Duality, Multilevel Optimization, and Game Theory: Algorithms and Applications

Ted Ralphs[1]
Joint work with Sahar Tahernajad[1], Scott DeNegre[3],
Menal Güzelsoy[2], Anahita Hassanzadeh[4]

[1]COR@L Lab, Department of Industrial and Systems Engineering, Lehigh University [2]SAS Institute, Advanced Analytics, Operations Research R & D
[3]The Hospital for Special Surgery
[4]Climate Corp

Virginia Tech University, Blacksburg, Virginia, 12 Febrary 2017

Industrial and Systems Engineering

COMPUTATIONAL OPTIMIZATION RESEARCH AT LEHIGH

Outline

Outline

General Setting

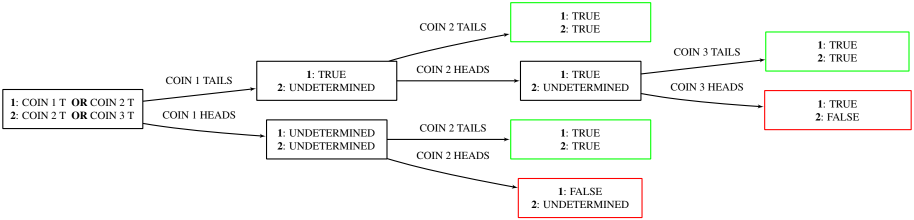
- In game theory terminology, the problems we address are known as *finite extensive-form games*, sequential games involving n players.

Loose Definition

- The game is specified on a tree with each node corresponding to a move and the outgoing arcs specifying possible choices.
- The leaves of the tree have associated payoffs.
- Each player's goal is to maximize payoff.
- There may be *chance* players who play randomly according to a probability distribution and do not have payoffs (*stochastic games*).

- All players are rational and have perfect information.
- The problem faced by a player in determining the next move is a *multilevel/multistage* optimization problem.
- The move must be determined by taking into account the *responses of the other players*.
- We are interested in games in which the number of options for each move is enormous, so we'll only be able to evaluate one or two moves.

Example Game Tree



Analyzing Games

- Categories
  - Multi-round vs. single-round
  - Zero sum vs. Non-zero sum
  - Winner take all vs. individual outcomes
- Goal of analysis
  - Find an equilibrium
  - Determine the optimal first move.

Multilevel and Multistage Games

- We use the term *multilevel* for competitive games in which there is no chance player.
- We use the term *multistage* for cooperative games in which all players receive the same payoff, but there are chance players.
- A *subgame* is the part of a game that remains after some moves have been made.

Stackelberg Game

- A Stackelberg game is a game with two players who make one move each.
- The goal is to find a *subgame perfect Nash equilibrium*, i.e., the move by each player that ensures that player's best outcome.

Recourse Game

- A cooperative game in which play alternates between cooperating players and chance players.
- The goal is to find a *subgame perfect Markov equilibrium*, i.e., the move that ensures the best outcome in a probabilistic sense.

Outline

Example: Coin Flip Game

Coin Flip Game

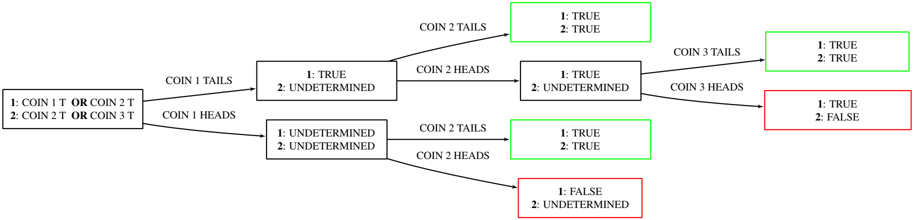
- k players take turns placing a set of coins heads or tails.
- In round i, player i places his/her coins.
- We have one or more logical expression that are of the form

  COIN 1 is heads OR *COIN 2 is tails* OR *COIN 3 is tails* OR ...
- With even (resp. odd) k, "even" (resp. "odd") players try to make all expressions true, while "odd" (resp. even) players try to prevent this.

Examples

- k = 1: Player looks for a way to place coins so that all expressions are true.
- k = 2: The first player tries to flip her coins so that no matter how the second player flips his coins, some expression will be false.
- k = 3: The first player tries to flip his coins such that the second player cannot flip her coins in a way that will leave the third player without any way to flip his coins to make the expressions true.

Coin Flip Game Tree

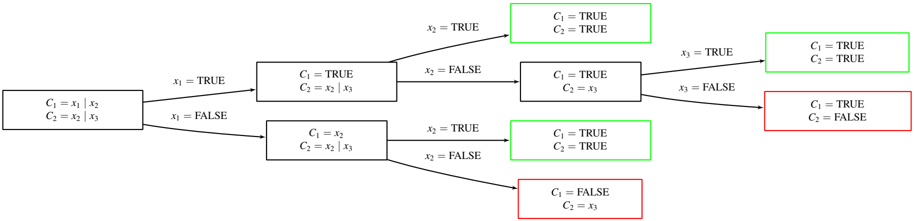


Example: Stochastic Variant

- The coin flip game can be modified to a recourse problem if we make the even player a "chance player".
- In this variant, there is only one "cognizant" player (the odd player) who first chooses heads or tails for an initial set of coins.
- The even player is a chance player who randomly flips some of the remaining coins.
- Finally, the odd player tries to flip the remaining coins so as to obtain a positive outcome.
- The objective of the odd player's first move could then be, e.g., to maximize the probability of a positive outcome across all possible scenarios.
- Note that we still need to know what happens in all scenarios in order to make the first move optimally.

The QBF Problem

- When expressed in terms of Boolean (TRUE/FALSE) variables, the problem is a special case of the so-called *quantified Boolean formula problem* (QBF).
- The case of $k = 1$ is the well-known Satisfiability Problem.
- This figure below illustrates the search for solutions to the problem as a tree.
- The nodes in green represent settings of the truth values that satisfy all the given clauses; red represents non-satisfying truth values.
  - With one player, the solution is any path to one of the green nodes.
  - With two players, the solution is a subtree in which there are no red nodes.
- The latter requires knowledge of *all* leaf nodes (important!).



