIP: [Carøe and Tind, 1998]

Continuous first stage, pure integer second stage.

- Solve the second stage problem using Gomory cuts to optimality for each x, ω
- Construct the optimal subadditive dual function $\mathcal{C}_\omega$ (Chvàtal function - nonlinear and nonconvex)
  $\mathcal{C}_\omega(d) = V\lceil M_t \lceil M_{t-1} \cdots \lceil M_2 \lceil M_1 d \rceil \rceil \cdots \rceil$, where M_j, V are rational matrices
- First-stage optimality cuts:

$$\eta \geq \sum_{\omega \in \Omega} p_\omega \mathcal{C}_\omega((r(\omega) - T(\omega)x))$$

- It is possible to represent $\mathcal{C}_\omega$ using integer variables - growth is explosive even for a single x

Gomory Cuts for SMIP: [Carøe and Tind, 1998]

Continuous first stage, pure integer second stage.

- Solve the second stage problem using Gomory cuts to optimality for each x, ω
- Construct the optimal subadditive dual function $\mathcal{C}_\omega$ (Chvàtal function - nonlinear and nonconvex)
  $\mathcal{C}_\omega(d) = V\lceil M_t \lceil M_{t-1} \cdots \lceil M_2 \lceil M_1 d \rceil \rceil \cdots \rceil$, where M_j, V are rational matrices
- First-stage optimality cuts:

$$\eta \geq \sum_{\omega \in \Omega} p_\omega \mathcal{C}_\omega((r(\omega) - T(\omega)x))$$

- It is possible to represent $\mathcal{C}_\omega$ using integer variables - growth is explosive even for a single x
- Conceptual algorithm, computationally unattractive

Gomory Cuts for SMIP: [Carøe and Tind, 1998]

Continuous first stage, pure integer second stage.

- Solve the second stage problem using Gomory cuts to optimality for each x, ω
- Construct the optimal subadditive dual function $\mathcal{C}_\omega$ (Chvàtal function - nonlinear and nonconvex)
  $\mathcal{C}_\omega(d) = V\lceil M_t \lceil M_{t-1} \cdots \lceil M_2 \lceil M_1 d \rceil \rceil \cdots \rceil$, where M_j, V are rational matrices
- First-stage optimality cuts:

$$\eta \geq \sum_{\omega \in \Omega} p_\omega \mathcal{C}_\omega((r(\omega) - T(\omega)x))$$

- It is possible to represent $\mathcal{C}_\omega$ using integer variables - growth is explosive even for a single x
- Conceptual algorithm, computationally unattractive

Research Question: Can we use Gomory cuts to develop a computationally amenable L-shaped algorithm for SIP?

Decomposition-based cutting plane approximations - Strategy

- Partition first stage and second stage

Decomposition-based cutting plane approximations - Strategy

- Partition first stage and second stage
- Solve the master problem (first stage) and obtain $\bar{x}$

Decomposition-based cutting plane approximations - Strategy

- Partition first stage and second stage
- Solve the master problem (first stage) and obtain $\bar{x}$
- Solve LP relaxation of second stage sub-problems for given $\bar{x}$ for each $\omega \in \Omega$

Decomposition-based cutting plane approximations - Strategy

- Partition first stage and second stage
- Solve the master problem (first stage) and obtain $\bar{x}$
- Solve LP relaxation of second stage sub-problems for given $\bar{x}$ for each $\omega \in \Omega$
- If the sub-problem is non-integral, generate violated cut(s) $\pi(\omega)^\top y \geq \pi_0$ and re-solve sub-LP

Decomposition-based cutting plane approximations - Strategy

- Partition first stage and second stage
- Solve the master problem (first stage) and obtain $\bar{x}$
- Solve LP relaxation of second stage sub-problems for given $\bar{x}$ for each $\omega \in \Omega$
- If the sub-problem is non-integral, generate violated cut(s) $\pi(\omega)^\top y \geq \pi_0$ and re-solve sub-LP
- Lift this inequality to obtain $\pi(\omega)^\top y \geq \pi_0(x, \omega)$
- Add Benders optimality cut to the master problem

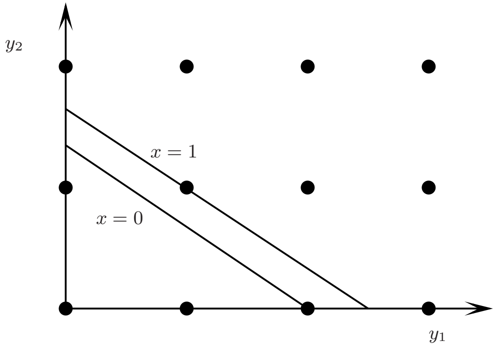
Decomposition-based cutting plane approximations - Strategy

- Partition first stage and second stage
- Solve the master problem (first stage) and obtain $\bar{x}$
- Solve LP relaxation of second stage sub-problems for given $\bar{x}$ for each $\omega \in \Omega$
- If the sub-problem is non-integral, generate violated cut(s) $\pi(\omega)^\top y \geq \pi_0$ and re-solve sub-LP
- Lift this inequality to obtain $\pi(\omega)^\top y \geq \pi_0(x, \omega)$
- Add Benders optimality cut to the master problem
  - For mixed binary second stage, and disjunctive cuts, $\pi_0(\cdot, \omega)$ is piecewise linear concave [Sen and Higle, 2005]
  - What about general integers and Gomory cuts?

Lifting Gomory Cuts for Second Stage

$$\min\{-x + h(x) : x \in \{0,1\}\}$$

$$h(x) = \min\{-y_1 : 2y_1 + 3y_2 = 4 + x, y_1, y_2 \in \mathbb{Z}_+\}$$

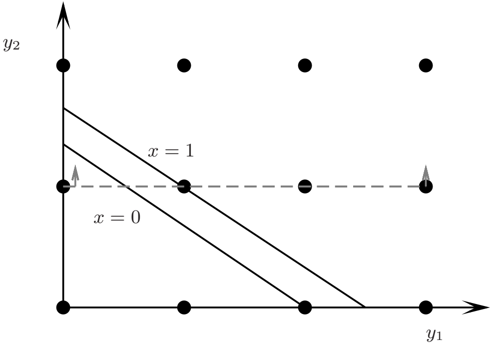


Lifting Gomory Cuts for Second Stage

- $\min\{-x + h(x) : x \in \{0, 1\}\}$, where $h(x) = \min\{-y_1 : 2y_1 + 3y_2 = 4 + x, y_1, y_2 \in \mathbb{Z}_+\}$
- First-stage solution: $x = 1$
- Source row: $y_1 + \frac{3}{2}y_2 = \frac{5}{2}$

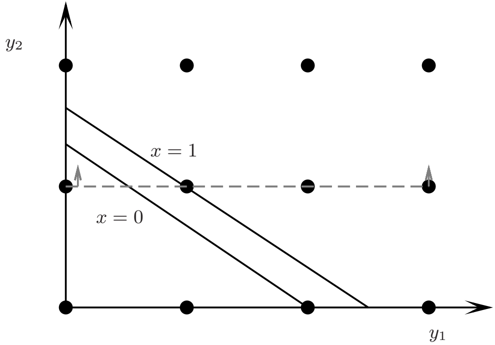
Lifting Gomory Cuts for Second Stage

- $\min\{-x + h(x) : x \in \{0,1\}\}$, where $h(x) = \min\{-y_1 : 2y_1 + 3y_2 = 4 + x, y_1, y_2 \in \mathbb{Z}_+\}$
- First-stage solution: $x = 1$
- Source row: $y_1 + \frac{3}{2}y_2 = \frac{5}{2}$
- Gomory Cut: $\frac{1}{2}y_2 \geq \frac{1}{2}$ (Not valid for $x = y_2 = 0, y_1 = 2$)



Lifting Gomory Cuts for Second Stage

- $\min\{-x + h(x) : x \in \{0, 1\}\}$, where $h(x) = \min\{-y_1 : 2y_1 + 3y_2 = 4 + x, y_1, y_2 \in \mathbb{Z}_+\}$
- First-stage solution: $x = 1$
- Source row: $y_1 + \frac{3}{2}y_2 = \frac{5}{2}$
- Gomory Cut: $\frac{1}{2}y_2 \geq \frac{1}{2}$ (Not valid for $x = y_2 = 0, y_1 = 2$)



- Carøe and Tind approach: $\frac{1}{2}y_2 \geq \lceil \frac{x}{2} \rceil - \frac{x}{2}$ (Nonlinear)

Desiderata

- A second-stage cut that is valid for all x.
- A first-stage cut that is affine in x.
- Finite convergence

Lifting Gomory Cuts for Second Stage

Want the cut to be valid for all x. Let $x' := 1 - x$. Write source row as:

$$y_1 + \frac{3}{2}y_2 = 2 + \frac{(1 - x')}{2}$$

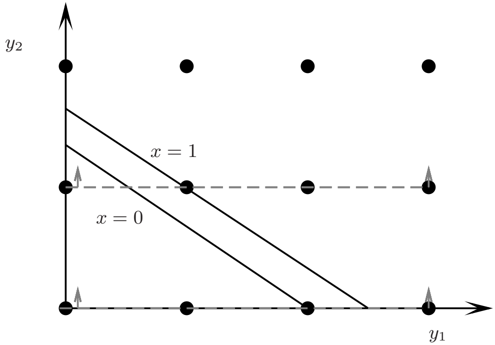
$$\text{Gomory Cut: } \frac{1}{2}x' + \frac{1}{2}y_2 \geq \frac{1}{2} \equiv y_2 \geq 1 - x' = x$$

Lifting Gomory Cuts for Second Stage

Want the cut to be valid for all x. Let $x' := 1 - x$. Write source row as:

$$y_1 + \frac{3}{2}y_2 = 2 + \frac{(1 - x')}{2}$$

Gomory Cut: $\frac{1}{2}x' + \frac{1}{2}y_2 \geq \frac{1}{2} \equiv y_2 \geq 1 - x' = x$ (affine in x)



Gomory Fractional Cuts - RHS as functions of x

- Assume w.l.o.g (by complementation, if necessary) that $\bar{x}_j = 0, \forall j = 1, \ldots, n_1$.

Gomory Fractional Cuts - RHS as functions of x

- Assume w.l.o.g (by complementation, if necessary) that $\bar{x}_j = 0, \forall j = 1, \dots, n_1$.
- Fix $\bar{x} \in X, \bar{\omega} \in \Omega$. $B, N, \mathcal{B}, \mathcal{N}$ - Basis, nonbasic columns, basic and non-basic index sets of LP $h_\ell(\bar{x}, \bar{\omega})$. Re-write second stage constraints $Wy = r - T\bar{x}$:

$$y_{\mathcal{B}} + \underbrace{B^{-1}N}_{\bar{w}_{ij}} y_{\mathcal{N}} = \underbrace{B^{-1}r}_{\rho} - \underbrace{B^{-1}T}_{\Gamma} \bar{x} =: \nu.$$

Gomory Fractional Cuts - RHS as functions of x

- Assume w.l.o.g (by complementation, if necessary) that $\bar{x}_j = 0, \forall j = 1, \ldots, n_1$.
- Fix $\bar{x} \in X, \bar{\omega} \in \Omega$. $B, N, \mathcal{B}, \mathcal{N}$ - Basis, nonbasic columns, basic and non-basic index sets of LP $h_\ell(\bar{x}, \bar{\omega})$. Re-write second stage constraints $Wy = r - T\bar{x}$:

$$y_{\mathcal{B}} + \underbrace{B^{-1}N}_{\bar{w}_{ij}} y_{\mathcal{N}} = \underbrace{B^{-1}r}_{\rho} - \underbrace{B^{-1}T}_{\Gamma} \bar{x} =: \nu.$$

- Re-write source row, with $\nu_i \notin \mathbb{Z}$, in terms of x as

$$y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j + \sum_{j=1}^{n_1} \gamma_{ij} x_j = \nu_i,$$

Gomory Fractional Cuts - RHS as functions of x

- Assume w.l.o.g (by complementation, if necessary) that $\bar{x}_j = 0, \forall j = 1, \ldots, n_1$.
- Fix $\bar{x} \in X, \bar{\omega} \in \Omega$. $B, N, \mathcal{B}, \mathcal{N}$ - Basis, nonbasic columns, basic and non-basic index sets of LP $h_\ell(\bar{x}, \bar{\omega})$. Re-write second stage constraints $Wy = r - T\bar{x}$:

$$y_{\mathcal{B}} + \underbrace{B^{-1}N}_{\bar{w}_{ij}} y_{\mathcal{N}} = \underbrace{B^{-1}r}_{\rho} - \underbrace{B^{-1}T}_{\Gamma}\bar{x} =: \nu.$$

- Re-write source row, with $\nu_i \notin \mathbb{Z}$, in terms of x as

$$y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j + \sum_{j=1}^{n_1} \gamma_{ij} x_j = \nu_i,$$

- Let $\xi(\beta) := \lceil \beta \rceil - \beta$. Derive a parametric GFC in the space of (x, y)-variables

$$\sum_{j \in \mathcal{N}} \xi(\bar{w}_{ij}) y_j \geq \xi(\nu_i) - \sum_{j=1}^{n_1} \xi(\gamma_{ij}) x_j.$$

Gomory Fractional Cuts - RHS as functions of x

- Assume w.l.o.g (by complementation, if necessary) that $\bar{x}_j = 0, \forall j = 1, \ldots, n_1$.
- Fix $\bar{x} \in X, \bar{\omega} \in \Omega$. $B, N, \mathcal{B}, \mathcal{N}$ - Basis, nonbasic columns, basic and non-basic index sets of LP $h_\ell(\bar{x}, \bar{\omega})$. Re-write second stage constraints $Wy = r - T\bar{x}$:

$$y_{\mathcal{B}} + \underbrace{B^{-1}N}_{\bar{w}_{ij}} y_{\mathcal{N}} = \underbrace{B^{-1}r}_{\rho} - \underbrace{B^{-1}T}_{\Gamma}\bar{x} =: \nu.$$

- Re-write source row, with $\nu_i \notin \mathbb{Z}$, in terms of x as

$$y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j + \sum_{j=1}^{n_1} \gamma_{ij} x_j = \nu_i,$$

- Let $\xi(\beta) := \lceil \beta \rceil - \beta$. Derive a parametric GFC in the space of (x, y)-variables

$$\sum_{j \in \mathcal{N}} \xi(\bar{w}_{ij}) y_j \geq \xi(\nu_i) - \sum_{j=1}^{n_1} \xi(\gamma_{ij}) x_j.$$

- When $x = \bar{x}$ we recover the original GFC. This GFC is valid for all binary x-variables.

Gomory Fractional Cuts - RHS as functions of x

- Assume w.l.o.g (by complementation, if necessary) that $\bar{x}_j = 0, \forall j = 1, \ldots, n_1$.
- Fix $\bar{x} \in X, \bar{\omega} \in \Omega$. $B, N, \mathcal{B}, \mathcal{N}$ - Basis, nonbasic columns, basic and non-basic index sets of LP $h_\ell(\bar{x}, \bar{\omega})$. Re-write second stage constraints $Wy = r - T\bar{x}$:

$$y_{\mathcal{B}} + \underbrace{B^{-1}N}_{\bar{w}_{ij}} y_{\mathcal{N}} = \underbrace{B^{-1}r}_{\rho} - \underbrace{B^{-1}T}_{\Gamma}\bar{x} =: \nu.$$

- Re-write source row, with $\nu_i \notin \mathbb{Z}$, in terms of x as

$$y_{\mathcal{B}_i} + \sum_{j \in \mathcal{N}} \bar{w}_{ij} y_j + \sum_{j=1}^{n_1} \gamma_{ij} x_j = \nu_i,$$

- Let $\xi(\beta) := \lceil \beta \rceil - \beta$. Derive a parametric GFC in the space of (x, y)-variables

$$\sum_{j \in \mathcal{N}} \xi(\bar{w}_{ij}) y_j \geq \xi(\nu_i) - \sum_{j=1}^{n_1} \xi(\gamma_{ij}) x_j.$$

- When $x = \bar{x}$ we recover the original GFC. This GFC is valid for all binary x-variables.
- Furthermore, $\pi(\bar{\omega})^\top y \geq \pi_0(x, \bar{\omega})$, $\pi_0(\cdot, \omega)$ is affine.

Gomory Driven Decomposition Algorithm - Notation

- Second-stage linear approximations at the beginning of iteration k

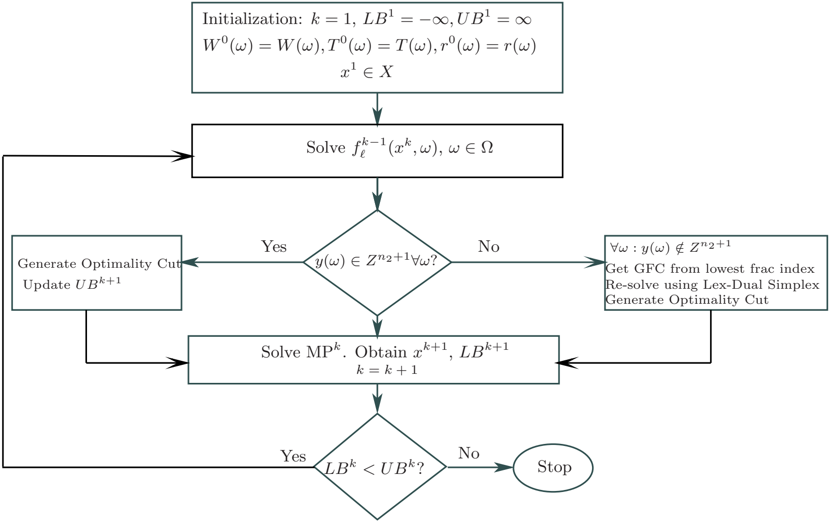
$$
\begin{aligned}
h_\ell^{k-1}(x,\omega) = \min\ & y_0 \\
& y_0 - g(\omega)^\top y = 0 \\
& W^{k-1}(\omega) y \geq r^{k-1}(\omega) - T^{k-1}(\omega) x \\
& y_0 \in \mathbb{R}, y \in \mathbb{R}_+^{n_2}.
\end{aligned}
$$

- $\psi^k(\omega)$: Dual multipliers of second-stage LP at iteration k
- $y^k(x,\omega)$: Lex-smallest solution to second-stage LP at iteration k, given x,ω
- Lower bounding Master Problem MP^k

$$
\begin{aligned}
& \min c^\top x + \eta \\
& Ax \geq b \\
& \eta \geq \sum_{\omega\in\Omega} p_\omega (\psi_\omega^t)^\top (r^t(\omega) - T^t(\omega)x), t = 1,\dots,k \\
& x \in \mathbb{B}^{n_1}, \eta \in \mathbb{R}.
\end{aligned}
$$

- LB^k, UB^k Lower and upper bounds on the SIP optimal solution

Gomory Driven Decomposition Algorithm is finitely convergent [Gade, , and Sen, 2014]



Proof of Convergence - Sketch

- Let $x^k = \bar{x}$ and $x^t = \bar{x}$, $t > k$
- Let $\alpha_k(\bar{x}, \omega) := \left(y_0^{k-1}(\bar{x}, \omega), y_1^{k-1}(\bar{x}, \omega), \ldots, y_{i_k-1}^{k-1}(\bar{x}, \omega), \lceil y_{i_k}^{k-1}(\bar{x}, \omega) \rceil, 0, \ldots, 0 \right)^\top$.

Proof of Convergence - Sketch

- Let $x^k = \bar{x}$ and $x^t = \bar{x}$, $t > k$
- Let $\alpha_k(\bar{x}, \omega) := \left(y_0^{k-1}(\bar{x}, \omega), y_1^{k-1}(\bar{x}, \omega), \ldots, y_{i_k - 1}^{k-1}(\bar{x}, \omega), \lceil y_{i_k}^{k-1}(\bar{x}, \omega) \rceil, 0, \ldots, 0\right)^\top$.
- Gomory cut for first fractional y_{i_k} and lex-dual simplex gives $y^k(\bar{x}, \omega) \succeq \alpha_k(\bar{x}, \omega)$

Proof of Convergence - Sketch

- Let $x^k = \bar{x}$ and $x^t = \bar{x}$, $t > k$
- Let $\alpha_k(\bar{x},\omega) := \left(y_0^{k-1}(\bar{x},\omega), y_1^{k-1}(\bar{x},\omega), \ldots, y_{i_k-1}^{k-1}(\bar{x},\omega), \lceil y_{i_k}^{k-1}(\bar{x},\omega)\rceil, 0, \ldots, 0\right)^\top$.
- Gomory cut for first fractional y_{i_k} and lex-dual simplex gives $y^k(\bar{x},\omega) \succeq \alpha_k(\bar{x},\omega)$
- Gomory cuts added during iterations $k+1, \ldots, t-1$ are all valid for $Y(\bar{x},\omega)$.

Proof of Convergence - Sketch

- Let $x^k = \bar{x}$ and $x^t = \bar{x}$, $t > k$
- Let $\alpha_k(\bar{x},\omega) := \left(y_0^{k-1}(\bar{x},\omega), y_1^{k-1}(\bar{x},\omega), \ldots, y_{i_k-1}^{k-1}(\bar{x},\omega), \lceil y_{i_k}^{k-1}(\bar{x},\omega)\rceil, 0, \ldots, 0\right)^\top$.
- Gomory cut for first fractional y_{i_k} and lex-dual simplex gives $y^k(\bar{x},\omega) \succeq \alpha_k(\bar{x},\omega)$
- Gomory cuts added during iterations $k+1, \ldots, t-1$ are all valid for $Y(\bar{x},\omega)$.
- So $y^{t-1}(\bar{x},\omega) \succeq y^k(\bar{x},\omega) \succeq \alpha_k(\bar{x},\omega)$.

