Projection Cuts for Two-Stage Stochastic Mixed-Integer Programs

Ruiwei Jiang

Industrial and Operations Engineering, University of Michigan

Joint work with Haoming Shen (IOE, Michigan)

Suported by the NSF (ECCS-1845980)

Outline

Optimization under Uncertainty

- Decision making in an uncertain environment
  - Crucial decisions made before uncertainty is realized.
  - Two-stage decision making: here-and-now + wait-and-see.
  - Key safety constraints.

Optimization under Uncertainty

- Decision making in an uncertain environment
  - Crucial decisions made before uncertainty is realized.
  - Two-stage decision making: here-and-now + wait-and-see.
  - Key safety constraints.
- Examples:
  - Day-Ahead Unit Commitment
    - ★ Uncertainty: renewable.
    - ★ Decision: UC + economic dispatch.
    - ★ Safety: transmission dispatchable.

Optimization under Uncertainty

- Decision making in an uncertain environment
  - Crucial decisions made before uncertainty is realized.
  - Two-stage decision making: here-and-now + wait-and-see.
  - Key safety constraints.
- Examples:
  - Day-Ahead Unit Commitment
    - ★ Uncertainty: renewable.
    - ★ Decision: UC + economic dispatch.
    - ★ Safety: transmission dispatchable.
  - Transmission Expansion Planning
    - ★ Uncertainty: load.
    - ★ Decision: expansion + economic dispatch.
    - ★ Safety: no load shedding.

Two-Stage Stochastic Mixed Integer Programs (SMIP)

SMIP

$$\begin{aligned} \min_{y \geq 0} \ & c^\top y + \mathbb{E}_{\tilde{\xi}}[Q(y, \tilde{\xi})] \\ \text{s.t.} \ & Ay = b \\ & y \in Y. \end{aligned} \qquad Q(y, \tilde{\xi}) := \min_{x \geq 0} \ q^\top x \quad \text{s.t. } Wx = h(\tilde{\xi}) + Ty.$$

- y: here-and-now (may be mixed-integer).
- x: wait-and-see.
- $h(\tilde{\xi})$: random right-hand-side.

Two-Stage Stochastic Mixed Integer Programs (SMIP)

SMIP

$$\begin{aligned} \min_{y \geq 0} \ & c^\top y + \mathbb{E}_{\tilde{\xi}}[Q(y, \tilde{\xi})] \\ \text{s.t.} \ & Ay = b \\ & y \in Y. \end{aligned} \qquad Q(y, \tilde{\xi}) := \min_{x \geq 0} \ q^\top x \quad \text{s.t. } Wx = h(\tilde{\xi}) + Ty.$$

- y: here-and-now (may be mixed-integer).
- x: wait-and-see.
- $h(\tilde{\xi})$: random right-hand-side.

- Example: stochastic unit commitment
  - $y =$ UC, $x =$ power flow, $h(\tilde{\xi}) =$ renewable input.
  - Inequality constraints recast as equalities WLOG.

Two-Stage Stochastic Mixed Integer Programs (SMIP)

SMIP

$$\begin{aligned} \min_{y \geq 0} \ & c^\top y + \frac{1}{S}\sum_{i=1}^{S} \theta_i \\ \text{s.t.} \ & Ay = b \\ & y \in Y. \end{aligned}$$

$$\begin{aligned} \theta_i := \min_{x \geq 0} \ & q^\top x \\ \text{s.t.} \ & Wx = h_i + Ty. \end{aligned}$$

- Monte Carlo approximation, sample average approximation.

Benders Decomposition (BD)

SMIP

$$(\textbf{MP}): \min_{y \geq 0} \; c^\top y + \frac{1}{S}\sum_{i=1}^{S} \theta_i \quad \text{s.t. } Ay \leq b, \; y \in Y.$$

$$(\textbf{SP}_i): \; \theta_i := \min_{x \geq 0} \; q^\top x \quad \text{s.t. } Wx = h_i + Ty.$$