Mathematical Optimization

- The general form of a *mathematical optimization problem* is:

Form of a General Mathematical Optimization Problem

$$\begin{array}{rrcll} z_{MP} = \min & f(x) & & & \\ \text{s.t.} & g_i(x) & \leq & b_i, & 1 \leq i \leq m \\ & x & \in & X & \end{array} \tag{MP}$$

where $X \subseteq \mathbb{R}^n$ may be a discrete set.

- The function f is the *objective function*, while g_i is the *constraint function* associated with constraint i.
- Our primary goal is to compute the optimal value z_{MP}.
- However, we may want to obtain some auxiliary information as well.
- More importantly, we may want to develop parametric forms of (MP) in which the input data are the output of some other function or process.

Multilevel and Multistage Optimization

- A (standard) mathematical optimization problem models a (set of) decision(s) to be made *simultaneously* by a *single* decision-maker (i.e., with a *single* objective).
- Decision problems arising in real-world sequential games can often be formulated as optimization problems, but they involve
  - multiple, independent decision-makers (DMs),
  - sequential/multi-stage decision processes, and/or
  - multiple, possibly conflicting objectives.
- Modeling frameworks
  - Multiobjective Optimization ⇐ multiple objectives, single DM
  - Mathematical Optimization with Recourse ⇐ multiple stages, single DM
  - Multilevel Optimization ⇐ multiple stages, multiple objectives, multiple DMs
- *Multilevel optimization* generalizes standard mathematical optimization by modeling hierarchical decision problems, such as finite extensive-form games.

From QBF to Multilevel Optimization

- For $k = 1$, SAT can be formulated as the (feasibility) integer program

$$\exists x \in \{0,1\}^n : \sum_{i \in C_j^0} x_i + \sum_{i \in C_j^1} (1 - x_i) \geq 1 \ \forall j \in J.\} \tag{SAT}$$

- (SAT) can be formulated as the optimization problem

$$\max_{x \in \{0,1\}^n} \alpha$$
$$\text{s.t.} \sum_{i \in C_j^0} x_i + \sum_{i \in C_j^1} (1 - x_i) \geq \alpha \ \forall j \in J$$

- For $k = 2$, we then have

$$\min_{x_{I_1} \in \{0,1\}^{I_1}} \max_{x_{I_2} \in \{0,1\}^{I_2}} \alpha$$
$$\text{s.t.} \sum_{i \in C_j^0} x_i + \sum_{i \in C_j^1} (1 - x_i) \geq \alpha \ \forall j \in J$$

How Difficult is the QBF?

- In general, we will focus on solving player one's decision problem, since this subsumes the solution of every other player's problem.
- No "efficient" algorithm exists for even the (single player) satisfiability problem.
- It is not surprising that the *k*-player satisfiability game is even more difficult (this can be formally proved).
  - The k^{th} player to move is faced with a satisfiability problem.
  - The $(k-1)^{th}$ player is faced with a 2-player subgame in which she must take into account the move of the k^{th} player.
  - And so on . . .
- Each player's decision problem appears to be exponentially more difficult than the succeeding player's problem.
- This complexity is captured formally in the hierarchy of so-called *complexity classes* known as the *polynomial time hierarchy*.

Roadmap for the Rest of the Talk

- We'll focus on simple games with two players (one of which may be a chance player) and two decision stages.
- We assume the determination of each player's move involves solution of an optimization problem.
- The optimization problem faced by the first player involves implicitly knowing what the second player's reaction will be to all possible first moves.
- The need for complete knowledge of the second player's possible reactions is what puts the complexity of these problems beyond that of standard optimization.

Outline

Brief Overview of Practical Applications

- Hierarchical decision systems
  - Government agencies
  - Large corporations with multiple subsidiaries
  - Markets with a single "market-maker."
  - Decision problems with recourse
- Parties in direct conflict
  - Zero sum games
  - Interdiction problems
- Modeling "robustness": Chance player is external phenomena that cannot be controlled.
  - Weather
  - External market conditions
- Controlling optimized systems: One of the players is a system that is optimized by its nature.
  - Electrical networks
  - Biological systems

Example: Tunnel Closures [Bruglieri et al., 2008]

- The EU wishes to close certain international tunnels to trucks in order to increase security.
- The response of the trucking companies to a given set of closures will be to take the shortest remaining path.
- Each travel route has a certain "risk" associated with it and the EU's goal is to minimize the riskiest path used after tunnel closures are taken into account.
- This is a classical Stackelberg game.

Example: Robust Facility Location [Snyder, 2006])

- We wish to locate a set of facilities, but we want our decision to be robust with respect to possible disruptions.
- The disruptions may come from natural disasters or other external factors that cannot be controlled.
- Given a set of facilities, we will operate them according to the solution of an associated optimization problem.
- Under the assumption that at most k of the facilities will be disrupted, we want to know what the worst case scenario is.
- This is a Stackelberg game in which the leader is not a cognizant DM.

Example: Fibrilation Ablation [Finta and Haines, 2004]

- Atrial fibrilation is a common form of heart arrhythmia that may be the result of impulse cycling within macroreentrant circuits.
- AF ablation procedures are intended to block these unwanted impulses from reaching the AV node.
- This is done by surgically removing some pathways.
- Since electrical impulses travel via the path of lowest resistance, we can model their flow using a mathematical optimization problem.
- If we wish to determine the least disruptive strategy for ablation, this is a Stackelberg game.
- In this case, the follower is not a cognizant DM.

Example: Electricity Network [Bienstock and Verma, 2008]

- As we know, electricity networks operate according to principles of optimization.
- Given a network, determining the power flows is an optimization problem.
- Suppose we wish to know the minimum number of links that need to be removed from the network in order to cause a failure.
- This, too, can be viewed as a Stackelberg game.
- Note that neither the leader nor the follower is a cognizant DM in this case.