Proof of Convergence - Sketch

- Let $x^k = \bar{x}$ and $x^t = \bar{x}$, $t > k$
- Let $\alpha_k(\bar{x},\omega) := \left(y_0^{k-1}(\bar{x},\omega), y_1^{k-1}(\bar{x},\omega), \ldots, y_{i_k-1}^{k-1}(\bar{x},\omega), \lceil y_{i_k}^{k-1}(\bar{x},\omega)\rceil, 0, \ldots, 0\right)^\top$.
- Gomory cut for first fractional y_{i_k} and lex-dual simplex gives $y^k(\bar{x},\omega) \succeq \alpha_k(\bar{x},\omega)$
- Gomory cuts added during iterations $k+1,\ldots,t-1$ are all valid for $Y(\bar{x},\omega)$.
- So $y^{t-1}(\bar{x},\omega) \succeq y^k(\bar{x},\omega) \succeq \alpha_k(\bar{x},\omega)$.
- $\alpha_t(\bar{x},\omega) \succ y^{t-1}(\bar{x},\omega)$ by definition.

Proof of Convergence - Sketch

- Let $x^k = \bar{x}$ and $x^t = \bar{x}$, $t > k$
- Let $\alpha_k(\bar{x}, \omega) := \left(y_0^{k-1}(\bar{x}, \omega), y_1^{k-1}(\bar{x}, \omega), \ldots, y_{i_k - 1}^{k-1}(\bar{x}, \omega), \lceil y_{i_k}^{k-1}(\bar{x}, \omega) \rceil, 0, \ldots, 0 \right)^\top$.
- Gomory cut for first fractional y_{i_k} and lex-dual simplex gives $y^k(\bar{x}, \omega) \succeq \alpha_k(\bar{x}, \omega)$
- Gomory cuts added during iterations $k+1, \ldots, t-1$ are all valid for $Y(\bar{x}, \omega)$.
- So $y^{t-1}(\bar{x}, \omega) \succeq y^k(\bar{x}, \omega) \succeq \alpha_k(\bar{x}, \omega)$.
- $\alpha_t(\bar{x}, \omega) \succ y^{t-1}(\bar{x}, \omega)$ by definition.
- Hence $\alpha_t(\bar{x}, \omega) \succ \alpha_k(\bar{x}, \omega)$.

Proof of Convergence - Sketch

- Let $x^k = \bar{x}$ and $x^t = \bar{x}$, $t > k$
- Let $\alpha_k(\bar{x}, \omega) := \left(y_0^{k-1}(\bar{x}, \omega), y_1^{k-1}(\bar{x}, \omega), \ldots, y_{i_k-1}^{k-1}(\bar{x}, \omega), \lceil y_{i_k}^{k-1}(\bar{x}, \omega) \rceil, 0, \ldots, 0\right)^\top$.
- Gomory cut for first fractional y_{i_k} and lex-dual simplex gives $y^k(\bar{x}, \omega) \succeq \alpha_k(\bar{x}, \omega)$
- Gomory cuts added during iterations $k+1, \ldots, t-1$ are all valid for $Y(\bar{x}, \omega)$.
- So $y^{t-1}(\bar{x}, \omega) \succeq y^k(\bar{x}, \omega) \succeq \alpha_k(\bar{x}, \omega)$.
- $\alpha_t(\bar{x}, \omega) \succ y^{t-1}(\bar{x}, \omega)$ by definition.
- Hence $\alpha_t(\bar{x}, \omega) \succ \alpha_k(\bar{x}, \omega)$.
- In finitely many steps, we obtain integral solutions for a given $(\bar{x}, \omega)$ for all $k \geq K(\bar{x}, \omega)$.

Proof of Convergence - Sketch

- Let $x^k = \bar{x}$ and $x^t = \bar{x}$, $t > k$
- Let $\alpha_k(\bar{x}, \omega) := \left(y_0^{k-1}(\bar{x}, \omega), y_1^{k-1}(\bar{x}, \omega), \ldots, y_{i_k-1}^{k-1}(\bar{x}, \omega), \lceil y_{i_k}^{k-1}(\bar{x}, \omega) \rceil, 0, \ldots, 0 \right)^\top$.
- Gomory cut for first fractional y_{i_k} and lex-dual simplex gives $y^k(\bar{x}, \omega) \succeq \alpha_k(\bar{x}, \omega)$
- Gomory cuts added during iterations $k+1, \ldots, t-1$ are all valid for $Y(\bar{x}, \omega)$.
- So $y^{t-1}(\bar{x}, \omega) \succeq y^k(\bar{x}, \omega) \succeq \alpha_k(\bar{x}, \omega)$.
- $\alpha_t(\bar{x}, \omega) \succ y^{t-1}(\bar{x}, \omega)$ by definition.
- Hence $\alpha_t(\bar{x}, \omega) \succ \alpha_k(\bar{x}, \omega)$.
- In finitely many steps, we obtain integral solutions for a given $(\bar{x}, \omega)$ for all $k \geq K(\bar{x}, \omega)$.
- Finitely many $(x, \omega) \in X \times \Omega \Rightarrow$ in finitely many steps $h_\ell^k(x, \omega)$ gives integral solutions $\forall (x, \omega)$ with $k \geq K = \sup_{(x,\omega)} K(x, \omega)$ (worst case).

Proof of Convergence - Sketch

- Let $x^k = \bar{x}$ and $x^t = \bar{x}$, $t > k$
- Let $\alpha_k(\bar{x},\omega) := \left(y_0^{k-1}(\bar{x},\omega), y_1^{k-1}(\bar{x},\omega), \ldots, y_{i_k-1}^{k-1}(\bar{x},\omega), \lceil y_{i_k}^{k-1}(\bar{x},\omega)\rceil, 0, \ldots, 0\right)^\top$.
- Gomory cut for first fractional y_{i_k} and lex-dual simplex gives $y^k(\bar{x},\omega) \succeq \alpha_k(\bar{x},\omega)$
- Gomory cuts added during iterations $k+1,\ldots,t-1$ are all valid for $Y(\bar{x},\omega)$.
- So $y^{t-1}(\bar{x},\omega) \succeq y^k(\bar{x},\omega) \succeq \alpha_k(\bar{x},\omega)$.
- $\alpha_t(\bar{x},\omega) \succ y^{t-1}(\bar{x},\omega)$ by definition.
- Hence $\alpha_t(\bar{x},\omega) \succ \alpha_k(\bar{x},\omega)$.
- In finitely many steps, we obtain integral solutions for a given $(\bar{x},\omega)$ for all $k \geq K(\bar{x},\omega)$.
- Finitely many $(x,\omega) \in X \times \Omega \Rightarrow$ in finitely many steps $h_\ell^k(x,\omega)$ gives integral solutions $\forall(x,\omega)$ with $k \geq K = \sup_{(x,\omega)} K(x,\omega)$ (worst case).
- Then the dual polyhedra of sub-problems remain fixed. Obtain full reformulation of SIP in (x,η).

Example from Literature

Variations of this example appear in [Schultz et al., 1998], [Sen et al., 2003], [Ahmed et al., 2004]

$$\begin{aligned} \min \quad & -1.5x_1 - 4x_2 + \mathbb{E}[f(x, \tilde{\omega})] \\ \text{s.t.} \quad & x \in \{0,1\}^2 \end{aligned}$$

where

$$\begin{aligned} f(x,\omega) = \min \quad & y_0 \\ \text{s.t.} \quad & y_0 + 16y_1 + 19y_2 + 23y_3 + 28y_4 - 100R = 0 \\ & 2y_1 + 3y_2 + 4y_3 + 5y_4 - R \leq r_1(\omega) - x_1 \\ & 6y_1 + 1y_2 + 3y_3 + 2y_4 - R \leq r_2(\omega) - x_2 \\ & y_0 \in \mathbb{Z}, y_i \in \{0, \ldots, 5\}, i = 1, \ldots, 4, R \in \mathbb{Z}_+, \end{aligned}$$

$\Omega = \{1, 2\}, p_1 = p_2 = 0.5.$

$(r_1(1), r_2(1)) = (10, 4), (r_1(2), r_2(2)) = (13, 8).$

Example from Literature

Variations of this example appear in [Schultz et al., 1998], [Sen et al., 2003], [Ahmed et al., 2004]

$$\begin{aligned} \min \quad & -1.5x_1 - 4x_2 + \mathbb{E}[f(x,\tilde{\omega})] \\ \text{s.t.} \quad & x \in \{0,1\}^2 \end{aligned}$$

where

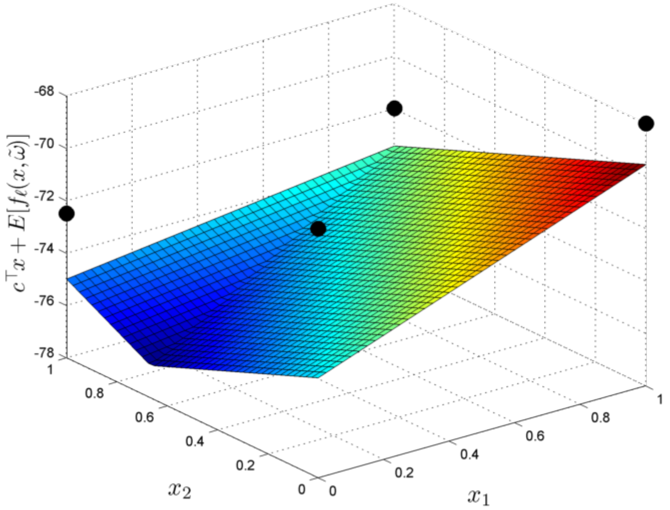
$$\begin{aligned} f(x,\omega) = \min \quad & y_0 \\ \text{s.t.} \quad & y_0 + 16y_1 + 19y_2 + 23y_3 + 28y_4 - 100R = 0 \\ & 2y_1 + 3y_2 + 4y_3 + 5y_4 - R \le r_1(\omega) - x_1 \\ & 6y_1 + 1y_2 + 3y_3 + 2y_4 - R \le r_2(\omega) - x_2 \\ & y_0 \in \mathbb{Z}, y_i \in \{0,\dots,5\}, i = 1,\dots,4, R \in \mathbb{Z}_+, \end{aligned}$$

$\Omega = \{1,2\}, p_1 = p_2 = 0.5.$

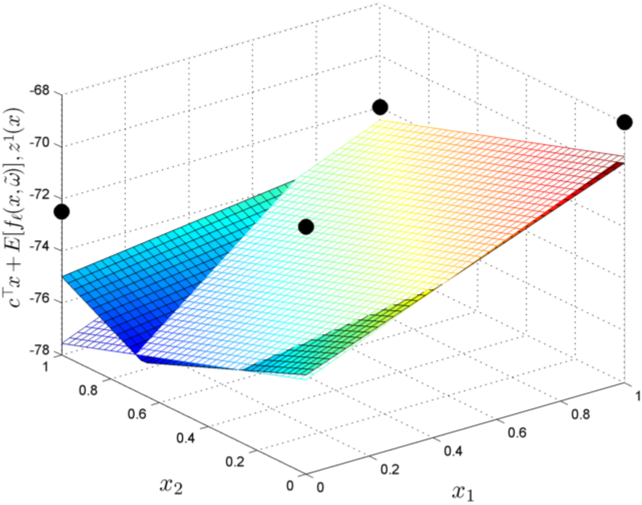
$(r_1(1), r_2(1)) = (10,4), (r_1(2), r_2(2)) = (13,8).$

$$z^k(x) := c^\top x + \max_{t=1,\dots,k} \left\{ \sum_{\omega\in\Omega} p_\omega (\psi_\omega^t)^\top (r^t(\omega) - T^t(\omega)x) \right\}.$$

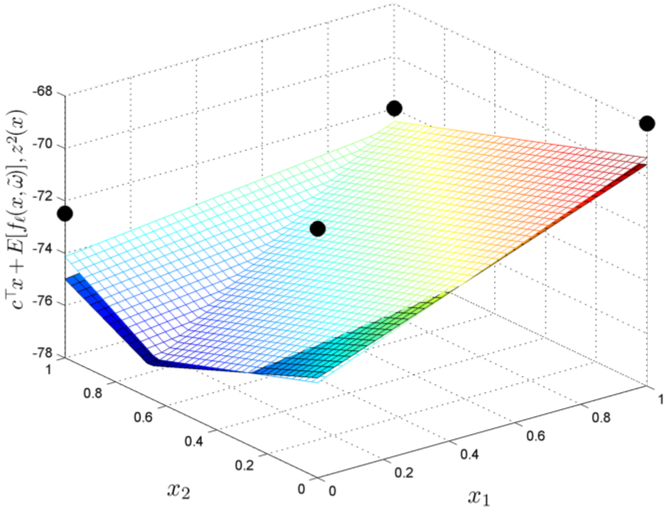
Best LP Approximation



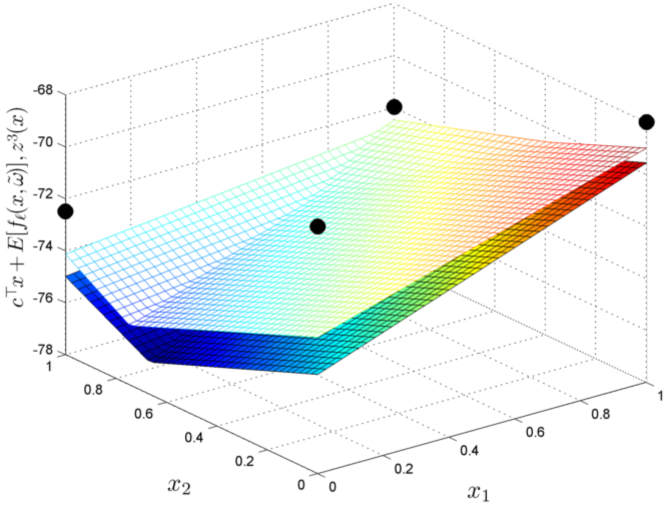
Approximation at $k = 1$



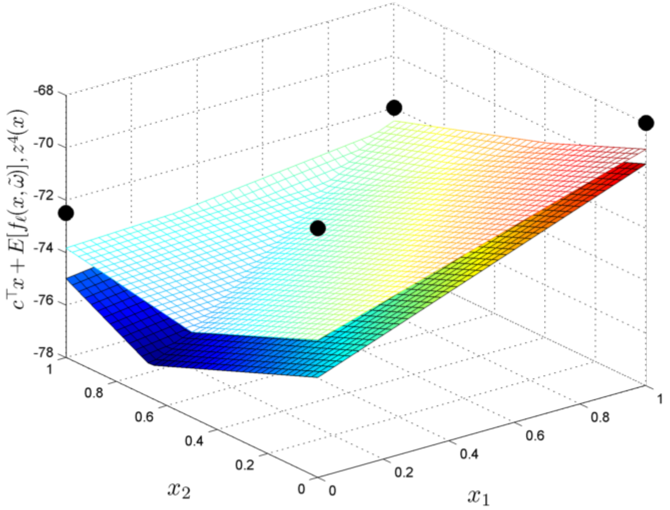
Approximation at $k = 2$



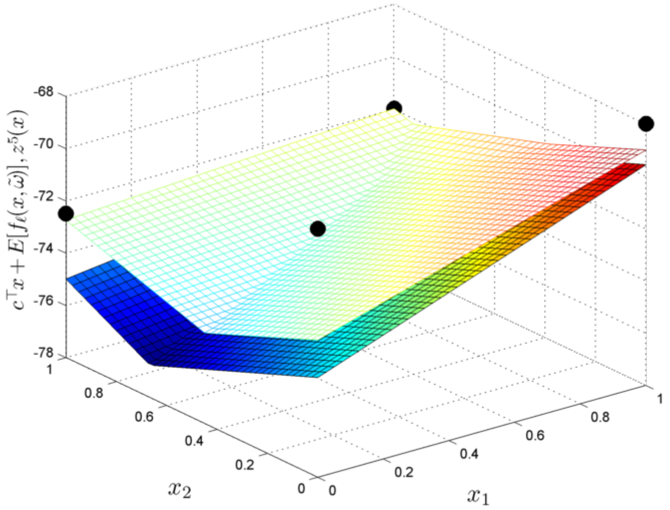
Approximation at $k = 3$



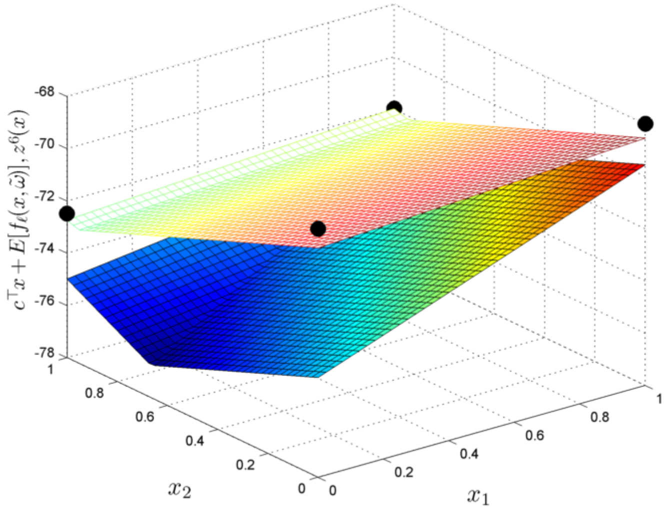
Approximation at $k = 4$



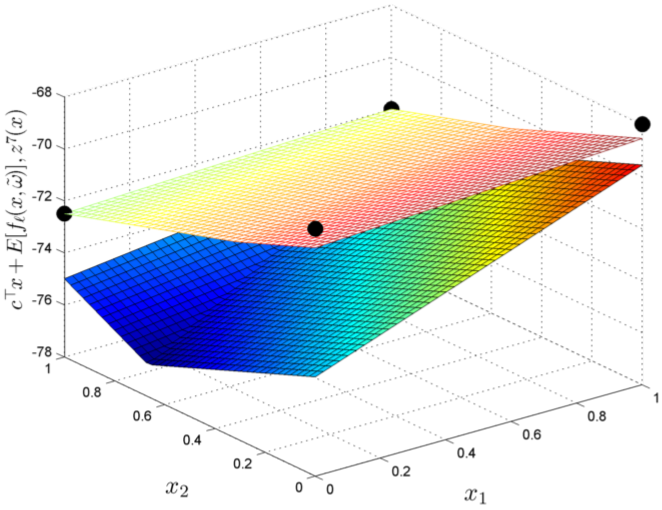
Approximation at $k = 5$



Approximation at $k = 6$



Approximation at $k = 7$



Deterministic Equivalent Comparison - SSLP Instances

Instances	DEF		Gomory	
	Time	Gap	Time	Gap
SSLP_5_25_50	2.03	0.00	0.18	0.00
SSLP_5_25_100	1.72	0.00	0.22	0.00
SSLP_5_50_50	1.06	0.00	0.27	0.00
SSLP_5_50_100	3.56	0.00	0.48	0.00
SSLP_5_50_1000	212.64	0.00	2.88	0.00
SSLP_5_50_2000	1020.54	0.00	5.73	0.00
SSLP_10_50_50	801.49	0.01	109.2	0.02
SSLP_10_50_100	*	0.10	218.42	0.02
SSLP_10_50_500	*	0.38	740.38	0.03
SSLP_10_50_1000	*	3.56	1615.42	0.02
SSLP_10_50_2000	*	18.59	2729.61	0.02

* 3600 second time limit

Alternative Implementations

- Single (η) vs. multi-optimality cuts (η_ω for each ω)

Alternative Implementations

- Single (η) vs. multi-optimality cuts (η_ω for each ω)

- Rounds of cuts in second stage

Alternative Implementations

- Single (η) vs. multi-optimality cuts (η_ω for each ω)
- Rounds of cuts in second stage
- Lexicography allows other structural cuts if known

Alternative Implementations

- Single (η) vs. multi-optimality cuts (η_ω for each ω)
- Rounds of cuts in second stage
- Lexicography allows other structural cuts if known
- Can also implement more efficient cut generation that maintains fixed recourse and fixed technology matrices

Alternative Implementations

- Single (η) vs. multi-optimality cuts (η_ω for each ω)
- Rounds of cuts in second stage
- Lexicography allows other structural cuts if known
- Can also implement more efficient cut generation that maintains fixed recourse and fixed technology matrices
- Partial branch-and-cut for binary second-stage variables

Summary - First Stage Binary, Second Stage Integer

- First computationally amenable incorporation of Gomory cuts into L-shaped decomposition algorithm

Summary - First Stage Binary, Second Stage Integer

- First computationally amenable incorporation of Gomory cuts into L-shaped decomposition algorithm
- In each iteration, solve at most two LP subproblems (not IP's to completion)

Summary - First Stage Binary, Second Stage Integer

- First computationally amenable incorporation of Gomory cuts into L-shaped decomposition algorithm
- In each iteration, solve at most two LP subproblems (not IP's to completion)
- Cost function vector, recourse & technology matrices and RHS are allowed to be random

Summary - First Stage Binary, Second Stage Integer

- First computationally amenable incorporation of Gomory cuts into L-shaped decomposition algorithm
- In each iteration, solve at most two LP subproblems (not IP's to completion)
- Cost function vector, recourse & technology matrices and RHS are allowed to be random
- All alternative implementations with lex-dual simplex are finite

Summary - First Stage Binary, Second Stage Integer

- First computationally amenable incorporation of Gomory cuts into L-shaped decomposition algorithm
- In each iteration, solve at most two LP subproblems (not IP's to completion)
- Cost function vector, recourse & technology matrices and RHS are allowed to be random
- All alternative implementations with lex-dual simplex are finite
- One can now integrate alternative classes of cuts: Disjunctive, Gomory, structural

First and Second Stages Integer [Zhang and K., 2014]

$B = D = \{1, 2\}, C = \emptyset$

First and Second Stages Integer [Zhang and K., 2014]

$B = D = \{1, 2\}, C = \emptyset$

- Second-stage problem is similar as before

First and Second Stages Integer [Zhang and K., 2014]

$B = D = \{1, 2\}, C = \emptyset$

- Second-stage problem is similar as before
- Use a more sophisticated lifting function relying on Gomory cuts for the first-stage

First and Second Stages Integer [Zhang and K., 2014]

$B = D = \{1, 2\}, C = \emptyset$

- Second-stage problem is similar as before
- Use a more sophisticated lifting function relying on Gomory cuts for the first-stage
- Finitely convergent, thanks to finite convergence of Gomory for deterministic pure integer programs

First and Second Stages Integer [Zhang and K., 2014]

$B = D = \{1, 2\}$, $C = \emptyset$

- Second-stage problem is similar as before
- Use a more sophisticated lifting function relying on Gomory cuts for the first-stage
- Finitely convergent, thanks to finite convergence of Gomory for deterministic pure integer programs

How about mixed-integer variables? Gomory (or Gomory Mixed-Integer) pure cutting plane method is no longer finitely convergent...

