An Algorithm for Two-stage Stochastic Programs with Mixed Integer Recourse

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

[1,2]Lehigh University, ISE Dept.

[3]SAS Institute

INFORMS Annual Meeting, October 2012

Outline

1. Introduction
2. MILP Value Function
3. MILP Duality
4. Integer Benders' Algorithm
5. Final Remarks

The Two-Stage ProblemProblem Formulation

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

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

where $X = \mathbb{R}_+^{n_1}$ and $Y = \mathbb{R}_+^{n_2}$.

We assume w follows a discrete distribution with a finite support.

The Big Picture

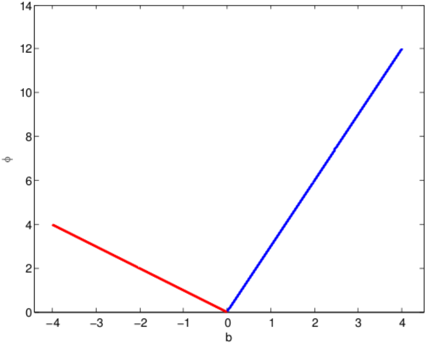
- Current methods cannot manage to solve the general mixed integer linear two-stage stochastic problem in practice.
- We believe a structured examination of the MILP value function can help to tackle the two-stage MILPs.
- The two-stage problem comes down to studying the optimal solution's behavior under some parameter perturbation. This is what value function provides.

LP Value function

Consider:

Example 1

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



LP Value function Structure In general:

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

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

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

or equivalently

$$\mathcal{L} := \{p : \nu_i^\top p \geq 0,\ \forall i = 1 \dots k\}$$

where $\nu_i^\top = c_i^\top A_i^{-1}, i = 1 \dots k$.

MILP Value function Consider the general MILP value function $\phi : \Lambda \rightarrow \mathbb{R} \cup \{\pm\infty\}$

$$\begin{aligned} \phi(b) = \min\ & c^\top x \\ \text{s.t. }\ & Ax = b \\ & x \in \mathbb{Z}_+^r \times \mathbb{R}_+^{n-r} \end{aligned} \tag{MILP}$$

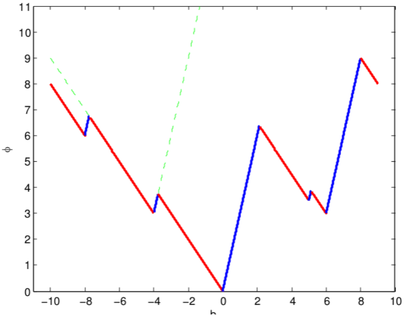
Let $\mathcal{S}(b) = \{x \in \mathbb{Z}_+^r \times \mathbb{R}_+^{n-r} : Ax = b\}$.
Also, $B = \{b : S(b) \neq \emptyset\}$.

Example: MILP Value function MILP Value function is in general a non-convex and discontinuous piecewise polyhedral function.

Example: MILP Value function MILP Value function is in general a non-convex and discontinuous piecewise polyhedral function.

Example 2

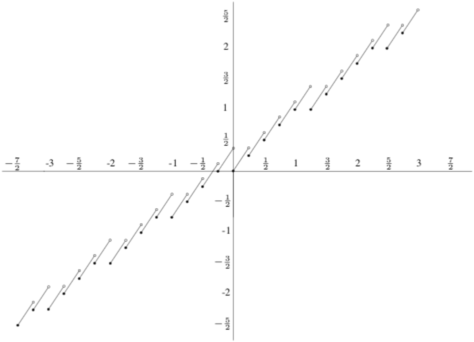
$$\phi(b) = \min\ 3x_1 + \frac{7}{2}x_2 + 3x_3 + 6x_4 + 7x_5 + 5x_6$$
$$\text{s.t. } 6x_1 + 5x_2 - 4x_3 + 2x_4 - 7x_5 + x_6 = b \qquad \text{(Ex1.MILP)}$$
$$x_1, x_2, x_3 \in \mathbb{Z}_+,\ x_4, x_5, x_6 \in \mathbb{R}_+$$