Outline

Setting: Two-Stage Mixed Integer Optimization

- We have the following general formulation:

2SMILP

$$z_{\text{2SMILP}} = \min_{x \in \mathcal{P}_1 \cap X} \Psi(x) = \min_{x \in \mathcal{P}_1 \cap X} \left\{ c^\top x + \Xi(x) \right\}, \quad \text{(2SMILP)}$$

where

$$\mathcal{P}_1 = \left\{ x \in \mathbb{R}^{n_1} \mid A^1 x = b^1 \right\}$$

is the *first-stage feasible region* with $X = \mathbb{Z}_+^{r_1} \times \mathbb{R}_+^{n_1 - r_1}$, $A^1 \in \mathbb{Q}^{m_1 \times n_1}$, and $b^1 \in \mathbb{R}^{m_1}$.

- Ξ is a "risk function" that represents the impact of future uncertainty.
- We'll refer to Ξ as the *second-stage risk function*.
- The uncertainty can arise either due to stochasticity or due to the fact that Ξ represents the reaction of a competitor.

Special Case I: Bilevel (Integer) Linear Optimization

We first consider the following well-known class of optimization problem.

Mixed Integer Bilevel Linear Optimization Problem (MIBLP)

$$\min \left\{ cx + d^1 y \mid x \in \mathcal{P}_1 \cap X, y \in \operatorname{argmin}\{d^2 y \mid y \in \mathcal{P}_2(b^2 - A^2 x) \cap Y\} \right\}, \tag{MIBLP}$$

where $A^2 \in \mathbb{Q}^{m_2 \times n_1}$, and $b^2 \in \mathbb{R}^{m_2}$, $\mathcal{P}_2(\beta) = \left\{ y \in \mathbb{R}_+ \mid G^2 y \geq \beta \right\}$, and $Y = \mathbb{Z}^{p_2} \times \mathbb{R}^{n_2 - p_2}$.

This problem is equivalent to (2SMILP) with the following risk function.

Bilevel Risk Function

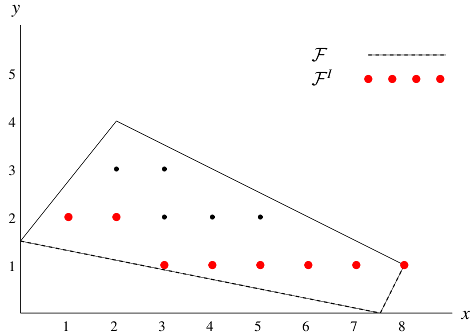
$$\Xi(x) = \min_{y \in \mathcal{P}_2(b^2 - A^2 x) \cap Y} \left\{ d^1 y \mid d^2 y = \phi(b^2 - A^2 x) \right\},$$

where ϕ is the so-called *second-stage value function* we'll define shortly.

Geometry of MIBLP

This well-known example from Moore and Bard [1990] illustrates the geometry of a simple MIBLP.

$$
\begin{aligned}
\max_{x \in X} \quad & x + 10y \\
\text{subject to} \quad & y \in \operatorname{argmin} \{ y : \quad -25x + 20y \leq 30 \\
& \qquad\qquad x + 2y \leq 10 \\
& \qquad\qquad 2x - y \leq 15 \\
& \qquad\qquad 2x + 10y \geq 15 \\
& \qquad\qquad y \in Y \, \}
\end{aligned}
$$



Special Case II: Recourse Problems

- Recourse problems are another special case in which the risk function has a different form.
- The canonical form of Ξ employed in the case of two-stage stochastic integer optimization is

Stochastic Risk Function

$$\begin{aligned} \Xi(x) &= \mathbb{E}_{\omega\in\Omega}\left[\phi(h_\omega - T_\omega x)\right] \\ &= \sum_{\omega\in\Omega} p_\omega \phi(h_\omega - T_\omega x), \end{aligned}$$

where ω is a random variable from a probability space $(\Omega, \mathscr{F}, \mathcal{P})$ with finite suport.

- For each $\omega \in \Omega$, $T_\omega \in \mathbb{Q}^{m_2\times n_1}$ and $h_\omega \in \mathbb{Q}^{m_2}$ is the realization of the input to the second-stage problem for scenario ω.
- ϕ is the value function of the *recourse problem*, to be defined shortly.

Other Special Cases

- Pure integer.
- Positive constraint matrix at second stage.
- Binary variables at the first and/or second stage.
- Zero sum and interdiction problems.

Mixed Integer Interdiction

$$\max_{x \in \mathcal{P}_1 \cap X} \min_{y \in \mathcal{P}_2(x) \cap Y} dy \tag{MIPINT}$$

where

$$\mathcal{P}_1 = \left\{ x \in X \mid A^1 x \leq b^1 \right\} \qquad X = \mathbb{B}^n$$

$$\mathcal{P}_2(x) = \left\{ y \in Y \mid G^2 y \geq b^2, y \leq u(e - x) \right\} \qquad Y = \mathbb{Z}^p \times \mathbb{R}^{n-p}$$

- The case where follower's problem has network structure is called the *network interdiction problem* and has been well-studied.
- The model above allows for second-stage systems described by general MILPs.

Economic Interpretation of Duality

- The economic viewpoint interprets the variables as representing possible *activities* in which one can engage at specific numeric levels.
- The constraints represent available *resources* so that $g_i(\hat{x})$ represents how much of resource i will be consumed at activity levels $\hat{x} \in X$.
- With each $\hat{x} \in X$, we associate a *cost* $f(\hat{x})$ and we say that $\hat{x}$ is *feasible* if $g_i(\hat{x}) \leq b_i$ for all $1 \leq i \leq m$.
- The space in which the vectors of activities live is the *primal space*.
- On the other hand, we may also want to consider the problem from the view point of the *resources* in order to ask questions such as
  - How much are the resources "worth" in the context of the economic system described by the problem?
  - What is the marginal economic profit contributed by each existing activity?
  - What new activities would provide additional profit?
- The *dual space* is the space of *resources* in which we can frame these questions.

Linear Optimization

- For this part of the talk, we focus on (single-level) mixed integer linear optimization problems (MILPs).

$$z_{IP} = \min_{x \in \mathcal{S}} c^\top x, \qquad \text{(MILP)}$$

where, $c \in \mathbb{R}^n$, $S = \{x \in \mathbb{Z}^r_+ \times \mathbb{R}^{n-r}_+ \mid Ax = b\}$ with $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{R}^m$.