Outline

1. Two-Stage Stochastic Integer Programming
   - Two-Stage Stochastic Linear Programming
   - Classification Scheme for Two-Stage Stochastic Mixed-Integer Programming
   - Two-Stage Stochastic Pure Integer Programming
   - Two-Stage Stochastic Mixed-Integer Programming

2. Chance-Constrained Programming
   - Static Joint Chance-Constrained Programming
   - Two-stage (Dynamic) Chance-Constrained Programming
   - Distributionally Robust Chance-Constrained Programming

Background: Deterministic 0-1 Mixed-Integer Linear Program (MILP)

$\min_{x \in X}\{c^T x | X = \{Ax \geq b, x \in \{0,1\}^{n_1} \times \mathbb{R}_+^{n-n_1}\}\}$.

- Let X_L be the LP relaxation of X.
- $P^-(j, \bar{X}) := \{x \in \bar{X} | x_j \leq 0\}$,
  $P^+(j, \bar{X}) := \{x \in \bar{X} | x_j \geq 1\}$,
- $\mathcal{H}_j(\bar{X}) := \text{clconv}(P^-(j, \bar{X}) \cup P^+(j, \bar{X}))$.

Theorem (Sequential convexification of 0-1 MILP [Balas, 1979])

$\text{clconv}(X) = \mathcal{H}_{n_1}(\mathcal{H}_{n_1-1}(\cdots(\mathcal{H}_1(X_L))\cdots))$.

Other finite characterizations: RLT [Sherali and Adams, 1990, 1994], SDP [Lovász and Schrijver, 1991], ...

Background: Deterministic 0-1 Mixed-Integer Linear Program (MILP)

$\min_{x \in X}\{c^T x | X = \{Ax \geq b, x \in \{0,1\}^{n_1} \times \mathbb{R}_+^{n-n_1}\}\}$.

- Let X_L be the LP relaxation of X.
- $P^-(j, \bar{X}) := \{x \in \bar{X} | x_j \leq 0\}$,
  $P^+(j, \bar{X}) := \{x \in \bar{X} | x_j \geq 1\}$,
- $\mathcal{H}_j(\bar{X}) := \text{clconv}(P^-(j, \bar{X}) \cup P^+(j, \bar{X}))$.

Theorem (Sequential convexification of 0-1 MILP [Balas, 1979])

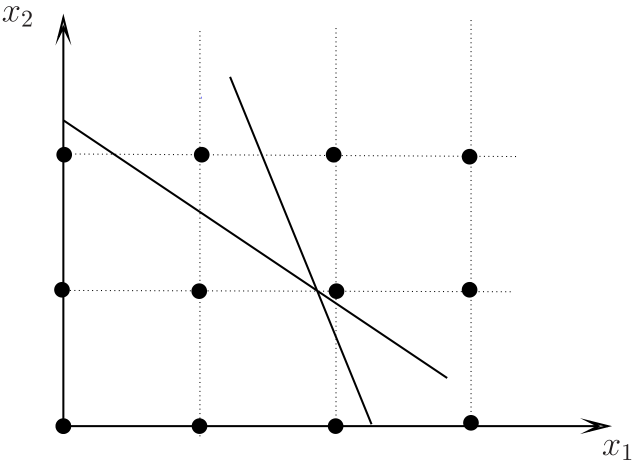
$\text{clconv}(X) = \mathcal{H}_{n_1}(\mathcal{H}_{n_1-1}(\cdots(\mathcal{H}_1(X_L))\cdots))$.

Other finite characterizations: RLT [Sherali and Adams, 1990, 1994], SDP [Lovász and Schrijver, 1991], ...

[Carøe and Tind, 1998] and [Sen and Higle, 2005] adapt this convexification scheme for two-stage stochastic mixed-binary optimization.

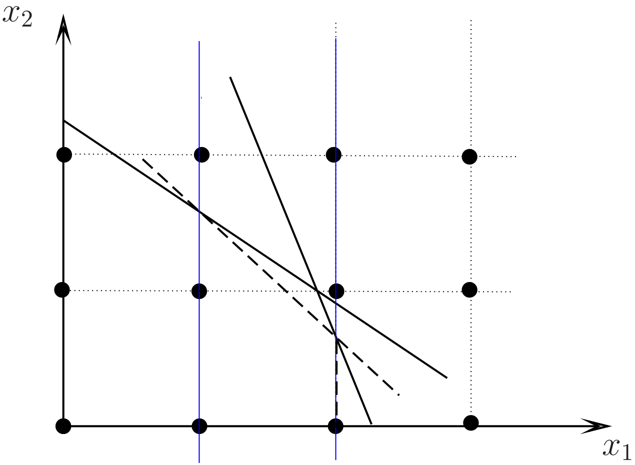
How about general MILP?

Example of [Owen and Mehrotra, 2001]



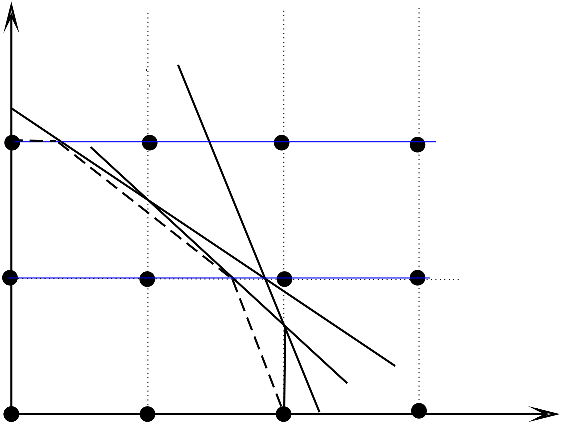
Convexification w.r.t x_1

Example of [Owen and Mehrotra, 2001]



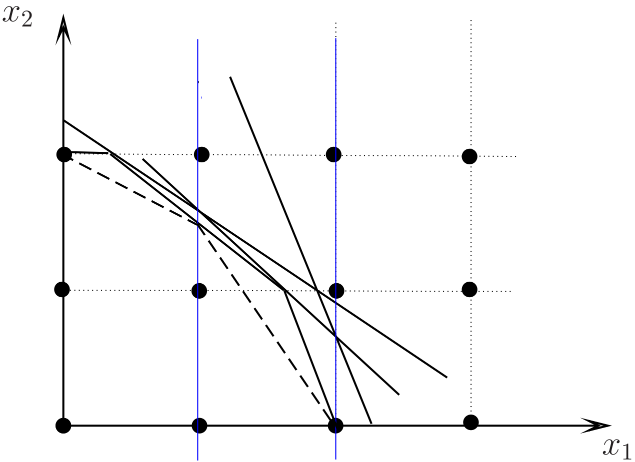
Convexification w.r.t first x_1, then $x_2 \neq \text{conv}(X)$!

Example of [Owen and Mehrotra, 2001]



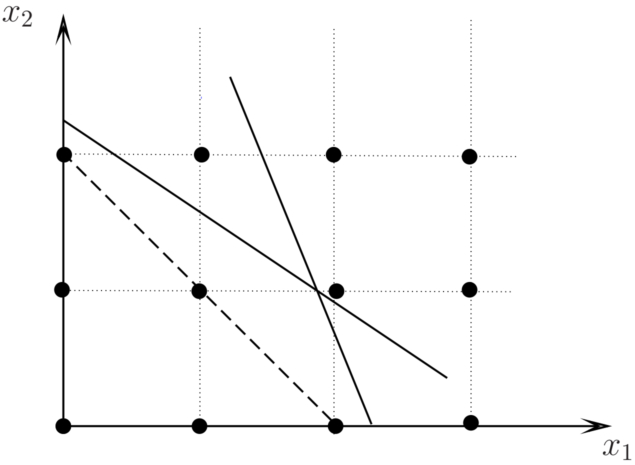
Convexification w.r.t first x_1, then x_2, then x_1

Example of [Owen and Mehrotra, 2001]



Ad infinitum

Example of [Owen and Mehrotra, 2001]



General MILP with bounded integer variables

$\min_{x \in X}\{c^T x | X = \{Ax \geq b, x \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n-n_1}\}\}$.

- Assume that all integer variables are bounded: $x_j \in [0, u_j]$ for all $j = 1, \ldots, n_1$.

General MILP with bounded integer variables

$\min_{x \in X}\{c^T x | X = \{Ax \geq b, x \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n-n_1}\}\}$.

- Assume that all integer variables are bounded: $x_j \in [0, u_j]$ for all $j = 1, \ldots, n_1$.
- One variable at a time convexification converges in the limit [Owen and Mehrotra, 2001]

General MILP with bounded integer variables

$\min_{x \in X}\{c^T x | X = \{Ax \geq b, x \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n-n_1}\}\}.$

- Assume that all integer variables are bounded: $x_j \in [0, u_j]$ for all $j = 1, \ldots, n_1$.
- One variable at a time convexification converges in the limit [Owen and Mehrotra, 2001]
- A disjunctive cutting plane algorithm using elementary disjunctions may not converge to the integral optimal solution [Owen and Mehrotra, 2001].

General MILP with bounded integer variables

$\min_{x \in X} \{c^T x | X = \{Ax \geq b, x \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n-n_1}\}\}.$

- Assume that all integer variables are bounded: $x_j \in [0, u_j]$ for all $j = 1, \ldots, n_1$.
- One variable at a time convexification converges in the limit [Owen and Mehrotra, 2001]
- A disjunctive cutting plane algorithm using elementary disjunctions may not converge to the integral optimal solution [Owen and Mehrotra, 2001].
- Binary expansion of bounded integer variables may not be effective in practice [Owen and Mehrotra, 2002]

General MILP with bounded integer variables

$\min_{x \in X}\{c^T x | X = \{Ax \geq b, x \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n-n_1}\}\}.$

- Assume that all integer variables are bounded: $x_j \in [0, u_j]$ for all $j = 1, \ldots, n_1$.
- One variable at a time convexification converges in the limit [Owen and Mehrotra, 2001]
- A disjunctive cutting plane algorithm using elementary disjunctions may not converge to the integral optimal solution [Owen and Mehrotra, 2001].
- Binary expansion of bounded integer variables may not be effective in practice [Owen and Mehrotra, 2002]
- [Adams and Sherali, 2005] give a finite RLT characterization using Lagrange interpolation polynomials

Questions

- Is there a finite disjunctive characterization of the convex hull of MILP solutions in the original space of general integer variables?

- Is there a finitely convergent cutting plane algorithm for a general MILP (with no assumptions on the integrality of the optimal objective)?

General MILP

$\min_{x \in X}\{c^T x | X = \{Ax \geq b, x \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n-n_1}\}\}.$

- Assume that all integer variables are bounded: $x_j \in [0, u_j]$ for all $j = 1, \ldots, n_1$.
- Let X_L be the LP relaxation of X.

General MILP

$\min_{x \in X} \{c^T x | X = \{Ax \geq b, x \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n-n_1}\}\}.$

- Assume that all integer variables are bounded: $x_j \in [0, u_j]$ for all $j = 1, \ldots, n_1$.
- Let X_L be the LP relaxation of X.
- Partition each interval $[0, u_j]$ into t_j sub-intervals $[\ell_{1j} := 0, u_{1j}]$, $[\ell_{2j}, u_{2j}], \ldots, [\ell_{t_j j}, u_{t_j j} := u_j]$

General MILP

$\min_{x \in X} \{c^T x | X = \{Ax \geq b, x \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n-n_1}\}\}$.

- Assume that all integer variables are bounded: $x_j \in [0, u_j]$ for all $j = 1, \ldots, n_1$.
- Let X_L be the LP relaxation of X.
- Partition each interval $[0, u_j]$ into t_j sub-intervals $[\ell_{1j} := 0, u_{1j}]$, $[\ell_{2j}, u_{2j}], \ldots, [\ell_{t_j j}, u_{t_j j} := u_j]$
- Given a partition $\mathcal{P}$, the collection of all n_1-tuples $\kappa := (\kappa_1, \ldots, \kappa_{n_1})$, where $\kappa_j \in \{1, \ldots, t_j\}$ for $j = 1, \ldots, n_1$, is denoted by $K(\mathcal{P})$.

General MILP

$\min_{x\in X}\{c^T x | X = \{Ax \geq b, x \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n-n_1}\}\}$.

- Assume that all integer variables are bounded: $x_j \in [0, u_j]$ for all $j = 1, \ldots, n_1$.
- Let X_L be the LP relaxation of X.
- Partition each interval $[0, u_j]$ into t_j sub-intervals $[\ell_{1j} := 0, u_{1j}]$, $[\ell_{2j}, u_{2j}], \ldots, [\ell_{t_j j}, u_{t_j j} := u_j]$
- Given a partition $\mathcal{P}$, the collection of all n_1-tuples $\kappa := (\kappa_1, \ldots, \kappa_{n_1})$, where $\kappa_j \in \{1, \ldots, t_j\}$ for $j = 1, \ldots, n_1$, is denoted by $K(\mathcal{P})$.
- A **unit partition**, $\mathcal{P}^*$, of all integer points is a partition for which $u_{\kappa_j j} - \ell_{\kappa_j j} \leq 1$, for all $\kappa_j = 1, \ldots, t_j$, and all $j = 1, \ldots, n_1$.

A Finite Disjunctive Characterization for General MILP

For a given vector $\kappa \in K(\mathcal{P}^*)$, an index j, and a polyhedron $\bar{X}$, let

$$P^-(\kappa, j, \bar{X}) := \{x \in \bar{X} | \ell_{\kappa_i i} \leq x_i \leq u_{\kappa_i i}, i = 1, \dots, n_1; x_j \leq \ell_{\kappa_j j}\},$$

$$P^+(\kappa, j, \bar{X}) := \{x \in \bar{X} | \ell_{\kappa_i i} \leq x_i \leq u_{\kappa_i i}, i = 1, \dots, n_1; x_j \geq u_{\kappa_j j}\}.$$

Also let $\mathcal{H}_j^\kappa(\bar{X}) := \text{clconv}(P^-(\kappa, j, \bar{X}) \cup P^+(\kappa, j, \bar{X}) \setminus \emptyset)$

A Finite Disjunctive Characterization for General MILP

For a given vector $\kappa \in K(\mathcal{P}^*)$, an index j, and a polyhedron $\bar{X}$, let

$$P^-(\kappa, j, \bar{X}) := \{x \in \bar{X} | \ell_{\kappa_i i} \le x_i \le u_{\kappa_i i}, i = 1, \ldots, n_1; x_j \le \ell_{\kappa_j j}\},$$

$$P^+(\kappa, j, \bar{X}) := \{x \in \bar{X} | \ell_{\kappa_i i} \le x_i \le u_{\kappa_i i}, i = 1, \ldots, n_1; x_j \ge u_{\kappa_j j}\}.$$

Also let $\mathcal{H}_j^\kappa(\bar{X}) := \text{clconv}(P^-(\kappa, j, \bar{X}) \cup P^+(\kappa, j, \bar{X}) \setminus \emptyset)$

Theorem (Sequential convexification of General MILP [Chen, K., and Sen, 2011])

Given a set $X = \{x \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n-n_1} | Ax \ge b\}$, $X \neq \emptyset$, with bounded integer variables, for any unit partition $\mathcal{P}^*$,

$$\text{clconv}(X) = \text{clconv}\{\cup_{\kappa \in K(\mathcal{P}^*)}[\mathcal{H}_{n_1}^\kappa(\mathcal{H}_{n_1-1}^\kappa(\cdots(\mathcal{H}_1^\kappa(X_L))\cdots)) \setminus \emptyset]\}.$$

A Finite Disjunctive Characterization for General MILP

For a given vector $\kappa \in K(\mathcal{P}^*)$, an index j, and a polyhedron $\bar{X}$, let

$$P^-(\kappa, j, \bar{X}) := \{x \in \bar{X} | \ell_{\kappa_i i} \leq x_i \leq u_{\kappa_i i}, i = 1, \dots, n_1; x_j \leq \ell_{\kappa_j j}\},$$

$$P^+(\kappa, j, \bar{X}) := \{x \in \bar{X} | \ell_{\kappa_i i} \leq x_i \leq u_{\kappa_i i}, i = 1, \dots, n_1; x_j \geq u_{\kappa_j j}\}.$$

Also let $\mathcal{H}_j^\kappa(\bar{X}) := \text{clconv}(P^-(\kappa, j, \bar{X}) \cup P^+(\kappa, j, \bar{X}) \setminus \emptyset)$

Theorem (Sequential convexification of General MILP [Chen, K., and Sen, 2011])

Given a set $X = \{x \in \mathbb{Z}_+^{n_1} \times \mathbb{R}_+^{n-n_1} | Ax \geq b\}$, $X \neq \emptyset$, with bounded integer variables, for any unit partition $\mathcal{P}^*$,

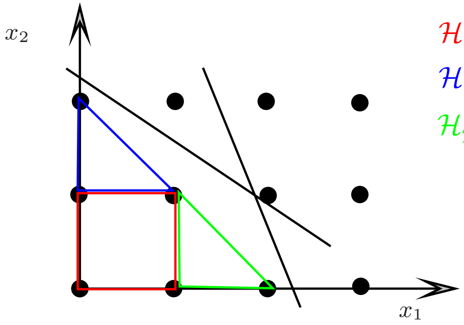
$$\text{clconv}(X) = \text{clconv}\{\cup_{\kappa \in K(\mathcal{P}^*)}[\mathcal{H}_{n_1}^\kappa(\mathcal{H}_{n_1-1}^\kappa(\cdots(\mathcal{H}_1^\kappa(X_L))\cdots)) \setminus \emptyset]\}.$$

Proof idea. The set $K(\mathcal{P}^*)$ decomposes the problem into boxes of at most unit size, each of which can be sequentially convexified.

Example (cont.)

A unit partition $\mathcal{P}^*$ is given by $x_j \in \{[0,1],[1,2],[2,3]\}$ for $j = 1,2$, $t_j = 3$ and $\kappa_j \in \{1,2,3\}$ for $j = 1,2$.

$$K(\mathcal{P}^*) = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)\}.$$



$\mathcal{H}_2^{(1,1)}(\mathcal{H}_1^{(1,1)}(X_L))$
$\mathcal{H}_2^{(1,2)}(\mathcal{H}_1^{(1,2)}(X_L))$
$\mathcal{H}_2^{(2,1)}(\mathcal{H}_1^{(2,1)}(X_L))$

How can we make this practical?

Unit partition contains exponentially many pieces.

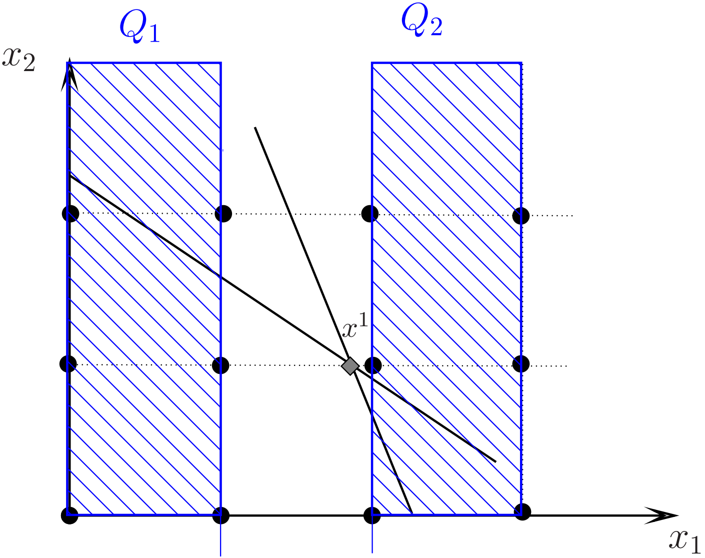
Overview of the Cutting plane tree (CPT) algorithm

Given a fractional point x, find and add a violated disjunctive cut, re-solve LP.

- Add one valid cut at a time from "box" disjunctions (Q_t's), using a cut generation LP (CGLP)
- Obtain Q_t's **on-the-fly** using a **cutting plane tree**
- CPT provides the memory needed for finite convergence.

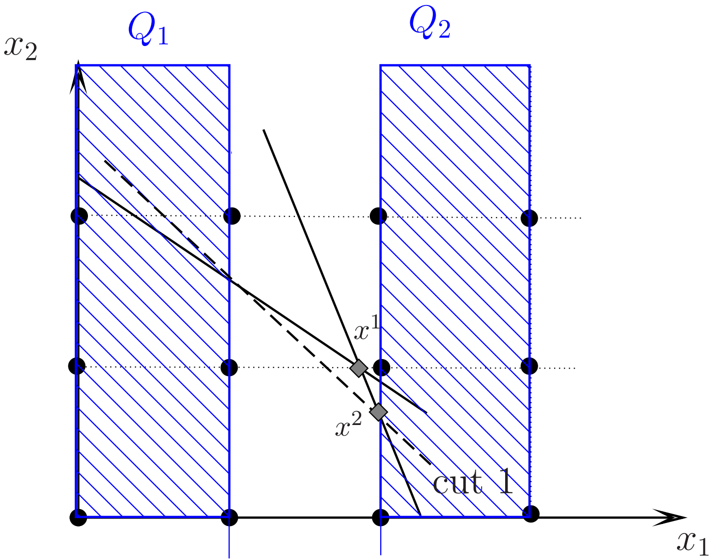
Example (cont.)

Cutting plane tree algorithm



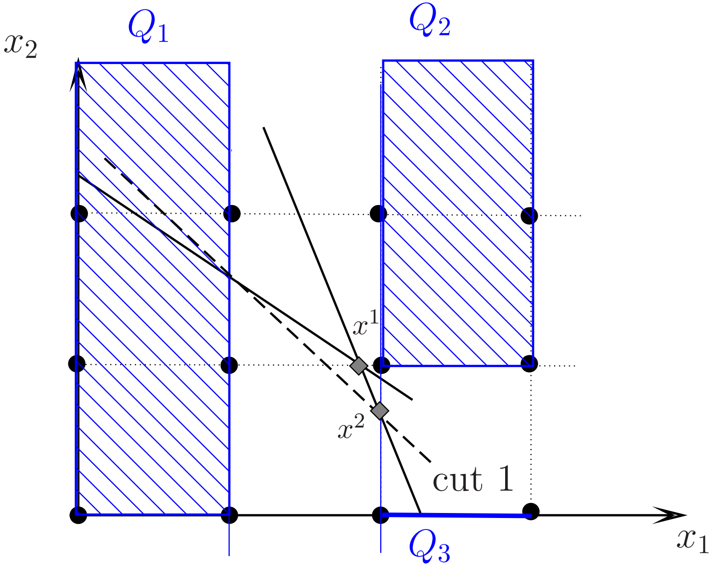
Example (cont.)