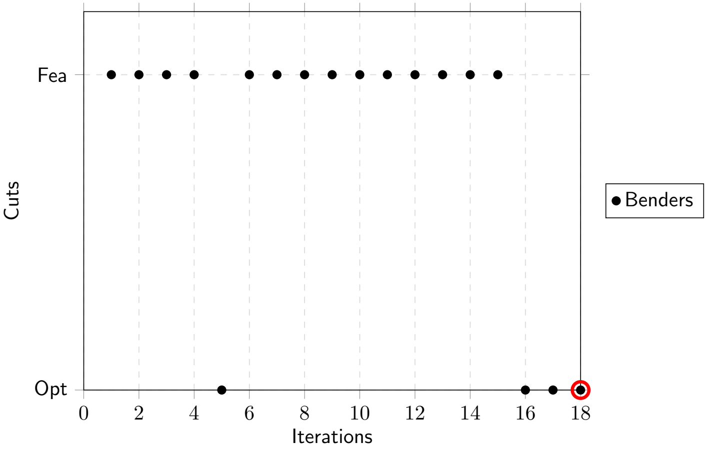
- Solve (MP) with an optimal solution $(\hat{y}, \hat{\theta})$; Plug $\hat{y}$ into (SP)'s.
- If an (SP_i) infeasible, add a feasibility cut.
- If ALL (SP_i) feasible, add an optimality cut.
- Iterate.

Benders Decomposition (BD)

- Challenge: slow; too many cuts needed to converge.

Benders Decomposition (BD)

- Challenge: slow; too many cuts needed to converge.
- Example: SUC, IEEE-57, #Fea = 14,000, #Opt = 2,681.



Very Brief Literature Review

- **Selectively cut adding**: Rei et al. (2009), Yang and Lee (2012).
- **Avoid feasibility cuts through valid inequalities**: Geoffrion and Graves (1974), de Sá et al. (2013).
- **Generate stronger feasibility / optimality cuts**: Codato and Fischetti (2006), Contreras et al. (2011), Fischetti et al. (2010), Magnanti and Wong (1981), Bodur et al. (2017), Bodur and Luedtke (2017), Rahmaniani et al. (2018).

Very Brief Literature Review

- **Selectively cut adding**: Rei et al. (2009), Yang and Lee (2012).
- **Avoid feasibility cuts through valid inequalities**: Geoffrion and Graves (1974), de Sá et al. (2013).
- **Generate stronger feasibility / optimality cuts**: Codato and Fischetti (2006), Contreras et al. (2011), Fischetti et al. (2010), Magnanti and Wong (1981), Bodur et al. (2017), Bodur and Luedtke (2017), Rahmaniani et al. (2018).

- Much less attention on feasibility cuts.
- Slack variables added and penalized; but not always desired.
- Typically application-dependent.
- **Our focus**: feasibility cuts, more general-purpose.

Why So Many Feasibility Cuts?

Second-Stage Problem

$$\theta_i := \min_{x \geq 0} \; q^\top x$$
$$\text{s.t. } Wx = h_i + Ty.$$

Observations

1. The i^{th} subproblem is feasible $\iff$ $(h_i + Ty) \in \text{pos}(W)$.
2. Oftentimes, $\text{pos}(W)$ is "thin," even a subspace.

Why So Many Feasibility Cuts?

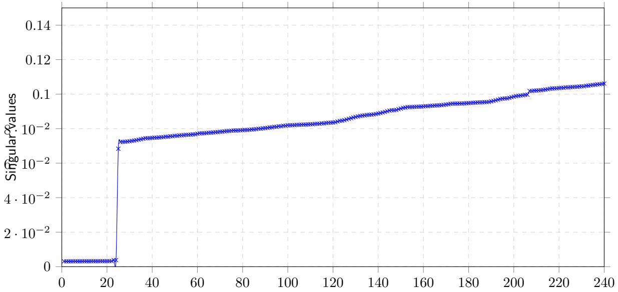


Figure: Smallest singular values of W in IEEE-118 system

- If not carefully guided, can take long for $(h_i + Ty)$ to enter pos(W).

Why So Many Feasibility Cuts?

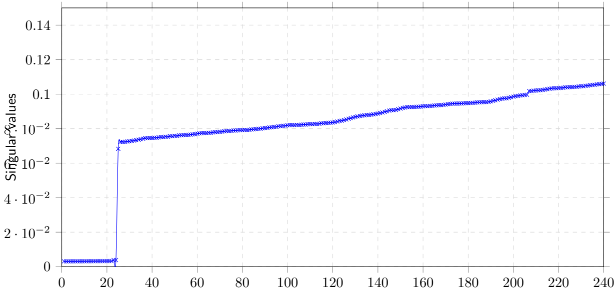


Figure: Smallest singular values of W in IEEE-118 system

- If not carefully guided, can take long for $(h_i + Ty)$ to enter pos(W).
- Idea: search for thin spreads and restrict $(h_i + Ty)$ accordingly.
- How to find thin spreads? One approach: Principal Component Analysis.