- In this context, we can consider the concepts outlined previously more concretely.
- We can think of each row of A as representing a resource and each each as representing an activity or product.
- For each activity, resource consumption is a linear function of activity level.
- We first consider the case $r = 0$, which is the case of the (continuous) linear optimization problem (LP).

The LP Value Function

- Of central importance in duality theory for linear optimization is the *value function*, defined by

$$\phi_{LP}(\beta) = \min_{x \in \mathcal{S}(\beta)} c^\top x, \tag{LPVF}$$

  for a given $\beta \in \mathbb{R}^m$, where $\mathcal{S}(\beta) = \{x \in \mathbb{R}^n_+ \mid Ax = \beta\}$.

- We let $\phi_{LP}(\beta) = \infty$ if $\beta \in \Omega = \{\beta \in \mathbb{R}^m \mid \mathcal{S}(\beta) = \emptyset\}$.
- The value function returns the optimal value as a parametric function of the right-hand side vector, which represents available resources.

Economic Interpretation of the Value Function

- What information is encoded in the value function?
  - Consider the gradient $u = \phi'_{LP}(\beta)$ at β for which ϕ_{LP} is continuous.
  - The quantity $u^\top \Delta b$ represents the marginal change in the optimal value if we change the resource level by Δb.
  - In other words, it can be interpreted as a vector of the *marginal costs of the resources*.
  - For reasons we will see shortly, this is also known as the *dual solution vector*.
- In the LP case, the gradient is a *linear under-estimator* of the value function and can thus be used to derive bounds on the optimal value for any $\beta \in \mathbb{R}^m$.

Small Example: Fractional Knapsack Problem

- We are given a set $N = \{1, \ldots n\}$ of items and a capacity W.
- There is a profit p_i and a size w_i associated with each item $i \in N$.
- We want a set of items that maximizes profit subject to the constraint that their total size does not exceed the capacity.
- In this variant of the problem, we are allowed to take a fraction of an item.
- For each item i, let variable x_i represent the fraction selected.

Fractional Knapsack Problem

$$
\begin{aligned}
\min \sum_{j=1}^{n} p_j x_j & \\
\text{s.t.} \sum_{j=1}^{n} w_j x_j \leq W & \\
0 \leq x_i \leq 1 & \qquad \forall i
\end{aligned}
\tag{1}
$$

- What is the optimal solution?

Generalizing the Knapsack Problem

- Let us consider the value function of a (generalized) knapsack problem.
- To be as general as possible, we allow sizes, profits, and even the capacity to be negative.
- We also take the capacity constraint to be an equality.
- This is a proper generalization.

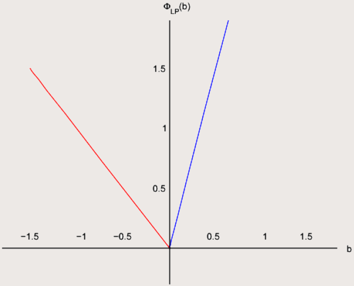
Example 1

$$\begin{aligned} \phi_{LP}(\beta) = \min\ & 6y_1 + 7y_2 + 5y_3 \\ s.t.\ & 2y_1 - 7y_2 + y_3 = \beta \\ & y_1, y_2, y_3, \in \mathbb{R}_+ \end{aligned}$$

Value Function of the (Generalized) Knapsack Problem

- Now consider the value function of the example from the previous slide.
- What do the gradients of this function represent?

Value Function for Example 1



The MILP Value Function

- We now generalize the notions seen so far to the MILP case.
- The *value function* associated with the base instance (MILP) is

MILP Value Function

$$\phi(\beta) = \min_{x \in \mathcal{S}(\beta)} c^\top x \tag{VF}$$

for $\beta \in \mathbb{R}^m$, where $\mathcal{S}(\beta) = \{x \in \mathbb{Z}^r_+ \times \mathbb{R}^{n-r}_+ \mid Ax = \beta\}$.

- Again, we let $\phi(\beta) = \infty$ if $\beta \in \Omega = \{\beta \in \mathbb{R}^m \mid \mathcal{S}(\beta) = \emptyset\}$.

Related Work on Value Function

Duality

- Johnson [1973, 1974, 1979]
- Jeroslow [1979]
- Wolsey [1981]
- Güzelsoy and R [2007], Güzelsoy [2009]

Structure and Construction

- Blair and Jeroslow [1977, 1982], Blair [1995]
- Kong et al. [2006]
- Hassanzadeh and R [2014b]

Sensitivity and Warm Starting

- R and Güzelsoy [2005, 2006], Güzelsoy [2009]
- Gamrath et al. [2015]

The (Mixed) Binary Knapsack Problem

- We now consider a further generalization of the previously introduced knapsack problem.
- In this problem, we must take some of the items either fully or not at all.
- In the example, we allow all of the previously introduced generalizations.

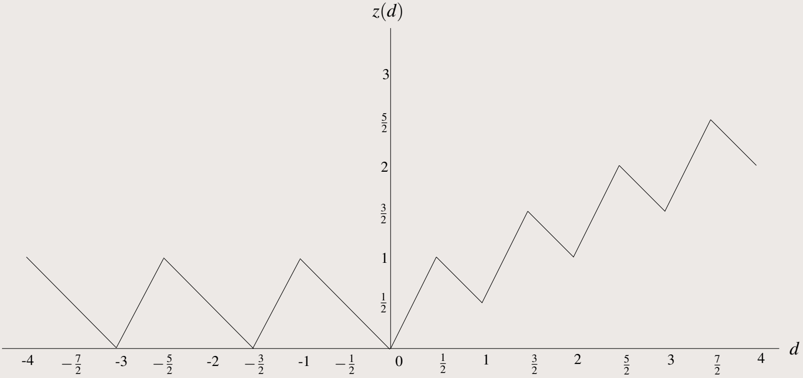
Example 2

$$\begin{aligned} \phi(\beta) = \min \quad & \tfrac{1}{2}x_1 + 2x_3 + x_4 \\ s.t \quad & x_1 - \tfrac{3}{2}x_2 + x_3 - x_4 = \beta \quad and \\ & x_1, x_2 \in \mathbb{Z}_+, x_3, x_4 \in \mathbb{R}_+. \end{aligned} \tag{2}$$

Value Function for (Generalized) Mixed Binary Knapsack

- Below is the value function of the optimization problem in Example 2.
- How do we interpret the structure of this function?