Cutting plane tree algorithm



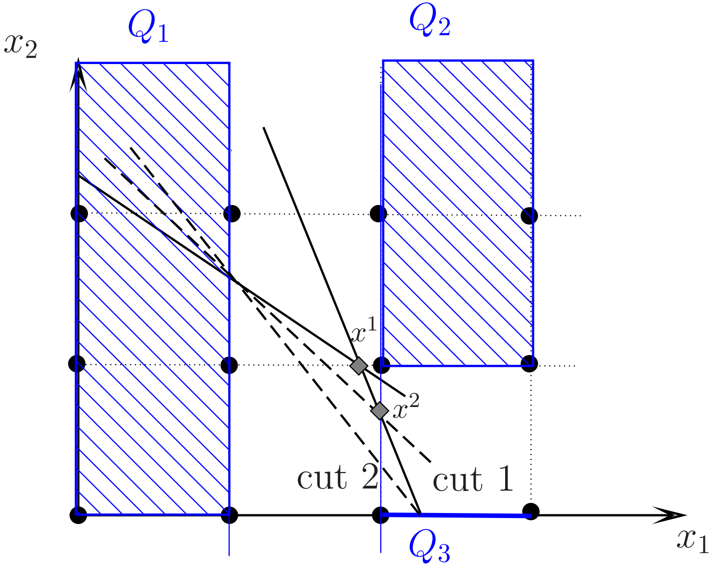
Example (cont.)

Cutting plane tree algorithm



Example (cont.)

Cutting plane tree algorithm



Example (cont.)

CPT algorithm

Iteration 1.

- Solve LP relaxation: $x^1 = (15/8, 1)$.

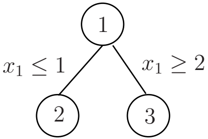
Example (cont.)

CPT algorithm

Iteration 1 (cont.)

- Create two branches in CPT: $x_1 \leq 1$ and $x_1 \geq 2$
- Solve the CGLP based on the two disjunctions (nodes 2&3) to generate a violated cut:

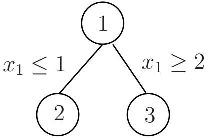
$$\frac{11}{12}x_1 + x_2 \leq \frac{5}{2}$$



Example (cont.)
CPT algorithm

Iteration 2.

- Solve LP relaxation: $x^2 = (2, 2/3)$.
- Search the current CPT to find where x^2 falls. (Node 3)



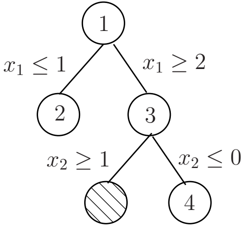
Example (cont.)

CPT algorithm

Iteration 2 (cont.)

- Create 2 branches for node 3: $x_2 \leq 0$ and $x_2 \geq 1$, remove infeasible nodes (crossed).
- Solve the CGLP based on the 2 disjunctions (nodes 2&4) to generate a violated cut:

$$x_1 + \frac{15}{19}x_2 \leq \frac{9}{4}$$

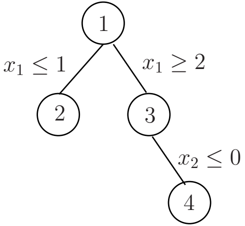


Example (cont.)

CPT algorithm

Iteration 3.

- Solve LP relaxation: $x^3 = (1, 19/12)$.
- Search the current CPT to find where x^3 falls. (Node 2)



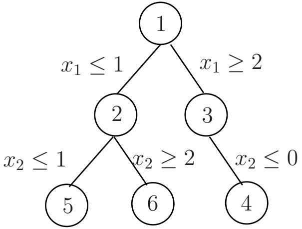
Example (cont.)

CPT algorithm

Iteration 3 (cont.)

- Create 2 branches for node 2: $x_2 \leq 1$ and $x_2 \geq 2$.
- Solve the CGLP based on the 3 disjunctions (nodes 4,5&6) to generate a violated cut:

$$x_1 + \frac{15}{16}x_2 \leq \frac{9}{4}$$

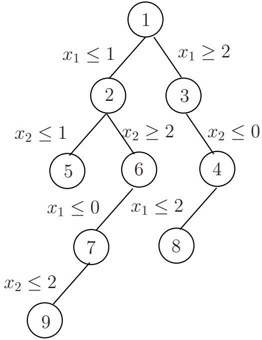


Example (cont.)

CPT algorithm

Iteration 7.

- Solve LP relaxation: $x^7 = (2, 0)$.



Finite convergence of CPT

Theorem ([Chen, K., and Sen, 2011])

For a general MILP with bounded integer variables, the cutting plane tree algorithm converges to an optimal solution in finitely many iterations.

Finite convergence of CPT

Theorem ([Chen, K., and Sen, 2011])

For a general MILP with bounded integer variables, the cutting plane tree algorithm converges to an optimal solution in finitely many iterations.

Proof sketch.

- The number of possible leaf nodes is finite. In the worst case, we reach a unit partition, $\mathcal{P}^*$.
- There are finitely many extreme points of the CGLP for clconv$\{\cup_{Q_t \in \mathcal{P}^*}(Q_t \cap X_{m_\sigma})\}$
- A node σ is visited finitely many times.
- The unique path from the root node to each leaf node defines a $\kappa \in K(\mathcal{P}^*)$.
- Now use General MILP Sequential Convexification Theorem.

Finite convergence of CPT

Theorem ([Chen, K., and Sen, 2011])

For a general MILP with bounded integer variables, the cutting plane tree algorithm converges to an optimal solution in finitely many iterations.

Proof sketch.

- The number of possible leaf nodes is finite. In the worst case, we reach a unit partition, $\mathcal{P}^*$.
- There are finitely many extreme points of the CGLP for $\text{clconv}\{\cup_{Q_t \in \mathcal{P}^*}(Q_t \cap X_{m_\sigma})\}$
- A node σ is visited finitely many times.
- The unique path from the root node to each leaf node defines a $\kappa \in K(\mathcal{P}^*)$.
- Now use General MILP Sequential Convexification Theorem.

[Chen, K., Sen, 2012] tests CPT algorithm on (deterministic) MIPLIB instances
[Qi and Sen, 2017, 2021+] leverage the CPT algorithm for two-stage stochastic MIPs

Discussion

- Successful adaptation of Benders-type approaches require
  - finite convexification in second stage,
  - tractable lifting of first-stage variables
- Extended formulations in second stage, e.g., [Kim and Mehrotra, 2015], [Bansal et al., 2018]
- Convex approximations, e.g., [Romeijnders et al., 2016], [van der Laan and Romeijnders, 2020+]
- Multi-stage stochastic MIP: SDDiP (JuMP) [Zou et al., 2019]
- Progressive hedging (Py-SP), e.g., [Rockafellar and Wets, 2004], [Watson et al., 2012], [Gade et al., 2016]
- Two-stage stochastic mixed-integer nonlinear programs, e.g., [Mehrotra and Özevin, 2009], [Li and Grossmann, 2018, 2019]

Outline

1. Two-Stage Stochastic Integer Programming
   - Two-Stage Stochastic Linear Programming
   - Classification Scheme for Two-Stage Stochastic Mixed-Integer Programming
   - Two-Stage Stochastic Pure Integer Programming
   - Two-Stage Stochastic Mixed-Integer Programming

2. Chance-Constrained Programming
   - Static Joint Chance-Constrained Programming
   - Two-stage (Dynamic) Chance-Constrained Programming
   - Distributionally Robust Chance-Constrained Programming

Risk-Averse Optimization

Modeling risk/reliability/quality-of-service restrictions

- Rare events with dire consequences
- Not every realization of uncertain data may lead to a feasible solution
- Using risk-neutral models (expectations) do not capture the risk involved with low probability events
- There exist multiple correlated risk criteria
- Supply chain disruptions, natural disasters, pandemic, etc.

Risk Models and Challenges

- Quantitative risk models
    - Models with (multivariate) conditional-value-at-risk (CVaR)
    - Stochastic multi-objective optimization: Efficient frontier stochastic
- Qualitative risk models
    - Models with joint chance-constraints
    - Feasible region highly non-convex
- A large number of samples (scenarios) needed to represent uncertainty

Preliminaries: Value-at-Risk (VaR)

Definition

For a univariate random variable X, with cumulative distribution function F_X, the *value-at-risk* (VaR) at confidence level $(1-\epsilon)$, also known as $(1-\epsilon)$-quantile, is given by:

$$\text{VaR}_{1-\epsilon}(X) = \min\{\eta \ : \ F_X(\eta) \geq 1-\epsilon\}. \tag{1}$$

- From (1), for any $x \in \mathbb{R}$, the inequalities $\text{VaR}_{1-\epsilon}(X) \leq \tau$ and $\mathbb{P}(X \leq \tau) \geq 1-\epsilon$ are equivalent.
- In optimization context, the r.v. X is dependent on the decision vector x and uncertain parameters ω
- In this context, a chance constraint on random variable X can be equivalently represented as a constraint on its VaR.
- Here, larger values of X are considered risky (e.g., losses).

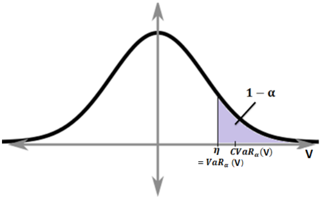
Preliminaries: Conditional Value-at-Risk (CVaR)

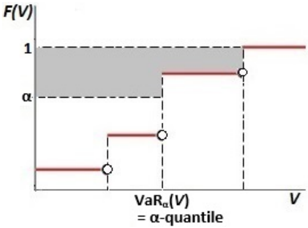
Definition ([Rockafellar and Uryasev, 2000,2002])

The *conditional value-at-risk* (CVaR) at confidence level $(1-\epsilon) \in (0,1]$ is given by

$$\mathrm{CVaR}_{1-\epsilon}(X) = \min\left\{\eta + \frac{1}{\epsilon}\mathbb{E}\left([X-\eta]_+\right) \ : \ \eta \in \mathbb{R}\right\}, \tag{2}$$

where $(a)_+ := \max\{0, a\}$.





Here $\alpha = 1 - \epsilon$.

Preliminaries: Alternative Representations of CVaR

- Suppose X is a r.v. with realizations $X_1, \ldots, X_N$ and probabilities $p_1, \ldots, p_N$.
- The optimization problem in (2) can equivalently be formulated as the linear program (LP):

$$\min \left\{ \eta + \frac{1}{\epsilon} \sum_{i \in [N]} p_i w_i \ : \ w_i \geq X_i - \eta, \ \forall \ i \in [N], \quad w \in \mathbb{R}_+^N \right\}. \tag{3}$$

- Let ρ denote an ordering of the realizations such that $X_{\rho_1} \leq X_{\rho_2} \leq \cdots \leq X_{\rho_N}$. Then, for a given confidence level $\epsilon \in (0,1]$ we have

$$\mathrm{VaR}_{1-\epsilon}(X) = X_{\rho_q}, \text{ where } q = \min \left\{ j \in [N] : \sum_{i \in [j]} p_{\rho_i} \geq 1 - \epsilon \right\}. \tag{4}$$

- CVaR provides a tractable approximation to an individual VaR constraint. (Replace $\mathrm{VaR}_{1-\epsilon}(X) \leq \tau$ with $\mathrm{CVaR}_{1-\epsilon}(X) \leq \tau$.)
- How about the multivariate case? [Prékopa, 1990], [K. and Noyan, 2016], [Meraklı and K., 2018]

Outline

Static Joint chance-constrained program (CCP)

- A linear joint chance-constrained program (CCP) with right-hand-side uncertainty is an optimization problem of the following form:

$$\min \left\{ c^\top x : \ \mathbb{P}\left[Ax \geq b(\omega) \right] \geq 1 - \epsilon, \ x \in X \right\} \tag{CCP}$$

  where
  - $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space,
  - X is a (polyhedral) domain,
  - $\epsilon \in (0, 1)$ is a risk level, and
  - $b(\omega)$ is the random right-hand-side vector that depends on the random variable $\omega \in \Omega$.

- Dates back to [Charnes et al., 1958], [Charnes and Cooper, 1959, 1963] (*individual* chance constraints), and [Miller and Wagner, 1965], [Prékopa,1973] (*joint* chance constraints)

- Why can't we handle $\mathbb{P}[f(x, \xi) \geq 0] \geq 1 - \epsilon$ directly?
  - Non-convex unless certain restrictive assumptions, e.g., [Prékopa, 1990], [Sen, 1992], [Dentcheva et al., 2000]
  - Evaluating $\mathbb{P}[f(x, \xi) \geq 0]$ is difficult (multidimensional integration).
  - In practice, $\mathbb{P}$ is often unknown. (We'll address this later.)

Static Joint chance-constrained program (CCP)

- A linear joint chance-constrained program (CCP) with right-hand-side uncertainty is an optimization problem of the following form:

$$\min\left\{c^\top x:\ \mathbb{P}\left[Ax \geq b(\omega)\right] \geq 1-\epsilon,\ x \in X\right\} \tag{CCP}$$

  where
  - $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space,
  - X is a (polyhedral) domain,
  - $\epsilon \in (0,1)$ is a risk level, and
  - $b(\omega)$ is the random right-hand-side vector that depends on the random variable $\omega \in \Omega$.

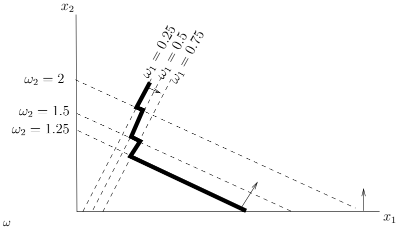
- Dates back to [Charnes et al., 1958], [Charnes and Cooper, 1959, 1963] (*individual* chance constraints), and [Miller and Wagner, 1965], [Prékopa,1973] (*joint* chance constraints)
- Used in modeling problems with "random supplies/demands".
- Why can't we handle $\mathbb{P}[f(x,\xi) \geq 0] \geq 1-\epsilon$ directly?
  - Non-convex unless certain restrictive assumptions, e.g., [Prékopa, 1990], [Sen, 1992], [Dentcheva et al., 2000]
  - Evaluating $\mathbb{P}[f(x,\xi) \geq 0]$ is difficult (multidimensional integration).
  - In practice, $\mathbb{P}$ is often unknown. (We'll address this later.)

Non-convex feasible region example adapted from [Sen, 1992]

$$
\begin{aligned}
\min \quad & x_1 + x_2 \\
\text{s.t.} \quad & \mathbb{P}\left\{ \begin{array}{ccc} 2x_1 - x_2 & \geq & \omega_1 \\ x_1 + 2x_2 & \geq & \omega_2 \end{array} \right\} \geq 0.6 \\
& x \geq 0,
\end{aligned}
$$

with joint probability density function of ω

Scenario	1	2	3	4	5	6	7	8	9
ω_1	0.75	0.5	0.5	0.25	0.25	0.25	0	0	0
ω_2	1.25	1.5	1.25	1.75	1.5	1.25	2	1.5	1.25
Probability	0.2	0.14	0.06	0.06	0.06	0.3	0.04	0.04	0.1



Finite sample space assumption

- We consider the setting where Ω is a finite sample space:

$$\Omega = \{\omega^1, \ldots, \omega^N\}$$

Finite sample space assumption

- We consider the setting where Ω is a finite sample space:

$$\Omega = \{\omega^1, \ldots, \omega^N\}$$

- Even if Ω is not finite, we can approximate (CCP) via an empirical distribution (see the theory of Sample Average Approximation (SAA), e.g., [Luedtke and Ahmed, 2008], [Calafiore and Campi, 2005, 2006]).

Finite sample space assumption

- We consider the setting where Ω is a finite sample space:

$$\Omega = \{\omega^1, \ldots, \omega^N\}$$

- Even if Ω is not finite, we can approximate (CCP) via an empirical distribution (see the theory of Sample Average Approximation (SAA), e.g., [Luedtke and Ahmed, 2008], [Calafiore and Campi, 2005, 2006]).
- Assuming that $\mathbb{P}\left[\omega = \omega^i\right] = p_i$ for $i \in [N]$,

$$\min\left\{c^\top x : \ \mathbb{P}\left[Ax \geq b(\omega)\right] \geq 1 - \epsilon, \ x \in X\right\} \tag{CCP}$$

can be rewritten as

$$\min\left\{c^\top x : \sum_{i\in[N]} p_i \mathbb{1}\left[Ax \geq b(\omega^i)\right] \geq 1 - \varepsilon, \ x \in X\right\}.$$

  - Also known as (ML) empirical risk, (stats) Monte Carlo.

Reformulation

- There is a deterministic reformulation: the problem can be reformulated as the following mixed-integer program [Ruszczyński, 2001],

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & Ax = y, \\
& y \geq b(\omega^i)(1 - z_i), \quad \forall i \in [N], \\
& \sum_{i \in [N]} p_i (1 - z_i) \geq 1 - \epsilon, \\
& x \in X, \; y \in \mathbb{R}^k_+, \; z \in \{0, 1\}^N,
\end{aligned}
$$

where

- we assume that $Ax \geq \mathbf{0}$ holds for all $x \in X$,
- $b(\omega^i) \geq \mathbf{0}$ for all i, i.e., $Ax \geq \mathbf{0}$ is satisfied for all $x \in X$,
- $1 - z_i \simeq \mathbb{1}\left[Ax \geq b(\omega^i)\right]$:

$$Ax \geq b(\omega^i) \text{ if } z_i = 0 \quad \text{and} \quad Ax \geq \mathbf{0} \text{ if } z_i = 1.$$

Big-M Reformulation

The problem can be reformulated as the following mixed-integer program:

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & Ax = y, \\
& y_j \geq w_{ij}(1 - z_i), \quad \forall i \in [N], \forall j \in [k], && \text{(big-M)} \\
& \sum_{i \in [N]} p_i z_i \leq \epsilon, && \text{(knapsack)} \\
& x \in X, \ y \in \mathbb{R}^k_+, \ z \in \{0,1\}^N,
\end{aligned}
$$

where $W = \{w_{ij}\} \in \mathbb{R}^{N \times k}_+$ is a nonnegative matrix.

Difficulties

- The MIP formulation is often difficult to solve.

Difficulties

- The MIP formulation is often difficult to solve.
- In fact, its LP relaxation is weak:

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & Ax = y, \\
& y_j \geq w_{ij}(1 - z_i), \quad \forall i \in [N], \forall j \in [k], && \text{(big-M)} \\
& \sum_{i \in [N]} p_i z_i \leq \epsilon, && \text{(knapsack)} \\
& x \in X,\ y \in \mathbb{R}^k_+,\ z \in [0,1]^N.
\end{aligned}
$$

Difficulties

- The MIP formulation is often difficult to solve.
- In fact, its LP relaxation is weak:

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & Ax = y, \\
& y_j \geq w_{ij}(1 - z_i), \quad \forall i \in [N], \forall j \in [k], && \text{(big-M)} \\
& \sum_{i \in [N]} p_i z_i \leq \epsilon, && \text{(knapsack)} \\
& x \in X, \ y \in \mathbb{R}^k_+, \ z \in [0,1]^N.
\end{aligned}
$$

- We will strengthen the formulation by integer programming techniques.

Known substructures

- We refer to the set

$$\left\{(y,z) \in \mathbb{R}^k_+ \times \{0,1\}^N : \; y_j \geq w_{ij}(1-z_i), \; \forall i \in [N], \forall j \in [k]\right\} \tag{Mix}$$

  as a (joint) mixing set (term coined by [Günlük and Pochet, 2001] for related set with general integer variables).

Known substructures

- We refer to the set
$$\left\{(y,z) \in \mathbb{R}^k_+ \times \{0,1\}^N : \; y_j \geq w_{ij}(1-z_i), \; \forall i \in [N], \forall j \in [k]\right\} \tag{Mix}$$
as a (joint) mixing set (term coined by [Günlük and Pochet, 2001] for related set with general integer variables).
- One can obtain the convex hull of (Mix) by adding the so-called mixing (or star) inequalities [Atamtürk, Nemhauser, Savelsbergh, 2000].

Known substructures

- We refer to the set

$$\left\{(y,z) \in \mathbb{R}^k_+ \times \{0,1\}^N : \ y_j \geq w_{ij}(1-z_i), \ \forall i \in [N], \forall j \in [k]\right\} \tag{Mix}$$

  as a (joint) mixing set (term coined by [Günlük and Pochet, 2001] for related set with general integer variables).
- One can obtain the convex hull of (Mix) by adding the so-called mixing (or star) inequalities [Atamtürk, Nemhauser, Savelsbergh, 2000].
- We call the set

$$\left\{(y,z) \in \text{(Mix)} : \ \sum_{i \in [N]} p_i z_i \leq \epsilon\right\} \tag{Mix-knapsack}$$

  as a (joint) mixing set with a knapsack constraint.

Known substructures

- We refer to the set

$$\left\{(y,z) \in \mathbb{R}^k_+ \times \{0,1\}^N : \; y_j \geq w_{ij}(1-z_i), \; \forall i \in [N], \forall j \in [k]\right\} \tag{Mix}$$

  as a (joint) mixing set (term coined by [Günlük and Pochet, 2001] for related set with general integer variables).
- One can obtain the convex hull of (Mix) by adding the so-called mixing (or star) inequalities [Atamtürk, Nemhauser, Savelsbergh, 2000].
- We call the set

$$\left\{(y,z) \in \text{(Mix)} : \sum_{i \in [N]} p_i z_i \leq \epsilon \right\} \tag{Mix-knapsack}$$

  as a (joint) mixing set with a knapsack constraint.
- Valid inequalities for (Mix-knapsack) are given in [Luedtke, Ahmed, Nemhauser, 2010], [K., 2012], [Abdi and Fukasawa, 2016], [Zhao, Huang, Zeng, 2017].