Example: MILP Value function

Example 3

$$\phi(b) = \min\ x_1 - \frac{3}{4}x_2 + \frac{3}{4}x_3$$
$$\text{s.t. } \frac{5}{4}x_1 - x_2 + \frac{1}{2}x_3 = b \qquad \text{(Ex2.MILP)}$$
$$x_1, x_2 \in \mathbb{Z}_+,\ x_3 \in \mathbb{R}_+$$



Linear and Integer Restriction of an MILP Consider

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

Linear 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 linear restriction of (**??**) 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 VF

- Define $\mathcal{S}_D = \{\nu : A_C^\top \nu \le c_C\}$.
- Let $\mathcal{E}$ denote the set of indices of the nonsingular square submatrices of A_C corresponding to dual feasible bases. Then $E \in \mathcal{E} \Leftrightarrow \exists \nu_E : \nu_E^\top = c_E^\top A_E^{-1}$
- We already know ϕ_C is a convex function that is fully defined by the set of ν_E, $E \in \mathcal{E}$.

Discrete Representation of VF

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

$$\begin{aligned} \phi(b) = \min\ & c_I x_I + \phi_C(b - A_I x_I) \\ \text{s.t. } & x_I \in \mathbb{Z}_+^r \end{aligned} \tag{1}$$

Note that for a fixed integer vector x_I, the value function is a transition of ϕ_C.

We aim to find the exact configuration of these transitions.

Properties of MILP Value Function

Proposition 2.1

The gradient of ϕ on a neighborhood of a differentiable point is a unique optimal dual feasible solution to (**??**).

Proposition 2.2

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

Properties of MILP Value Function

- Consider $epi(\phi_C)$.

Properties of MILP Value Function

- Consider *epi*(ϕ_C).
- We introduce points of an MILP value function reminiscent to the extreme point of *epi*(ϕ_C):

Definition 2.1

A point $\hat{b}$ is called a point of strict local convexity *of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ if for some $\epsilon > 0$ and $g \in \partial f(\hat{b})$*

$$f(b) > f(\hat{b}) + g^\top (b - \hat{b}) \text{ for all } b \in \mathcal{N}_\epsilon(\hat{b}),\ b \neq \hat{b}$$

Properties of MILP Value Function

- Consider $epi(\phi_C)$.
- We introduce points of an MILP value function reminiscent to the extreme point of $epi(\phi_C)$:

Definition 2.1

A point $\hat{b}$ is called a point of strict local convexity *of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ if for some $\epsilon > 0$ and $g \in \partial f(\hat{b})$*

$$f(b) > f(\hat{b}) + g^\top (b - \hat{b}) \text{ for all } b \in \mathcal{N}_\epsilon(\hat{b}),\ b \neq \hat{b}$$

- Note that the extreme point of $epi(\phi_C)$ is the only point of strict local convexity of ϕ_C.

Properties of MILP Value Function

- For a given $\hat{x}$, let $\phi_C(b, \hat{x}) = \phi_C(b - A_I\hat{x}_I)\ \forall b \in \mathbb{R}^m$
- Similarly, for given $\hat{x}$, let $\mathcal{P}(\hat{x})$ denote the strict point of convexity of $\phi_C(., \hat{x})$.

Define $\mathcal{P}$ as the set containing the strict local convexity points of ϕ.

Proposition 2.3

$\mathcal{P} \subseteq \cup_{x_I \in proj_z(S)} \mathcal{P}(x_I)$.

Properties of MILP Value Function

- For a given $\hat{x}$, let $\phi_C(b, \hat{x}) = \phi_C(b - A_I \hat{x}_I)\ \forall b \in \mathbb{R}^m$
- Similarly, for given $\hat{x}$, let $\mathcal{P}(\hat{x})$ denote the strict point of convexity of $\phi_C(., \hat{x})$.

