A Generalized Benders' Algorithm for the Two-Stage Stochastic Optimization Problems With Mixed Integer Recourse

Anahita Hassanzadeh[1] Ted Ralphs[2] Menal Güzelsoy[3]

[1,2]COR@L Lab, Department of Industrial and Systems Engineering, Lehigh University

[3]SAS Institute

INFORMS Annual Meeting Minneapolis, October 2013

Outline

Outline

Two-Stage Stochastic Program with Recourse

$$\begin{aligned} \min f(x) = \min\ & c^\top x + \mathbb{E}_{w\in\Omega}[\Phi(x,w)] \\ \text{s.t. } & x \in X \end{aligned} \tag{SP}$$

$$\begin{aligned} \Phi(x,w) = \min\ & q^\top y \\ \text{s.t. } & Wy = h(w) - T(w)x \\ & y \in Y \end{aligned} \tag{RP}$$

where X and Y are the feasible regions of the first and second stages and may be discrete sets. In this talk, we assume

- w follows a discrete distribution with a finite support, and
- W and q are fixed.
- X is non-empty and bounded.
- $\mathbb{E}_{w\in\Omega}[\Phi(x,w)]$ is finite for all $x \in X$.

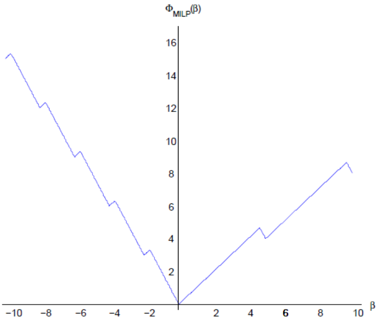
Overview

- We aim to develop an algorithm for solving two-stage stochastic mixed integer programs that can be implemented in practice.
- Solution of the problem requires analysis of how the solution to the second-stage problem varies as a function of the first stage solution.
- The first part of this talk will focus on properties of the *value function* of a mixed integer linear program.
- In the second part, we describe a Benders-like algorithm based on approximation of the value function.

MILP Value function

Example 1

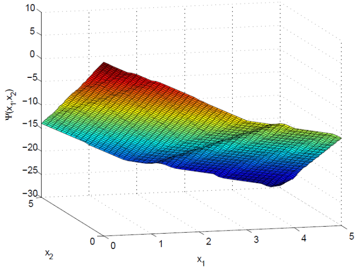
$$\Phi_{MILP}(b) = \min\ 6y_1 + 4y_2 + 3y_3 + 4y_4 + 5y_5 + 7y_6$$
$$\text{s.t. } 2y_1 + 5y_2 - 2y_3 - 2y_4 + 5y_5 + 5y_6 = b \qquad \text{(Ex.MILP)}$$
$$y_1, y_2, y_3 \in \mathbb{Z}_+, y_4, y_5, y_6 \in \mathbb{R}_+.$$



Example 2

$$\min f(x_1, x_2) = \min \ -3x_1 - 4x_2 + \mathbb{E}[\Phi_{MILP}(\omega - 2x_1 - 0.5x_2)]$$
$$\text{s.t. } x_1 \leq 5, x_2 \leq 5$$
$$x_1, x_2 \in \mathbb{R}_+, \qquad \text{(Ex.SMP)}$$

and $\omega \in \{6, 12\}$ with a uniform probability distribution.



The Role of the MILP Value Function

- Benders' method relies on finding effective functions to approximate the recourse value function from below.
- The Benders' method for solving the continuous two-stage problems does not apply directly when Y contains integer variables.
- We want to find functions that are approximating the MILP value function well, and are computationally implementable.
- The search for effective approximating functions begins with learning about the MILP value function structure.

Outline

1. Introduction
2. MILP Value Function
3. Algorithm
4. Conclusions

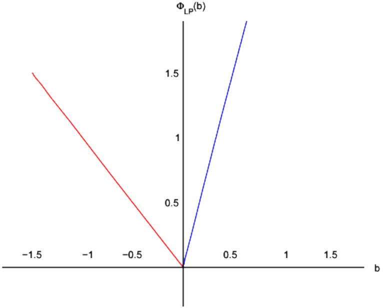
Discrete Structure of the Value Function

- Our goal is to answer some important questions about the MILP value function such as:
  - Is there any relation between LP value function and MILP value function? Why is the latter non-convex?
  - What is the role of continuous and integer variables in defining the structure of the MILP value function?
  - Where do the discontinuities occur?
- We show that these questions can be answered through exploiting the discrete structure of the value function.
- This structure arises as a combination of the discrete structures of two underlying value functions.
  - The continuous restriction.
  - The integer restriction.