Value Function for Example 2



Points of Strict Local Convexity (Finite Representation)

Theorem 1 *[Hassanzadeh and R, 2014b]*
Under the assumption that $\{\beta \in \mathbb{R}^{m_2} \mid \phi_I(\beta) < \infty\}$ *is finite, there exists a finite set* $\mathcal{S} \subseteq Y$ *such that*

$$\phi(\beta) = \min_{x_I \in \mathcal{S}} \{c_I^\top x_I + \phi_C(\beta - A_I x_I)\},$$

where, for $I = \{1, \ldots, p_2\}$ *and* $C = \{p_2 + 1, \ldots, n_2\}$*, we have*

$$\begin{aligned} \phi_C(\beta) = \min\ & c_C^\top x_C \\ s.t.\ & A_C x_C = \beta, \\ & x_C \in \mathbb{R}_+^{n_2 - r_2} \end{aligned} \tag{CR}$$

and the similarly defined integer restriction:

$$\begin{aligned} \phi_I(\beta) = \min\ & c_I^\top x_I \\ s.t.\ & A_I x_I = \beta \\ & x_I \in \mathbb{Z}_+^{r_2} \end{aligned} \tag{IR}$$

Dual Problems

- A *dual function* $F : \mathbb{R}^m \rightarrow \mathbb{R}$ is one that satisfies $F(\beta) \leq \phi(\beta)$ for all $\beta \in \mathbb{R}^m$.
- The problem of finding a dual function for which $F(b) \approx \phi(b)$ is the *dual problem* associated with the base instance (MILP).

$$\max \left\{F(b) : F(\beta) \leq \phi(\beta),\ \beta \in \mathbb{R}^m, F \in \Upsilon^m\right\} \tag{D}$$

where $\Upsilon^m \subseteq \{f \mid f : \mathbb{R}^m \rightarrow \mathbb{R}\}$

- We call F^* *strong* for this instance if F^* is a *feasible* dual function and $F^*(b) = \phi(b)$.
- This dual instance always has a solution F^* that is strong if the value function is bounded and $\Upsilon^m \equiv \{f \mid f : \mathbb{R}^m \rightarrow \mathbb{R}\}$. Why?

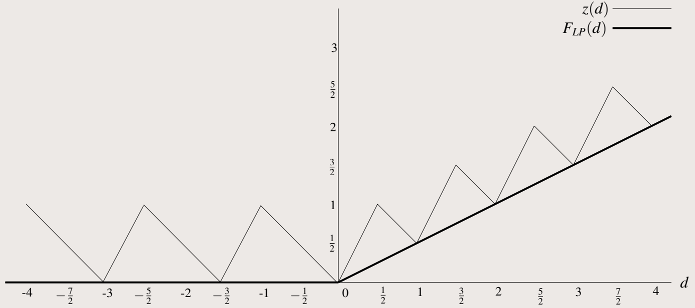
Example: LP Relaxation Dual Function

Example 3

$$F_{LP}(d) = \min \quad vd, \quad s.t \quad 0 \geq v \geq -\tfrac{1}{2}, \; and \quad v \in \mathbb{R}, \tag{3}$$

which can be written explicitly as

$$F_{LP}(\beta) = \begin{cases} 0, & \beta \leq 0 \\ -\frac{1}{2}\beta, & \beta > 0 \end{cases}.$$



What is the Importance in This Context?

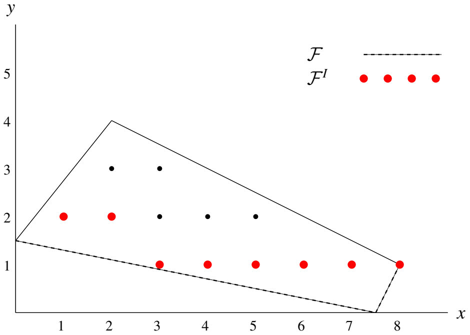
- The dual problem is important is because it gives us a set of *optimality conditions*.
- For a given $b \in \mathbb{R}^m$, whenever we have
  - $x^* \in \mathcal{S}(\beta) \cup X$,
  - $F \in \Upsilon^m$, and
  - $c^\top x^* = F(b)$,

  then x^* is optimal.
- This means we can write down a set of constraints involving the value function that ensure optimality.
- This set of constraints can then be embedded inside another optimization problem.

Outline

Value Function Reformulation [R, 2016]

More generally, we can reformulate (MIBLP) as



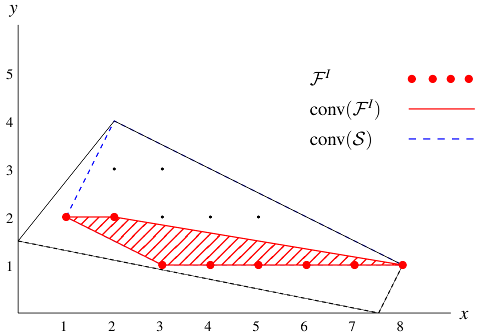
$$
\begin{aligned}
\min \quad & c^1 x + d^1 y \\
\text{subject to} \quad & A^1 x \leq b^1 \\
& G^2 y \geq b^2 - A^2 x \\
& d^2 y \leq \phi(b^2 - A^2 x) \\
& x \in X, y \in Y,
\end{aligned}
$$

where ϕ is the value function of the second-stage problem.

- This is, in principle, a standard mathematical optimization problem.
- Note that the second-stage variables need to appear in the formulation in order to enforce feasibility.

Polyhedral Reformulation [DeNegre and R, 2009]

Convexification considers the following conceptual reformulation.



$$\begin{aligned} \min \quad & c^1x + d^1y \\ \text{s.t.} \quad & (x,y) \in \text{conv}(\mathcal{F}^I) \end{aligned}$$

where $\mathcal{F}^I = \{(x,y) \mid x \in \mathcal{P}_1 \cap X, y \in \text{argmin}\{d^2y \mid y \in \mathcal{P}_2(x) \cap Y\}\}$

- To get bounds, we'll optimize over a relaxed feasible region.
- We'll iteratively approximate the true feasible region with linear inequalities.