Define $\mathcal{P}$ as the set containing the strict local convexity points of ϕ.

Proposition 2.3

$\mathcal{P} \subseteq \cup_{x_I \in proj_z(S)} \mathcal{P}(x_I)$.

Now how to generate points in $\mathcal{P}$?

Properties of MILP Value FunctionThe Jeroslow Formula

- Let:
  $\lfloor b \rfloor_E = A_E \left\lfloor A_E^{-1} b \right\rfloor, \ \mathcal{T}_E = \{b : A_E^{-1} b \in \mathbb{Z}^m\}, \ \mathcal{T} = \bigcap_{E \in \mathcal{E}} \mathcal{T}_E$

Jeroslow Formula (?)

The value function (**??**) can be constructed by

$$\phi(b) = \min_{E \in \mathcal{E}} G(\lfloor b \rfloor_E) + \nu_E^\top (b - \lfloor b \rfloor_E)$$

where $M \in \mathbb{Z}_+$ such that $MA_E^{-1} A_I^j$ is a vector of integers for all $E \in \mathcal{E}, j = 1 \ldots r$, and G is the value function of a related pure integer problem.

Properties of MILP Value Function From (**?**), we know for $b \in \mathcal{T}$ it suffices to solve:

$$\begin{aligned} \min\ & c_I^\top x_I + \frac{1}{M} c_C^\top x_C \\ \text{s.t. } & A_I x_I + \frac{1}{M} A_C x_C = b \\ & (x_I, x_C) \in \mathbb{Z}_+^r \times \mathbb{Z}_+^{n-r} \end{aligned} \tag{2}$$

Theorem 2.1

$\mathcal{P} \subseteq \mathcal{T}$

Theorem 2.2

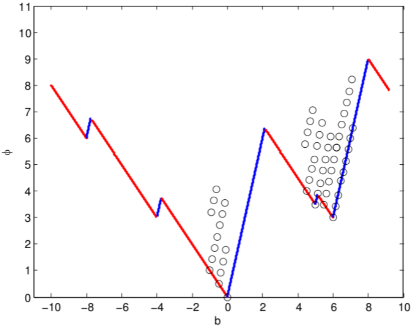
The solution to $\phi(b)$ with $b \in \mathcal{P}$ is the solution to a pure integer problem; that is, $x_C \in \mathbb{Z}^{n-r}$ in the optimal solution.

Scaled IP Counterpart of an MILP

For the example:

$$\begin{aligned} \phi(b) = \min\ & 3x_1 + \frac{7}{2}x_2 + 3x_3 + \frac{3}{7}x_4 + \frac{1}{2}x_5 \\ \text{s.t. } & 6x_1 + 5x_2 - 4x_3 + \frac{1}{7}x_4 - \frac{1}{2}x_5 = b \\ & x_1, x_2, x_3, x_4, x_5 \in \mathbb{Z}_+ \end{aligned} \qquad \text{(Ex.Scaled)}$$

The solutions to the above problem for $\{[-1,0] \cup [4.4, 7.1]\}$



Properties of MILP Value Function

Theorem 2.3

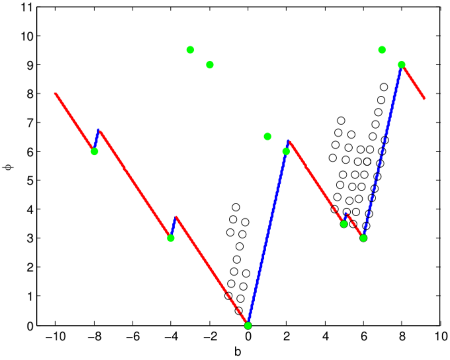
There exists an integer solution $x_I \in \mathbb{Z}^r$ *to* $\phi(b)$ *with* $b \in \mathcal{P}$ *such that* $A_I x_I = b$.