Consider:

Example 3

$$\phi_{LP}(b) = \min\ 6x_1 + 7x_2 + 5x_3$$
$$\text{s.t. } 2x_1 - 7x_2 + x_3 = b$$
$$x_1, x_2, x_3 \in \mathbb{R}_+$$



LP Value Function Structure

$$\Phi_{LP}(b) = \min \ c^\top x \quad \text{s.t. } Ax = b, \quad x \in \mathbb{R}^n_+ \tag{LP}$$

- Assume the dual of (LP) is feasible.
- The epigraph of Φ_{LP} is a convex cone, call it $\mathcal{L}$:

$$\mathcal{L} := cone\{(A_1, c_1), (A_2, c_2), \ldots, (A_n, c_n), (0, 1)\}$$

- Let $u_1, \ldots, u_k$ be extreme points of the feasible region of the dual of (LP) and $d_1, \ldots, d_p$ be its extreme directions. Then

$$\mathcal{L} := \{(b, z) : z \geq u_i^\top b, i = 1, \ldots, k, d_j^\top b \leq 0, j = 1, \ldots, p\}.$$

- Note that the value function has an underlying discrete structure.

MILP Value Function

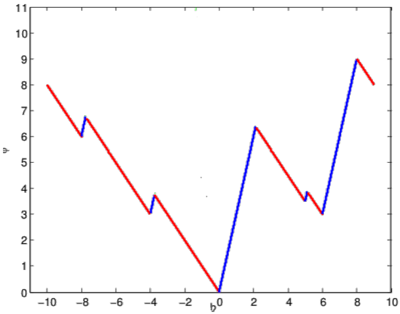
MILP value function is non-convex and discontinuous piecewise polyhedral.

Example 4

$$\phi(b) = \min\ 3x_1 + \frac{7}{2}x_2 + 3x_3 + 6x_4 + 7x_5 + 5x_6$$

$$\text{s.t. } 6x_1 + 5x_2 - 4x_3 + 2x_4 - 7x_5 + x_6 = b$$

$$x_1, x_2, x_3 \in \mathbb{Z}_+,\ x_4, x_5, x_6 \in \mathbb{R}_+$$



Continuous and Integer Restriction of an MILP

Consider

$$\begin{aligned}\Phi(b) = \min\, & c_I^\top x_I + c_C^\top x_C \\ \text{s.t. } & A_I x_I + A_C x_C = b, \\ & x \in \mathbb{Z}_+^r \times \mathbb{R}_+^{n-r}\end{aligned} \tag{MILP}$$

Define the *continuous restriction* of (MILP) as

$$\begin{aligned}\Phi_C(b) = \min\, & c_C^\top x_C \\ \text{s.t. } & A_C x_C = b, \\ & x \in \mathbb{R}_+^{n-r}\end{aligned} \tag{CR}$$

and its *integer restriction* as

$$\begin{aligned}\Phi_I(b) = \min\, & c_I^\top x_I \\ \text{s.t. } & A_I x_I = b \\ & x_I \in \mathbb{Z}_+^r\end{aligned} \tag{IR}$$

Discrete Representation of the Value Function

For $b \in \mathbb{R}^m$, we have that

$$\begin{aligned} \Phi(b) = \min\ & c_I^\top x_I + \Phi_C(b - A_I x_I) \\ \text{s.t. } & x_I \in \mathbb{Z}^r_+ \end{aligned} \tag{1}$$

- From this we see that the value function is comprised of the minimum of a set of shifted copies of Φ_C.
- The set of shifts, along with Φ_C describe the value function exactly.
- For $\hat{x}_I \in \mathbb{Z}^r_+$, let

$$\Phi_C(b, \hat{x}_I) = c_I^\top \hat{x}_I + \Phi_C(b - A_I \hat{x}_I)\ \forall b \in \mathbb{R}^m.$$

- Then we have that $\Phi(b) = \min_{x_I \in \mathbb{Z}^r_+} \Phi_C(b, \hat{x}_I)$.

Properties of MILP Value Function

We define

- $\mathcal{S}_D = \{\nu : A_C^\top \nu \leq c_C\}$.
- $\mathcal{E} = \{E \in \mathbb{R}^n : E \text{ is the index set of a dual feasible basis of (CR)}\}$.
- For $E \in \mathcal{E}$, $\nu_E^\top = c_E^\top A_E^{-1}$ (extreme points of $\mathcal{S}_D$).