Outline

Overview of Algorithms

- There are two main classes of algorithms

Dual

- Generalized Benders approach
- Approximate the value function from **below**.
- "Benders cuts" are (non-linear, non-convex) "dual functions".
- Can be combined with branching to get "local convexity".

Primal

- Generalized branch-and-cut approach
- As usual, convexify the feasible region and generate valid inequalities
- Approximate the value function from **above** with (linear) "optimality cuts".

- Naturally, we can also have hybrids.
- Any convergent algorithm for bilevel optimization must somehow construct an approximation of the value function, usually by intelligent "sampling."

Related Work On Bilevel Optimization

General Nonconvex

- Mitsos [2010]
- Kleniati and Adjiman [2014a,b]

Discrete Linear

- Moore and Bard [1990]
- DeNegre [2011], DeNegre and R [2009], DeNegre et al. [2016a]
- Xu [2012]
- Caramia and Mari [2013]
- Caprara et al. [2014]
- Fischetti et al. [2016]
- Hemmati and Smith [2016], Lozano and Smith [2016]

Related Work on Stochastic Optimization with Recourse

	First Stage			Second Stage			Stochasticity			
	$\mathbb{R}$	$\mathbb{Z}$	$\mathbb{B}$	$\mathbb{R}$	$\mathbb{Z}$	$\mathbb{B}$	$\mathbf{W}$	$\mathbf{T}$	$\mathbf{h}$	$\mathbf{q}$
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]		*	*		*	*			*	
Hassanzadeh and R [2014a]	*	*	*	*	*	*		*	*	

Generalized Benders [Hassanzadeh and R, 2014a]

Benders' Master Problem

$$\begin{aligned} \min \quad & c'x + w \\ \text{subject to} \quad & A'x \leq b' \\ & w \geq \underline{\Xi}(x) \\ & x \in X \end{aligned}$$

- $\underline{\Xi}$ is a lower approximation of the risk function Ξ.
- This lower approximation can be obtained, in turn from a lower approximation $\underline{\phi}$ of ϕ, as follows:

$$\underline{\Xi}(x) = \sum_{\omega \in \Omega} p_\omega \underline{\phi}(h_\omega - T_\omega x) \qquad \text{(U-2S-VF)}$$

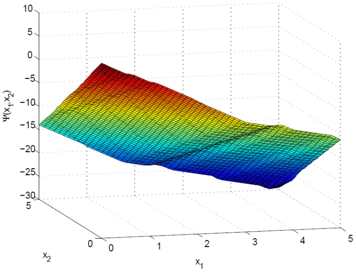
- In iteration t, we solve the master problem to obtain x^t.
- If $\underline{\phi}(h_\omega - T_\omega x^t) = \phi(h_\omega - T_\omega x^t)$, then x^t is optimal.
- Otherwise, we update $\underline{\phi}$ using information obtained while evaluating ϕ.

Example

Example 4

$$\begin{aligned} \min \Psi(x_1, x_2) = \min\ & -3x_1 - 4x_2 + \mathbb{E}[\phi(\omega - 2x_1 - 0.5x_2)] \\ s.t.\ & x_1 \leq 5, x_2 \leq 5 \\ & x_1, x_2 \in \mathbb{R}_+, \end{aligned} \tag{Ex.SMP}$$

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



Quick Example (cont'd)

where

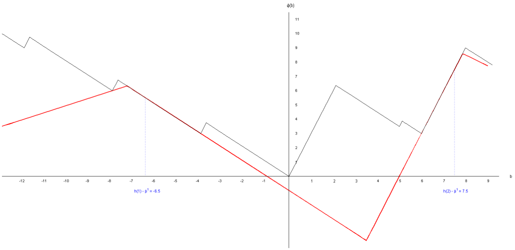
$$\phi(\beta) = \min\ 3y_1 + \frac{7}{2}y_2 + 3y_3 + 6y_4 + 7y_5$$
$$\text{s.t. } 6y_1 + 5y_2 - 4y_3 + 2y_4 - 7y_5 = \beta$$
$$y_1, y_2, y_3 \in \mathbb{Z}_+,\ y_4, y_5 \in \mathbb{R}_+ \tag{4}$$

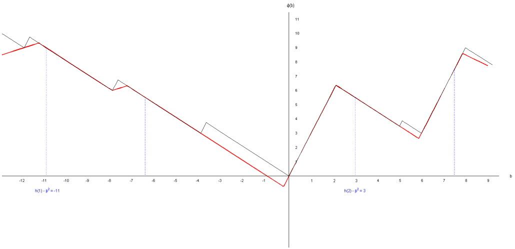
The master problem is

$$\min\ -3x_1 - 4x_2 + \sum_{\omega=1}^{2} 0.5\underline{\phi}(2x_1 - 0.5x_2)$$
$$x_1 \leq 5, x_2 \leq 5$$
$$x \in \mathbb{Z}_+ \tag{5}$$

and $\underline{\phi}$ looks as follows in the first two iterations.

Example





Outline

Algorithm for General MIBLP [DeNegre et al., 2016a]

- The second major class of algorithms take a "primal" approach.
- An important tool will be *convexification* by which we obtain convex relaxations that can be used for bounding.
- The value function of the second-stage problem still plays a role here, but we generally won't bound it globally.
- We propose a branch-and-bound approach.

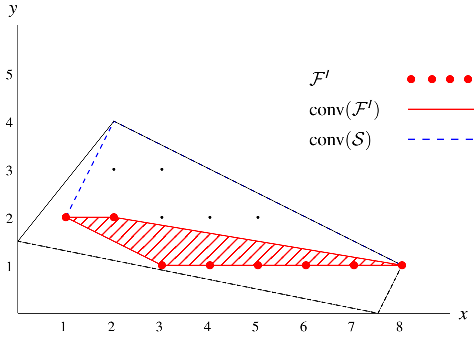
Components of Branch and Cut

- Bounding
- Branching
- Feasibility checking
- Search strategies
- Preprocessing methods
- Primal heuristics

- In the remainder of the talk, we address development of these components, focusing mainly on bounding.