Theorem 2.4

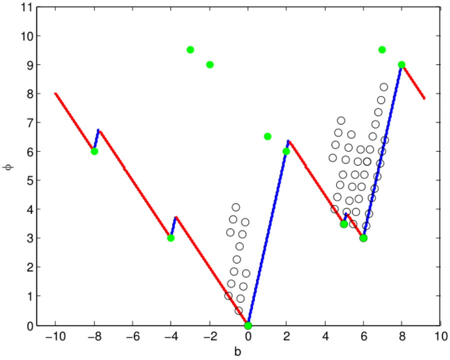
The solutions to the problems $\phi_I(b)$ *and* $\phi(b)$ *are equal if* $b \in \mathcal{P}$, *where*

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

Integer Restriction of an MILP



Integer Restriction of an MILP



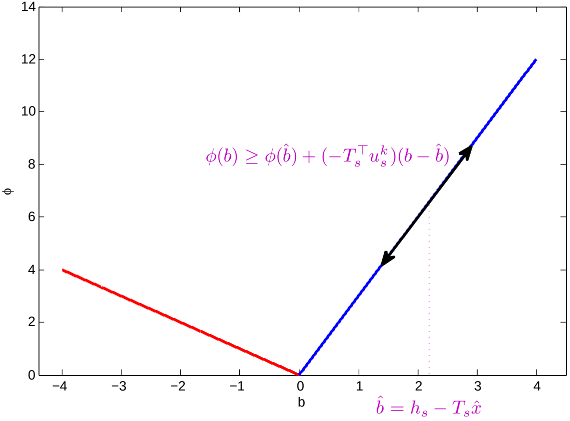
Theorem 2.5

The set of strict local minima of ϕ is a subset of $\mathcal{P}$.

Two-Stage Problem and Value Function

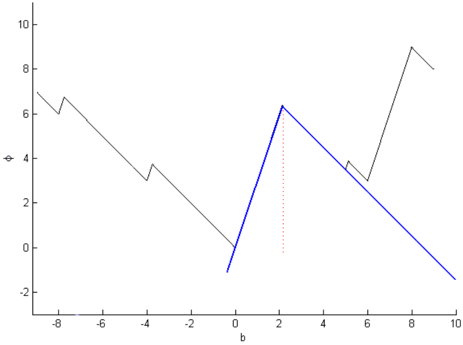
- When Y is continuous, the two-stage problem is solved by Benders' decomposition.
- This requires approximating $\mathbb{E}_{w \in \Omega}[Q(x, w)]$.
- This is due to the fact that the value function of an LP is convex and piece-wise polyhedral.

Optimality Cuts



Two-Stage Problem and Value Function

- Benders' method does not apply when Y contains integer variables.
- To generalize it, we need lower bounding functions to approximate the MILP value function.



Dual Functions

A dual function $\varphi : \Lambda \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is

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

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

$$\phi_D = max\{\varphi(\hat{b}) : \varphi(b) \leq \phi(b) \quad \forall b \in \Lambda, \; \varphi : \Lambda \rightarrow \mathbb{R} \cup \{\pm\infty\}\}$$

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:

$$\begin{aligned} \min\ & c^\top x \\ \text{s.t. } & Ax = b, \\ & l^t \leq x \leq u^t, x \geq 0 \end{aligned} \tag{3}$$

The dual at node t:

$$\begin{aligned} \max\ & \{\pi^t b + \underline{\pi}^t l^t + \bar{\pi}^t u^t\} \\ \text{s.t. } & \pi^t A + \underline{\pi}^t + \bar{\pi}^t \leq c^\top \\ & \underline{\pi} \geq 0, \bar{\pi} \leq 0 \end{aligned} \tag{4}$$