Principal Component Analysis (PCA) on pos(W)

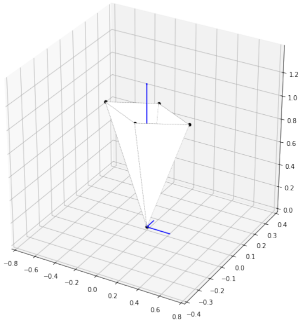


Figure: PCA on a polyhedral cone

Blue lines: singular vectors;
Length $\propto$ singular value.

Theorem (Levin and Shashua, 2002)

The covariance matrix Cov(W) of the entire polytope is the same as the covariance matrix over the vertices of W.

Principal Component Analysis (PCA) on pos(W)

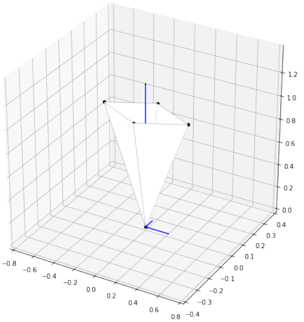


Figure: PCA on a polyhedral cone

Blue lines: singular vectors;
Length $\propto$ singular value.

Last k principal components of pos(W):

1. Normalize the polyhedral cone $\{W\alpha \mid \|W\alpha\|_1 \leq 1, \alpha \geq 0\}$ and get its extreme points.
2. Center these extreme points.
3. Perform PCA and get $\mathcal{U} := \{u_{n-k+1}, \ldots, u_n\}$, along which pos($W$) has thinnest spreads.

Projection Cuts

For any $u \in \mathcal{U}$, we hope that the projection of $(h_i + Ty)$ onto u is small in order to stay in pos(W):

$$\underline{\epsilon} \leq \frac{u^\top (h_i + Ty)}{\|h_i + Ty\|_1} \leq \bar{\epsilon}, \tag{7}$$

where
$$\bar{\epsilon} := \max_{\lambda} \{u^\top W\lambda \mid \|W\lambda\|_1 \leq 1, \lambda \geq 0\},$$
$$\underline{\epsilon} := \min_{\lambda} \{u^\top W\lambda \mid \|W\lambda\|_1 \leq 1, \lambda \geq 0\}.$$

$\bar{\epsilon}$, $\underline{\epsilon}$ found by solving 2 LPs.

Assumptions on SMIP

Assumptions

A1. The uncertainty and the hear-and-now decisions are separable, i.e., there exist $d(\tilde{\xi})$ and T' such that

$$h(\tilde{\xi}) + Ty = \begin{bmatrix} d(\tilde{\xi}) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ T' \end{bmatrix} y,$$

A2. All entries of T' are nonnegative.

Assumptions on SMIP

- Projection cuts can be linearized in more general ways.
- Assumptions A1–A2: simple and exact linearization.
- Valid for a wide range of SMIP counterparts:
  - Unit commitment, transmission expansion planning.
  - Multicommodity network design, production routing.

Projection Cuts

For any $u \in \mathcal{U}$, we hope that the projection of $(h_i + Ty)$ onto u is small in order to stay in pos(W):

$$\underline{\epsilon} \leq \frac{u^\top (h_i + Ty)}{\|h_i + Ty\|_1} \leq \bar{\epsilon}$$

Linearization:

$$\begin{aligned} \|h_i + Ty\|_1 &= \left\| \begin{bmatrix} d_i \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ T' \end{bmatrix} y \right\|_1 \quad (\because \text{Separable } d(\tilde{\xi}) \text{ and } y) \\ &= \|h_i\|_1 + \|Ty\|_1 \\ &= \|h_i\|_1 + \mathbf{1}^\top Ty \qquad (\because T \geq 0) \end{aligned}$$

Projection Cuts

Projection Cuts

For any u in $\mathcal{U}$, the following inequalities are valid for SMIP:

$$\max_{i \in [S]} \{u^\top h_i - \bar{\epsilon}\, \|h_i\|_1\} \leq \left(u^\top - \bar{\epsilon}\, \mathbf{1}^\top\right) Ty,$$

$$\left(u^\top - \underline{\epsilon}\, \mathbf{1}^\top\right) Ty \leq \min_{i \in [S]} \{u^\top h_i - \underline{\epsilon}\, \|h_i\|_1\}.$$

Two-Stage SUC

Unit commitment problem

1. Minimizes system operating costs.
2. Guarantees all system constraints are satisfied.
3. Needs to be solved day-ahead and every day.
4. System load is subject to uncertainty.