Proposition 2.1

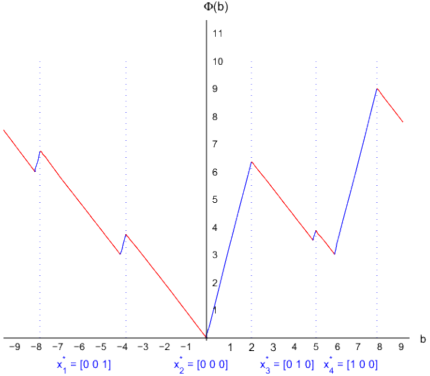
The gradient of Φ on a neighborhood of a differentiable point is a unique optimal dual feasible solution to (CR).

Proposition 2.2

Consider $\mathcal{N} \subseteq B$ over which Φ is differentiable. Then, there exist an integral part of the solution $x_I^* \in \mathbb{Z}^r$ and $E \in \mathcal{E}$ such that $\Phi(b) = c_I^\top x_I^* + \nu_E^\top (b - A_I x_I^*)$ for all $b \in \mathcal{N}$.

Proposition 2.3

Consider $\mathcal{N} \subseteq B$ *over which* Φ *is convex. Then, there exist an integral part of the solution* $x_I^* \in \mathbb{Z}^r$ *such that* $\Phi(b) = c_I^\top x_I^* + \nu_E^\top (b - A_I x_I^*)$ *for all* $b \in \mathcal{N}$ *for some* $E \in \mathcal{E}$ *.*



Properties of MILP Value Function

Bottom Line: The value function of a MILP has discrete structure arising from a pure integer problem and can be constructed without solving the original MILP.

Outline

Related Algorithms

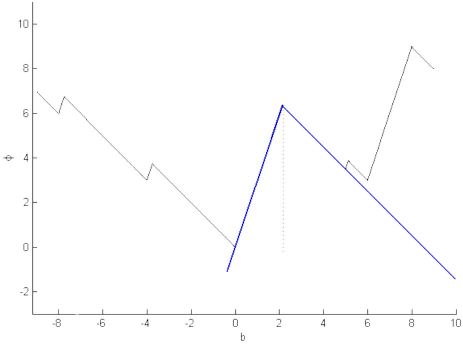
The algorithmic framework we utilize builds on a number of previous works.

- Modification to the L-shaped framework [Laporte and Louveaux, 1993, Carøe and Tind, 1998, Sen and Higle, 2005]
    - Linear cuts in first stage for binary first stage
    - Optimality cuts from B&B and cutting plane, applied to pure integer second stage
    - Disjunctive programming approaches and cuts in the second stage
- Value function approaches: Pure integer case [Ahmed et al., 2004, Kong et al., 2006]
- Scenario decomposition [Carøe and Schultz, 1998]
- Enumeration/Gröbner basis reduction [Schultz et al., 1998]

	First Stage			Second Stage			Stochasticity			
	$\mathbb{R}$	$\mathbb{Z}$	$\mathbb{B}$	$\mathbb{R}$	$\mathbb{Z}$	$\mathbb{B}$	**W**	**T**	**h**	**q**
[Laporte and Louveaux, 1993]			*	*	*	*	*	*	*	
[Carøe and Tind, 1997]	*		*	*		*	*	*	*	*
[Carøe and Tind, 1998]	*	*	*		*	*		*	*	
[Carøe and Schultz, 1998]	*	*	*	*	*	*		*	*	*
[Schultz et al., 1998]	*				*	*			*	
[Sherali and Fraticelli, 2002]			*	*		*	*	*	*	*
[Ahmed et al., 2004]	*	*	*		*	*	*		*	*
[Sen and Higle, 2005]			*	*		*		*	*	
[Sen and Sherali, 2006]			*	*	*	*		*	*	
[Sherali and Zhu, 2006]	*		*	*		*	*	*	*	
[Kong et al., 2006]		*	*		*	*	*	*	*	*
[Sherali and Smith, 2009]			*	*		*	*	*	*	*
[Yuan and Sen, 2009]			*	*		*		*	*	*
[Ntaimo, 2010]			*	*		*	*			*
[Gade et al., 2012]			*		*	*	*	*	*	*
[Trapp et al., 2013]		*	*		*	*			*	
Current work	*	*	*	*	*	*		*	*	

Two-Stage Problem and Value Function

We already observed that for an effective integer Benders' method, we need effective lower bounding functions to approximate the MILP value function.