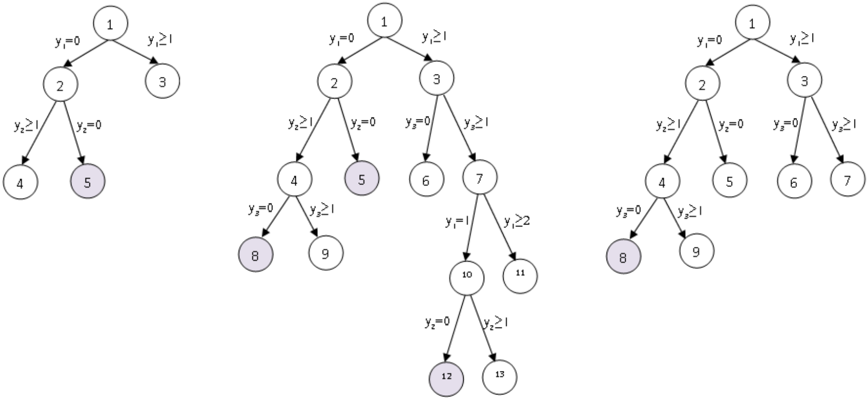
By-product of a B&B:

$$\min_{t \in T} \{\pi^t b + \underline{\pi}^t l^t + \bar{\pi}^t u^t\} \tag{5}$$

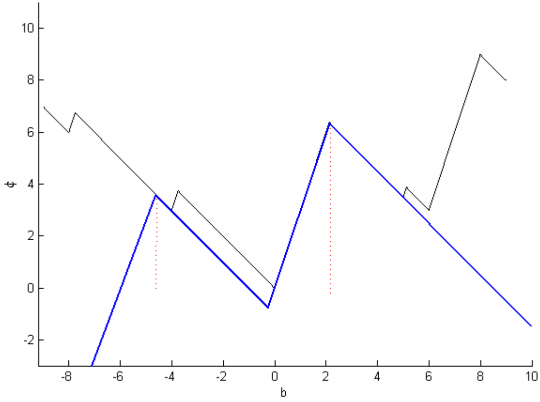
Strong at the given right hand side and a valid dual function.

MILP Duals from Branch-and-Bound

Figure: Dual Functions from B&B for right hand sides 1, 2.125, 3.5



MILP Duals from Branch-and-Bound



Current Algorithms for the Two-Stage Problem

- Modification to the L-shaped framework (**??**)
  - Linear cuts in first stage for binary first stage
  - Disjunctive programming approaches and cuts in the second stage
- Value function approaches: Pure integer case (**??**)
- Scenario decomposition (**?**)
- Enumeration/Gröbner basis reduction (**?**)

Value Function Reformulation of the Two-Stage Problem

$$
\begin{aligned}
f(\beta) &= \{\psi(\beta) + \min \mathbb{E}_s[\phi(h_s - \beta)] : \beta \in \mathcal{B}\} \\
\psi(\beta) &= \min\{c^\top x : x \in S_1(\beta)\} \\
S_1(\beta) &= \{Tx = \beta,\ x \in X\} \\
\phi(\beta) &= \{q^\top y : y \in S_2(\beta)\} \\
S_2(\beta) &= \{Wy = \beta,\ y \in \mathbb{Z}_+^{r_2} \times \mathbb{R}_+^{n_2 - r_2}\} \\
\mathcal{B} &= \{\beta : \beta = Tx, x \in X\}
\end{aligned}
\qquad \text{(SP.VF)}
$$

Assumptions:

i) q, T, and W are fixed.

ii) $pos(W_C) = \mathbb{R}^{m_2}$

iii) The dual of the LP relaxation of the recourse problem is feasible. That is, $\{\nu \in \mathbb{R}^{m_2} : W_I^\top \nu \leq q_I, W_C^\top \nu \leq q_C\} \neq \emptyset$

iv) X is non-empty and compact.