Known substructures

- We refer to the set

$$\left\{(y,z) \in \mathbb{R}_+^k \times \{0,1\}^N : \ y_j \geq w_{ij}(1-z_i), \ \forall i \in [N], \forall j \in [k]\right\} \tag{Mix}$$

  as a (joint) mixing set (term coined by [Günlük and Pochet, 2001] for related set with general integer variables).
- One can obtain the convex hull of (Mix) by adding the so-called mixing (or star) inequalities [Atamtürk, Nemhauser, Savelsbergh, 2000].
- We call the set

$$\left\{(y,z) \in \text{(Mix)} : \ \sum_{i \in [N]} p_i z_i \leq \epsilon\right\} \tag{Mix-knapsack}$$

  as a (joint) mixing set with a knapsack constraint.
- Valid inequalities for (Mix-knapsack) are given in [Luedtke, Ahmed, Nemhauser, 2010], [K., 2012], [Abdi and Fukasawa, 2016], [Zhao, Huang, Zeng, 2017].
- Random technology matrix and right-hand-side extensions [Tanner and Ntaimo, 2010], [Luedtke, 2014]

Known substructures

- We refer to the set

$$\left\{(y, z) \in \mathbb{R}_+^k \times \{0,1\}^N : \ y_j \geq w_{ij}(1 - z_i), \ \forall i \in [N], \forall j \in [k]\right\} \qquad \text{(Mix)}$$

  as a (joint) mixing set (term coined by [Günlük and Pochet, 2001] for related set with general integer variables).
- One can obtain the convex hull of (Mix) by adding the so-called mixing (or star) inequalities [Atamtürk, Nemhauser, Savelsbergh, 2000].
- We call the set

$$\left\{(y, z) \in \text{(Mix)} : \ \sum_{i \in [N]} p_i z_i \leq \epsilon\right\} \qquad \text{(Mix-knapsack)}$$

  as a (joint) mixing set with a knapsack constraint.
- Valid inequalities for (Mix-knapsack) are given in [Luedtke, Ahmed, Nemhauser, 2010], [K., 2012], [Abdi and Fukasawa, 2016], [Zhao, Huang, Zeng, 2017].
- Random technology matrix and right-hand-side extensions [Tanner and Ntaimo, 2010], [Luedtke, 2014]
- It is harder to convexify (Mix-knapsack) due to the knapsack structure.

Binary mixing (star) inequalities

- The basic mixing set for given $j \in [k]$:

$$\left\{(y_j, z) \in \mathbb{R} \times \{0,1\}^N : \; y_j \geq w_{ij}(1 - z_i), \; \forall i \in [N]\right\}$$

Binary mixing (star) inequalities

- The basic mixing set for given $j \in [k]$:

$$\left\{(y_j, z) \in \mathbb{R} \times \{0,1\}^N : \ y_j \geq w_{ij}(1 - z_i), \ \forall i \in [N]\right\}$$

- The mixing inequality for a given subset $\Pi_j = \{j_1, \ldots, j_\tau\}$ with $w_{j_1 j} \geq \cdots \geq w_{j_\tau j}$ is:

$$y_j + \sum_{s \in [\tau]} (w_{j_s j} - w_{j_{s+1} j}) z_{j_s} \geq w_{j_1 j}$$

where $w_{j_{\tau+1} j} := 0$.

Binary mixing (star) inequalities

- The basic mixing set for given $j \in [k]$:

$$\left\{(y_j, z) \in \mathbb{R} \times \{0,1\}^N : \ y_j \geq w_{ij}(1 - z_i), \ \forall i \in [N]\right\}$$

- The mixing inequality for a given subset $\Pi_j = \{j_1, \dots, j_\tau\}$ with $w_{j_1 j} \geq \cdots \geq w_{j_\tau j}$ is:

$$y_j + \sum_{s \in [\tau]} (w_{j_s j} - w_{j_{s+1} j}) z_{j_s} \geq w_{j_1 j}$$

  where $w_{j_{\tau+1} j} := 0$.

- For example, the convex hull of

$$\left\{ (y_1, z) \in \mathbb{R}_+ \times \{0,1\}^3 \quad : \quad \begin{array}{l} y_1 \geq 8(1 - z_1) \\ y_1 \geq 6(1 - z_2) \\ y_1 \geq 13(1 - z_3) \end{array} \right\}$$

  is

$$\left\{ (y_1, z) \in \mathbb{R}_+ \times [0,1]^3 \quad : \quad \begin{array}{l} y_1 \geq 13 - 6z_2 - 7z_3 \\ y_1 \geq 13 - 13z_3 \\ y_1 \geq 13 - 8z_1 - 5z_3 \\ y_1 \geq 13 - 2z_1 - 6z_2 - 5z_3 \end{array} \right\}$$

$$= \left\{(y_1, z) \in \mathbb{R}_+ \times [0,1]^3 : \ \text{the mixing inequalities for } y_1\right\}.$$

How about the knapsack constraint?

- Typically, $p_i = \frac{1}{N}$ due to i.i.d. sampling
- In this case, the knapsack constraint is a cardinality constraint:

$$\sum_{i \in [N]} z_i \leq \lfloor N\epsilon \rfloor =: q$$

- Suppose $w_{1j} \geq \cdots \geq w_{Nj}$, then we must have

$$y_j \geq w_{(q+1)j}$$

- Use this to strengthen the formulation as

$$\left\{ (y_j, z) \in \mathbb{R} \times \{0,1\}^N : \; y_j + (w_{ij} - w_{(q+1)j}) z_i \geq w_{ij}, \; \forall i \in [q], \sum_{i \in [N]} z_i \leq q \right\}$$

- Apply mixing inequalities to the strengthened formulation [Luedtke et al., 2010].

Quantile cuts

- We can exploit the knapsack structure "indirectly" by the quantile cuts [Luedtke, 2014], [Xie and Ahmed, 2018].

Quantile cuts

- We can exploit the knapsack structure "indirectly" by the quantile cuts [Luedtke, 2014], [Xie and Ahmed, 2018].

- A quantile cut is of the following form: for some $h \in \mathbb{R}_+^k$,

$$h^\top y \geq \min \left\{ h^\top y : \ (y, z) \in \text{(Mix-knapsack)} \right\}.$$

Quantile cuts

- We can exploit the knapsack structure "indirectly" by the quantile cuts [Luedtke, 2014], [Xie and Ahmed, 2018].
- A quantile cut is of the following form: for some $h \in \mathbb{R}_+^k$,

$$h^\top y \geq \min \left\{ h^\top y : \ (y, z) \in \text{(Mix-knapsack)} \right\}.$$

- Quantile cuts are valid for (Mix-knapsack), and thus, for the formulation.

Quantile cuts

- We can exploit the knapsack structure "indirectly" by the quantile cuts [Luedtke, 2014], [Xie and Ahmed, 2018].
- A quantile cut is of the following form: for some $h \in \mathbb{R}_+^k$,

$$h^\top y \geq \min \left\{ h^\top y : \ (y, z) \in \text{(Mix-knapsack)} \right\}.$$

- Quantile cuts are valid for (Mix-knapsack), and thus, for the formulation.
- We replace/relax the knapsack constraint by the quantile cut

$$y_1 + \cdots + y_k \geq \varepsilon$$

Mixing set with lower bounds

- Consider the set

$$\left\{ (y,z) : \begin{array}{ll} y_j \geq w_{ij}(1 - z_i), & \forall i \in [N], \forall j \in [k], \\ y_1 + \cdots + y_k \geq \varepsilon, & \\ y \in \mathbb{R}^k_+, \ z \in \{0,1\}^N & \end{array} \right\} \tag{Mix-lb}$$

referred to as a (joint) mixing set with lower bounds.

Mixing set with lower bounds

- Consider the set

$$\left\{ (y,z): \begin{array}{ll} y_j \geq w_{ij}(1-z_i), & \forall i \in [N], \forall j \in [k], \\ y_1 + \cdots + y_k \geq \varepsilon, & \\ y \in \mathbb{R}^k_+,\ z \in \{0,1\}^N & \end{array} \right\} \tag{Mix-lb}$$

referred to as a (joint) mixing set with lower bounds.

- Our goal is to understand the polyhedral structure of (Mix-lb) to generate strong valid inequalities.

Example 1

- The convex hull of

$$\left\{ (y,z) \in \mathbb{R}^2_+ \times \{0,1\}^3 \;:\; \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \end{array} \;,\; \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \end{array} \right\}$$

  is

$$\left\{ (y,z) \in \mathbb{R}^2_+ \times [0,1]^3 \;:\; \begin{array}{l} y_1 \geq 13 - 6z_2 - 7z_3 \\ y_1 \geq 13 - 13z_3 \\ y_1 \geq 13 - 8z_1 - 5z_3 \\ y_1 \geq 13 - 2z_1 - 6z_2 - 5z_3 \end{array} \;,\; \begin{array}{l} y_2 \geq 4 - z_1 - z_2 - 2z_3 \\ y_2 \geq 4 - 2z_2 - 2z_3 \\ y_2 \geq 4 - 3z_1 - z_2 \\ y_2 \geq 4 - 4z_2 \end{array} \right\}$$

$$= \left\{ (y,z) \in \mathbb{R}^2_+ \times [0,1]^3 : \text{the mixing inequalities for } y_1, y_2 \right\}.$$

- This was shown by [Atamtürk, Nemhauser, Savelsbergh '00].

Example 1

- The convex hull of

$$\left\{ (y,z) \in \mathbb{R}^2_+ \times \{0,1\}^3 \;:\; \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \end{array} \;,\; \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \end{array} \right\}$$

is

$$\left\{ (y,z) \in \mathbb{R}^2_+ \times [0,1]^3 \;:\; \begin{array}{l} y_1 \geq 13 - 6z_2 - 7z_3 \\ y_1 \geq 13 - 13z_3 \\ y_1 \geq 13 - 8z_1 - 5z_3 \\ y_1 \geq 13 - 2z_1 - 6z_2 - 5z_3 \end{array} \;,\; \begin{array}{l} y_2 \geq 4 - z_1 - z_2 - 2z_3 \\ y_2 \geq 4 - 2z_2 - 2z_3 \\ y_2 \geq 4 - 3z_1 - z_2 \\ y_2 \geq 4 - 4z_2 \end{array} \right\}$$

$$= \left\{ (y,z) \in \mathbb{R}^2_+ \times [0,1]^3 : \text{ the mixing inequalities for } y_1, y_2 \right\}.$$

Example 1

- The convex hull of

$$\left\{ (y,z) \in \mathbb{R}^2_+ \times \{0,1\}^3 \; : \; \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \end{array} \; , \; \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \end{array} \right\}$$

  is

$$\left\{ (y,z) \in \mathbb{R}^2_+ \times [0,1]^3 \; : \; \begin{array}{l} y_1 \geq 13 - 6z_2 - 7z_3 \\ y_1 \geq 13 - 13z_3 \\ y_1 \geq 13 - 8z_1 - 5z_3 \\ y_1 \geq 13 - 2z_1 - 6z_2 - 5z_3 \end{array} \; , \; \begin{array}{l} y_2 \geq 4 - z_1 - z_2 - 2z_3 \\ y_2 \geq 4 - 2z_2 - 2z_3 \\ y_2 \geq 4 - 3z_1 - z_2 \\ y_2 \geq 4 - 4z_2 \end{array} \right\}$$

$$= \left\{ (y,z) \in \mathbb{R}^2_+ \times [0,1]^3 : \text{ the mixing inequalities for } y_1, y_2 \right\}.$$

- This was shown by [Atamtürk, Nemhauser, Savelsbergh '00].

Example 2

The convex hull of

$$\left\{ \begin{array}{l} (y,z) \in \\ \mathbb{R}^2_+ \times \{0,1\}^3 \end{array} : \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \\ y_1 \geq (1-z_4) \\ y_1 \geq 4(1-z_5) \end{array}, \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \\ y_2 \geq 2(1-z_4) \\ y_2 \geq (1-z_5) \end{array}, y_1 + y_2 \geq 7 \right\}$$

is

Example 2

The convex hull of

$$\left\{ \begin{array}{c} (y,z) \in \\ \mathbb{R}^2_+ \times \{0,1\}^3 \end{array} : \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \\ y_1 \geq (1-z_4) \\ y_1 \geq 4(1-z_5) \end{array}, \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \\ y_2 \geq 2(1-z_4) \\ y_2 \geq (1-z_5) \end{array}, y_1 + y_2 \geq 7 \right\}$$

is

$$\left\{ \begin{array}{c} (y,z) \in \\ \mathbb{R}^2_+ \times [0,1]^3 \end{array} : \begin{array}{l} \text{the mixing inequalities for } y_1, y_2 \\ y_1 + y_2 \geq 17 - z_1 - z_2 - 8z_3 \\ y_1 + y_2 \geq 17 - 2z_2 - 8z_3 \\ y_1 + y_2 \geq 17 - 3z_2 - 7z_3 \\ y_1 + y_2 \geq 17 - 2z_1 - 3z_2 - 5z_3 \\ y_1 + y_2 \geq 17 - 4z_1 - z_2 - 5z_3 \end{array} \right\}$$

$$= \left\{ \begin{array}{c} (y,z) \in \\ \mathbb{R}^2_+ \times [0,1]^3 \end{array} : \begin{array}{l} \text{the mixing inequalities for } y_1, y_2 \\ \text{the ``aggregated'' mixing inequalities for ``} y_1 + y_2 \text{''} \end{array} \right\}$$

Example 2

The convex hull of

$$\left\{ \begin{array}{l} (y,z) \in \\ \mathbb{R}^2_+ \times \{0,1\}^3 \end{array} : \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \\ y_1 \geq (1-z_4) \\ y_1 \geq 4(1-z_5) \end{array}, \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \\ y_2 \geq 2(1-z_4) \\ y_2 \geq (1-z_5) \end{array}, y_1 + y_2 \geq 7 \right\}$$

is

$$\left\{ \begin{array}{l} (y,z) \in \\ \mathbb{R}^2_+ \times [0,1]^3 \end{array} : \begin{array}{l} \text{the mixing inequalities for } y_1, y_2 \\ y_1 + y_2 \geq 17 - z_1 - z_2 - 8z_3 \\ y_1 + y_2 \geq 17 - 2z_2 - 8z_3 \\ y_1 + y_2 \geq 17 - 3z_2 - 7z_3 \\ y_1 + y_2 \geq 17 - 2z_1 - 3z_2 - 5z_3 \\ y_1 + y_2 \geq 17 - 4z_1 - z_2 - 5z_3 \end{array} \right\}$$

$$= \left\{ \begin{array}{l} (y,z) \in \\ \mathbb{R}^2_+ \times [0,1]^3 \end{array} : \begin{array}{l} \text{the mixing inequalities for } y_1, y_2 \\ \text{the ``aggregated'' mixing inequalities for ``} y_1 + y_2 \text{''} \end{array} \right\}$$

Are the mixing and the aggregated mixing inequalities enough to describe the convex hull of (Mix-lb)?

Example 3

- The convex hull of

$$\left\{ \begin{array}{c} (y,z) \in \\ \mathbb{R}_+^2 \times \{0,1\}^3 \end{array} : \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \\ y_1 \geq (1-z_4) \\ y_1 \geq 4(1-z_5) \end{array} , \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \\ y_2 \geq 2(1-z_4) \\ y_2 \geq (1-z_5) \end{array} , y_1 + y_2 \geq 9 \right\}$$

is

Example 3

- The convex hull of

$$\left\{ \begin{array}{c} (y,z) \in \\ \mathbb{R}^2_+ \times \{0,1\}^3 \end{array} : \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \\ y_1 \geq (1-z_4) \\ y_1 \geq 4(1-z_5) \end{array} , \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \\ y_2 \geq 2(1-z_4) \\ y_2 \geq (1-z_5) \end{array} , y_1 + y_2 \geq 9 \right\}$$

is

$$\left\{ \begin{array}{c} (y,z) \in \\ \mathbb{R}^2_+ \times [0,1]^3 \end{array} : \begin{array}{l} \text{the mixing inequalities for } y_1, y_2 \\ \text{the aggregated mixing inequalities for } y_1 + y_2 \\ 7y_1 + 6y_2 \geq 115 - 12z_2 - 49z_3 \\ 6y_1 + 5y_2 \geq 98 - 10z_2 - 42z_3 - z_4 \\ 3y_1 + 2y_2 \geq 47 - 4z_2 - 21z_3 - z_4 - 3z_5 \\ 3y_1 + 2y_2 \geq 47 - 4z_2 - 21z_3 - 4z_5 \\ 2y_1 + 3y_2 \geq 38 - 6z_2 - 14z_3 \\ y_1 + 2y_2 \geq 21 - 4z_2 - 7z_3 - z_5 \end{array} \right\}$$

Example 4

- The convex hull of

$$\left\{ \begin{array}{l} (y,z) \in \\ \mathbb{R}^2_+ \times \{0,1\}^3 \end{array} : \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \\ y_1 \geq (1-z_4) \\ y_1 \geq 4(1-z_5) \end{array} , \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \\ y_2 \geq 3(1-z_4) \\ y_2 \geq (1-z_5) \end{array} , y_1 + y_2 \geq 7 \right\}$$

is

Example 4

- The convex hull of

$$\left\{ \begin{array}{l}(y,z)\in \\ \mathbb{R}^2_+\times\{0,1\}^3\end{array} : \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \\ y_1 \geq (1-z_4) \\ y_1 \geq 4(1-z_5) \end{array} , \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \\ y_2 \geq 3(1-z_4) \\ y_2 \geq (1-z_5) \end{array} , y_1+y_2\geq 7 \right\}$$

is

$$\left\{ \begin{array}{l}(y,z)\in \\ \mathbb{R}^2_+\times[0,1]^3\end{array} : \begin{array}{l} \text{the mixing inequalities for } y_1, y_2 \\ \text{the aggregated mixing inequalities for } y_1+y_2 \\ 2y_1+3y_2 \geq 38-3z_2-18z_3-3z_4 \\ 2y_1+y_2\geq 30-z_2-21z_3-z_4 \\ 2y_1+y_2\geq 30-z_2-18z_3-z_4-3z_5 \\ y_1+2y_2\geq 21-2z_2-9z_3-2z_4-z_5 \end{array} \right\}$$

Submodularity in joint mixing sets

- When are the mixing and the aggregated mixing inequalities sufficient?

Submodularity in joint mixing sets

- When are the mixing and the aggregated mixing inequalities sufficient?
- We discover an underlying submodularity in (Mix-lb)!

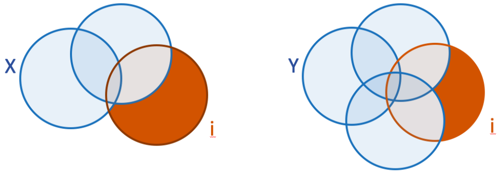
Submodularity in joint mixing sets

- When are the mixing and the aggregated mixing inequalities sufficient?
- We discover an underlying submodularity in (Mix-lb)!
- A function $f \in \{0,1\}^N \to \mathbb{R}$ is submodular if

$$f(A) + f(B) \geq f(A \cap B) + f(A \cup B) \quad \forall A, B \subseteq [N].$$

- Alternatively, a function $f \in \{0,1\}^N \to \mathbb{R}$ is submodular if

$$f(X \cup \{i\}) - f(X) \geq f(Y \cup \{i\}) - f(Y) \quad \forall X \subset Y \subseteq [N], i \notin Y.$$



Submodularity in joint mixing sets

- (Mix) can be written as

$$\left\{(y,z):\ y_j \geq \max_{i\in[N]}\left\{w_{ij}(1-z_i)\right\},\ \forall j\in[k]\right\}$$
$$=\left\{(y,z):\ y_j \geq f_j(\mathbf{1}-z),\ \forall j\in[k]\right\}$$

where

$$f_j(z)=\max_{i\in[N]}\left\{w_{ij}z_i\right\} \quad \text{for } z\in\{0,1\}^N.$$

Submodularity in joint mixing sets

- (Mix) can be written as

$$\left\{(y,z):\ y_j \geq \max_{i\in[N]}\left\{w_{ij}(1-z_i)\right\},\ \forall j\in[k]\right\}$$
$$=\left\{(y,z):\ y_j \geq f_j(\mathbf{1}-z),\ \forall j\in[k]\right\}$$

where

$$f_j(z)=\max_{i\in[N]}\left\{w_{ij}z_i\right\}\quad \text{for } z\in\{0,1\}^N.$$

Remark

Each f_j is a submodular function:

$$\max_{i\in A}\left\{w_{ij}\right\}+\max_{i\in B}\left\{w_{ij}\right\}\geq \max_{i\in A\cup B}\left\{w_{ij}\right\}+\max_{i\in A\cap B}\left\{w_{ij}\right\}$$

for any $A, B\subseteq[N]$.

Submodularity and polymatroid inequalities

- Given a submodular (set) function $f : 2^{[N]} \to \mathbb{R}$, the extended polymatroid of f is

$$EP_f := \{\pi \in \mathbb{R}^n : \ \pi(V) \leq f(V), \ \forall V \subseteq [N]\} .$$

Submodularity and polymatroid inequalities

- Given a submodular (set) function $f : 2^{[N]} \to \mathbb{R}$, the extended polymatroid of f is

$$EP_f := \{\pi \in \mathbb{R}^n : \ \pi(V) \le f(V), \ \forall V \subseteq [N]\} .$$

- Given a submodular function $f : \{0,1\}^N \to \mathbb{R}$, consider

$$Q_f := \left\{(y,z) \in \mathbb{R} \times \{0,1\}^N : \ y \ge f(\mathbf{1} - z)\right\}$$

Submodularity and polymatroid inequalities

- Given a submodular (set) function $f : 2^{[N]} \to \mathbb{R}$, the extended polymatroid of f is
$$EP_f := \{\pi \in \mathbb{R}^n : \ \pi(V) \le f(V), \ \forall V \subseteq [N]\}.$$
- Given a submodular function $f : \{0,1\}^N \to \mathbb{R}$, consider
$$Q_f := \left\{(y, z) \in \mathbb{R} \times \{0,1\}^N : \ y \ge f(\mathbf{1} - z)\right\}$$

Theorem [Lovász, 1983, Atamtürk and Narayanan 2008]

The convex hull of Q_f is given by
$$\left\{(y, z) \in \mathbb{R} \times [0,1]^N : \ y \ge \pi^\top(\mathbf{1} - z) + f(\emptyset), \ \forall \pi \in EP_{f - f(\emptyset)}\right\}.$$

Theorem [Edmonds, 1970]

Let $f : \{0,1\}^n \to \mathbb{R}$ be a submodular function. Then $\pi \in \mathbb{R}^n$ is an extreme point of EP_f if and only if there exists a permutation σ of $[N]$ such that $\pi_{\sigma(t)} = f(V_t) - f(V_{t-1})$, where $V_t = \{\sigma(1), \ldots, \sigma(t)\}$ for $t \in [N]$ and $V_0 = \emptyset$.