Dual Functions

A dual function $\varphi : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}$ is

$$\varphi(b) \leq \Phi(b) \quad \forall b \in \Lambda$$

For a particular instance $\hat{b}$, the dual problem is

$$\Phi_D = \max\{\varphi(\hat{b}) : \varphi(b) \leq \Phi(b) \quad \forall b \in \mathbb{R}^m, \; \varphi : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}\}$$

Dual Functions

A dual function $\varphi : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}$ is

$$\varphi(b) \leq \Phi(b) \quad \forall b \in \Lambda$$

For a particular instance $\hat{b}$, the dual problem is

$$\Phi_D = \max\{\varphi(\hat{b}) : \varphi(b) \leq \Phi(b) \quad \forall b \in \mathbb{R}^m,\ \varphi : \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}\}$$

Let $\mathcal{F}_s$ be the set of all dual functions to subproblem s.
Then the master Benders' problem is

$$\begin{aligned} \min\ & c^\top x + \theta \\ & \theta \geq \mathbb{E}_s[\max_{f_s \in \mathcal{F}_s} f_s(h(s) - T(s)x)] \\ & x \in X \end{aligned} \qquad \text{(MP)}$$

MILP Duals from Branch-and-Bound

Let T be set of the terminating nodes of the tree. Then in a terminating node $t \in T$ we solve:

$$\begin{aligned} \min\ & c^\top x \\ \text{s.t. } & Ax = b, \\ & l^t \leq x \leq u^t, x \geq 0 \end{aligned} \tag{2}$$

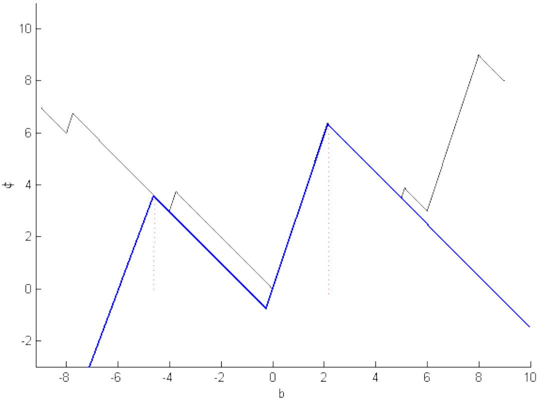
The dual at node t:

$$\begin{aligned} \max\ & \{\pi^t b + \underline{\pi}^t l^t + \bar{\pi}^t u^t\} \\ \text{s.t. } & \pi^t A + \underline{\pi}^t + \bar{\pi}^t \leq c^\top \\ & \underline{\pi} \geq 0, \bar{\pi} \leq 0 \end{aligned} \tag{3}$$

We obtain the following strong dual function:

$$\min_{t \in T}\{\pi^t b + \underline{\pi}^t l^t + \bar{\pi}^t u^t\} \tag{4}$$

MILP Duals from Branch-and-Bound



Master Problem Formulation

Notation:

- $s, r \in \{1, \ldots, S\}$ where S is the number of scenarios
- $p \in \{1, \ldots, k\}$ where k is the iteration number
- $n \in \{1, \ldots, N(p, r)\}$ where $N(p, r)$ is the number of terminating nodes in the B&B tree solved for scenario r at iteration p.
- $\theta_s = \mathcal{F}(h(s) - \beta)$
- $t_{spr} = F_r^p(h(s) - \beta)$ the approximation of scenario s's recourse obtained from the optimal dual function of iteration p and scenario r.
- ν_{prn}, a_{prn} respectively, the dual vector and intercept obtained from node n of the B&B tree solved for scenario r in iteration p.
- p_s probability of scenario s
- $M > 0$ an appropriate large number

Master Problem Formulation

$$
\begin{aligned}
f^k = \min\ & c^\top x + \sum_{s=1}^{S} p_s \theta_s \\
\text{s.t. } & \theta_s \geq t_{spr} && \forall s, p, r \\
& t_{spr} \leq a_{prn} + \nu_{prn}^\top (h(s) - T(s)x) && \forall s, r, p, n \\
& t_{spr} \geq a_{prn} + \nu_{prn}^\top (h(s) - T(s)x) - M u_{sprn} && \forall s, p, r, n \\
& \sum_{n=1}^{N} u_{sprn} = N(p, r) - 1 && \forall s, p, r \\
& x \in X, u_{sprn} \in \mathbb{B} && \forall s, p, r, n
\end{aligned}
\qquad \text{(master)}
$$