Convexification

Convexification considers the following conceptual reformulation.



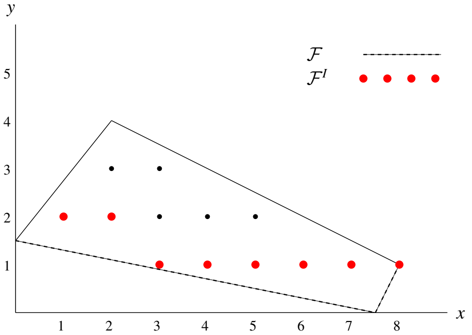
$$\begin{aligned} \min \quad & c^1 x + d^1 y \\ \text{s.t.} \quad & (x, y) \in \text{conv}(\mathcal{F}^I) \end{aligned}$$

where $\mathcal{F}^I = \{(x, y) \mid x \in \mathcal{P}_1 \cap X, y \in \text{argmin}\{d^2 y \mid y \in \mathcal{P}_2(x) \cap Y\}\}$

- To get bounds, we'll optimize over a relaxed feasible region.
- We'll iteratively approximate the true feasible region with linear inequalities.

Dual Bounds

Dual bounds for the MIBLP can be obtained by relaxing the value function constraint.



$$
\begin{aligned}
\min \quad & c^1x + d^1y \\
\text{subject to} \quad & A^1x \leq b^1 \\
& G^2y \geq b^2 - a^2x \\
& H^ix + H^2y \leq h \\
& x \in X, y \in Y,
\end{aligned}
$$

- Note that in practice, we may further relax integrality conditions.
- The additional inequalities are valid inequalities that serve to approximate the value function.
- The algorithm is very similar to branch-and-cut for solving traditional mathematical optimization problems.

Bilevel Feasibility Check

- Let $(\hat{x}, \hat{y}) \in X \times Y$ be a solution to the dual bounding relaxation problem.
- We fix $x = \hat{x}$ and solve the second-stage problem

$$\min_{y \in \mathcal{P}_2(\hat{x}) \cap X} d^2 y \tag{6}$$

with the fixed first-stage solution $\hat{x}$.
- Let y^* be the solution to (6).
  - $(\hat{x}, y^*)$ is bilevel feasible $\Rightarrow c^1\hat{x} + d^1 y^*$ is a valid primal bound on the optimal value of the original MIBLP
  - Either
    1. $d^2\hat{y} = d^2 y^* \Rightarrow (\hat{x}, \hat{y})$ is bilevel feasible.
    2. $d^2\hat{y} > d^2 y^* \Rightarrow (\hat{x}, \hat{y})$ is bilevel infeasible.
- What do we do in the case of bilevel infeasibility?
  - Generate a valid inequality violated by $(\hat{x}, \hat{y})$ (improve our approximation of the value function).
  - Branch on a disjunction violated by $(\hat{x}, \hat{y})$.

Optimality Cuts

- Strong cuts can be obtained by exploiting the **bound information obtained during the feasibility check**.
- Implicitly, we will impose the constraint

$$d^2 y \leq \bar{\phi}(b^2 - A^2 x)$$

by adding a set of linear cuts (which may be locally or globally valid).

- In order to accomplish this, we need to do it in tandem with branching—impose cuts that are locally valid to overcome nonconvexity.
- After checking bilevel feasibility of $(x, y) \in (\mathcal{P}_1 \cap X) \times (\mathcal{P}_2(x) \cap Y)$, we know that

$$\hat{y} \in \mathcal{P}_2(x) \Rightarrow d^2 y \leq d^2 \hat{y}$$

- There are a number of ways to impose this logic.
  - Generate intersection cuts [Fischetti et al., 2016].
  - Impose the logic with integer variables [Mitsos, 2010]

Branching Scheme

- The branching scheme is similar to that in the MILP case except that we branch only on first-stage variables that appear in the second-stage problem.
- This is because once these variables are fixed, the problem reduces to a standard MILP.
- This may require branching on variables with integer values.

SYMPHONY and MibS

- The *Mixed Integer Bilevel Solver* (MibS) is a solver for bilevel integer programs available open source on github (`http://github.com/tkralphs/MibS`).
- It depends on a number of other projects available through the COIN-OR repository (`http://www.coin-or.org`).

COIN-OR Components Used

- The COIN High Performance Parallel Search (CHiPPS) framework to manage the global branch and bound.
- The SYMPHONY framework for checking bilevel feasibility..
- The COIN LP Solver (CLP) framework for solving the LPs arising in the branch and cut.
- The Cut Generation Library (CGL) for generating cutting planes within both SYMPHONY and MibS itself.
- The Open Solver Interface (OSI) for interfacing with SYMPHONY and CLP.

- SYMPHONY implements the procedures for constructing and exporting dual functions from branch and bound.

Conclusions

- This general class of problems is extremely challenging computationally.
- There are special cases (interdiction/zerp sum) that are substantially easier and much progress has been made on solving these.
- Both the theory and methodology for the general case is maturing slowly.
- Many of the computational tools necessary for experimentation now also exist.
- We are currently focusing on the general case, developing the software and methodology necessary.
- There are many avenues for contribution, so please join us!

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.

D. Bienstock and A. Verma. The n-k problem in power grids: New models, formulations and computation. Available at `http://www.columbia.edu/~dano/papers/nmk.pdf`, 2008.

C.E. Blair. A closed-form representation of mixed-integer program value functions. *Mathematical Programming*, 71:127–136, 1995.

C.E. Blair and R.G. Jeroslow. The value function of a mixed integer program: I. *Discrete Mathematics*, 19:121–138, 1977.

C.E. Blair and R.G. Jeroslow. The value function of an integer program. *Mathematical Programming*, 23:237–273, 1982.

M. Bruglieri, R. Maja, G. Marchionni, and P.L. Da Vinci. Safety in hazardous material road transportation: State of the art and emerging problems. In *Advanced Technologies and Methodologies for Risk Management in the Global Transport of Dangerous Goods*, chapter 6, pages 88–129. IOS Press, 2008.