Submodularity and polymatroid inequalities

- Given a submodular (set) function $f : 2^{[N]} \to \mathbb{R}$, the extended polymatroid of f is

$$EP_f := \{\pi \in \mathbb{R}^n : \ \pi(V) \leq f(V), \ \forall V \subseteq [N]\}.$$

- Given a submodular function $f : \{0,1\}^N \to \mathbb{R}$, consider

$$Q_f := \left\{(y,z) \in \mathbb{R} \times \{0,1\}^N : \ y \geq f(\mathbf{1} - z)\right\}$$

Theorem [Lovász, 1983, Atamtürk and Narayanan 2008]

The convex hull of Q_f is given by

$$\left\{(y,z) \in \mathbb{R} \times [0,1]^N : \ y \geq \pi^\top(\mathbf{1} - z) + f(\emptyset), \ \forall \pi \in EP_{f - f(\emptyset)}\right\}.$$

Theorem [Edmonds, 1970]

Let $f : \{0,1\}^n \to \mathbb{R}$ be a submodular function. Then $\pi \in \mathbb{R}^n$ is an extreme point of EP_f if and only if there exists a permutation σ of $[N]$ such that $\pi_{\sigma(t)} = f(V_t) - f(V_{t-1})$, where $V_t = \{\sigma(1), \ldots, \sigma(t)\}$ for $t \in [N]$ and $V_0 = \emptyset$.

- The inequalities $y \geq \pi^\top(\mathbf{1} - z) + f(\emptyset)$ for $\pi \in EP_{f-f(\emptyset)}$ are referred to as the polymatroid inequalities of f.

Submodularity and polymatroid inequalities

- Given a submodular (set) function $f : 2^{[N]} \to \mathbb{R}$, the extended polymatroid of f is

$$EP_f := \{\pi \in \mathbb{R}^n : \ \pi(V) \leq f(V), \ \forall V \subseteq [N]\}.$$

- Given a submodular function $f : \{0,1\}^N \to \mathbb{R}$, consider

$$Q_f := \left\{(y,z) \in \mathbb{R} \times \{0,1\}^N : \ y \geq f(\mathbf{1} - z)\right\}$$

Theorem [Lovász, 1983, Atamtürk and Narayanan 2008]

The convex hull of Q_f is given by

$$\left\{(y,z) \in \mathbb{R} \times [0,1]^N : \ y \geq \pi^\top(\mathbf{1} - z) + f(\emptyset), \ \forall \pi \in EP_{f - f(\emptyset)}\right\}.$$

Theorem [Edmonds, 1970]

Let $f : \{0,1\}^n \to \mathbb{R}$ be a submodular function. Then $\pi \in \mathbb{R}^n$ is an extreme point of EP_f if and only if there exists a permutation σ of $[N]$ such that $\pi_{\sigma(t)} = f(V_t) - f(V_{t-1})$, where $V_t = \{\sigma(1), \ldots, \sigma(t)\}$ for $t \in [N]$ and $V_0 = \emptyset$.

- The inequalities $y \geq \pi^\top(\mathbf{1} - z) + f(\emptyset)$ for $\pi \in EP_{f-f(\emptyset)}$ are referred to as the polymatroid inequalities of f.
- Separating the polymatroid inequalities can be done in $O(N \log N)$ time.

Example 1 (revisited)

- The convex hull of

$$\left\{ (y_1, z) \in \mathbb{R}_+ \times \{0,1\}^3 \quad : \quad \begin{array}{l} y_1 \geq 8(1 - z_1) \\ y_1 \geq 6(1 - z_2) \\ y_1 \geq 13(1 - z_3) \end{array} \right\},$$

is

$$\left\{ (y_1, z) \in \mathbb{R}_+ \times [0,1]^3 \quad : \quad \begin{array}{l} y_1 \geq 13 - 6z_2 - 7z_3 \\ y_1 \geq 13 - 13z_3 \\ y_1 \geq 13 - 8z_1 - 5z_3 \\ y_1 \geq 13 - 2z_1 - 6z_2 - 5z_3 \end{array} , \right\}$$

$$= \left\{ (y_1, z) \in \mathbb{R}_+ \times [0,1]^3 : \text{ the mixing inequalities for } y_1 \right\}.$$

- Consider $\sigma = \{2, 3, 1\}$.

Example 1 (revisited)

- The convex hull of

$$\left\{ (y_1, z) \in \mathbb{R}_+ \times \{0,1\}^3 \quad : \quad \begin{array}{l} y_1 \geq 8(1 - z_1) \\ y_1 \geq 6(1 - z_2) \\ y_1 \geq 13(1 - z_3) \end{array} \right\},$$

  is

$$\left\{ (y_1, z) \in \mathbb{R}_+ \times [0,1]^3 \quad : \quad \begin{array}{l} y_1 \geq 13 - 6z_2 - 7z_3 \\ y_1 \geq 13 - 13z_3 \\ y_1 \geq 13 - 8z_1 - 5z_3 \\ y_1 \geq 13 - 2z_1 - 6z_2 - 5z_3 \end{array} , \right\}$$
$$= \left\{ (y_1, z) \in \mathbb{R}_+ \times [0,1]^3 : \text{ the mixing inequalities for } y_1 \right\}.$$

- Consider $\sigma = \{2, 3, 1\}$.

Joint mixing sets and mixing inequalities

- Recall the basic mixing set:

$$\left\{(y,z) \in \mathbb{R} \times \{0,1\}^N : \; y_j \geq f_j(\mathbf{1}-z), \; \forall j \in [k]\right\}$$

where

$$f_j(z) = \max_{i \in [N]} \left\{w_{ij} z_i\right\} \quad \text{for } z \in \{0,1\}^N.$$

Joint mixing sets and mixing inequalities

- Recall the basic mixing set:

$$\left\{(y,z) \in \mathbb{R} \times \{0,1\}^N : \ y_j \geq f_j(\mathbf{1}-z), \ \forall j \in [k]\right\}$$

where

$$f_j(z) = \max_{i \in [N]} \left\{w_{ij} z_i\right\} \quad \text{for } z \in \{0,1\}^N.$$

- The mixing inequality from a subset $\Pi_j = \{j_1, \cdots, j_\tau\}$ with $w_{j_1 j} \geq \cdots \geq w_{j_\tau j}$ is:

$$y_j + \sum_{s \in [\tau]} (w_{j_s j} - w_{j_{s+1} j}) z_{j_s} \geq w_{j_1 j}$$

where $w_{j_{\tau+1} j} := 0$.

Joint mixing sets and mixing inequalities

- Recall the basic mixing set:

$$\left\{(y, z) \in \mathbb{R} \times \{0,1\}^N : \ y_j \geq f_j(\mathbf{1} - z), \ \forall j \in [k]\right\}$$

where

$$f_j(z) = \max_{i \in [N]} \left\{w_{ij} z_i\right\} \quad \text{for } z \in \{0,1\}^N.$$

- The mixing inequality from a subset $\Pi_j = \{j_1, \cdots, j_\tau\}$ with $w_{j_1 j} \geq \cdots \geq w_{j_\tau j}$ is:

$$y_j + \sum_{s \in [\tau]} (w_{j_s j} - w_{j_{s+1} j}) z_{j_s} \geq w_{j_1 j}$$

where $w_{j_{\tau+1} j} := 0$.

Theorem [Kılınç-Karzan, Küçükyavuz, Lee, 2019+]

The polymatroid inequalities of f_j of the form

$$y_j \geq \pi^\top (\mathbf{1} - z) + f_j(\emptyset) \text{ for } \pi \in EP_{f_j - f_j(\emptyset)}$$

are mixing inequalities.

Multiple submodular constraints

Theorem [Baumann et al., 2013]

Given submodular functions $f_1, \ldots, f_k : \{0,1\}^N \to \mathbb{R}$, the convex hull of

$$\Big\{(y, z) \in \mathbb{R}^k \times \{0,1\}^N : \; y_j \geq f_j(\mathbf{1} - z), \; \forall j \in [k]\Big\}$$

is given by

Multiple submodular constraints

Theorem [Baumann et al., 2013]

Given submodular functions $f_1, \ldots, f_k : \{0,1\}^N \to \mathbb{R}$, the convex hull of

$$\left\{(y,z) \in \mathbb{R}^k \times \{0,1\}^N : \ y_j \geq f_j(\mathbf{1} - z), \ \forall j \in [k]\right\}$$

is given by

$$\left\{(y,z) \in \mathbb{R}^k \times [0,1]^N : \ y_j \geq \pi^\top(\mathbf{1} - z) + f_j(\emptyset), \ \forall \pi \in EP_{f_j - f_j(\emptyset)}, \forall j \in [k]\right\}.$$

Multiple submodular constraints

Theorem [Baumann et al., 2013]

Given submodular functions $f_1, \dots, f_k : \{0,1\}^N \to \mathbb{R}$, the convex hull of

$$\left\{(y,z) \in \mathbb{R}^k \times \{0,1\}^N : \; y_j \geq f_j(\mathbf{1} - z), \; \forall j \in [k]\right\}$$

is given by

$$\left\{(y,z) \in \mathbb{R}^k \times [0,1]^N : \; y_j \geq \pi^\top(\mathbf{1} - z) + f_j(\emptyset), \; \forall \pi \in EP_{f_j - f_j(\emptyset)}, \forall j \in [k]\right\}.$$

Theorem [Kılınç-Karzan, K., Lee, 2019+]

Let $f_1, \dots, f_\ell : \{0,1\}^N \to \mathbb{R}$ be submodular. If $h_1, \dots, h_\ell \in \mathbb{R}^k$ are weakly independent, then

$$\left\{(y,z) \in \mathbb{R}^k \times \{0,1\}^N : \; h_j^\top y \geq f_j(\mathbf{1} - z), \; \forall j \in [\ell]\right\}$$

is given by

Multiple submodular constraints

Theorem [Baumann et al., 2013]

Given submodular functions $f_1, \ldots, f_k : \{0,1\}^N \to \mathbb{R}$, the convex hull of

$$\left\{(y,z) \in \mathbb{R}^k \times \{0,1\}^N : \; y_j \geq f_j(\mathbf{1}-z), \; \forall j \in [k]\right\}$$

is given by

$$\left\{(y,z) \in \mathbb{R}^k \times [0,1]^N : \; y_j \geq \pi^\top(\mathbf{1}-z) + f_j(\emptyset), \; \forall \pi \in EP_{f_j - f_j(\emptyset)}, \forall j \in [k]\right\}.$$

Theorem [Kılınç-Karzan, K., Lee, 2019+]

Let $f_1, \ldots, f_\ell : \{0,1\}^N \to \mathbb{R}$ be submodular. If $h_1, \ldots, h_\ell \in \mathbb{R}^k$ are weakly independent, then

$$\left\{(y,z) \in \mathbb{R}^k \times \{0,1\}^N : \; h_j^\top y \geq f_j(\mathbf{1}-z), \; \forall j \in [\ell]\right\}$$

is given by

$$\left\{(y,z) \in \mathbb{R}^k \times [0,1]^N : \; h_j^\top y \geq \pi^\top(\mathbf{1}-z) + f_j(\emptyset), \; \forall \pi \in EP_{f_j - f_j(\emptyset)}, \forall j \in [\ell]\right\}.$$

Submodularity in joint mixing sets with lower bounds

- Now consider (Mix-lb):

$$\left\{(y, z) \in \text{(Mix)} : \; y_1 + \cdots + y_k \geq \varepsilon\right\}.$$

Submodularity in joint mixing sets with lower bounds

- Now consider (Mix-lb):

$$\{(y,z) \in \text{(Mix)} : \; y_1 + \cdots + y_k \geq \varepsilon\}.$$

- Then (Mix-lb) can be written as

$$\left\{(y,z) : \; y_j \geq f_j(\mathbf{1} - z), \; \forall j \in [k], \quad y_1 + \cdots + y_k \geq g(\mathbf{1} - z)\right\}$$

where

$$f_j(z) = \max_{i \in [N]} \left\{w_{ij} z_i\right\}, \quad g(z) = \max\left\{\varepsilon, \sum_{j \in [k]} f_j(z)\right\} \quad \text{for } z \in \{0,1\}^N.$$

Submodularity in joint mixing sets with lower bounds

- Now consider (Mix-lb):

$$\{(y,z) \in \text{(Mix)} : \; y_1 + \cdots + y_k \geq \varepsilon\}.$$

- Then (Mix-lb) can be written as

$$\left\{(y,z) : \; y_j \geq f_j(\mathbf{1}-z), \; \forall j \in [k], \quad y_1 + \cdots + y_k \geq g(\mathbf{1}-z)\right\}$$

where

$$f_j(z) = \max_{i \in [N]} \left\{w_{ij} z_i\right\}, \quad g(z) = \max\left\{\varepsilon, \; \sum_{j \in [k]} f_j(z)\right\} \quad \text{for } z \in \{0,1\}^N.$$

- In contrast to f_j, the function g is not always submodular.

Submodularity in joint mixing sets with lower bounds

- Now consider (Mix-lb):

$$\{(y,z) \in \text{(Mix)} : \; y_1 + \cdots + y_k \geq \varepsilon\}.$$

- Then (Mix-lb) can be written as

$$\left\{(y,z) : \; y_j \geq f_j(\mathbf{1}-z), \; \forall j \in [k], \quad y_1 + \cdots + y_k \geq g(\mathbf{1}-z)\right\}$$

where

$$f_j(z) = \max_{i\in[N]} \left\{w_{ij} z_i\right\}, \quad g(z) = \max\left\{\varepsilon, \sum_{j\in[k]} f_j(z)\right\} \quad \text{for } z \in \{0,1\}^N.$$

- In contrast to f_j, the function g is not always submodular.

- Can we characterize when g is submodular?

Submodularity in joint mixing sets with lower bounds

- Let $\bar{I}(\varepsilon) \subseteq [N]$ be a collection of scenarios defined as follows:

$$\bar{I}(\varepsilon) := \left\{ i \in [N] : \sum_{j \in [k]} w_{ij} \leq \varepsilon \right\}$$

Submodularity in joint mixing sets with lower bounds

- Let $\bar{I}(\varepsilon) \subseteq [N]$ be a collection of scenarios defined as follows:

$$\bar{I}(\varepsilon) := \left\{ i \in [N] : \sum_{j \in [k]} w_{ij} \leq \varepsilon \right\}$$

- $\bar{I}(\varepsilon)$ collects a set of scenarios with small coefficients.

Submodularity in joint mixing sets with lower bounds

- Let $\bar{I}(\varepsilon) \subseteq [N]$ be a collection of scenarios defined as follows:

$$\bar{I}(\varepsilon) := \left\{ i \in [N] : \sum_{j \in [k]} w_{ij} \leq \varepsilon \right\}$$

- $\bar{I}(\varepsilon)$ collects a set of scenarios with small coefficients.

- In Example 1, $\bar{I}(\varepsilon) = \{4, 5\}$.

$$\left\{ \begin{array}{l} (y,z) \in \\ \mathbb{R}^2_+ \times \{0,1\}^3 \end{array} : \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \\ y_1 \geq (1-z_4) \\ y_1 \geq 4(1-z_5) \end{array}, \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \\ y_2 \geq 2(1-z_4) \\ y_2 \geq (1-z_5) \end{array}, y_1 + y_2 \geq 7 \right\}$$

Submodularity in joint mixing sets with lower bounds

- We say that $\bar{I}(\varepsilon)$ is ε-negligible if $\bar{I}(\varepsilon) = \emptyset$

Submodularity in joint mixing sets with lower bounds

- We say that $\bar{I}(\varepsilon)$ is ε-negligible if $\bar{I}(\varepsilon) = \emptyset$ or $\bar{I}(\varepsilon) \neq \emptyset$ and

Submodularity in joint mixing sets with lower bounds

- We say that $\bar{I}(\varepsilon)$ is ε-negligible if $\bar{I}(\varepsilon) = \emptyset$ or $\bar{I}(\varepsilon) \neq \emptyset$ and $\bar{I}(\varepsilon)$ satisfies
  (1) $\sum_{j \in [k]} \max_{i \in \bar{I}(\varepsilon)} \{w_{ij}\} \leq \varepsilon$,
  (2) $\max_{i \in \bar{I}(\varepsilon)} \{w_{ij}\} \leq w_{ij}$ for every $i \in [N] \setminus \bar{I}(\varepsilon)$ and $j \in [k]$.

Submodularity in joint mixing sets with lower bounds

- We say that $\bar{I}(\varepsilon)$ is ε-negligible if $\bar{I}(\varepsilon) = \emptyset$ or $\bar{I}(\varepsilon) \neq \emptyset$ and $\bar{I}(\varepsilon)$ satisfies
  (1) $\sum_{j \in [k]} \max_{i \in \bar{I}(\varepsilon)} \{w_{ij}\} \leq \varepsilon$,
  (2) $\max_{i \in \bar{I}(\varepsilon)} \{w_{ij}\} \leq w_{ij}$ for every $i \in [N] \setminus \bar{I}(\varepsilon)$ and $j \in [k]$.

Theorem [Kılınç-Karzan, K., Lee, 2019+]

g is submodular if and only if ε satisfies

Submodularity in joint mixing sets with lower bounds

- We say that $\bar{I}(\varepsilon)$ is ε-negligible if $\bar{I}(\varepsilon) = \emptyset$ or $\bar{I}(\varepsilon) \neq \emptyset$ and $\bar{I}(\varepsilon)$ satisfies
  (1) $\sum_{j \in [k]} \max_{i \in \bar{I}(\varepsilon)} \{w_{ij}\} \leq \varepsilon$,
  (2) $\max_{i \in \bar{I}(\varepsilon)} \{w_{ij}\} \leq w_{ij}$ for every $i \in [N] \setminus \bar{I}(\varepsilon)$ and $j \in [k]$.

Theorem [Kılınç-Karzan, K., Lee, 2019+]

g is submodular if and only if ε satisfies

1. $\bar{I}(\varepsilon)$ is ε-negligible,
2. $\varepsilon \leq L_W(\varepsilon) := \begin{cases} \min\limits_{p,q \in [N] \setminus \bar{I}(\varepsilon)} \left\{ \sum\limits_{j \in [k]} \min\{w_{pj}, \ w_{qj}\} \right\}, & \text{if } \ \bar{I}(\varepsilon) \neq [N], \\ +\infty, & \text{if } \ \bar{I}(\varepsilon) = [N] \end{cases}.$

- Now we know when (Mix-lb) has a submodularity structure.

Aggregated mixing inequalities

- (Mix-lb) can be written as

$$\left\{(y,z):\ y_j \geq f_j(\mathbf{1}-z),\ \forall j \in [k], \quad y_1 + \cdots + y_k \geq g(\mathbf{1}-z)\right\}$$

where

$$f_j(z) = \max_{i \in [N]} \left\{w_{ij} z_i\right\}, \quad g(z) = \max\left\{\varepsilon,\ \sum_{j \in [k]} f_j(z)\right\} \quad \text{for } z \in \{0,1\}^N.$$

Aggregated mixing inequalities

- (Mix-lb) can be written as

$$\left\{(y,z):\ y_j \geq f_j(\mathbf{1}-z),\ \forall j \in [k], \quad y_1 + \cdots + y_k \geq g(\mathbf{1}-z)\right\}$$

where

$$f_j(z) = \max_{i \in [N]} \left\{w_{ij} z_i\right\}, \quad g(z) = \max\left\{\varepsilon,\ \sum_{j \in [k]} f_j(z)\right\} \quad \text{for } z \in \{0,1\}^N.$$

Theorem [Kılınç-Karzan, K., Lee, 2019+]

The polymatroid inequalities of g of the form

$$y_1 + \cdots + y_k \geq \pi^\top(\mathbf{1}-z) + g(\emptyset) \text{ for } \pi \in EP_{g-g(\emptyset)}$$

are aggregated mixing inequalities. They can be separated in $O(kN \log N)$ time.

Example 2 (revisited)

The convex hull of

$$\left\{ \begin{array}{c} (y,z) \in \\ \mathbb{R}^2_+ \times \{0,1\}^3 \end{array} : \begin{array}{l} y_1 \geq 8(1-z_1) \\ y_1 \geq 6(1-z_2) \\ y_1 \geq 13(1-z_3) \\ y_1 \geq (1-z_4) \\ y_1 \geq 4(1-z_5) \end{array} , \begin{array}{l} y_2 \geq 3(1-z_1) \\ y_2 \geq 4(1-z_2) \\ y_2 \geq 2(1-z_3) \\ y_2 \geq 2(1-z_4) \\ y_2 \geq (1-z_5) \end{array} , y_1 + y_2 \geq 7 \right\}$$

is

$$\left\{ \begin{array}{c} (y,z) \in \\ \mathbb{R}^2_+ \times [0,1]^3 \end{array} : \begin{array}{l} \text{the mixing inequalities for } y_1, y_2 \\ y_1 + y_2 \geq 17 - z_1 - z_2 - 8z_3 \\ y_1 + y_2 \geq 17 - 2z_2 - 8z_3 \\ y_1 + y_2 \geq 17 - 3z_2 - 7z_3 \\ y_1 + y_2 \geq 17 - 2z_1 - 3z_2 - 5z_3 \\ y_1 + y_2 \geq 17 - 4z_1 - z_2 - 5z_3 \end{array} \right\}$$

$$= \left\{ \begin{array}{c} (y,z) \in \\ \mathbb{R}^2_+ \times [0,1]^3 \end{array} : \begin{array}{l} \text{the mixing inequalities for } y_1, y_2 \\ \text{the ``aggregated'' mixing inequalities for ``} y_1 + y_2 \text{''} \end{array} \right\}$$

Consider $\sigma = \{2, 3, 1, 4, 5\}$.