Example

Consider

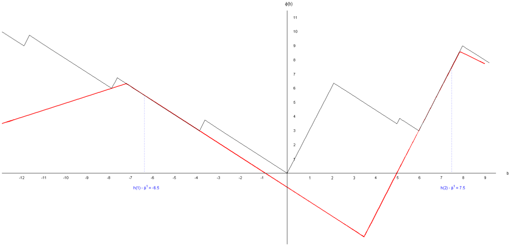
$$\min f(x) = \min \ -3x_1 - 4x_2 + \sum_{s=1}^{2} 0.5Q(x,s)$$
$$\text{s.t. } x_1 + x_2 \leq 5$$
$$x \in \mathbb{Z}_+ \tag{5}$$

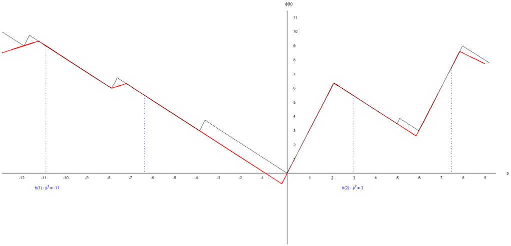
where

$$Q(x,s) = \min \ 3y_1 + \frac{7}{2}y_2 + 3y_3 + 6y_4 + 7y_5$$
$$\text{s.t. } 6y_1 + 5y_2 - 4y_3 + 2y_4 - 7y_5 = h(s) - 2x_1 - \frac{1}{2}x_2 \tag{6}$$
$$y_1, y_2, y_3 \in \mathbb{Z}_+, \ y_4, y_5 \in \mathbb{R}_+$$

with $h(s) \in \{-4, 10\}$.

Example

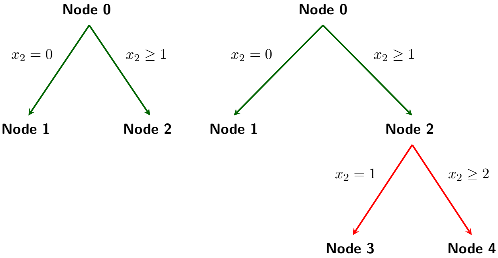


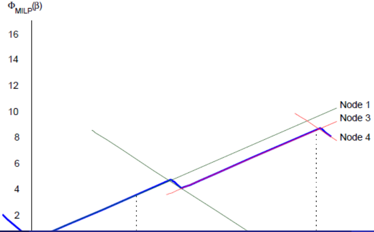


B&B Dual Functions

Adding cuts to the master problem blindly makes it very hard.

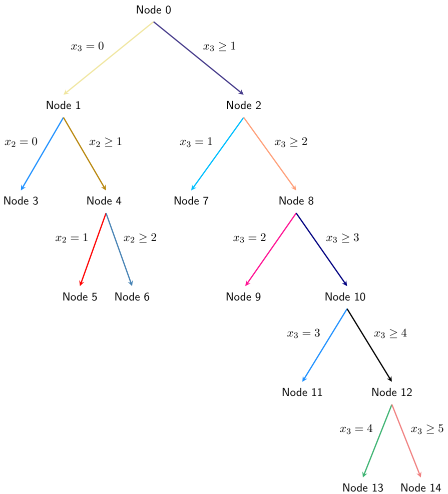
But B&B dual functions offer desirable properties:

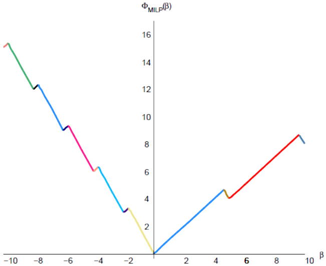




B&B Dual Functions

In fact, the value function of a MILP can be constructed by the dual function of a single B&B tree.