References II

A. Caprara, M. Carvalho, A. Lodi, and G.J. Woeginger. Bilevel knapsack with interdiction constraints. Technical Report OR-14-4, University of Bologna, 2014.

M. Caramia and R. Mari. Enhanced exact algorithms for discrete bilevel linear problems. *Optimization Letters*, ??:??–??, 2013.

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.

S DeNegre. *Interdiction and Discrete Bilevel Linear Programming*. Phd, Lehigh University, 2011. URL `http://coral.ie.lehigh.edu/{~}ted/files/papers/ScottDeNegreDissertation11.pdf`.

References III

S DeNegre and T K R. A branch-and-cut algorithm for bilevel integer programming. In *Proceedings of the Eleventh INFORMS Computing Society Meeting*, pages 65–78, 2009. doi: 10.1007/978-0-387-88843-9_4. URL `http://coral.ie.lehigh.edu/~ted/files/papers/BILEVEL08.pdf`.

S.T. DeNegre, T.K. R, and S.A. Tahernejad. Mibs: An open source solver for mixed integer bilevel optimization problems. Technical report, COR@L Laboratory, Lehigh University, 2016a.

S.T DeNegre, T.K R, and S.A. Tahernejad. Mibs version 0.9, 2016b.

B. Finta and D.E Haines. Catheter ablation therapy for atrial fibrillation. *Cardiology Clinics*, 22(1):127–145, 2004.

M. Fischetti, I. Ljubic, M. Monaci, and M. Sinnl. Intersection cuts for bilevel optimization. In *Proceedings of the 18th Conference on Integer Programming and Combinatorial Optimization*, 2016.

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 IV

G. Gamrath, B. Hiller, and J. Witzig. Reoptimization techniques for mip solvers. In *Proceedings of the* 14^{th} *International Symposium on Experimental Algorithms*, 2015.

M Güzelsoy. *Dual Methods in Mixed Integer Linear Programming*. Phd, Lehigh University, 2009. URL `http://coral.ie.lehigh.edu/{~}ted/files/papers/MenalGuzelsoyDissertation09.pdf`.

M Güzelsoy and T K R. Duality for mixed-integer linear programs. *International Journal of Operations Research*, 4:118–137, 2007. URL `http://coral.ie.lehigh.edu/~ted/files/papers/MILPD06.pdf`.

A Hassanzadeh and T K R. A generalized benders' algorithm for two-stage stochastic program with mixed integer recourse. Technical report, COR@L Laboratory Report 14T-005, Lehigh University, 2014a. URL `http://coral.ie.lehigh.edu/~ted/files/papers/SMILPGenBenders14.pdf`.

References V

A Hassanzadeh and T K R. On the value function of a mixed integer linear optimization problem and an algorithm for its construction. Technical report, COR@L Laboratory Report 14T-004, Lehigh University, 2014b. URL `http://coral.ie.lehigh.edu/~ted/files/papers/MILPValueFunction14.pdf`.

Mehdi Hemmati and Cole Smith. A mixed integer bilevel programming approach for a competitive set covering problem. Technical report, Clemson University, 2016.

Robert G Jeroslow. Minimal inequalities. *Mathematical Programming*, 17(1):1–15, 1979.

Ellis L Johnson. Cyclic groups, cutting planes and shortest paths. *Mathematical programming*, pages 185–211, 1973.

Ellis L Johnson. On the group problem for mixed integer programming. In *Approaches to Integer Programming*, pages 137–179. Springer, 1974.

Ellis L Johnson. On the group problem and a subadditive approach to integer programming. *Annals of Discrete Mathematics*, 5:97–112, 1979.

References VI

P. Kleniati and C. Adjiman. Branch-and-sandwich: A deterministic global optimization algorithm for optimistic bilevel programming problems. part i: Theoretical development. *Journal of Global Optimization*, 60:425–458, 2014a.

P. Kleniati and C. Adjiman. Branch-and-sandwich: A deterministic global optimization algorithm for optimistic bilevel programming problems. part ii: Convergence analysis and numerical results. *Journal of Global Optimization*, 60: 459–481, 2014b.

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.

L. Lozano and J.C. Smith. A backward sampling framework for interdiction problems with fortification. Technical report, Clemson University, 2016.

References VII

A. Mitsos. Global solution of nonlinear mixed integer bilevel programs. *Journal of Global Optimization*, 47:557–582, 2010.

J.T. Moore and J.F. Bard. The mixed integer linear bilevel programming problem. *Operations Research*, 38(5):911–921, 1990.

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

T K R and M Güzelsoy. The symphony callable library for mixed-integer linear programming. In *Proceedings of the Ninth INFORMS Computing Society Conference*, pages 61–76, 2005. doi: 10.1007/0-387-23529-9_5. URL `http://coral.ie.lehigh.edu/~ted/files/papers/SYMPHONY04.pdf`.

T K R and M Güzelsoy. Duality and warm starting in integer programming. In *The Proceedings of the 2006 NSF Design, Service, and Manufacturing Grantees and Research Conference*, 2006. URL `http://coral.ie.lehigh.edu/~ted/files/papers/DMII06.pdf`.

T K R, M Guzelsoy, and A Mahajan. Symphony version 5.6, 2016.

References VIII

T.K. R. Multistage discrete optimization. In S Dempe, V Kalashnikov, and B Mordukhovich, editors, *Bilevel Programming: Theory and Applications*, chapter Multistage. Bantham Science Publishers, 2016.

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 C3 theorem and a D2 algorithm for large scale stochastic mixed-integer programming: Set convexification. *Mathematical Programming*, 104 (1):1–20, 2005. ISSN 0025-5610.

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.

References IX

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.

L. V. Snyder. Facility location under uncertainty: A review. *IIE Transactions*, 38(7): 537–554, 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.

L.A. Wolsey. Integer programming duality: Price functions and sensitivity analysis. *Mathematical Programming*, 20(1):173–195, 1981. ISSN 0025-5610.

P. Xu. *Three Essays on Bilevel Optimization and Applications*. PhD thesis, Iowa State University, 2012.

References X

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.