Two-stage decision-making

1. UC. (binary)
2. Economic dispatch. (continuous)

Experimental Design

Compare vanilla Benders Decomposition (BD) and Benders Decomposition equipped with Projection Cuts (ProjBD).

- Testcases: standard IEEE testcases with $9, 14, 30, 57, 118$ buses.
- System demands: sampled from Gaussian distribution with 80% of the nominal system demand and standard deviation is 0.1.
- $\mathcal{U}$: The shortest 5% principal component vectors of pos(W).
- Time limit: 3600 seconds.
- ns : number of samples (realizations) of $\tilde{\xi}$.

Experiment Results

Table: BD vs. ProjBD, `ns`=500

	BD				ProjBD			
Testcases	time (sec)	gap	#feascuts	#optcuts	time (sec)	gap	#feascuts	#optcuts
9 bus	233.48	0.0	186408	176550	49.66	0.0	52256	90414
14 bus	299.70	0.0	86210	74197	59.53	0.0	24742	16020
30 bus	382.0	0.0	42920	14642	69.95	0.0	16598	9381
57 bus	826.18	0.0	106415	22082	118.91	0.0	22598	13677
118 bus	NaN	inf	NaN	NaN	430.57	0.0	140714	425786

Experiment Results

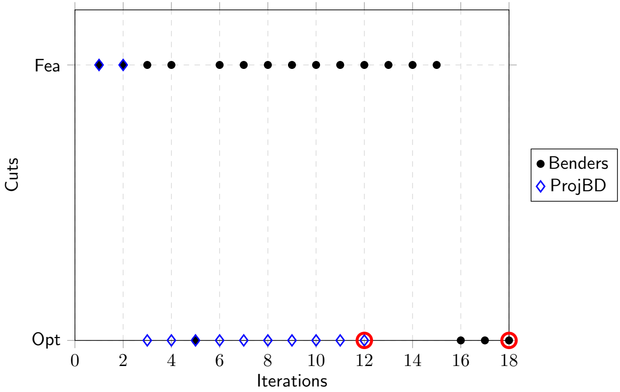
Table: BD vs. ProjBD, `ns=1000`

	BD				ProjBD			
Testcases	time (sec)	gap	#feascuts	#optcuts	time (sec)	gap	#feascuts	#optcuts
9 bus	505.04	0.0	259305	524608	271.75	0.0	102353	283946
14 bus	1102.58	0.0	74200	60675	76.43	0.0	20098	11466
30 bus	330.29	0.0	4170	4685	297	0.0	7890	4987
57 bus	957.33	0.0	99398	19164	208.21	0.0	21598	12740
118 bus	NaN	inf	NaN	NaN	807.805	0.0	134414	418954

- The numbers of **both** feasibility and optimality cuts reduced.
- Total time significantly reduced.
- Similar observations from multicommodity network design and production routing case studies.

Revisit

- Challenge: slow; too many cuts needed to converge.
- BD: #Fea = 14,000, #Opt = 2,681.
- ProjBD: #Fea = 2,000, #Opt = 1,198.



Root-Node Performance

Table: Root node performace of BD and ProjBD

Instances	ProjBD			BD			lbd ratio
	lbd	gap	ubd	lbd	gap	ubd	
9 bus	116,414.20	3.52%	100.00%	3,730.54	N/A	0.00%	32.73
14 bus	198,392.79	5.45%	100.00%	19,254.23	N/A	0.00%	214.40
30 bus	20,238.12	46.25%	100.00%	5,196.32	N/A	0.00%	4.85
57 bus	930,093.01	0.00%	100.00%	189,103.97	N/A	0.00%	205.71
118 bus	2,039,338.84	10.54%	100.00%	48,154.25	N/A	0.00%	1,596.91

How Many Singular Values to Involve?

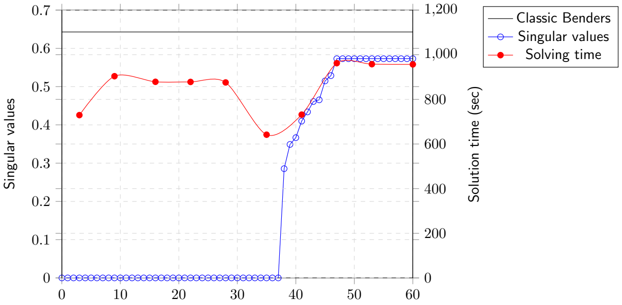


Figure: # of Singular Values to Involve vs. Solving Time