Convex hull of (Mix-lb)

Theorem [Kılınç-Karzan, K., Lee, 2019+]

The following statements are equivalent:

(i) the convex hull of (Mix-lb) is obtained after adding the mixing and the aggregated mixing inequalities,

(ii) $f_1, \ldots, f_k, g$ are submodular.

(iii) ε satisfies the following 2 conditions:

1. $\bar{I}(\varepsilon)$ is ε-negligible,
2. $\varepsilon \leq L_W(\varepsilon) := \begin{cases} \min\limits_{p,q \in [N] \setminus \bar{I}(\varepsilon)} \left\{ \sum\limits_{j \in [k]} \min\{w_{pj},\ w_{qj}\} \right\}, & \text{if } \bar{I}(\varepsilon) \neq [N], \\ +\infty, & \text{if } \bar{I}(\varepsilon) = [N] \end{cases}.$

Outline

Two-stage (dynamic) chance-constrained problem (2CCP)

- Order of events:

Two-stage (dynamic) chance-constrained problem (2CCP)

- Order of events: $x \rightarrow$

Two-stage (dynamic) chance-constrained problem (2CCP)

- Order of events: $x \to \omega$

Two-stage (dynamic) chance-constrained problem (2CCP)

- Order of events: $x \rightarrow \omega \rightarrow$

Two-stage (dynamic) chance-constrained problem (2CCP)

- Order of events: $x \rightarrow \omega \rightarrow y(\omega)$

Two-stage (dynamic) chance-constrained problem (2CCP)

- Order of events: $x \to \omega \to y(\omega)$

- $y(\omega) \in \mathbb{R}_+^{n_2}$: second-stage decision vector for each $\omega \in \Omega$

Two-stage (dynamic) chance-constrained problem (2CCP)

- Order of events: $x \to \omega \to y(\omega)$
- $y(\omega) \in \mathbb{R}_+^{n_2}$: second-stage decision vector for each $\omega \in \Omega$

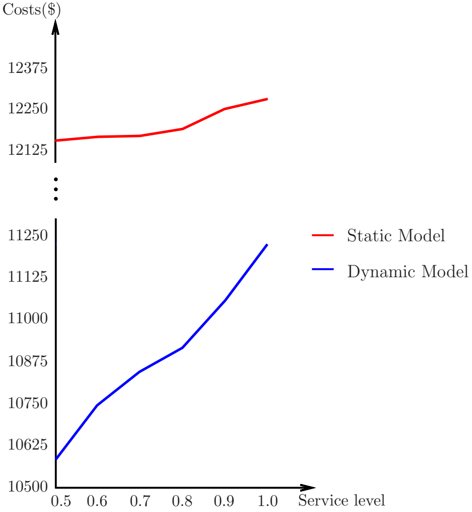
A two-stage chance-constrained program:

$$\begin{aligned} \min \quad & c^\top x + \mathbb{E}_\omega(g(\omega)^\top y(\omega)) \\ s.t. \quad & \mathbb{P}\{W(\omega)x + T(\omega)y(\omega) \geq r(\omega)\} \geq 1 - \epsilon \\ & x \in X \cap \mathcal{X}, y(\omega) \in \mathbb{R}_+^{n_2}, \omega \in \Omega. \end{aligned}$$

- Assume (wlog) i.i.d sample ($\mathbb{P}(\omega) = \frac{1}{N}$) and $g(\omega) \geq \mathbf{0}$.

Static vs Dynamic Decisions

Multi-stage inventory control problem with a service level constraint [Zhang, K., Goel, 2014]



- Significant cost savings by dynamic model.
- Higher service level gives rise to higher cost.
- Static model: limited flexibility

  Dynamic model: large cost savings with small decrease in service level

Deterministic Equivalent Formulation (DEF)

$$
\begin{array}{lllllll}
\min\limits_{x,y,z} & c^\top x & +\frac{1}{N}(g(\omega^1)^\top y(\omega^1) z_1 + g(\omega^2)^\top y(\omega^2) z_2 & \ldots + g(\omega^N)^\top y(\omega^N) z_N) & & & \\
 & T(\omega^1)x & +W(\omega^1)y(\omega^1) & & & +\bar{M}_1 z_1 & \geq r(\omega^1) \\
 & T(\omega^2)x & & +W(\omega^2)y(\omega^2) & & +\bar{M}_2 z_2 & \geq r(\omega^2) \\
 & \vdots & & & \ddots & & \vdots \\
 & T(\omega^N)x & & & +W(\omega^N)y(\omega^N) & +\bar{M}_N z_N & \geq r(\omega^N)
\end{array}
$$

$$
\sum_{k=1}^{N} z_k \leq \lfloor N\epsilon \rfloor = p; \quad x \in X \cap \mathcal{X},\ y(\omega) \in \mathbb{R}_+^{n_2}, \omega \in \Omega, z \in \mathbb{B}^N,
$$

where $\bar{M}_i$ is a vector of very large numbers, $\omega^i \in \Omega$, and

$$
z_i = \begin{cases} 0 & \text{if scenario } \omega^i \text{ is satisfied} \\ 1 & \text{otherwise.} \end{cases}
$$

Let $g(\omega^i) = g_i$, $T(\omega^i) = T_i$, $W(\omega^i) = W_i$, $r(\omega^1) = r_i$.

Decomposition algorithm for 2CCP

If there are second stage costs, and only a subset of scenarios are satisfied, then the traditional Benders feasibility and optimality cuts are no longer valid.

Goal: Develop valid feasibility and optimality cuts to the master problem of 2CCP.

Decomposition algorithm for 2CCP

If there are second stage costs, and only a subset of scenarios are satisfied, then the traditional Benders feasibility and optimality cuts are no longer valid.

Goal: Develop valid feasibility and optimality cuts to the master problem of 2CCP.

- First, the algorithm requires solving a master problem (MP):

$$\begin{aligned} \mathrm{MP}(C,B) = \min_{x,z,\eta} \quad & c^\top x + \frac{1}{N}\sum_{i\in[N]} \eta_i \\ \text{s.t.} \quad & \sum_{i\in[N]} z_i \le q \\ & z \in \mathbb{B}^N \\ & x \in X \cap \mathcal{X},\ \eta \in \mathbb{R}_+^N \\ & (x,z) \in \mathcal{F},\ (x,z,\eta) \in \mathcal{O}, \end{aligned}$$

- $\mathcal{F}$ represents the collection of feasibility cuts and
- $\mathcal{O}$ represents the collection of optimality cuts.
- Let $P_i = \{x \in X \cap \mathcal{X} | \exists y \ge \mathbf{0} : T_i x + W_i y \ge r_i\}$, $i \in [N]$.

Subproblem 1 (SP1): Optimality Cut Generation (Basic)

- SP1 is used to cut off a feasible solution $(\hat{x}, \hat{z})$ which has incorrect second stage value $\hat{\eta}$.
- If the solution $(\hat{x}, \hat{z})$ is feasible, then $\forall \hat{z}_i = 0$, we solve single scenario linear optimization problem (SP1$_i$):

$$Y_i = \min_{y \in \mathbb{R}_+^{n_2}} g_i^\top y$$
$$s.t. \quad W_i y \geq r_i - T_i \hat{x} \qquad (\psi_i)$$

where ψ_i is the vector of dual variables for kth scenario subproblem.

- If SP1$_i$ is feasible, then compare $\hat{\eta}_i$ with Y_i. If $\hat{\eta}_i < Y_i$, then add the modified Benders optimality cut to $\mathcal{O}$:

$$\eta_i + M_i z_\omega \geq \psi_i^\top (r_i - T_i x)$$

M_i : big-M

- If SP1$_i$ (or equivalently $(\hat{x}, \hat{z})$) is infeasible, then go to the second subproblem (feasibility cut generation). [Luedtke, 2014]

Computations

A call center staffing problem

Instances		DEF		Basic Decomposition	
(N, ϵ)	(n_1, d)	`Time (slvd)`	Gap(%)	`Time (slvd)`	Gap(%)
(300, 0.05)	(5,10)	55.8 (5)	0	**54.6 (5)**	0
	(10,20)	258.3 (4)	0.1	**134.2 (5)**	0
(300, 0.1)	(5,10)	**126.0 (5)**	0	258.3 (4)	0.1
	(10,20)	1294.7 (4)	1.3	**483.7 (3)**	0.3
(400, 0.05)	(5,10)	**83.6 (5)**	0	133.8 (5)	0
	(10,20)	781(3)	2.3	**233.2 (5)**	0
(400, 0.1)	(5,10)	243 (5)	0	**220 (3)**	0.0
	(10,20)	>3600 (0)	3.4	**909.8 (5)**	0
(500, 0.05)	(5,10)	**170.6 (5)**	0	221(5)	0
	(10,20)	>3600 (0)	2.9	**313.2(5)**	0
(500, 0.1)	(5,10)	730 (2)	1.3	**166 (3)**	0.3
	(10,20)	>3600 (0)	5.8	**142.7 (3)**	0.3
Avg (Sum)	(n, m)	916.2 (38)	3.2	**276.1 (51)**	0.2

n_1: number of first stage variables (servers); d: number of customers.

Improved optimality cuts [Liu, K., Luedtke, 2016]

- For a given $\alpha \in \mathbb{R}^{n_1}$ and each $i \in [N]$, let

$$v_i(\alpha) = \min\{\alpha^\top x : x \in P_i\}$$

- Note $v_i(\alpha) \leq \alpha^\top x$ for all feasible x

- Then an improved optimality cut with $\phi = \psi_i^\top T_i$ is:

$$\eta_i + \left(\psi_i^\top r_i - v_i(\phi)\right) z_i \geq \psi_i^\top (r_i - T_i x).$$

Improved optimality cuts [Liu, K., Luedtke, 2016]

- For a given $\alpha \in \mathbb{R}^{n_1}$ and each $i \in [N]$, let

$$v_i(\alpha) = \min\{\alpha^\top x : x \in P_i\}$$

- Note $v_i(\alpha) \leq \alpha^\top x$ for all feasible x
- Then an improved optimality cut with $\phi = \psi_i^\top T_i$ is:

$$\eta_i + \left(\psi_i^\top r_i - v_i(\phi)\right) z_i \geq \psi_i^\top (r_i - T_i x).$$

For $z_i = 0$, this is the traditional Benders cut, so it is valid.

For $z_i = 1$, we get $\underbrace{\eta_i}_{\geq 0} \geq \underbrace{v_i(\phi) - \phi x}_{\leq 0}$, so it is valid.

Improved optimality cuts [Liu, K., Luedtke, 2016]

- For a given $\alpha \in \mathbb{R}^{n_1}$ and each $i \in [N]$, let

$$v_i(\alpha) = \min\{\alpha^\top x : x \in P_i\}$$

- Note $v_i(\alpha) \leq \alpha^\top x$ for all feasible x
- Then an improved optimality cut with $\phi = \psi_i^\top T_i$ is:

$$\eta_i + \left(\psi_i^\top r_i - v_i(\phi)\right) z_i \geq \psi_i^\top (r_i - T_i x).$$

For $z_i = 0$, this is the traditional Benders cut, so it is valid.

For $z_i = 1$, we get $\underbrace{\eta_i}_{\geq 0} \geq \underbrace{v_i(\phi) - \phi x}_{\leq 0}$, so it is valid.

- We also give another class of strong optimality cuts

Computational results with strong decomposition

Instances		DEF	Basic Decomp.	Strong Decomp.
(N, ϵ)	(n_1, d)	Time(slvd) / gap	Time / gap	Time(slvd) / gap
(2000, 0.05)	(5,10)	120	1.8%	133
	(10,20)	9.0%	1.8%	1012
	(15,30)	14.6%	3.8%	343
(2500, 0.05)	(5,10)	165(2) / 6.5%	3.0%	131
	(10,20)	9.5%	2.8%	1246
	(15,30)	–	3.3%	1246
(3000, 0.05)	(5,10)	262(1) / 5.9%	1.8%	273
	(10,20)	17.4%	2.2%	2030
	(15,30)	–	3.2%	1207(2) / 0.4%

- " -" : failed to find solution.
- If the algorithm hits the time or memory limit, we report the end gap, otherwise we report time.
- For DEP (3000,0.05) (5,10), CPLEX successfully solved 1 instance in 262 seconds, and failed to solve the other 2 instances, with 5.9% end gap.

Do we really know $\mathbb{P}$?

- So far we discussed two-stage stochastic MIPs and chance-constrained programs with a given (finite) $\mathbb{P}$.

- Do we really know $\mathbb{P}$?

Outline

Chance-constrained program (CCP)

Consider **chance-constrained programs** in the general form:

$$\begin{aligned} \min_{x} \quad & c^\top x \\ \text{s.t.} \quad & \mathbb{P}^*[f(x,\xi) \geq 0] \geq 1-\epsilon, \\ & x \in \mathcal{X}. \end{aligned} \qquad \text{(CCP)}$$

Often, we do not know $\mathbb{P}^*$ precisely.

Sample average approximation (SAA)

- **Sample average approximation**: draw i.i.d. samples $\{\xi_i\}_{i\in[N]}$ from $\mathbb{P}^*$.

$$\mathbb{P}^*[f(x,\xi) \geq 0] \approx \mathbb{P}_N[f(x,\xi) \geq 0] := \frac{1}{N}\sum_{i\in[N]} \mathbb{1}(f(x,\xi_i) \geq 0).$$

- Focus on constraint functions $f(x,\xi)$ in piecewise linear form

$$f(x,\xi) := \min_{p\in[P]} \left\{ (b_p - A^\top x)^\top \xi + (d_p - a_p^\top x) \right\}.$$

Sample average approximation (SAA)

Approximate , (CCP) by

$$
\begin{aligned}
\min_x \quad & c^\top x \\
\text{s.t.} \quad & \underbrace{\frac{1}{N}\sum_{i\in[N]} \mathbb{1}(f(x,\xi_i)\geq 0) \geq 1-\epsilon,}_{\text{MIP-representable}} \qquad \text{(SAA)} \\
& x \in \mathcal{X}.
\end{aligned}
$$

Essentially, we need to ensure that that at least $N(1-\epsilon)$ samples satisfy $f(x,\xi_i)\geq 0$.

Sample average approximation (SAA)

Approximate , (CCP) by

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.} \quad & \underbrace{\frac{1}{N} \sum_{i \in [N]} \mathbb{1}(f(x, \xi_i) \geq 0) \geq 1 - \epsilon,}_{\text{MIP-representable}} \\ & x \in \mathcal{X}. \end{aligned} \tag{SAA}$$

Essentially, we need to ensure that that at least $N(1-\epsilon)$ samples satisfy $f(x, \xi_i) \geq 0$.

The out-of-sample performance of the solution from (SAA) is often poor, particularly for small N.

- Just because $\mathbb{P}_N[f(x, \xi) \geq 0] \geq 1 - \epsilon$ does **not** mean that $\mathbb{P}^*[f(x, \xi) \geq 0] \geq 1 - \epsilon$.
- The so-called "Optimizer's Curse" [Smith and Winkler, 2006].

Improving out-of-sample performance

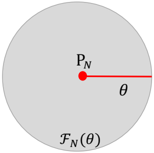
- Distributionally robust chance constrained program:

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{s.t.} \quad & \mathbb{P}[f(x,\xi) \geq 0] \geq 1-\epsilon \;\; \forall\, \mathbb{P} \in \mathcal{F}_N(\theta), \\ & x \in \mathcal{X}, \end{aligned} \tag{DR-CCP}$$

where $\mathcal{F}_N(\theta)$: an ambiguity set of distributions on $\mathbb{R}^K$ that contains the empirical distribution $\mathbb{P}_N$:

$$\mathcal{F}_N(\theta) := \{\mathbb{P} : d(\mathbb{P}_N, \mathbb{P}) \leq \theta\}, \quad \text{w.h.p. } \mathbb{P}^* \in \mathcal{F}_N(\theta).$$

- **Intuition:** $\mathbb{P}_N$ will be (w.h.p.) close to $\mathbb{P}^*$, so make sure $\mathbb{P}[f(x,\xi) \geq 0] \geq 1-\epsilon$ **for all** $\mathbb{P}$ in a radius θ ball around $\mathbb{P}_N$.



- When N large, make the radius θ smaller.
- When N small, we are not as confident that $\mathbb{P}_N$ is close to $\mathbb{P}^*$, so make the radius θ larger.

Ambiguity set

Wasserstein ambiguity set with radius θ:

$$\mathcal{F}_N(\theta) := \{\mathbb{P} : d_W(\mathbb{P}_N, \mathbb{P}) \leq \theta\}$$

where

$$d_W(\mathbb{P}, \mathbb{P}') := \inf_{\Pi} \left\{\mathbb{E}_{(\xi,\xi')\sim\Pi}[\|\xi - \xi'\|] : \Pi \text{ has marginal distributions } \mathbb{P}, \mathbb{P}'\right\}.$$

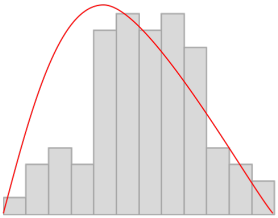


Figure 2: Wasserstein distance $d_W(\mathbb{P}_N, \mathbb{P})$: minimum distance required to transport grey bars to red curve.

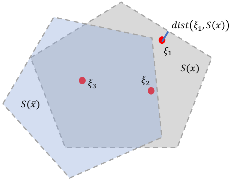
Has recently become very popular in optimization and machine learning [Mohajerin Esfahani and Kuhn, 2018].

Distance to violation

- For a given parameter ξ and decision x, define the **distance to violation**:

$$\mathrm{dist}(\xi, x) := \inf_{\Delta} \left\{ \|\Delta\| : f(x, \xi + \Delta) < 0 \right\}.$$

- Safe set $\mathcal{S}(x) = \{\xi : f(x, \xi) \geq 0\}$



Reformulation of (DR-CCP)

We now need to reformulate semi-infinite constraint $\mathbb{P}[f(x,\xi) \geq 0] \geq 1-\epsilon \ \ \forall\, \mathbb{P} \in \mathcal{F}_N(\theta)$.

- [Blanchet and Murthy, 2019], [Gao and Kleywegt, 2016], [Xie, 2019] show that for Wasserstein ambiguity

$$\mathbb{P}[f(x,\xi) \geq 0] \geq 1-\epsilon \ \ \forall\, \mathbb{P} \in \mathcal{F}_N(\theta) \iff \mathrm{CVaR}^{\mathbb{P}_N}_{1-\epsilon}(\mathrm{dist}(\xi,x)) \geq \frac{\theta}{\epsilon}$$

$$\begin{aligned}\mathrm{CVaR}^{\mathbb{P}_N}_{1-\epsilon}(\mathrm{dist}(\xi,x)) &:= \text{take the lowest } \epsilon N \text{ distances amongst } \{\mathrm{dist}(\xi_i,x)\}_{i\in[N]},\\ &\quad\ \text{then take their average}\\ &= \max_{t,r}\left\{ t - \frac{1}{\epsilon N}\sum_{i\in[N]} r_i : \begin{array}{l} r_i \geq 0,\ i \in [N] \\ t - r_i \leq \mathrm{dist}(\xi_i,x),\ i \in [N] \end{array}\right\}.\end{aligned}$$

Here larger distances are preferred, so distances are acceptability functionals rather than risk. CVaR definition is adapted accordingly.

Reformulation of (DR-CCP)

We now need to reformulate semi-infinite constraint $\mathbb{P}[f(x,\xi) \geq 0] \geq 1-\epsilon \ \ \forall\, \mathbb{P} \in \mathcal{F}_N(\theta)$.

- [Blanchet and Murthy, 2019], [Gao and Kleywegt, 2016], [Xie, 2019] show that for Wasserstein ambiguity

$$\mathbb{P}[f(x,\xi) \geq 0] \geq 1-\epsilon \ \ \forall\, \mathbb{P} \in \mathcal{F}_N(\theta) \iff \mathrm{CVaR}^{\mathbb{P}_N}_{1-\epsilon}(\mathrm{dist}(\xi,x)) \geq \frac{\theta}{\epsilon}$$

$$\begin{aligned}\mathrm{CVaR}^{\mathbb{P}_N}_{1-\epsilon}(\mathrm{dist}(\xi,x)) &:= \text{take the lowest } \epsilon N \text{ distances amongst } \{\mathrm{dist}(\xi_i,x)\}_{i\in[N]},\\ &\quad\ \text{then take their average}\\ &= \max_{t,r}\left\{ t - \frac{1}{\epsilon N}\sum_{i\in[N]} r_i : \begin{array}{l} r_i \geq 0,\ i \in [N] \\ t - r_i \leq \mathrm{dist}(\xi_i,x),\ i\in[N] \end{array}\right\}.\end{aligned}$$

  Here larger distances are preferred, so distances are acceptability functionals rather than risk. CVaR definition is adapted accordingly.

- Usual SAA-CCP formulation implies $\mathrm{VaR}^{\mathbb{P}_N}_{1-\epsilon}(\mathrm{dist}(\xi,x)) \geq 0$. Its (conservative) CVaR approximation gives $\mathrm{CVaR}^{\mathbb{P}_N}_{1-\epsilon}(\mathrm{dist}(\xi,x)) \geq 0$. Compare with (DR-CCP).

Reformulation of (DR-CCP)

This implies that (DR-CCP) can be reformulated as

$$\begin{aligned} \min_{x,t,r} \quad & c^\top x \\ \text{s.t.} \quad & \epsilon t \geq \theta + \frac{1}{N} \sum_{i \in [N]} r_i, \\ & t - r_i \leq \text{dist}(\xi_i, x), \quad i \in [n] \\ & r_i \geq 0, \quad i \in [n] \\ & x \in \mathcal{X}. \end{aligned} \tag{DR-CCP-f}$$

The last step is to reformulate the constraint $t - r_i \leq \text{dist}(\xi_i, x)$.

- This depends on how we define $f(x, \xi)$.

Linear constraints

- For simple presentation, we focus on a single linear function with right-hand side uncertainty (no bilinear term):

$$f(x, \xi) := \xi + d - a^\top x,$$

  for given a, d.
- Distance to violation:

$$\mathrm{dist}(\xi, x) = \max\{0, \xi + d - a^\top x\} = \max\{0, f(x, \xi)\}.$$

- Our results extend to polyhedral structures of the form

$$f(x, \xi) := \min_{p \in [P]} \left\{ (b_p - A^\top x)^\top \xi + (d_p - a_p^\top x) \right\} \geq 0.$$

  - The *only* condition we impose is that the bilinear term $(A^\top x)^\top \xi$ is *the same for all* $p \in [P]$.