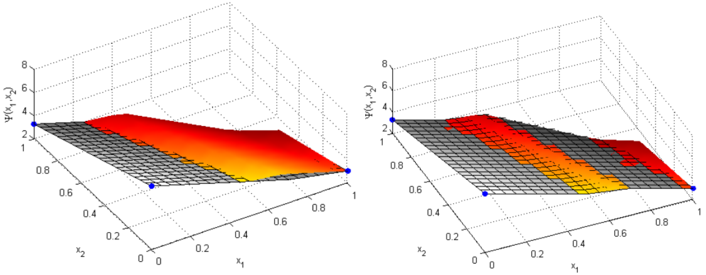
B&B Dual Functions

- We need to examine local areas of this tree to get local approximations of the value function.
- This is done through warm-starting individual subtrees and sensitivity analysis on their terminating nodes.
- Local optimality cuts are generated as we keep branching on the subtrees.
- A single tree can be optimal for multiple scenarios.
- We benefit from sensitivity analysis with scenario bunching.

Outline

Conclusions

Non-convex optimality cuts are ugly. But they may be worthwhile!



- We have developed an algorithm for the two-stage problem with general mixed integer in both stages.
- The algorithm uses the Benders' framework with B&B dual functions as the optimality cuts.
- Such cuts have computationally desirable properties such as warm-starting.
- We need to keep the size of approximations small. This can be done through warm-starting trees and scenario bunching.

Future Work

- We have implemented the algorithm using SYMPHONY as our mixed-integer linear optimization solver.
- Warm-starting a B&B tree is possible in the solver.
- The choice of a "good" warm-starting tree is not known a priori. We still need to choose between many non-optimal trees.
- We also need to develop a scenario bunching scheme. Doing this, we decide on the local area of the tree to examine.

References I

S. Ahmed, M. Tawarmalani, and N.V. Sahinidis. A finite branch-and-bound algorithm for two-stage stochastic integer programs. *Mathematical Programming*, 100(2):355–377, 2004.

C.C. Carøe and R. Schultz. Dual decomposition in stochastic integer programming. *Operations Research Letters*, 24(1):37–46, 1998.

C.C. Carøe and J. Tind. A cutting-plane approach to mixed 0-1 stochastic integer programs. *European Journal of Operational Research*, 101(2): 306–316, 1997.

C.C. Carøe and J. Tind. L-shaped decomposition of two-stage stochastic programs with integer recourse. *Mathematical Programming*, 83(1): 451–464, 1998.

Dinakar Gade, Simge Küçükyavuz, and Suvrajeet Sen. Decomposition algorithms with parametric gomory cuts for two-stage stochastic integer programs. *Mathematical Programming*, pages 1–26, 2012.

References II

N. Kong, A.J. Schaefer, and B. Hunsaker. Two-stage integer programs with stochastic right-hand sides: a superadditive dual approach. *Mathematical Programming*, 108(2):275–296, 2006.

G. Laporte and F.V. Louveaux. The integer l-shaped method for stochastic integer programs with complete recourse. *Operations research letters*, 13(3):133–142, 1993.

Lewis Ntaimo. Disjunctive decomposition for two-stage stochastic mixed-binary programs with random recourse. *Operations research*, 58 (1):229–243, 2010.

R. Schultz, L. Stougie, and M.H. Van Der Vlerk. Solving stochastic programs with integer recourse by enumeration: A framework using Gröbner basis. *Mathematical Programming*, 83(1):229–252, 1998.

S. Sen and J.L. Higle. The C 3 theorem and a D 2 algorithm for large scale stochastic mixed-integer programming: Set convexification. *Mathematical Programming*, 104(1):1–20, 2005. ISSN 0025-5610.

References III

S. Sen and H.D. Sherali. Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. *Mathematical Programming*, 106(2):203–223, 2006. ISSN 0025-5610.

Hanif D Sherali and J Cole Smith. Two-stage stochastic hierarchical multiple risk problems: models and algorithms. *Mathematical programming*, 120(2):403–427, 2009.

H.D. Sherali and B.M.P. Fraticelli. A modification of Benders' decomposition algorithm for discrete subproblems: An approach for stochastic programs with integer recourse. *Journal of Global Optimization*, 22(1):319–342, 2002.

H.D. Sherali and X. Zhu. On solving discrete two-stage stochastic programs having mixed-integer first-and second-stage variables. *Mathematical Programming*, 108(2):597–616, 2006.

Andrew C Trapp, Oleg A Prokopyev, and Andrew J Schaefer. On a level-set characterization of the value function of an integer program and its application to stochastic programming. *Operations Research*, 61 (2):498–511, 2013.

Yang Yuan and Suvrajeet Sen. Enhanced cut generation methods for decomposition-based branch and cut for two-stage stochastic mixed-integer programs. *INFORMS Journal on Computing*, 21(3): 480–487, 2009.