Step 0. Initialize

a) Set $\beta^1 = Tx^1$ where $x^1 \in \text{argmin}\{Ax \geq b,\ x \in \mathbb{Z}^{r_1} \times \mathbb{R}^{n_1 - r_1}\}$

b) Initialize the dual function lists $\mathcal{F}_1 = \emptyset, \mathcal{F}_s = \emptyset$.

c) Set $k = 1$.

Step 0. Initialize

a) Set $\beta^1 = Tx^1$ where $x^1 \in \text{argmin}\{Ax \geq b,\ x \in \mathbb{Z}^{r_1} \times \mathbb{R}^{n_1 - r_1}\}$

b) Initialize the dual function lists $\mathcal{F}_1 = \emptyset, \mathcal{F}_s = \emptyset$.

c) Set $k = 1$.

Step 1. Lower bound the problem and check for termination

a) Find optimal dual functions F_1^k and F_s^k for each $s \in 1 \ldots S$ to $\psi(\beta^k)$ and $\phi(h_s - \beta^k)$ respectively.

b) If

$$\max_{f_1 \in \mathcal{F}_1, f_s \in \mathcal{F}_s} \{f_1(\beta^k) + \mathbb{E}[f_s(h_s - \beta^k)]\} = F_1^k(\beta^k) + \mathbb{E}_s[F_s^k(h_s - \beta^k)]$$

then stop, $x^* = \text{argmin}\{c^\top x : Ax \geq b, Tx = \beta^k,\ x \in \mathbb{Z}^{r_1} \times \mathbb{R}^{n_1 - r_1}\}$ is an optimal solution.

Integer Benders' Algorithm

Step 2. Update the lower bound

a) Update the dual functions lists: $\mathcal{F}_1 = \mathcal{F}_1 \cup F_1^k$ and let $\mathcal{F}_s = \mathcal{F}_s \cup_{s \in \Omega} F_s^k$.

b) Solve the problem

$$z^k = \min_{\beta \in B} \max_{f_1 \in \mathcal{F}_1, f_s \in \mathcal{F}_s} \{ f_1(\beta^k) + \mathbb{E}_s[f_s(h_s - \beta^k)] \}$$

and set its optimal solution to β^{k+1}.

c) Go to Step 1.

Implementation Challenges

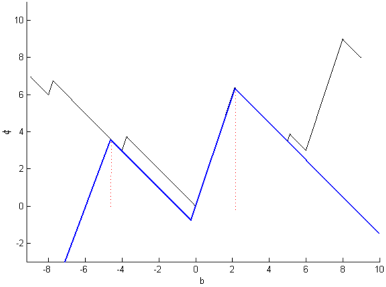
- To make the algorithm practical, several issues need to be solved.
- The master problem includes a piecewise linear function which grows in dimensions.
- Keeping track of the pieces of this function requires an appropriate database.
- The examined right hand sides and their corresponding dual functions also need to be stored in an efficient manner.

Implementation for a Single Constrained Recourse

- For storing dual functions, a "nested hash table" is used.
- The first level of hashing consists of pairs (key = r.h.s, value = linear pieces of dual function).
- The value itself consists of pairs (key = slope, value = intercept).
- Therefore, look ups are cheap.
- A linear piece of B&B tree node is only added if it is stronger than the previously found ones.

Right Hand Side Modification

- Can we do better than blindly solving the master problem to get candidate right hand sides β^k ?
- The pieces obtained from B&B are shifted pieces of z_C.
- Sensitivity analysis on each node of the B&B tree gives us information about $\{\beta : \beta = \lfloor \beta^k \rfloor_{E^*}\}$.
- This allows us to build pieces of the value function locally around the examined right hand side.



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.E. Blair. A closed-form representation of mixed-integer program value functions. *Mathematical Programming*, 71(2):127–136, 1995.

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

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.

V. Laporte François et al. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, 13(3):133–142, 1993.

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.