Reformulation of (DR-CCP)

However, $t - r_i \leq \text{dist}(\xi_i, x) = \max\{0, f(x, \xi_i)\}$

$$\iff t - r_i \leq 0 \quad \text{OR} \quad t - r_i \leq f(x, \xi_i).$$

is a non-convex constraint.

- We can model this with a binary variable and big-M constants:

$$\begin{aligned} & z_i \in \{0,1\}, \\ & t - r_i \leq f(x, \xi_i) + M_i z_i \\ & t - r_i \leq M_i(1 - z_i) \end{aligned}$$

  $z_i = 1$ indicates when $t - r_i \leq 0$, and $z_i = 0$ indicates when $t - r_i \leq f(x, \xi_i)$.

Reformulation of (DR-CCP)

However, $t - r_i \leq \mathrm{dist}(\xi_i, x) = \max\{0, f(x, \xi_i)\}$

$$\iff t - r_i \leq 0 \quad \text{OR} \quad t - r_i \leq f(x, \xi_i).$$

is a non-convex constraint.

- We can model this with a binary variable and big-M constants:

$$\begin{aligned} & z_i \in \{0, 1\}, \\ & t - r_i \leq f(x, \xi_i) + M_i z_i \\ & t - r_i \leq M_i(1 - z_i) \end{aligned}$$

  $z_i = 1$ indicates when $t - r_i \leq 0$, and $z_i = 0$ indicates when $t - r_i \leq f(x, \xi_i)$.

- M_i is a sufficiently large constant. For some fixed *optimal* decision x of (DR-CCP), we need

$$M_i \geq |f(x, \xi_i)| \quad \forall i \in [N].$$

  Choosing in this way requires understanding the structure of optimal solutions, which is not easy, and can still result in large values.

The basic MIP reformulation of (DR-CCP)

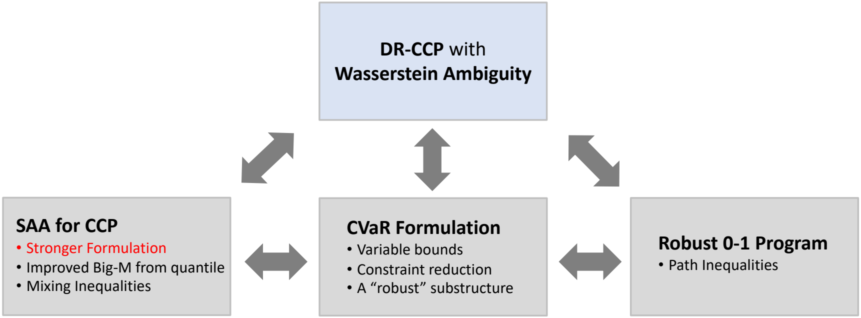
[Chen et al., 2018], [Xie, 2019] gave the following MIP reformulation for (DR-CCP):

$$\begin{aligned}
\min_{z,r,t,x} \quad & c^\top x \\
\text{s.t.} \quad & z \in \{0,1\}^N,\ t \geq 0,\ r \geq \mathbf{0},\ x \in \mathcal{X}, \\
& \epsilon\, t \geq \theta + \frac{1}{N} \sum_{i\in[N]} r_i, \\
& M_i(1-z_i) \geq t - r_i, \quad i \in [N], \\
& f(x,\xi_i) + M_i z_i \geq t - r_i, \quad i \in [N].
\end{aligned} \tag{DR-CCP-MIP}$$

Difficult to solve, especially for small θ even for $N = 100$.
In [Ho-Nguyen,Kılınç-Karzan, K., Lee, 2021a], we scale this up to $N = 1000 \sim 3000$.

Improvements to (DR-CCP-MIP) [Ho-Nguyen,Kılınç-Karzan, K., Lee, 2021a+]

Our key insight finds a link between (SAA) and (DR-CCP). This leads to a number of enhancements.



Connection to (SAA)

Denote the feasible regions of (SAA) and (DR-CCP) as

$$
\begin{aligned}
\mathcal{X}_{\mathrm{SAA}} &:= \{x \in \mathcal{X} : \mathbb{P}_N[f(x,\xi) \geq 0] \geq 1-\epsilon\}, \\
&= \left\{ x \in \mathcal{X} : \begin{array}{l} \frac{1}{N}\sum_{i \in [N]} w_i \leq \epsilon, \quad w \in \{0,1\}^N \\ f(x,\xi_i) + M_i w_i \geq 0, \quad i \in [N] \end{array} \right\} \\
\mathcal{X}_{\mathrm{DR}} &:= \left\{ x \in \mathcal{X} : \inf_{\mathbb{P} \in \mathcal{F}_N(\theta)} \mathbb{P}[f(x,\xi) \geq 0] \geq 1-\epsilon \right\} \\
&= \left\{ x \in \mathcal{X} : \begin{array}{l} \epsilon\, t \geq \theta + \frac{1}{N}\sum_{i \in [N]} r_i, \quad z \in \{0,1\}^N \\ M_i(1-z_i) \geq t - r_i, \quad i \in [N], \\ f(x,\xi_i) + M_i z_i \geq t - r_i, \quad i \in [N] \end{array} \right\}.
\end{aligned}
$$

Connection to (SAA)

Denote the feasible regions of (SAA) and (DR-CCP) as

$$\begin{aligned}
\mathcal{X}_{\mathrm{SAA}} &:= \{x \in \mathcal{X} : \ \mathbb{P}_N[f(x,\xi) \geq 0] \geq 1-\epsilon\}, \\
&= \left\{ x \in \mathcal{X} : \begin{array}{l} \frac{1}{N} \sum_{i \in [N]} w_i \leq \epsilon, \quad w \in \{0,1\}^N \\ f(x,\xi_i) + M_i w_i \geq 0, \quad i \in [N] \end{array} \right\} \\
\mathcal{X}_{\mathrm{DR}} &:= \left\{ x \in \mathcal{X} : \inf_{\mathbb{P} \in \mathcal{F}_N(\theta)} \mathbb{P}[f(x,\xi) \geq 0] \geq 1-\epsilon \right\} \\
&= \left\{ x \in \mathcal{X} : \begin{array}{l} \epsilon\, t \geq \theta + \frac{1}{N} \sum_{i \in [N]} r_i, \quad z \in \{0,1\}^N \\ M_i(1-z_i) \geq t - r_i, \quad i \in [N], \\ f(x,\xi_i) + M_i z_i \geq t - r_i, \quad i \in [N] \end{array} \right\}.
\end{aligned}$$

Observation: in general $\mathcal{F}_N(0) = \{\mathbb{P}_N\} \subseteq \mathcal{F}_N(\theta)$ for any $\theta \geq 0$, so $\mathcal{X}_{\mathrm{DR}} \subseteq \mathcal{X}_{\mathrm{SAA}}$.

Naïvely, BLUE constraints are valid for $\mathcal{X}_{\mathrm{DR}}$, but require **different** binary variables (w vs. z).

Stronger formulation

Key result 1: for both RED and BLUE constraints, **the same binary variables** z **can be used.**

$$
\begin{aligned}
\min_{z,r,t,x} \quad & c^\top x \\
\text{s.t.} \quad & z \in \{0,1\}^N,\ t \geq 0,\ r \geq \mathbf{0},\ x \in \mathcal{X}, \\
& \epsilon\, t \geq \theta + \frac{1}{N} \sum_{i \in [N]} r_i, \\
& M_i(1 - z_i) \geq t - r_i, \quad i \in [N], \\
& f(x, \xi_i) + M_i z_i \geq t - r_i, \quad i \in [N], \\
& \frac{1}{N} \sum_{i \in [N]} z_i \leq \epsilon, \\
& f(x, \xi_i) + M_i z_i \geq 0, \quad i \in [N].
\end{aligned}
$$

Big-M reduction via the mixing procedure

Key result 2: we gain much more from the SAA constraints

$$\sum_{i\in[N]} z_i \le \epsilon N, \qquad f(x,\xi_i) + M_i z_i \ge 0,\ \forall i \in [N].$$

(Mixing procedure) [Luedtke et al., 2010] showed that we can **drastically** reduce M_i to

$$\sum_{i\in[N]} z_i \le \epsilon N, \qquad f(x,\xi_i) + m_i z_i \ge 0,\ \forall i \in [N].$$

Big-M reduction via the mixing procedure

Key result 2: we gain much more from the SAA constraints

$$\sum_{i \in [N]} z_i \leq \epsilon N, \qquad f(x, \xi_i) + M_i z_i \geq 0, \ \forall i \in [N].$$

(Mixing procedure) [Luedtke et al., 2010] showed that we can **drastically** reduce M_i to

$$\sum_{i \in [N]} z_i \leq \epsilon N, \qquad f(x, \xi_i) + m_i z_i \geq 0, \ \forall i \in [N].$$

- For each $i \in [N]$, we have the inequalities

$$t - r_i \leq M_i(1 - z_i), \quad t - r_i \leq f(x, \xi_i) + M_i z_i$$
$$0 \leq f(x, \xi_i) + m_i z_i.$$

- It is easily checked that these imply

$$t - r_i \leq f(x, \xi_i) + m_i z_i.$$

- These can replace the inequalities $t - r_i \leq f(x, \xi_i) + M_i z_i$ in (DR-CCP-MIP).

Compact formulation of (DR-CCP-MIP) via CVaR interpretation

Key result 3: recall that the DR-CCP is

$$\mathrm{CVaR}_{1-\epsilon}^{\mathbb{P}_N}(\mathrm{dist}(\xi, x)) = \max_{t,r} \left\{ t - \frac{1}{\epsilon N} \sum_{i \in [N]} r_i : \begin{array}{l} r_i \geq 0,\ \ i \in [N] \\ t - r_i \leq \mathrm{dist}(\xi_i, x),\ \ i \in [N] \end{array} \right\} \geq \frac{\theta}{\epsilon}.$$

Compact formulation of (DR-CCP-MIP) via CVaR interpretation

Key result 3: recall that the DR-CCP is

$$\mathrm{CVaR}_{1-\epsilon}^{\mathbb{P}_N}(\mathrm{dist}(\xi, x)) = \max_{t,r} \left\{ t - \frac{1}{\epsilon N} \sum_{i \in [N]} r_i : \begin{array}{l} r_i \geq 0,\ i \in [N] \\ t - r_i \leq \mathrm{dist}(\xi_i, x),\ i \in [N] \end{array} \right\} \geq \frac{\theta}{\epsilon}.$$

- There always exists an optimal solution to the program such that

$$t = (\lfloor \epsilon N \rfloor + 1)\text{-th smallest value amongst } \left\{ \mathrm{dist}(\xi_i, x) = (\xi_i + d - a^\top x)_+ \right\}_{i \in [N]}$$

$$q = (\lfloor \epsilon N \rfloor + 1)\text{-th smallest value amongst } \{\xi_i\}_{i \in [N]}.$$

Compact formulation of (DR-CCP-MIP) via CVaR interpretation

Key result 3: recall that the DR-CCP is

$$\mathrm{CVaR}^{\mathbb{P}_N}_{1-\epsilon}(\mathrm{dist}(\xi,x)) = \max_{t,r}\left\{ t - \frac{1}{\epsilon N}\sum_{i\in[N]} r_i : \begin{array}{l} r_i \geq 0,\ i\in[N] \\ t - r_i \leq \mathrm{dist}(\xi_i,x),\ i\in[N] \end{array} \right\} \geq \frac{\theta}{\epsilon}.$$

- There always exists an optimal solution to the program such that

$$t = (\lfloor \epsilon N \rfloor + 1)\text{-th smallest value amongst } \left\{\mathrm{dist}(\xi_i,x) = (\xi_i + d - a^\top x)_+\right\}_{i\in[N]}$$
$$q = (\lfloor \epsilon N \rfloor + 1)\text{-th smallest value amongst } \{\xi_i\}_{i\in[N]}.$$

- Suppose $\xi_i \geq q$. Then immediately $t \leq \mathrm{dist}(\xi_i,x)$. But then

$$t - r_i \leq \mathrm{dist}(\xi_i,x) \iff 0 \leq r_i + (\mathrm{dist}(\xi_i,x) - t).$$

Therefore when $\xi_i \geq q$, this constraint is **vacuous**, so we can **remove** $N - \lfloor \epsilon N \rfloor$ **constraints.**

Strengthened compact formulation of (DR-CCP-MIP)

$$
\begin{aligned}
\min_{z,r,t,x} \quad & c^\top x \\
\text{s.t.} \quad & z \in \{0,1\}^N,\ t \geq 0,\ r \geq \mathbf{0},\ x \in \mathcal{X}, \\
& \epsilon\, t \geq \theta + \frac{1}{N} \sum_{i \in [N]} r_i, \\
& M_i(1 - z_i) \geq t - r_i, \quad i \in [N], \\
& f(x, \xi_i) + (q - \xi_i) z_i \geq 0, \quad i \in [N], \\
& \frac{1}{N} \sum_{i \in [N]} z_i \leq \epsilon, \\
& f(x, q) - t \geq 0 \\
& f(x, \xi_i) + m_i z_i \geq t - r_i, \quad i \in [N] \text{ s.t. } q > \xi_i.
\end{aligned}
$$

Valid inequalities for (DR-CCP-MIP)

Key result 4: classes of valid inequalities can be derived by analysing different substructures in the formulation.

- Consider again the so-called mixing substructure from the (SAA) constraints:

$$\text{MIX} = \left\{ (x,z) : \begin{array}{ll} f(x,\xi_i) + m_i z_i \geq 0, & i \in [N] \\ z \in \{0,1\}^N & \end{array} \right\}$$

$$\text{conv}(\text{MIX}) = \text{MIX} \cap \{\text{mixing inequalities}\}.$$

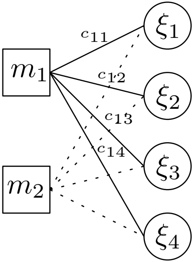
- There is also a substructure arising from robust 0-1 programming [Bertsimas and Sim, 2003]:

$$\text{ROB} = \left\{ (x,z,r,t) : \begin{array}{l} f(x,\xi_i) + m_i z_i \geq t - r_i, \ \ i \in [N] \text{ s.t } q > \xi_i \\ z \in \{0,1\}^N \end{array} \right\}$$

$$\text{conv}(\text{ROB}) = \text{ROB} \cap \{\text{path inequalities [Atamtürk, 2006]}\}.$$

Computational study

A distributionally robust chance-constrained transportation problem [Chen et al., 2018].



Given a set of factories $[F]$ with capacities $m_f, f \in [F]$, a set of distribution centers $[D]$ must meet the random demands ξ_d, $d \in [D]$ with high probability at minimum cost.

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & \mathbb{P}\left[\sum_{f\in[F]} x_{fd} \geq \xi_d, \quad \forall d \in [D]\right] \geq 1-\epsilon, \quad \mathbb{P} \in \mathcal{F}(\theta), \\ & \sum_{d\in[D]} x_{fd} \leq m_f, \quad f \in [F], \\ & x_{fd} \geq 0, \quad f \in [F],\ d \in [D]. \end{aligned}$$

$F = 5,\ D = 50,\ \epsilon = 0.1,\ \theta_1 = 0.001,\ \theta_j = \frac{j-1}{10}\theta_{\max}\ j = 2, \ldots, 10$

Performance analysis

We compare the following formulations (1 hour time limit)

- `Basic`: the basic formulation
- `Improved`: the strengthened compact formulation
- `Mixing+Path`: the strengthened compact formulation with both mixing and path inequalities.

Metrics:

- Time: recorded in seconds if instance is solved to optimality within one hour.
- Gap: if instance not solved in one hour, the final optimality gap as a percentage.

Summary of computational results

$N = 100$

	Basic	Improved	Mixing+Path	
	Time(Gap)$^{\text{Fnd}}$	Time	Time	M/P Cuts
θ_1	*(1.16)^{10}	4.29	8.40	41.7/274.6
θ_2	26.58(*)	0.04	0.06	0.3/88.2
θ_3	4.27(*)	0.04	0.05	0.0/73.8

$N = 3000$

	Basic	Improved	Mixing+Path	
	Time(Gap)$^{\text{Fnd}}$	Time(Gap)$^{\text{Fnd}}$	Time(Gap)$^{\text{Fnd}}$	M/P Cuts
θ_1	n/a^{0}	*(0.78)^{10}	*(0.48)^{10}	1470.3/4228.1
θ_2	*(69.56)^{5}	*(0.49)^{10}	*(0.41)^{10}	0.0/6102.2
θ_3	*(48.65)^{4}	17.89(*)	18.29(*)	0.0/200.8
θ_4	*(15.01)^{4}	13.74(*)	13.94(*)	0.0/94.1
θ_5	*(1.11)^{10}	12.75(*)	13.55(*)	0.0/88.3

Summary of computational results

$N = 3000$

	Basic		Improved		Mixing+Path	
	R.time	R.gap	R.time	R.gap	R.time	R.gap
θ_1	n/a	n/a	72.08	0.80	3601.05	0.48
θ_2	3144.09	70.41	134.46	0.55	3600.22	0.41
θ_3	2952.26	51.31	17.89	0.01	18.29	0.01
θ_4	2684.77	15.72	13.74	0.01	13.94	0.01
θ_5	3181.43	1.14	12.75	0.00	13.55	0.00
θ_6	3176.11	0.63	12.29	0.00	12.68	0.00
θ_7	2958.81	0.55	12.28	0.01	12.95	0.01
θ_8	2876.49	0.47	12.48	0.01	12.65	0.01
θ_9	2781.77	0.45	11.96	0.01	12.52	0.01
θ_{10}	2439.69	0.41	8.04	0.01	8.94	0.01

Discussion

- Strong reformulation of (DR-CCP) that exploits connections with various other models for uncertainty
  - nominal (SAA) relaxation
  - conditional value-at-risk (CVaR) interpretation
  - a substructure that arises in robust 0-1 programming.

  Using these connections we provided two classes of valid inequalities for (DR-CCP).
- Extended to more general polyhedral safety sets involving multiple linear constraints and left-hand side uncertainty. [Ho-Nguyen,Kılınç-Karzan, K., Lee, 2021b+]
- Left-hand side uncertainty case involves conic constraints in the form

$$\|Ax\|_p \leq t.$$

  - [Xie, 2019] use polymatroid inequalities to strengthen the formulation when x is a pure binary decision vector, using submodularity of $\|Ax\|_p$.
  - [Kılınç-Karzan, K., and Lee, 2020+] extend the polymatroid inequalities to obtain valid inequalities when x is mixed-binary. (MIP Workshop, May 25, 2021)
  - Submodularity can also be exploited for distributionally robust pure binary optimization problems under moment-based ambiguity sets, e.g., [Zhang et al., 2018].

Parting thoughts

- Stochastic optimization problems often give rise to large-scale MIPs
- Opportunities for theoretical, methodological, and computational MIP research
- Wide range of applications with broad impact (disaster logistics, energy, healthcare, and more).

Selected References

- B. Chen, S. Küçükyavuz, and S. Sen. Finite disjunctive programming characterizations for general mixed-integer linear programs, **Operations Research**, 59:202-210, 2011.
- B. Chen, S. Küçükyavuz, and S. Sen. A computational study of the cutting plane tree algorithm for general mixed-integer linear programs, **Operations Research Letters**, 40:15-19, 2012.
- D. Gade, S. Küçükyavuz, and S. Sen. Decomposition algorithms with parametric Gomory cuts for two-stage stochastic integer programs, **Mathematical Programming**, 144(1-2):39-64, 2014.
- S. Küçükyavuz, and R. Jiang, Chance-Constrained Optimization: A Review of Mixed-Integer Conic Formulations and Applications, arXiv:2101.08746, 2021. (Survey)
- N. Ho-Nguyen, F. Kılınç-Karzan, S. Küçükyavuz, and D. Lee, Distributionally Robust Chance-Constrained Programs with Right-Hand Side Uncertainty under Wasserstein Ambiguity, *forthcoming*, **Mathematical Programming**, 2021a.
- N. Ho-Nguyen, F. Kılınç-Karzan, S. Küçükyavuz, and D. Lee, Strong Formulations for Distributionally Robust Chance-Constrained Programs with Left-Hand Side Uncertainty under Wasserstein Ambiguity, arXiv:2007.06750, 2021b.
- F. Kılınç-Karzan, S. Küçükyavuz, and D. Lee, Joint Chance-Constrained Programs and the Intersection of Mixing Sets through a Submodularity Lens, arXiv:1910.01353, 2019.
- S. Küçükyavuz, and S. Sen, An Introduction to Two-Stage Stochastic Mixed-Integer Programming, **2017 INFORMS TutORials in Operations Research** (eds. R. Batta and J. Peng), 1-27, 2017. (Tutorial)
- S. Küçükyavuz. On mixing sets arising in chance-constrained programming, **Mathematical Programming**, 132:31-56, 2012.
- S. Küçükyavuz, and N. Noyan, Cut Generation for Optimization Problems with Multivariate Risk Constraints, **Mathematical Programming**, 159(1), 165-199, 2016.
- X. Liu, F. Kılınç-Karzan, and S. Küçükyavuz, On Intersection of Two Mixing Sets with Applications to Joint Chance-Constrained Programs, **Mathematical Programming**, 175(1-2), 29-68, 2019.
- X. Liu, S. Küçükyavuz, and J. Luedtke. Decomposition algorithms for two-stage chance-constrained programs, **Mathematical Programming**, 157(1):219-243, 2016.
- M. Meraklı and S. Küçükyavuz, Vector-Valued Multivariate Conditional Value-at-Risk, **Operations Research Letters**, 46(3), 300-305, 2018.
- M. Zhang and S. Küçükyavuz. Finitely convergent decomposition algorithms for two-stage stochastic pure integer programs, **SIAM Journal on Optimization**, 24(4):1933-1951, 2014.
- M. Zhang, S. Küçükyavuz, and S. Goel. A branch-and-cut method for dynamic decision making under joint chance constraints, **Management Science**, 60(5):1317-1333, 2014.